Project acronym:	BYTE
Project title:	Big data roadmap and cross-disciplinarY community for addressing socieTal Externalities
Grant number:	619551
Programme:	Seventh Framework Programme for ICT
Objective:	ICT-2013.4.2 Scalable data analytics
Contract type:	Co-ordination and Support Action
Start date of project:	01 March 2014
Duration:	36 months
Website:	www.byte-project.eu

Deliverable D7.5:
Standards and methodologies big data guidance

Author(s):	Jarl Magnusson, DNV GL AS Erik Stensrud, DNV GL AS Tore Hartvigsen, DNV GL AS Lorenzo Bigagli, National Research Council of Italy

Dissemination level:	Public
Deliverable type:	Final
Version:	1.1
Submission date:	26 July 2017

Table of Contents

Preface

The BYTE project will assist European science and industry in capturing the positive externalities and diminishing the negative externalities associated with big data in order to gain a greater share of the big data market by 2020. The project comprises three phases of work: a preliminary investigation, an exploration of present and future societal impacts, and the future agenda for big data.

This deliverable captures the results of part of the activity performed in phase three, within Work Package 7 – The big data community (WP7), namely in Task 7.5 – Standards and methodologies, big data guidance.

The overall objectives of WP7 are:

- To design and form the big data community, including drafting founding texts
- To prepare a BYTE project final report, including a series of guidelines supported by community members
- To input BYTE findings and guidelines into relevant networks.

Task 7.5 Description

1. Identify touchpoints between roadmap and vision (policy externalities and research topics), and standards and methodologies.
2. Produce guideline/s for developers and publishers of standards and methodologies to influence their future development.
3. Identify relevant standards for the roadmap items

Executive summary

A standard is a formal document that establishes uniform criteria, methods, processes and practices. In contrast, a custom, convention, company product, corporate standard, and so forth that becomes generally accepted and dominant is often called a de facto standard. A standard may be developed by a project, any type of organization or standards organization (SDO) [8].

BYTE recognizes the use of standards and methodologies as a vital source of knowledge when addressing the challenges with big data usage, development and implementation. Big Data covers so many different topics and situations, so it is not feasible to recommend one comprehensive selection of standards. Standards should instead be regarded as an open and current solution, possible to change, to integrate with other standards and solutions, and a source for further development.

BYTE will be replaced by the BYTE Big Data Community (BBDC), a sustainable, cross-disciplinary platform that will implement the roadmap identified within the project, and will assist the European stakeholders in identifying and meeting the big data challenges, to finally achieve the scenario envisioned by BYTE for 2020.

BBDC would therefore need to continuously refine and update a list of Big Data standards and methodologies, that includes potential solutions, tools and guidelines for policy externalities and research topics.

Developing new Big Data standards and solutions are a good thing. A widely accepted standard, formal or informal, disseminates the knowledge and encourages other researchers and user communities to build upon its already accepted solution.

BBDC should organize itself, dedicate a function to be continuously updated on the current and future list of potential and applicable Big Data standards and methodologies. Just upkeeping and providing information on relevant "big data standards" to the EU, member nations and other communities is a formidable task.

1 Introduction

From the ISO website[1] one can read; *"A standard is a document that provides requirements, specifications, guidelines or characteristics that can be used consistently to ensure that materials, products, processes and services are fit for their purpose. We published over 21000 International Standards that can be purchased from the ISO store or from our members."*

It is almost impossible to create a comprehensive list over "Big Data" standards. New science creates demands for new answers/solutions and as solutions evolve, the need for harmonization and standards increases. We would therefore, just see a list of standards as a snapshot in time, partly relevant, but not all inclusive. This report must therefore be regarded as a partial answer to the tasking of D7.5.

The big data issue can be viewed from many different angles. But, we have focused our search on the data-aspect of big data, and eventually came up with a list of standards from ISO/IEC JTC 1 [1], that in 2015 represented a snapshot or "a current list" of big data standards. We have also used a DNVGL Recommended Practice, DNVGL-RP-0497, "Data quality assessment framework" [9], that lists standards and documents relevant for data quality issues, to show a more detailed segment and its reference standards.

We have done a mapping of JTC1 understanding of big data standards towards policy externalities (economic, social and ethical, legal and political) and research topics (data management, data processing, data analytics, data protection, data visualization, non-technical priorities and other). In order to simplify the mapping, we have detailed externalities by its explanations and clarifications and research topics by its subtopics and requirements. It was possible to map a majority of the standards to the map, due to inclusive reasoning, "in order to address this topic, one must at least address this standard". The standard could then represent a solution or being fit for purpose.

[1] http://www.iso.org/iso/home/standards.htm

2 Big Data Standards Organizations

There are international, regional and national standards organizations (SDO), that participates in the development of big data standards.

https://en.wikipedia.org/wiki/Standards_organization *"There are many international standards organizations. The three largest and most well-established such organizations are the International Organization for Standardization (ISO), the International Electrotechnical Commission (IEC), and the International Telecommunication Union (ITU), which have each existed for more than 50 years (founded in 1947, 1906, and 1865, respectively) and are all based in Geneva, Switzerland.*

In addition to these, a large variety of independent international standards organizations such as the ASME, the ASTM International, the IEEE, the Internet Engineering Task Force (IETF), SAE International, TAPPI, the World Wide Web Consortium (W3C), and the Universal Postal Union (UPU) develop and publish standards for a variety of international uses.

Regional standards bodies also exist, such as the European Committee for Standardization (CEN), the European Committee for Electrotechnical Standardization (CENELEC), the European Telecommunications Standards Institute (ETSI), and the Institute for Reference Materials and Measurements (IRMM) in Europe, the Pacific Area Standards Congress (PASC), the Pan American Standards Commission (COPANT), the African Organization for Standardization (ARSO), the Arabic industrial development and mining organization (AIDMO), and others.

In general, each country or economy has a single recognized national standards body (NSB). A national standards body is likely the sole member from that economy in ISO; ISO currently has 161 members. National standards bodies usually do not prepare the technical content of standards, which instead is developed by national technical societies."

The list of SDOs below are referenced from the [1] ISO/IEC JTC 1 report, http://www.iso.org/iso/big_data_report-jtc1.pdf.

SDO/Consortium	**Interests area on standardization**	**Main deliverables**
[1] ISO/IEC JTC 1/SC 32	Data management and interchange, including database languages, multimedia object management, metadata management, and e-Business.	e-Business standards, including role negotiation; metadata repositories, model specification, metamodel definitions; SQL; and object libraries and application packages built on (using) SQL.
[2] ISO/IEC JTC 1/SC 38	Standardization for interoperable Distributed Application Platform and Services including Web Services, Service Oriented Architecture (SOA), and Cloud Computing	Cloud Data Management Interfaces, Open Virtualization Format, Web Services Interoperability
[3] ITU-T SG13	Cloud computing for Big Data	Cloud computing based big data requirements, capabilities, and use cases.
[4] W3C	Web and Semantic related standards for markup, structure, query, semantics, and interchange.	Multiple standards including ontology specification standards, data markup, query, access control, and interchange.

