

ICS 01.140.20; 35.240.30

**Taxonomie zu Regelwerken bei Smart Data;
Text in Englisch**Taxonomy of Rules and Regulations in Smart Data;
Text in EnglishTaxonomie des règles et règlements sur le smart data;
Texte en anglais

Zur Erstellung einer DIN SPEC können verschiedene Verfahrensweisen herangezogen werden:
Das vorliegende Dokument wurde nach den Verfahrensregeln einer PAS erstellt.

Gesamtumfang 40 Seiten

Dieses Dokument wurde durch die im Vorwort genannten Verfasser erarbeitet und verabschiedet.

Contents

	Page
Foreword	4
Introduction.....	5
1 Scope	5
2 Terms and definitions.....	6
3 Fields of Standardization.....	8
3.1 General	8
3.2 Core Fields of Standardization.....	9
3.2.1 General	9
3.2.2 Technical fields.....	9
3.2.3 Economic fields.....	9
3.2.4 Legal fields	9
3.3 Additional fields.....	10
4 Data Value Chain	10
4.1 General	10
4.2 Steps of the Data Value Chain.....	10
4.3 Common aspects	11
5 Representation and refinement.....	11
5.1 General	11
5.2 Basic representation	11
5.3 Limitation to fewer sections	11
5.4 Refinement.....	11
6 Template for the description of rules and regulations.....	12
6.1 General	12
6.2 Properties in the template	12
6.3 Further specification of the properties	14
6.4 Representation	15
7 Application scenarios and example use cases	16
7.1 General	16
7.2 Cataloguing rules and regulations.....	16
7.3 Finding relevant rules and regulations	16
7.4 Identifying gaps in standardization.....	17
Annex A (informative) Related activities.....	18
A.1 European Union/European Commission.....	18
A.2 Big Data Value Association.....	18
A.3 NIST Big Data Interoperability Framework.....	18
A.4 Other activities	19
Annex B (informative) Examples of Fields of Standardization.....	20
Annex C (informative) Examples of the Data Value Chain.....	22
C.1 Variation on data types	22
C.2 Example of a Data Value Chain for textual data.....	22
C.3 Example of a Data Value Chain on sensor data	23
Annex D (informative) Examples of representations.....	24

Annex E (normative) Types for the description of rules and regulations.....	25
E.1 Basic types.....	25
E.2 Types given by standards.....	25
E.3 Structured types.....	25
E.4 Controlled vocabulary	26
E.4.1 Status.....	26
E.4.2 Field of standardization	26
E.4.3 Step of the Data Value Chain.....	27
Annex F (informative) Examples of the template for the description of rules and regulations.....	28
F.1 General	28
F.2 Resource Description Framework (RDF)	28
F.3 Web Ontology Language (OWL)	30
F.4 MapReduce.....	33
F.5 EU General Data Protection Regulation (EU-GDPR)	35
F.6 ISO SQUARE standard (ISO 25012)	36
Annex G (informative) Examples of application	38
G.1 Setting.....	38
G.2 Use of rules and regulations.....	38
G.2.1 Technical fields	38
G.2.2 Economic fields	38
G.2.3 Legal fields	38
G.3 Classification	39

Foreword

This DIN SPEC has been developed according to the PAS procedure. The development of a DIN SPEC according to the PAS procedure is carried out in workshops and does not require the participation of all stakeholders.

This document has been developed and adopted by the initiator(s) and authors named below:

- FZI Forschungszentrum für Informatik
Stefan Jähnichen, Luise Kranich, Leif Bonorden
- Aeternam ISÆN
David Robert
- BMWi Bundesministerium für Wirtschaft und Energie
Volker Genetzky
- DFKI Deutsches Forschungszentrum für Künstliche Intelligenz
Felix Sasaki, Jack Thoms
- GI Gesellschaft für Informatik e. V.
Daniel Krupka
- IESE Fraunhofer-Institut für Experimentelles Software Engineering
Jörg Dörr, Michael Kläs
- InfAI Institut für Angewandte Informatik e. V.
Markus Freudenberg, Kay Müller
- ITSO IT Service Omikron GmbH
Stephan Drooff

At present, there are no standards covering this topic.

DIN SPECS are not part of the body of German Standards.

Despite great efforts to ensure the accuracy, reliability and precision of technical and non-technical information, the Workshop cannot give any explicit or implicit assurance or warranty in respect of the accuracy of the document. Users of this document are hereby made aware that the Workshop cannot be held liable for any damage or loss. The application of this DIN SPEC does not release users from the responsibility for their own actions and is applied at their own risk.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. DIN [and/or DKE] shall not be held responsible for identifying any or all such patent rights.

