ISO/WD TR 8344 Information and documentation- Issues and considerations for managing records in structured data environments

Versions	Date	Contributions	Contributors
Version 1	2021-12-04	Outline and contents	Xiaomi An Jill Hudson Wenlin Bai
Version 2	2021-12-05	Changes made are in red color or in remarks Comments on Clause 1 (Introduction), Clause 2 (Scope), Clause 3 (partially structured data, semi-structured data, record), Clause 4.1. 4.4 Concept system of structured data 5.4 Managing records in structured data environments and the challenges 5.4.1 Key factors of managing structured data 7.3 Types of emerging technology and the records concerns 8.1 Data governance framework and the records policy 8.2 Digital continuity plan and the records controls 8.3 Data models and the trustworthy interoperability	Adrie Spruit Xiaomi An Wenlin Bai
Version 3	2021-12-08	redefine scope, outline of contents, core concepts mapping with ISO 30300:2020	Xiaomi An Hudson, Jill Wenlin Bai Poole, Roger Shipman, Alan Khramtsovsky, Natasha Spruit, A.M.J. Bustelo Ruesta, Carlota
Version 4	2022-01-02	Updated contents according to 1st meeting minutes, the proposed outline	Xiaomi An Hudson, Jill Wenlin Bai
Version 5	2022-01-08	Comments from Adrie and feedbacks by Xiaomi and Jill	Adrie Spruit Xiaomi An Jill Hudson
Version 6	2022-01-12	Comments from 2nd WG22 meeting （Annex B in Doc. 7 meeting minutes of the 2nd meeting and Doc. 10 list of sources for managing records in structured data environment could be foundations for checking）	Xiaomi An (Convener) Jill Hudson (Project leader) Roger Poole Adrie Spruit Alan Shipman Jan Huta Andrew Khramtsovsky

Version 7	2022-01-28	Updated Clause 4, Clause 5 and Clause 6 according to comments from the 2nd WG22 meeting. ● Adding different unit of thinking about 'record' from different perspectives	Xiaomi An Wenlin Bai
Version 8	2022-02-01	Updated according to comments from sspchramnatasha@gamil.com, January 18 and previous comments from Adrie Spruit Adding records management by design in Future Directions	Adrie Spruit Natasha Khramtsovsky Xiaomi An Wenlin Bai
Version 9	2022-02-07	Adding source of NSW data strategy and data governance toolkit, example of records management by design into data management lifecycle and level of risk control of trust and safe use of data	Adrie Spruit Xiaomi An
Version 10	2022-02-16	Adding a concept diagrams about record from different domain to Clause 4, adding special issues of records management in structured data environment to Clause 5. Deleting Clause 7 and records by design'	Xiaomi An Jill Hudson Wenlin Bai
Version 11	2022-03-01	Adding issues and considerations specific to database records based on Adrie's advice and his updated version	Adrie Spruit Xiaomi An Wenlin Bai
Version 12	2022-03-05	Checking the document against Final Report by ISO/TC46/SC11/AHG3 N20 Final Report (2021-05-15), adding 4.5 Concepts related to structured data, unstructured data, structured records and unstructured records, Clause 5 are updated checking with ISO/TC46/SC11/AHG3 N20 Final Report (2021-05-15) in responding to Adrie Spruit's contributions and comments	Xiaomi An Adrie Spruit Wenlin Bai
Version13	2022-03-19	Checking the document against meeting minutes ISO/TC46/SC11/WG22 N6 and updated Clause 4 , Clause 5 and Clause 6	Xiaomi An (Xiaomi'steam: Rong Shen, Yu Qi, Huanxin Liu, Qing Han) Wenlin Bai
Version 14	2022-04-05	Comments by Adrie Spruit and proposed disposition of comments by Xiaomi ISO/TC46/SC11/WG22 N17 Taking place of ISO/TC46/SC11/WG22 N15	Adrie Spruit Xiaomi An