Open Geospatial Consortium (OGC)	Geospatial related standards for the specification, structure, query, and processing of location related data.	Multiple standards related to the encoding, processing, query, and access control of geospatial data.
[5] Organization for the Advancement of Structured Information Standards (OASIS)	Information access and exchange.	A set of protocols for interacting with structured data content such as OData, standards for security, Cloud computing, SOA, Web services, the Smart Grid, electronic publishing, emergency management, and other areas
[6] Transaction Processing Performance Council	Benchmarks for Big Data Systems	Specification of TPC Express, BenchmarkTM for Hadoop system and the related kit
[7] TM Forum	Enable enterprises, service providers and suppliers to continuously transform in order to succeed in the digital economy	Share experiences to solve critical business challenges including IT transformation, business process optimization, big data analytics, cloud management, and cyber security.

3 Big Data Standards

This listing shows the status of key technologies and relevant standards, models, studies, use cases and scenarios for Big Data from JTC 1, ISO, IEC and other standard setting organizations.

This list of standards is based on the reports from [1] ISO/IEC JTC 1, http://www.iso.org/iso/big_data_report-jtc1.pdf

ISO/IEC JTC 1/SC 32, "Data management and interchange", currently works in several distinct, but related, areas of Big Data technology.

- ISO/IEC JTC 1/SC 32, SQL
- ISO/IEC JTC 1/SC 32, Metadata

ISO/IEC JTC 1/SC 38, titled "Distributed application platforms and services (DAPS)", currently works in several areas related to areas of the Big Data Paradigm:

- ISO/IEC JTC 1/SC 38 Cloud Data Management Interfaces;
- ISO/IEC JTC 1/SC 38 Open Virtualization Format;
- ISO/IEC JTC 1/SC 38 Web Services Interoperability.

[3] ITU-T SG13with the title of "Requirements and capabilities for cloud computing based big data" in July 2013. The scope is:

- ITU-T SG13 Overview of cloud computing based big data;
- ITU-T SG13 Cloud computing based big data requirements;
- ITU-T SG13 Cloud computing based big data capabilities;
- ITU-T SG13 Cloud computing based big data use cases and scenarios.

W3C [4]-work revolves around the standardization of Web technologies. Given that one of the primary contributors to the growth of Big Data has been the growth of the Internet and World Wide Web (WWW) many of the developing standards around web technologies must deal with the challenges inherent in Big Data. Currently the following examples of W3C standard efforts relate to big data technologies and interests:

- W3C Model for Tabular Data and Metadata on the Web
- W3C Delivery Context Ontology
- W3C Efficient XML Interchange
- W3C Linked Data
- W3C Mathematical Markup Language
- W3C OWL Web Ontology Language
- W3C Platform for Privacy Preferences
- W3C Protocol for Web Description Resources (POWDER)
- W3C Provenance
- W3C Relational Database to Resource Description Framework (RDB2RDF)
- W3C Resource Description Framework (RDF)

- W3C Rule Interchange Format (RIF)
- W3C Service Modelling Language
- W3C Sparse Query Language (SPARQL)

W3C Extensible Markup Language (XML) and associated technologies (XQuery, XPath, etc.). Furthermore, W3C has the following data activities which may relate to Big Data:

- W3C Big Data CG
- W3C ETL Markup Language CG
- W3C Resource Description Framework (RDF) WG
- W3C Linked Data Platform (LDP) WG
- W3C Government Linked Data (GLD) WG
- W3C CSV (comma-separated values) on the Web WG
- W3C Provenance WG

Open Geospatial Consortium (OGC). The Open Geospatial Consortium is an international industry consortium of companies, government agencies, research institutes and universities participating to develop standards for interoperable "geo-enable" solutions on the Web, wireless and location-based services. Currently the following examples of OGC standard efforts relate to big data technologies and interfaces:

- OGC Data Model Extension standards (e.g. netCDF, HDF, GeoTIFF, geolSON)
- OGC Registry Services
- OGC Metadata Profiles
- OGC GeoAPI Implementation Standard
- OGC Geospatial eXtensible Access Control Markup Language Encoding Standard (GeoXACML)
- OGC GeoSPARQL – A geographic query language for RDF
- OGC Web Coverage Processing Service (WCPS) an Interface Standard

[5] Organization for the Advancement of Structured Information Standards (OASIS). OASIS is a non-profit consortium that drives the development, convergence and adoption of open standards for the global information society. The following OASIS technical committees and activities are relevant to Big Data:

- OASIS Advanced Message Queuing Protocol (AMQP) TC: Defining a ubiquitous, secure, reliable and open internet protocol for handling business messaging.
- OASIS Key-Value Database Application Interface (KVDB) TC: Defining an open application programming interface for managing and accessing data from database systems based on a key-value model
- OASIS Message Queuing Telemetry Transport (MQTT) TC: Providing a lightweight publish/subscribe reliable messaging transport protocol suitable for communication in M2M/IoT contexts where a small code footprint is required and/or network bandwidth is at a premium.

- OASIS XML Interchange Language (XMILE) for System Dynamics TC: Defining an open XML protocol for sharing interoperable system dynamics models and simulations.
- OASIS Cross-Enterprise Security and Privacy Authorization (XSPA)
- OASIS Darwin Information Typing Architecture (DITA)
- OASIS Directory Services Markup Language (DSML)
- OASIS eXtensible Access Control Markup Language (XACML)
- OASIS XRD (Extensible Resource Descriptor)
- OASIS Open Data Protocol (OData)
- OASIS Search Web Services (SWS)
- OASIS Service Provisioning Markup Language (SPML)
- OASIS Topology and Orchestration Specification for Cloud Applications (TOSCA)
- OASIS Universal Description, Discovery and Integration (UDDI)
- OASIS Unstructured Information Management Architecture (UIMA)

[6] Transaction Processing Performance Council (TPC). The TPC defines transaction processing and database benchmarks to provide objective, verifiable TPC performance data to industry. The lack of easily verifiable performance claims and the absence of a neutral industrywide benchmark for Big Data has led the TPC to create a Big Data Working Group (TPCBDWG) tasked with developing industry standards for benchmarking Big Data systems.

- TPC Big Data Working Group (TPCBDWG)

[7] TM Forum is a global trade association to help enterprises, service providers and suppliers continuously transform to succeed in the digital economy. They are creating the tools, best practice guidance for success in Big Data Analytics. The following TM Forum activities are relevant to Big Data:

- TM Forum The Framework provides operational standards, best practices and tools for leveraging Big Data and Analytics.
- TM Forum's IPDR provides the protocol for collecting and managing large volumes of usage data operating across any digital service infrastructure.

Open Data Platform initiative (ODPi) is a nonprofit organization committed to simplification & standardization of the big data ecosystem with common reference specifications and test suites.

- ODPi Defines, integrates, tests, and certifies a standard "ODPi Core" of compatible versions of select Big Data open source projects.

4 Big Data Quality Standards

Big Data usefulness and trustworthiness is dependent on our ability to verify its data quality. A comprehensive study by DNVGL [9], resulting in a The Recommended Practice for Data Quality (a standard), lists the following standards, regulations and papers that describes how to manage and apply data quality.