A draft of this DIN SPEC has not been published.

Introduction

In the field of *Big Data*, technology has evolved more rapidly than respective formal standardization. Rather, some solutions became de-facto standards although they had not been discussed and documented as such and often come in forms other than formal specifications. When related fields such as economic systems and legal regulations are included – resulting in *Smart Data* – the variety becomes unmanageable with additions such as best practice collections and industry's own commitments – especially for smaller enterprises and players new to the field.

Possible scenarios where an overview on rules and regulations is needed include assistance for developers in creating new products, supporting communication with stakeholders, making research results compatible and simplifying the work of reviewers.

Systems for the classification of formal standards do not fit the needs of Smart Data activities, as the items considered do not always meet the requirements for formal standards. Currently there is neither an abstract framework for their classification nor a supporting software solution.

DIN SPEC 91349 introduces a two-dimensional classification schema for these rules and regulations. The *Fields of Standardization* (Clause 3) are seen in terms of their respective point of direct influence, while the distinction in the steps of the *Data Value Chain* (Clause 4) considers the point of application in a data process. The classification is accompanied by a *Template for the description of rules and regulations* (Clause 6). Classification and template cover various types of rules and regulations in the field of Smart Data. The specification is suitable both as a framework and as a software solution used for implementation.

In Clause 7, examples of the taxonomy's usage are presented for the following application scenarios, while other applications are also possible:

- Cataloguing rules and regulations;
- Finding relevant rules and regulations;
- Identifying gaps in standardization.

The relation to other activities and standardization efforts in Big Data and Smart Data is detailed in Annex A "Related activities".

1 Scope

This DIN SPEC presents a classification framework for technical rules, specifications and solutions, economic models and guidelines, laws and legal regulations, as well as other documents describing requirements or guidance in technical, economic or legal areas for products and services within the field of Smart Data.

Products and services considered in Smart Data are all applications which use Big Data technology and do not serve the sole purpose of technically processing large amounts of data, and are considered embedded in an economic and societal environment. Products and services in the core field of Big Data may also be classified, but do not benefit to the same degree.

This DIN SPEC is applicable to all industry sectors using Smart Data.

This DIN SPEC focuses on, but is not limited to, the scenarios in small and medium enterprises.

2 Terms and definitions

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

2.1

access

obtainment of the use of a resource

[SOURCE : ISO/IEC 2382:2015, 2121274, modified - To be compatible with the scope of this DIN SPEC.]

2.2

algorithm

finite ordered set of well-defined rules for the solution of a problem

[SOURCE: ISO/IEC 2382:2015, 2121376, modified - Notes deleted]

2.3

business model

organizations' approach to operating in its environment

[SOURCE: ISO 30400:2016, 3.4]

2.4

business process

partially ordered set of enterprise activities

[SOURCE: ISO 15531-31:2004, 3.5.2, modified - To be compatible with the scope of this DIN SPEC.]

2.5

computing platform

service to host applications while providing fundamental functionality

2.6

contract design

methods and processes of drafting conditions of contract coordinated with economic interests either specific to contracting parties or as fundamental terms for an offered service

2.7

data

recorded information

[SOURCE: ISO 22005:2007, 3.11]

2.8

data format

arrangement of data in a file or stream

[SOURCE: ISO/TS 27790:2009, 3.18]

2.9

data governance

system of directing and controlling the use of data

[SOURCE : ISO/IEC 38500:2015, 2.8, modified - To be compatible with the scope of this DIN SPEC.]

2.10

data licensing

authorization for usage of data under specified conditions

[NOTE: See ISO 5127:2001, 7.2.1.10, modified - To be compatible with the scope of this DIN SPEC]

2.11

data quality

degree to which the characteristics of data satisfy stated and implied needs when used under specified conditions

[SOURCE: ISO/IEC 25012:2008, 4.3]

2.12

data storage

<Smart Data> unit into which data can be placed, in which they can be retained, and from which they can be retrieved

[SOURCE: ISO/IEC 2382:2015, 2121280, modified - To be compatible with the scope of this DIN SPEC; storage device was replaced by data storage.]

2.13

internal policies

policies mandatory for a respective unit that do not necessarily apply to others

Note 1 to entry: Examples of such policies are conditions set by a higher organizational structure (company-wide regulation, which has to be applied by a specific unit) or on a greater scope (previously made commitment which is not permitted to be changed for the current situation)

2.14

laws and regulations

mandatory instructions issued by a legislative body and other conditions with an accepted legally binding status

2.15

management

coordinated activities to direct and control an organization

[SOURCE: ISO 24511:2007, 2.20, modified - Notes deleted.]