Version 15	2022-04-05	Comments on Clause 4, Clause 5 and Clause 6, actions to be taken for improvement are in yellow color	Adrie Spruit, Lijuan Yu, Natasha, Roger Poole, Xiaomi An Wenlin Bai, Lilin Du
Version 16	2022-04-07	Modify concept relationships between data, information knowledge; data, information and records and the diagrams, move original Table 2 and Figure 3 and Figure 5 to Annex, taking place of original Annex A, Annex B and Annex C, as they are not citied in this document.	Xiaomi An, Wenlin Bai
Version 17	2022-04-23	Modify contents according to Adrie, Lijuan Yu's comments	Xiaomi An, Wenlin Bai Adrie Spruit, Lijuan Yu
Version 18	2022-05-29	Modify Clause 4, Clause 5 based on meeting held on May 11 and comments from Adrie and Xiaomi's team (ISO/TC46/SC11WG22N29, N31)	Adrie Spruit, Xiaomi An, Wenlin Bai
Version 19	2022-08-16	Modify Clause 5 and Clause 6 based on meeting held on July 5 and contributions from Andrew, Adrie, Johan and Xiaomi's team, move the original Annex B to Clause 4 and add Adrie's contribution (N51) as new Annex taking place of the original Annex B, add Xiaomi's team and Jinghua Zhao's contribution (N46) as new Annex D, add Adrie's contribution(N57) as new Annex E.	Adrie Spruit, Andrew Khramtsovsky, Johan, Xiaomi An（Xiaomi's team: Wenlin Bai, Rong Shen, Huanxin Liu, Qing Han, Yu Qi, Jinghua Zhao）
Version 20	2022-08-22	Page 17, new contributions to 5.4.3	Michelle Tolliday
Version 21	2022-08-19	Clause 5.4.3 is updated Clause 4 is modified	Michelle Tolliday
Version 22	2022-09-05	Contributions to the introduction, Clause 4, Comments on N 66, N 68, N 69	Michelle Tolliday Adrie Spriut
Version 23	2022-11-19	Contributions from Michelle: N68 taking place of introduction, 4.1 N70 taking place of N68 and others. Contributions from Adrie: Replace the Clause 5 (N77), 5.1(N84), according to meeting minutes (N73, meeting held on October 4, 2022) Replace the original Annex D and Annex E with the latest version of the use case template provided by Adrie to become the new Annex	Wenlin Bai putting all the updated versions together

		D. （N79 taking place[October 28, 2022] of N80[September 2, 2022]） Modify clause 6 based on meeting held on November 10, AHG8 (Records management by design, N89) and November 17 and contributions from Adrie and Xiaomi, move the introduction Records management by design approach to Annex E (N75, N81, N85 updated by N90)	
Version24	2022-12-05	Adding ISO15489 to normative reference, updated contents, adding titles to Annexes, checked language, change “must” “shall” to “has to”, change “should “ to “can”. Modify Clause 6 and Annex E, reduce repeated parts. Major changes are in yellow color	Wenlin Bai and Xiaomi An
Version 25	2022-12-13	Modify English wording and languages in Clause 4 and 5	Michelle Tolliday
Version26	2022-12-21	Modify clause 6 and moved description of by design approach to Annex	Wenlin Bai and Xiaomi An
Version27	2023-01-14	Modify Clause 6, add challenges mentioned in Clause 5 and make connections with Clause 5	Wenlin Bai and Xiaomi An
Version 28	2023-01-26	Disposition of comments from RU on N104 (N 105), accepted all the comments Checked and updated the English	Contributions from Andrew Khramtsovsky, Michelle Tolliday Xiaomi An and Wenlin Bai
Version 29	2023-02-01	Modify the language that does not conform to TR according to advice given by technical programme manager of ISO, Patricia Cook, such as“recommendations、 must、 may、 require、 should、 need to”， delete some unnecessary reference and move the bibliography to the end.	Xiaomi An and Wenlin Bai
Version 30	2023-02-02	Adding (informative) to Annex in the contents	Wenlin Bai
Version 31	2023-04-01	Disposion of 23 commens from RU and 16 comments from CN（three experts）	Wenlin Bai and Xiaomi An
Version32	2023-04-11	Update the WD based on disposition of comments which had been discussed	Wenlin Bai Xiaomi An

		(WG22 N138, WG22 N139)	Andrew Khramtsovsky Michelle Tolliday Tony Leviston May Robertson Lilin Du
Version33	2023-05-23	Update the draft based on disposition of comments from RUWG22 N146 taking place of N141	Wenlin Bai Xiaomi An
Version 34	2023-06-01	checked all the comments received and update the CD draft based on agreement achieved at WG22 meeting held on May 25, 2023 changing “by-design” to “by design”, by-default” to “by default” for consistency of usages.	Xiaomi An Wenlin Bai

ISO/AWI TR 8344 Information and documentation- Issues and considerations for managing records in structured data environments

Introduction

With the digital transformation of government, business, and society, records are increasingly being created in structured data environments such as databases, or in records and business systems that are underpinned by databases. These changes mean that the structured data environments now play an important role as trusted data, digital evidence, digital memory, digital identity, and digital assets of business transactions that meet respective legal, regulatory, and business requirements; and enable data-based, data-driven, data-powered decision making and operations. As the basis for this decision making and operations, the data becomes the auditable evidence that is subject to e-discovery requirements. If not properly managed, their business, legal, evidential, and information value could diminish and adversely impact the organization's productivity, compliance, trustworthiness, transparency, accountability and reputation.

Building the capability to manage records properly in structured data environments has become essential to the governance and management of organizations and communities, and there is a growing business need for guidance and recommendations around the design and implementation of adequate policies and procedures across organizational systems or communities in data environments to help ensure that records in structured data environment have the attributes of authenticity, reliability, integrity and usability.

Other high-level messages could include the following:

- Securing top management's support is essential;
- Management of records in structured data environments is a multidisciplinary effort, so cooperation with other information professionals and IT professionals is very important;|
- Engagement of records managers and archivists at early stages of design and development of business and IT systems is highly desirable;
- In the constantly changing regulatory, technological, and business environments, records managers and archivists will often face challenges of managing new types of records (or existing types becoming obsolete) in the absence of established regulation and/or guidance. In many cases, waiting for guidance, standards and best practices to become available is not an option. In such circumstances, records managers and archivists have to demonstrate due diligence, do the best they can and consider relying on tried and tested ISO Management Systems approaches.