- ISO 8000-8, Information and Data Quality. Information and data quality: Concepts and measuring
- ISO 9000, Quality Management Systems
- ISO/IEC 27000, Information Security Management Systems
- ISO 31000, Risk Management.
- W3C [4] Data On The Web Best Practice, W3C. http://www.w3.org/TR/dwbp/
- Data Management Maturity Model. CMMI institute 2014.
- The Practitioner's Guide to Data Quality Improvements, by David Loshin. MK OMG press 2011.
- ISO 55000, Asset Management
- ISO 13379 Condition monitoring and diagnostics of machines - Data interpretation and diagnostics techniques - Part 1: General guidelines
- ISO/IEC/IEEE 15288 Systems and software engineering - System life cycle processes. 2015
- BS ISO/IEC 25010:2011 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models
- The Effect of Data Quality on Data Mining - Improving Prediction Accuracy by Generic Data Cleansing, Stang & al, Proceedings of the 14th International Conference on Information Quality, 2009
- The Effects and Interactions of Data Quality and Problem Complexity on Data Mining, Blake & al, Proceedings of the 13th International Conference on Information Quality, 2008
- DNV GL RP-0496 Cyber Security Resilience Management for ships and mobile offshore units in operation, September 2016
- The Data Management Body of Knowledge (DAMA-DMBOK Guide) First edition 2010.
- Data Quality Assessment - For Sensor Systems and Time Series Data. DNV GL report 2016-0782.
- Understanding sensor systems reliability. DNV GL position paper 2-2016.
- Datiris profiler software. http://www.datiris.com/metrics.html
- A data quality framework applied to e-government metadata. Per Myrseth, Jørgen Stang and Vibeke Dalberg. 2011 International Conference on E-Business and E-Government (ICEE), Shanghai.

- ISO/IEC 11179 Metadata registry.
- Journey to Data Quality. Yang et al. MIT Press 2006.

5 Mapping

Here is the mapping between policy externalities and research topics in the A and B-columns and a limited selection of "big data" standards in the first row. The purpose with the mapping is to verify the applicability of these standards towards detailed externalities or research topics. The table below is further detailed on following pages.

The following pages contain an easier-to-read representation to the above table.

Standards mapped towards Policy Externalities, Economic

Policy Externalities	Externality Group (incl. Explanations, clarifications)	STANDARDS AND METHODS	ISO/IEC JTC 1/SC 32 SQL	ISO/IEC JTC 1/SC 32 METADATA	ISO/IEC JTC 1/SC 38 Cloud Data Management Interfaces;	ISO/IEC JTC 1/SC 38 Open Virtualization Format;	ISO/IEC JTC 1/SC 38 Web Services Interoperability.	ITU-T SG13 cloud computing based big data;	ITU-T SG13 Cloud computing based big data requirements;	ITU-T SG13 Cloud computing based big data capabilities;	ITU-T SG13 big data use cases and scenarios.	W3C Tabular Data and Metadata on the Web	W3C Delivery Context Ontology	W3C Efficient xML Interchange	W3C Linked Data	W3C Mathematical Markup Language	W3C OWL Web Ontology Language	W3C Platform for Privacy Preferences	W3C Protocol for Web Description Resources (POWDER)	W3C Provenance	W3C Relational Database to Resource Description Framework (RDB2RDF)	W3C Resource Description Framework (RDF)	W3C Rule Interchange Format (RIF)	W3C Service Modelling Language	W3C Sparse Query Language (SPARQL)	W3C Extensible Markup Language (xML)	W3C associated xML technologies (xQuery, xPath, etc.)	W3C Big Data CG	W3C ETL Markup Language CG	W3C Resource Description Framework (RDF) WG	W3C Linked Data Platform (LDP) WG	W3C Government Linked Data (GLD) WG	W3C CSV (comma-separated values) on the Web WG	W3C Provenance WG	OGC Data Model Extension standards (e.g. netCDF, HDF, GeoTIFF, geoISON)	OGC GeoAPI Implementation Standard	OGC Geospatial extensible Access Control Markup Language Encoding Standard (GeoxACML)	OGC GeoSPARQL – A geographic query language for RDF	OGC Web Coverage Processing Service (WCPS)	OASIS Advanced Message Queuing Protocol (AMQP)	OASIS Key-Value Database Application Interface (KVDB)	OASIS Message Queuing Telemetry Transport (MQTT)	OASIS xML Interchange Language (xMILE)	OASIS Cross-Enterprise Security and Privacy Authorization (xSPA)	OASIS Darwin Information Typing Architecture (DITA)	OASIS Directory Services Markup Language (DSML)	OASIS extensible Access Control Markup Language (xACML)	OASIS Open Data Protocol (OData)	OASIS Search Web Services (SWS)	OASIS Service Provisioning Markup Language (SPML)	OASIS Topology and Orchestration Specification for Cloud Applications (TOSCA)	OASIS Universal Description, Discovery and Integration (UDDI)	OASIS Unstructured Information Management Architecture (UIMA)
Economic																																																
	Improved efficiency																					x	x	x	x	x	x	x	x	x	x			x															
	* targeted services through data sharing, analysis and profiling		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	* public insight by identifying social trends and statistics		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x											x	x													x	x			x
	Innovation																					x	x	x	x	x	x	x	x	x	x			x															
	* new value chains or major transformation of existing ones		x	x	x	x	x	x	x	x	x	x		x	x	x	x			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	* Reduced innovation due to restrictive legislation		x	x	x	x	x	x	x	x	x	x	x		x					x														x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x
	Changing business models																																																	
	* Challenge of traditional non-digital services												x	x		x			x			x	x	x	x	x	x	x	x	x	x			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	* new services based on a new use of data or by specialisation in specific services		x	x	x	x	x	x	x	x	x	x	x	x	x				x	x													x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	Employment																																																	
	* Data-driven employment offerings. Employment losses for certain job categories																																																	x
	Dependency on public funding																																																	
	* Innovative business models through community building																																																	
Social and ethical																																																	

Standards mapped towards Policy Externalities, Social and Ethical

Social and ethical																																																				
	Improved efficiency and innovation																				x	x	x	x	x	x	x	x	x	x			x																			x
	* Tracking environmental challenges.			x	x	x	x	x	x	x	x			x					x		x	x	x	x	x	x	x	x	x	x	x		x																			x
	* services through data sharing, analysis and profiling			x	x	x	x	x	x	x	x			x				x	x												x								x	x	x	x		x			x	x		x	x	
	Improved awareness and improved decision-making																				x	x	x	x	x	x	x	x	x	x			x																			x
	* Transparency and accountability of the public sector			x	x	x	x	x	x	x	x			x				x	x		x	x	x	x	x	x	x	x	x	x	x		x							x				x								x
	* identifying social trends and statistics	x	x	x	x	x	x	x	x	x	x			x				x	x												x	x								x												
	Participation																				x	x	x	x	x	x	x	x	x	x			x																			
	* Increased citizen participation. Support communities			x	x	x	x	x	x	x	x			x					x												x									x											x	x
	Equality																																																			
	* lack of skills, access, infrastructure or language																																													x						x
	Discrimination																				x	x	x	x	x	x	x	x	x	x			x																			
	* use of data for profiling, concerns about political abuse and prosecuting specific groups	x	x	x	x	x	x	x	x	x	x			x					x		x	x	x	x	x	x	x	x	x	x	x	x	x																			x
	* Private data misuse, especially sharing with third parties without consent			x	x	x	x	x	x	x	x			x					x												x												x									x
	Trust																				x	x	x	x	x	x	x	x	x	x			x																			
	* concerns about exploitation and manipulation, privacy violation or data abuse			x	x	x	x	x	x	x	x			x					x		x	x	x	x	x	x	x	x	x	x	x		x										x									x
	* Public reluctance to provide information. Consumer manipulation. Lack of context or incomplete data. Market manipulation.			x	x	x	x	x	x	x	x			x					x												x												x									x
Legal																																																				