2.16

management control

verification that acceptance criteria of management activities are met

[SOURCE: ISO 22716:2007, 2.13, modified - Term expanded by “management”.]

2.17

metadata

data about data or data elements, possibly including their data descriptions, and data about data ownership, access paths, access rights and data volatility

[SOURCE: ISO/IEC 2382:2015, 2121505, modified - Notes deleted.]

2.18

organizational agreement

instruction agreed upon by a community of different parties

2.19

performance

degree to which a system or component accomplishes its designated functions within given constraints, such as speed, accuracy, or memory usage

[SOURCE: ISO/IEC/IEEE 24765:2010, 3.2107]

2.20

protection

implementation of administrative, technical, or physical measures to guard against the unauthorized access to or usage of data

[SOURCE : ISO/IEC 2382:2015, 2126371, modified - To be compatible with the scope of this DIN SPEC.]

2.21

provenance

information on the origin or derivation of data

[SOURCE: ISO 19153:2014, 4.39, modified - To be compatible with the scope of this DIN SPEC.]

2.22

semantics

information on the meaning of data

2.23

service interaction

communication between systems

2.24

validation

confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled

[SOURCE: ISO 9000:2015, 3.8.13, modified - Notes deleted.]

2.25

verification

confirmation, through the provision of objective evidence, that specified requirements have been fulfilled

[SOURCE: ISO 9000:2015, 3.8.12, modified - Notes deleted.]

3 Fields of Standardization

3.1 General

Rules and regulations apply to different areas of a product or service. While often only technical standards are considered, further aspects are of importance in Smart Data. *Fields of Standardization* classify the direct content of a rule or regulation – in contrast to the usage in a process or the objective of standardization.

These fields are given for the context of Smart Data in the following. Examples of the fields are given in Annex B “Examples of *Fields of Standardization*”.

3.2 Core Fields of Standardization

3.2.1 General

The core fields are divided into three groups: “Technical” for the concrete processing of data, “Economic” for self-chosen conditions and procedures, and “Legal” for rules mandatory in the considered situation.

3.2.2 Technical fields

Technical fields correspond to aspects regarding the data itself as well as storage and processing of data and processes’ results.

The following fields consider aspects of *Infrastructure and Platform*:

- T1 Data storage;
- T2 Computing platforms;
- T3 Performance;
- T4 Access;
- T5 Protection.

The following fields consider aspects of *Data*:

- T6 Data formats;
- T7 Semantics;
- T8 Data Quality;
- T9 Further Metadata.

The following fields consider aspects of *Process*:

- T10 Algorithms;
- T11 Validation and verification;
- T12 Provenance;
- T13 Process quality;
- T14 Service interaction.

3.2.3 Economic fields

Economic fields cover the planning of businesses and activities and their connection to partners, as well as their evaluation. Conditions in these fields are mostly set by the respective units themselves.

- E1 Data governance and licensing;
- E2 Business models and processes;
- E3 Contract design;
- E4 Management;
- E5 Management control.

3.2.4 Legal fields

Legal fields cover rules that are mandatory for the situation considered.

- L1 Laws and Regulations;
- L2 Organizational Agreements;
- L3 Internal Policies.

3.3 Additional fields

An optional item for fields not covered above is usable for each group:

T15	Further technical information;
E6	Further economic information;
L4	Further legal information.

4 Data Value Chain

4.1 General

The concept of physical value chains in industrial processes has been transferred to digital data. Processes on data and information, and the related increase in value, can be modelled as a sequence of steps.

Several largely similar models exist for the *Data Value Chain*. In the following, a model consisting of six steps is presented for Smart Data, which is generally compatible with existing models in related fields.

The steps of the *Data Value Chain* cover the following: *Preparation* (Step I), the *Core process* (Steps II–V) and the *Effect* (Step VI). Examples of the *Data Value Chain* are given in Annex C “Examples of the *Data Value Chain*”.

4.2 Steps of the Data Value Chain

I. Data generation

Data is generated, measured or otherwise initially stored, and made available for access and further processing. In general, this is independent of future processing or analysis. Data sources are identified and selected. The connection between the processing system and the data sources is enabled and established.

II. Data acquisition

Actual data is collected from the data sources. The data is stored together with supporting information for further processing.

III. Data preparation

Data is adapted and transformed for further analysis. Data from different sources and of different types is brought in line or joined. A subset of the data may be identified and selected.

IV. Data analysis

The prepared data is processed according to the desired objectives. Information is explored, generated, highlighted or extracted.

V. Presentation

Results from the previous steps are made available directly or in a visualized form. The presentation is readable by humans or machines. Visualization does not include further interpretation.