Standards mapped towards Policy Externalities, Legal

Legal																																																				
	Data protection and Privacy																				x	x	x	x	x	x	x	x	x	x			x																			
	* Private data misuse. Threats to data protection and personal privacy. Private data accumulation and ownership			x	x	x	x	x	x	x	x			x					x												x												x									x
	* Continuous and invisible surveillance																																										x									x
	Intellectual Property Rights (IPR)																				x	x	x	x	x	x	x	x	x	x			x																			
	* Private data accumulation and ownership. Reduced innovation due to restrictive legislation.			x	x	x	x	x	x	x	x			x					x												x																					x
	* Need to reconcile different laws and agreements.																				x	x	x	x	x	x	x	x	x	x			x											x								x
	Liablility and accountability			x	x	x	x	x	x	x	x			x					x												x																					
	* Lack of norms for data storage and processing																																	x	x	x	x	x														
Polictical																																																				

Standards mapped towards Policy Externalities, Political

Polictical																																																				
	Relations between private vs. public and non-profit sector																				x	x	x	x	x	x	x	x	x	x			x																			
	* Dependency on external data sources, platforms and services			x	x	x	x	x	x	x	x			x					x												x			x	x	x	x	x						x					x			
	Losing control to actors abroad																				x	x	x	x	x	x	x	x	x	x			x																			
	* Need to reconcile different laws and agreements			x	x	x	x	x	x	x	x			x					x												x																					x
	Improved decision-making and participation																				x	x	x	x	x	x	x	x	x	x			x																			
	* Transparency and accountability of the public sector			x	x	x	x	x	x	x	x			x					x												x																					x
	Political abuse and surveillance																				x	x	x	x	x	x	x	x	x	x			x																			
	* Political tensions due to surveillance out of the boundaries of states.			x	x	x	x	x	x	x	x			x					x												x																					x

Standards mapped towards Research topics, Data Management

Research topics	Detailed research topics (incl. Examples of subtopics or requirements)																																																			
Data Management																																																				
	Handling unstructured and semi-structured data																				x	x	x	x	x	x	x	x	x	x			x																			
	* multilingualism			x	x	x	x	x	x	x	x	x		x		x			x		x	x	x	x	x	x	x	x	x	x	x		x																			x
	* easy-to-use reporting tools (healthcare)	x	x	x	x	x	x	x	x	x	x			x	x	x			x												x	x											x									x
	Semantic interoperability																				x	x	x	x	x	x	x	x	x	x			x																			
	* Data integration	x	x	x	x	x	x	x	x	x	x	x		x	x	x			x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x						x		x
	* Data Fusion	x	x	x	x	x	x	x	x	x	x	x		x	x	x			x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x				x								x
	* semantic search, schema matching and mapping, and ontology alignment	x	x	x	x	x	x	x	x	x	x	x		x	x	x			x												x	x		x	x	x	x	x						x						x		x
	Measuring and assuring data quality																				x	x	x	x	x	x	x	x	x	x			x																			
	* transparency on the data collection and meta info on data context and purpose	x	x	x	x	x	x	x	x	x	x			x					x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x		x								x
	* syntactic, semantic and pragmatic data quality management. algorithms to validate and annotate data	x	x	x	x	x	x	x	x	x	x			x		x			x												x	x		x	x	x	x	x														x
	Data lifecycle																				x	x	x	x	x	x	x	x	x	x			x																			
	* data detection (search and retrieval), capture, ranking, storage	x	x	x	x	x	x	x	x	x	x			x		x			x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x						x			x	x		x		x
	* data citation, curation and preservation for the future			x	x	x	x	x	x	x	x			x		x			x												x			x	x	x	x	x						x						x		x
	Data provenance, control and IPR																				x	x	x	x	x	x	x	x	x	x			x																			
	* Data licensing and ownership, permissions and digital rights, integrity and confidentiality			x	x	x	x	x	x	x	x			x					x		x	x	x	x	x	x	x	x	x	x	x		x										x	x								x
	* provide context through data provenance. Nanopublications			x	x	x	x	x	x	x	x			x					x												x													x								x
	Data-as-a-service model and paradigm.																				x	x	x	x	x	x	x	x	x	x			x																			
	* licensing, ownership, value from data (present and future) and marketplace			x	x	x	x	x	x	x	x			x					x		x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x						x								x
	* open data practices			x	x	x	x	x	x	x	x			x					x												x													x			x	x				x
Data processing																																																				

Standards mapped towards Research topics, Data Processing

Data processing																																																				
	Architectures for data-at-rest and data-in-motion																				x	x	x	x	x	x	x	x	x	x			x																			
	Techniques and tools for processing real-time heterogeneous data	x	x	x	x	x	x	x	x	x	x			x	x				x		x	x	x	x	x	x	x	x	x	x	x	x	x																			
	* new tools for sensor data processing and social media mining	x	x	x	x	x	x	x	x	x	x			x	x				x												x	x																				
	Scalable algorithms and techniques for real-time analytics																				x	x	x	x	x	x	x	x	x	x			x																			
	* high volume of stream data mining		x	x	x	x	x	x	x	x	x			x	x				x												x																					
	Decentralised and distributed architectures																				x	x	x	x	x	x	x	x	x	x			x																			
	* efficient and scalable cryptographic mechanisms for the cloud and attribute-based encryption		x	x	x	x	x	x	x	x	x			x					x												x			x	x	x	x	x					x									
	Efficient mechanisms for storage and processing																				x	x	x	x	x	x	x	x	x	x			x																			
	* automate complex tasks, make them scalable, hybrid human-algorithmic data curation, energy-efficient data storage methods		x	x	x	x	x	x	x	x	x			x	x				x												x			x	x	x	x	x														
Data analytics																																																				

Standards mapped towards Research topics, Data Analytics

Data analytics																																																				
	Improved models and simulations																				x	x	x	x	x	x	x	x	x	x			x																			
	* integration between algorithmic and human computation approaches	x	x	x	x	x	x	x	x	x	x			x	x				x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x									x	x				
	* catchment techniques, recommendations and customer tendency research	x	x	x	x	x	x	x	x	x	x			x	x				x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x														x
	* simulations for the crisis informatics and smart city resource allocation	x	x	x	x	x	x	x	x	x	x			x	x				x												x	x																				x
	Semantic analysis																				x	x	x	x	x	x	x	x	x	x			x																			
	* sentiment analysis			x	x	x	x	x	x	x	x			x	x				x		x	x	x	x	x	x	x	x	x	x	x		x																			
	* using natural language for search, query and analytics, supporting data integration			x	x	x	x	x	x	x	x			x		x			x												x			x	x	x	x	x														x
	Event and pattern discovery																				x	x	x	x	x	x	x	x	x	x			x																			
	* predictive and prescriptive analytics, evidence-driven, bottom-up approach			x	x	x	x	x	x	x	x			x	x				x		x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x														x
	* anomaly detection, pattern recognition			x	x	x	x	x	x	x	x			x	x	x			x												x			x	x	x	x	x														x
	Multimedia (unstructured) data mining																				x	x	x	x	x	x	x	x	x	x			x																			
	* multilingual sentiment analysis			x	x	x	x	x	x	x	x			x					x												x																					x
	Machine learning techniques, deep learning for BI, predictive and prescriptive analytics																				x	x	x	x	x	x	x	x	x	x			x																			
	* visualisation and end-user usability			x	x	x	x	x	x	x	x			x					x		x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x														x
	* validated methodologies and standards behind decision making analytics			x	x	x	x	x	x	x	x			x	x				x		x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x														x
	* assess the representativity of data and possible data biases			x	x	x	x	x	x	x	x			x					x												x			x	x	x	x	x														x
	Context-aware analytics																																																			
Data protection																																																				

Standards mapped towards Research topics, Data Protection

Data protection																																																				
	Complete data protection framework																				x	x	x	x	x	x	x	x	x	x			x																			
	* security challenges are now in non-relational data stores			x	x	x	x	x	x	x	x			x					x		x	x	x	x	x	x	x	x	x	x	x		x										x									
	* allow sharing data on a fine-grained level			x	x	x	x	x	x	x	x			x					x		x	x	x	x	x	x	x	x	x	x	x		x						x		x	x	x	x							x	x
	* New legal means to access and permit use of data, ensuring data protection is not an obstacle for big data practices			x	x	x	x	x	x	x	x			x					x												x								x		x	x									x	x
	Data minimization																																																			
	Privacy-preserving mining algorithms																				x	x	x	x	x	x	x	x	x	x			x																			
	* mining algorithms of social media			x	x	x	x	x	x	x	x			x					x												x																					
	Robust anonymisation algorithms																				x	x	x	x	x	x	x	x	x	x			x																			
	* novel algorithms such as k-anonymity			x	x	x	x	x	x	x	x			x	x				x												x																					
	Protection against reversibility																				x	x	x	x	x	x	x	x	x	x			x																			
	* understand misuse of data, how it needs to be protected and integrated in big data storage solutions			x	x	x	x	x	x	x	x			x					x		x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x					x									x
	Pattern hiding mechanism			x	x	x	x	x	x	x	x			x					x		x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x					x									x
	Secure multiparty mining mechanism			x	x	x	x	x	x	x	x			x					x												x			x	x	x	x	x														x
Data visualisation																																																				

Standards mapped towards Research topics, Data Visualization

Data visualisation																																																				
	End user visualisation and analytics																					x	x	x	x	x	x	x	x	x	x			x																			
	* Natural language interfaces, interactive and easy-to-use data access and transformation methods			x	x	x	x	x	x	x	x			x					x												x			x	x	x	x	x														x
	Dynamic clustering of information																					x	x	x	x	x	x	x	x	x	x			x																			
	* data summarisation and visualisation, and user interfaces for parallel exploration			x	x	x	x	x	x	x	x			x					x												x			x	x	x	x	x														
	New visualisation for geospatial data																					x	x	x	x	x	x	x	x	x	x			x																			
	* user interfaces for parallel exploration			x	x	x	x	x	x	x	x			x					x												x			x	x	x	x	x														
	Interrelated data and semantics relationships																					x	x	x	x	x	x	x	x	x	x			x																			
	* semantic search			x	x	x	x	x	x	x	x	x		x		x			x												x			x	x	x	x	x														x
	Qualitative analysis at a high semantic level																					x	x	x	x	x	x	x	x	x	x			x																			
	* Intuitive data transformation interfaces			x	x	x	x	x	x	x	x			x					x		x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x									x	x				
	Real-time and collaborative 3-D visualisation			x	x	x	x	x	x	x	x			x	x				x		x	x	x	x	x	x	x	x	x	x	x		x																			
	Time dimension of big data			x	x	x	x	x	x	x	x			x	x				x		x	x	x	x	x	x	x	x	x	x	x		x																			
	Real-time adaptable and interactive visualisation			x	x	x	x	x	x	x	x			x	x				x												x																					
Non-technical priorities																																																				

Standards mapped towards Research topics, Non-technical priorities

Non-technical priorities																																																				
	Data standardisation																					x	x	x	x	x	x	x	x	x	x			x																			
	Establish and increase trust			x	x	x	x	x	x	x	x			x		x			x		x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x														
	* method to increase trust and transparency			x	x	x	x	x	x	x	x			x					x												x													x								x
	Privacy-by-design, security-by-design, anti-discrimination-by-design																					x	x	x	x	x	x	x	x	x	x			x																			
	* privacy-by-design and similar by-design approaches are vital			x	x	x	x	x	x	x	x			x					x		x	x	x	x	x	x	x	x	x	x	x		x																	x		
	* transparent handling of data and algorithms and company audits should be considered			x	x	x	x	x	x	x	x			x	x				x												x																					
	Ethical issues																					x	x	x	x	x	x	x	x	x	x			x																			
	* research that analyses human data should fall within the regulations of research based on human subjects			x	x	x	x	x	x	x	x			x					x		x	x	x	x	x	x	x	x	x	x	x		x																			
	* scientific community to access data owned by companies for research purposes			x	x	x	x	x	x	x	x			x					x		x	x	x	x	x	x	x	x	x	x	x		x							x												
	* quantify the risks posed by data science practices that rely on big data			x	x	x	x	x	x	x	x			x					x		x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x														
	* account for and mitigate the risks of sharing datasets that can be later combined with auxiliary datasets			x	x	x	x	x	x	x	x			x					x		x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x														
	* Usage of publicly available, although illicitly obtained data sets			x	x	x	x	x	x	x	x			x					x												x													x								x
	* Bias is also a relevant ethical issue that needs further research			x	x																																																	x
	Develop new business models																					x	x	x	x	x	x	x	x	x	x			x																			
	* moving beyond the Open Data Initiative to an interoperable data scheme to process data from heterogeneous sources			x	x	x	x	x	x	x	x			x		x			x		x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x		x				x								x
	* pre-competitive partnerships, that do not affect competitive advantage			x	x	x	x	x	x	x	x			x		x			x												x																					
	* public-private partnerships			x	x											x																												x								
	Citizen research																					x	x	x	x	x	x	x	x	x	x			x																			
	* Crowdsourcing may be used to increase data accuracy			x	x	x	x	x	x	x	x			x	x				x		x	x	x	x	x	x	x	x	x	x	x		x											x								
	* to better understand the social engagement mechanisms			x	x	x	x	x	x	x	x			x					x												x																						x
	Discrimination discovery and prevention																					x	x	x	x	x	x	x	x	x	x			x																			
	* research on legal informatics and algorithm accountability is needed			x	x	x	x	x	x	x	x			x					x												x			x	x	x	x	x														x
Other																						x	x	x	x	x	x	x	x	x	x			x																			
	Data-intensive engineers	x	x	x	x	x	x	x	x	x	x			x	x				x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x		x
	Data scientist	x	x	x	x	x	x	x	x	x	x			x	x	x			x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x		x
	Data-intensive business experts	x	x	x	x	x	x	x	x	x	x			x		x			x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	Technology standardisation		x	x	x	x	x	x	x	x	x			x	x	x			x												x								x	x	x	x	x	x	x		x	x	x	x		x

6 Standard development guideline/s

The ISO Standard Development Process, and its associated key principles for development of standards (applicable for "Big Data" standards), is presented at its website, [10].