VI. Utilization

Data and information produced in the previous steps are used for further interaction and interpretation. Interaction with the results includes the adjustment of the process for further iterations. Results may be interpreted for use by humans or are used as input for other processes.

4.3 Common aspects

While some rules and regulations affect only one of the defined steps of the *Data Value Chain*, others may also be classified into two or more steps. In particular, rules and regulations may affect the transition of one step to the next; hence, they correspond to both of them. An item may be classified into all steps across the *Data Value Chain*.

5 Representation and refinement

5.1 General

The obvious visual representation for a two-dimensional classification is given in two dimensions. While other representations may be possible in certain situations, a basic representation is given in the following.

Additionally, the need for further refinement needs to be considered, since the classification may be applied in specific situations in which most rules and regulations are categorized into only a few of the given header items.

Examples of a basic and a refined representation are given in Annex D “Examples of representations”.

5.2 Basic representation

For a tabular representation of the classification of rules and regulations, the two dimensions *Fields of Standardization* and *Data Value Chain* are considered as rows and columns, respectively, whereas rules and regulations are noted in or are otherwise linked to cells. *Fields of Standardization* shall be used as row headers, steps in the *Data Value Chain* shall be used as column headers. *Additional fields* should only be shown if needed.

The dimensions should be labelled “*Fields of Standardization*” and “*Data Value Chain*”, respectively, and the groups of fields should be mentioned. The order of the items shall follow the order given in Clauses 3 and 4.

Colours and additional symbols for clarification may be displayed.

5.3 Limitation to fewer sections

A representation may be limited to fewer sections of either or both dimensions. In the dimension of *Fields of Standardization*, entire groups or single fields may be omitted. In the dimension of the *Data Value Chain*, steps may be omitted. Empty rows or columns do not need to be shown.

In case of refinement individual sub-items may be omitted.

5.4 Refinement

Further details in classification are achieved through the creation of sub-items. Sub-items to the *Fields of Standardization* and the steps in the *Data Value Chain* are not given in general, but need to be defined according to the scenario in which the taxonomy is used.

Each field or step may receive refinement. Each sub-item shall be given in relation to the corresponding superior item. The sub-itemization of an item is independent of the sub-itemization of other items. The order of sub-items shall be fixed.

Sub-items are used instead of, or in addition to, their superior item in all representations described in this document and the template for the description of rules and regulations.

6 Template for the description of rules and regulations

6.1 General

Working effectively with rules and regulations in Smart Data requires a common description format which helps to store different rules and regulations in a common template. While instruments to describe formal standards exist, the template introduced in this clause widens the field of applicable items to all kinds of rules and regulations covered by this document.

Existing work on the description of resources has been taken as the basis. The Metadata Terms by the Dublin Core Metadata Initiative have served as the primary example. The standard describing template in the study "The Standardisation Environment for Cloud Computing"¹ has also been used. In this clause, the template properties are introduced and then specified in terms of the possible values. Additional definitions on the types used in the template are given in Annex E "Types for the description of rules and regulations". Examples of the description of rules and regulations are given in the Annex F "Examples of the Template for the Description of Rules and Regulations".

6.2 Properties in the template

The properties to describe rules and regulations are divided into four main sections: Taxonomy, General information, Provenance, and Restrictions. The sections and the respective properties are introduced in the following.

1. Taxonomy

1.1. Field of standardization

Field of standardization to which this rule or regulation is related.

1.2. Step of the Data Value Chain

Step of the *Data Value Chain* in which the rule or regulation may be used.

2. General information

2.1. Identifier

Unambiguous reference to the rule or regulation in a given context.

2.2. Standards body

Organization formally issuing the rule or regulation.

2.3. Official specification

Reference to the official document in which this rule or regulation is issued. Usually a link to a document on the Internet.

2.4. Official identifier

Official identifier used by the issuer of the rule or regulation.

2.5. Title

Full name of the rule or regulation as given in the official specification or commonly used.

2.6. Description

Textual description of the rule or regulation.

2.7. Keywords

Set of terms describing the main contents of the rule or regulation.

2.8. Type of rule or regulation

Description of the way in which the rule or regulation is presented.

¹ booz&co., FZI Forschungszentrum Informatik: "The Standardisation Environment for Cloud Computing", 2012. Available at <http://www.digitale-technologien.de/DT/Redaktion/DE/Downloads/Publikation/Trusted-Cloud/trustedcloud-studie-standardisation-environment-for-cloud-computing.html> (Short version in English), <http://www.digitale-technologien.de/DT/Redaktion/DE/Downloads/Publikation/Trusted-Cloud/trustedcloud-studie-normungs-und-standardisierungsumfeld-lang.html> (Long version in German only)