"ISO standards respond to a need in the market

- *ISO does not decide when to develop a new standard, but responds to a request from industry or other stakeholders such as consumer groups. Typically, an industry sector or group communicates the need for a standard to its national member who then contacts ISO. Contact details for national members can be found in the list of members.*

ISO standards are based on global expert opinion

- *ISO standards are developed by groups of experts from all over the world, that are part of larger groups called technical committees. These experts negotiate all aspects of the standard, including its scope, key definitions and content. Details can be found in the list of technical committees.*

3. ISO standards are developed through a multi-stakeholder process

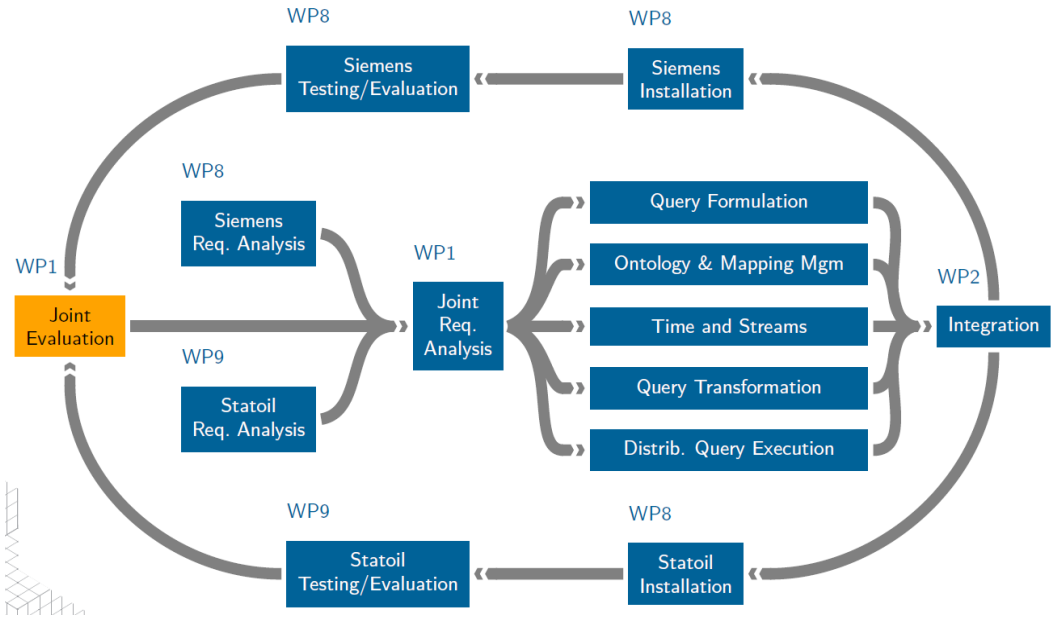
- *The technical committees are made up of experts from the relevant industry, but also from consumer associations, academia, NGOs and government. Read more about who develops ISO standards.*

4. ISO standards are based on a consensus

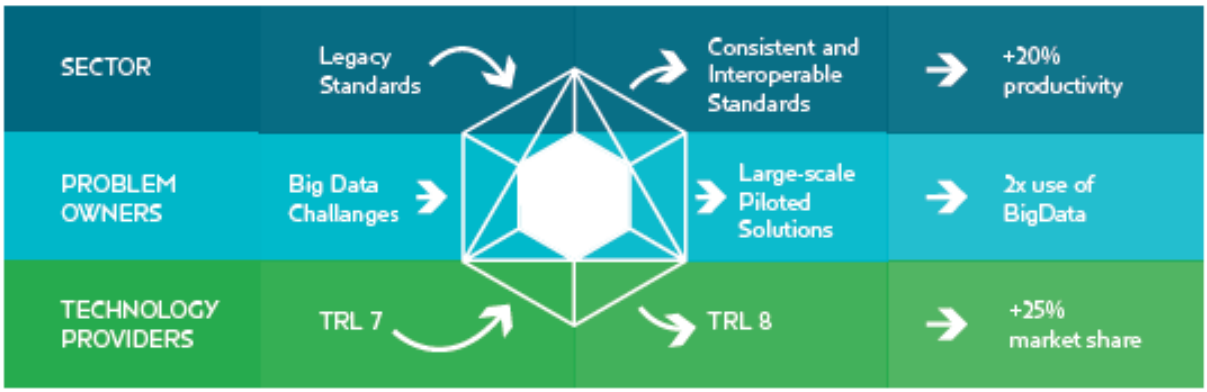
- *Developing ISO standards is a consensus-based approach and comments from all stakeholders are considered."*

The development process used in EU project Optique have gained recognition. It is an iterative process that could be applied as a standard development guideline:

1. Joint evaluation (WP1), refine the desired outcome, allocate resources
2. Requirements break-down and analysis (WP8, WP9), identify changes and/or new development, identify tasks, specify solutions and deliverables
3. Joint requirement analysis (WP1), plan tasks
4. Develop solutions, execute changes
5. Integrate solutions (WP2) into new and/or updated standards
6. Install, test and evaluate (WP8), verify towards requirements
7. Goto 1.



EU Project Optique's follow-on project notes that legacy standards will be used in the project as a source of current knowledge, and during the development process will the standards be improved to become more consistent and interoperable.



7 Conclusions and recommendations

7.1 Conclusions

There are numerous standards, specifications and recommended practices that potentially affect how we structure, represent, manage and make good use of Big Data. But, even more laws, rules, regulations, and standards, are required to manage and control Big Data when it is being used for specific business domains; such as Defense, Energy, Automotive, Space, Public Sector and others.

General Big Data standards are issued by numerous bodies, including, but not limited to, the following organizations: ISO/IEC JTC 1, ITU-T SG13, W3C, OGC, OASIS, TPC, TM Forum, and ODPi. In addition to general Big Data standards, we also deemed as relevant Big Data *Quality* standards from the following bodies ISO, W3C, CMMI, OMG, DAMA-DMBOK, as well as from DNV GL. From all these sources of standards, and with the support from the ISO Big Data Report [1], a subset of standards was deemed relevant for the BYTE Big Data Roadmaps and therefore included in the document.

The purpose of using a mapping table was to identify a potential relationship (or as it was stated in the Task 7.5 Description - *Identify touchpoints*) between "Big Data standards" and the BYTE Roadmaps, externalities and research topics, respectively. Mapping was only possible due to *inclusive reasoning*. By this, we mean that for every BYTE Roadmap item in the mapping tables, we used our expert opinion to select the *minimum set* of standards that at least should be in scope for this particular Roadmap item. In other words, we went *row first*. The standard could then represent a set of requirements, a solution, or being fit for purpose.

As an illustration of our general method of identifying the relevance of a standard to a given BYTE Roadmap item, we provide the following case example on the mapping between "Policy externalities/economic/Innovation" and "W3C Provenance WG". The W3C (World Wide Web Consortium) listings are a mix of standards and working groups (WG) that might refer to, or result in, de-facto standards or best practices. Most W3C work revolves around the standardization of Web technologies. The "W3C Provenance WG" is a standard development working group, working to develop a standard, with working name "PROV", specifying how to represent provenance data for exchange online. "W3C Provenance WG" is not mapped towards the general term "Innovation", but to its two subtopics due to the need to understand, record and trace the place or source of original work. Every selected intersection in the mapping table has been evaluated in a similar fashion.

This approach resulted in the mapping table, which has 51 Big Data standards listed in columns and 136 BYTE Roadmap Policy Externalities and Research topics headlines (HL) and bullet points listed in rows, creating a total of 6936 potential intersections. The mapping was based on the question *"does this topic contain any Big Data technical challenges?",* resulting in 2246 intersections, or 32%:

- Policy Externalities, Economics: No standards mapped towards Changing Business Models (HL), Employment (HL) and Dependency on public funding (HL)
- Policy Externalities, Social and Ethical: Few standards mapped towards Equality (HL)
- Policy Externalities, Legal: All headlines and bullets had standards mapped
- Policy Externalities, Political: All headlines and bullets had standards mapped
- Research topics, Data Management: All headlines and bullets had standards mapped
- Research topics, Data Processing: All headlines and bullets had standards mapped
- Research topics, Data Analytics: All headlines and bullets had standards mapped
- Research topics, Data Protection: All headlines and bullets had standards mapped
- Research topics, Data Visualization: All headlines and bullets had standards mapped
- Research topics, Non-technical priorities: All headlines and bullets had standards mapped

We have looked at Social, Ethical and Legal topics. The excerpt table below details the coverage of standards for these externalities. Red is regarded as a non-acceptable coverage of 0-10 standards out of 51, Yellow-some coverage valued at 10-20 standards and Green-acceptable coverage at 20 standards and above):

Social and Ethical	# of Stds
Improved efficiency and innovation	12
• Tracking environmental challenges.	23
• Services through data sharing, analysis and profiling	21
Improved awareness and improved decision-making	12
• Transparency and accountability of the public sector	26
• identifying social trends and statistics	16
Participation	11
• Increased citizen participation. Support communities	14
Equality	0
• lack of skills, access, infrastructure or language	2
Discrimination	11
• Use of data for profiling, concerns about political abuse and prosecuting specific groups	26
• Private data misuse, especially sharing with third parties without consent	13
Trust	11
• Concerns about exploitation and manipulation, privacy violation or data abuse	24
• Public reluctance to provide information. Consumer manipulation. Lack of context or incomplete data. Market manipulation.	13
Legal	
Data protection and Privacy	11
• Private data misuse. Threats to data protection and personal privacy. Private data accumulation and ownership	13
• Continuous and invisible surveillance	2
Intellectual Property Rights (IPR)	11
• Private data accumulation and ownership. Reduced innovation due to restrictive legislation.	12
• Need to reconcile different laws and agreements.	13
Liablility and accountability	11
• Lack of norms for data storage and processing	5

Overall, *social and ethical issues* are well covered by standards from ISO and W3C, while standards from OGC and OASIS are not selected to cover social and ethical aspects of Big Data. Our findings show that the "Equality" issue is not well covered by any standards organizations, except for the *OASIS extensible Access Control Markup Language* (xACML). Furthermore,

from a DNV GL perspective, “Trust” is well covered by ISO standards, but not by W3C when it comes to “Public reluctance to provide information”.

As for *legal issues,* there are in general more gaps by all standards organizations. However, again, ISO and W3C have some coverage, but there are significant gaps related to the “Continuous and invisible surveillance” issue. Only two OASIS standards address this issue. There are also big gaps in standards related to “Lack of norms for data storage and processing”. Only five OGC standards cover this issue. The entire legal externality including privacy, IPR and liability topics needs further research.

To conclude Social, Ethical and Legal issues. These provides an indication that the “soft” has some momentum, and gladly we can note that the topic of personal data protection has been in the public focus more than ever. One of the reasons for this is the upcoming EU General Data Protection Regulation (GDPR) that will come into force in May 2018 and which will replace local personal data protection laws.

As a word on precaution on the applicability of the mapping table, it should be observed that *standards evolve over time*. We observe how digital content has been represented for technical documentation over time, for example from the 1960-70s with plain ASCII text to GML (Generalized Markup Language from IBM) and structured taxonomies, in the 1980s, to a more advanced approach through SGML (Standardized General Markup Language), to HTML (HyperText Markup Language) in the 1990s, for building Web Pages, to XML (Extensible Markup Language), which is by itself a whole new family of standards, and it is evolving as we speak with semantics and ontologies. Big Data standards have therefore time-, content-, and applicability constraints. Now in 2017, we are not using GML, SGML, or HTML anymore, nor the technology needed to implement these standards, nor its content descriptions or models.

Every moment in time has therefore its own set (a snapshot) of Big Data standards. If the lifecycle of data/information exceeds the actuality and usefulness of standards, then we create a compatibility issue, that might require us to access old standards, so we can use them for mapping and conversion, transforming the data from old to new representations.

In summary, we conclude:

- There is no such thing as a complete list of standards for BYTE. Standards are constantly “in flux”, being updated or initiating new standards development to reflect new science.
- We have not included *all* relevant standards. The listed standards apply mostly to the *technical challenges* of Big Data; we have not identified standards that are equally applicable to the “soft” social, ethical, economic and legal aspects of the Big Data Roadmap.
- Mapping was only possible due to inclusive reasoning, “in order to address a BYTE externality or research topic, one must at least take notice of what is in this standard”. The standard could then represent a set of requirements, a solution, or being fit for purpose.

- We have not selected to describe all 2246 intersections in the mapping table, which would be a formidable task and especially when the current selection might not be deemed correct from another person's point-of-view.
- We have not in the mapping table reflected the progression of standards, and the possible compatibility issues that it might create.

7.2 Recommendations

The BYTE project makes the following recommendations that could help developers and publishers of standards and methodologies to influence their future development.

- It is preferable for any organization to have a keen awareness over current Big Data standards. As an initial source, we do recommend to study the ISO/IEC JTC 1, http://www.iso.org/iso/big_data_report-jtc1.pdf
- This BYTE report summarizes 136 BYTE Roadmap Policy Externalities and Research topics headlines (HL) in the Column A of the mapping table, as a guideline for easier selection.
- This report lists 51 Big Data standards listed in Row 1 of the mapping table, as a guidance for easier selection.
- When applying Big Data to future BBDC projects for any BYTE externality or research topic, a survey of applicable standards and regulations should be done. **T**he mapping table in this report can serve as a guideline, example, or as a template. Do note, however, that there is no right number of standards, and most standards have a time, content and applicability constraints.
- Social, Ethical and Legal externalities need more research to develop technical and regulatory guidelines (laws, rules, regulations and standards). The need is especially critical for Legal topics, due to the strong dependency from legal prerequisites to other Policy externalities. Nevertheless, some issues like digital equality, the social impacts of data-based surveillance activities and the creation of norms for data storage and processing require immediate attention to provide some basic coverage.

A final, and probably the most important recommendation, the guidelines in this BYTE report is not a substitute for a certification process, carried out by certified bodies, like for example DNV GL.

8 Terms and definitions

API	Applications Programming Interface
Big Data Analytics	Analytical functions to support the integration of results derived in parallel across distributed pieces of one or more data sources. This is a rapidly evolving field both in terms of functionality and the underlying programming model
Big Data Engineering	Storage and data manipulation technologies that leverage a collection of horizontally coupled resources to achieve a nearly linear scalability in performance
Big Data Models	Logical data models (relational and non-relational) and processing/computation models (batch, streaming, and transactional) for the storage and manipulation of data across horizontally scaled resources
Big Data Paradigm	Distribution of data systems across horizontally-coupled independent resources to achieve the scalability needed for the efficient processing of extensive datasets
BSON	Binary JSON, representing simple data structures and associative arrays
CAGR	Compound Annual Growth Rate
Capability	Quality of being able to perform a given activity. [SOURCE: ISO 15531-1:2004]
Cloud computing	Paradigm for enabling network access to a scalable and elastic pool of shareable physical or virtual resources with self-service provisioning and administration on-demand. [SOURCE: Recommendation ITU-T Y.3500 | ISO/IEC 17788:2014]
DAPS	Distributed Application Platforms and Services
Framework	Structure expressed in diagrams, text, and formal rules which relates the components of a conceptual entity to each other. [SOURCE: ISO 17185-1:2014]
GPS	Global Positioning System
HTML	Hypertext Markup Language
ICT	Information and Communication Technology
IEC	International Electrotechnical Commission
Internet of Things	Integrated environment, inter-connecting anything, anywhere at anytime. [SOURCE: ISO/IEC JTC 1 SWG 5 Report:2013]
ISO	International Standards Organization

ISO/IEC JTC 1/SC 6	Telecommunications and information exchange between systems
ISO/IEC JTC 1/SG 1	Smart Cities
ISO/IEC JTC 1/SG 2	Big Data
ISO/IEC JTC 1/SWG 5	Internet of Things, (IoT)
ITU-T	International Telecommunications Union-Telecommunications Standardization Sector
JSON	Java Script Object Notation
JTC 1	Joint Technical Committee 1
Lifecycle	Evolution of a system, product, service, project or other human-made entity from conception through retirement. [SOURCE: ISO/IEC/TR 29110‑1:2011]
MPP	Massively Parallel Processing
NB	National Body
NIST	National Institute of Standards and Technology
Non-Relational Models	Logical data models such as document, graph, key value and others that are used to provide more efficient storage and access to non-tabular datasets
NoSQL	Not Only Structured Query Language. Data-stores and interfaces that are not tied to strict relational approaches. Note 1 to entry: Alternately called “no SQL” or “not only SQL”
OASIS	Organization for the Advancement of Structured Information Standards
OGC	Open Geospatial Consortium
Ownership	Legal right of possession, including the right of disposition, and sharing in all the risks and profits commensurate with the degree of ownership interest or shareholding, as demonstrated by an examination of the substance, rather than the form, of ownership arrangements. [SOURCE: ISO 10845‑5:2011]
POSIX	Portable Operating System Interface
Privacy	Right of individuals to control or influence what information related to them may be collected and stored and by whom and to whom that information may be disclosed. [SOURCE: ISO/TS 17574:2009]

Provenance	Information on the place and time of origin or derivation or a resource or a record or proof of authenticity or of past ownership. [SOURCE: ISO 19153:2014]
RA	Reference Architecture
Relational model	Data model whose structure is based on a set of relations. [SOURCE: ISO/IEC 2382‑17:1999]
Repository	Collection of all software-related artefacts belonging to a system or the location/format in which such a collection is stored. [SOURCE: ISO/IEC IEEE 24765:2010]
RFID	Radio-Frequency Identification
Role	Set of activities that serves a common purpose. [SOURCE: Recommendation ITU-T Y.3502 | ISO/IEC 17789:2014]
SC	Standards Committee
Schema-on-read	Big data is often stored in a raw form based on its production, with the schema, needed for organizing (and often cleansing) the data, is discovered and transformed as the data is queried
SDO	Standards Development Organization
Security	All aspects related to defining, achieving, and maintaining confidentiality, integrity, availability, non-repudiation, accountability, authenticity, and reliability of a system. [SOURCE: ISO/IEC 15288:2008]
Sensor	Device that observes and measures a physical property of a natural phenomenon or man-made process and converts that measurement into a signal. [SOURCE: ISO/IEC 29182‑2:2013]
SG	Study Group
Smart grid	Electric grid system, which is characterized by the use of communication networks and the control of grid components and loads. [SOURCE: ISO/IEC/TR 27019:2013]
SOA	Service-Oriented Architecture
SQL	SQL Query Language
SQL/MM	SQL Multimedia
Streaming data	Data passing across an interface from a source that is operating continuously. [SOURCE: ISO/IEC 19784‑4:2011]
SWG	Special Working Group

TPC	Transaction Processing Performance Council
Traceability	Property that allows the tracking of the activity of an identity, process, or an element throughout the supply chain. [SOURCE: ISO/IEC 27036-3:2013]
W3C	World Wide Web Consortium
XML	Extensible Markup Language

9 References

[1] ISO/IEC JTC 1, http://www.iso.org/iso/big_data_report-jtc1.pdf

[2] Cloud Security Alliance, https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Expanded_Top_Ten_Big_Data_Security_and_Privacy_Challenges.pdf

[3] ITU, http://www.itu.int/dms_pub/itu-t/oth/23/01/T23010000220001PDFE.pdf

[4] World Wide Web Consortium (W3C); http://www.w3.org

[5] Organization for the Advancement of Structured Information Standards (OASIS); https://www.oasis-open.org/standards

[6] Transaction Processing Performance Council (TPC). Additional information on these and other potentially applicable TPC standards can be found at http://www.tpc.org/default.asp.

[7] TM Forum is a global trade association. Additional information on these and other TM Forum standards can be found at http://www.tmforum.org.

[8] https://en.wikipedia.org/wiki/Technical_standard

[9] DNVGL Recommended Practice, DNVGL-RP-0497, "Data quality assessment framework", https://www.dnvgl.com/data-platform

[10] http://www.iso.org/iso/home/standards_development.htm