This document provides a landscape review of records management in a structured environment and gap analysis. It also identifies problems and provides high-level considerations for addressing them.

The primary audiences for this document will be data policy makers, systems designers, business system owners, data management professionals, database professionals, and the records management professionals working together to ensure the application of appropriate records management, records process, and records controls for a structured data environment.

1 Scope

This document identifies issues and considerations for managing records in a structured data environment.

2 Normative reference

There are no normative reference in this document.

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 30300 apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
database

collection of data organized according to a conceptual structure describing the characteristics of these data and the relationships among their corresponding entities, supporting one or more application areas

Note 1 to entry: database: term and definition standardized by ISO/IEC [ISO/IEC 2382-1:1993; ISO/IEC 2382-17:1999].

[SOURCE: ISO/IEC 2382:2015, 2121413, modified, Note 2 is deleted]

3.2

database management system

system, based on hardware and software, for defining, creating, manipulating, controlling, managing, and using databases';

Note: The software for using a database may be part of the database management system or may be stand-alone;

Note 1 to entry: The software for using a database may be part of the database management system or may be stand-alone.

Note 2 to entry: database management system; DBMS: term, abbreviation and definition standardized by ISO/IEC [ISO/IEC 2382-17:1999].

[SOURCE: ISO/IEC 2382:2015, 2121417, modified, Note 3 is deleted]

3.3

record

information created or received and maintained as evidence and as an asset by an organization，in pursuit of legal obligations or in the course of conducting business

Note 1 to entry: Records are normally used in plural.

Note 2 to entry: In a management system standard（MSS）implementation，the records created to conduct and direct the management system and to document its implementation are called documented information.

[SOURCE: ISO 30300:2020, 3.2.10]

3.4
records control

instrument for helping in the conduct of records processes

Note 1 to entry: Example of records control include *metadata schemas* for records，*business classification schemes*，*access and permission rules*，and *disposition authorities*.

[SOURCE: ISO 30300:2020, 3.5.6]

3.5
relational database

database in which the data are organized according to a relational model

Note 1 to entry: relational database: term and definition standardized by ISO/IEC [ISO/IEC 2382-17:1999].

[SOURCE: ISO/IEC 2382:2015, 17.04.05, Note 2 to entry is deleted]

3.6

relational database management system

database management system designed for relational databases

Note 1 to entry: In order to use relational data base management systems (RDBMS), it is necessary to represent relational model of data that organizes data (4.5) with specific characteristics (tables or relations, unique key, etc.) (see Table C.3.1).

[SOURCE: ISO/IEC 25024:2015, 4.34]

3.7
semi-structured data

aggregate datatype whose components' datatypes and their labels are not predetermined

Note 1 to entry: Semi-structured data are forms of structured data that do not follow structure of data models related to relational databases or other forms of databases.

Note 2 to entry: Examples of semi-structured data include the data that contain HTML tags or other markers to separate semantic elements and to represent hierarchies of records and fields within the data.

[SOURCEL ISO/IEC 38505-3:2021, 3.14]

3.8
structured data

data which are organized based on a pre-defined (applicable) set of rules.

Note 1 to entry: The predefined set of rules governing the basis on which the data is structured needs to be clearly stated and made known.

Note 2 to entry: A pre-defined data model is often used to govern the structuring of data.

Note 3 to entry: Example of structured data are data contained in relational databases.

[SOURCE: ISO/IEC 38505-3:2021, 3.15]

3.9
unstructured data

data which are characterized by not having any structure apart from that record or file level

Note 1 to entry: On the whole unstructured data is not composed of data elements.

EXAMPLE: An example of unstructured data is free text.

[SOURCE: ISO/IEC 20546:2019, 3.1.37]

4 Basic concepts

4.1 Understanding data environment

From a records management perspective, ISO 30300:2020 defines data as a “set of characters or symbols to which meaning is or could be assigned”, information as “data in context with a particular meaning” and records as “information created or received and maintained as evidence and as an asset by an organization, in pursuit of legal obligations or in the course of conducting business”.

From an ICT perspective, ISO/IEC 2382:2015, 2121272 and ISO 8000-8:2015, 3.1 define data as “reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing”. In an ICT environment, data is a digital representation of information. It is considered to be the result of how information has been recorded and consists of bits, bytes, characters and pixels. ISO/IEC 2382:2015, 21212 2 and ISO 8000-9:2015, 3.3 define information

as “knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning”. From ICT and artificial intelligence domain perspectives, ISO/IEC 2382:2015, 2123771 defines knowledge as a “collection of facts, events, beliefs, and rules, organized for systematic use”.

Data, information, and knowledge are closely related concepts. Figure 1 shows the concept relationships of data, information, and knowledge from unknown to known. Knowledge is what humans know, understand, and can apply, based on what one has perceived, discovered, and learned from processed, organized, contextualized, and meaningful data.

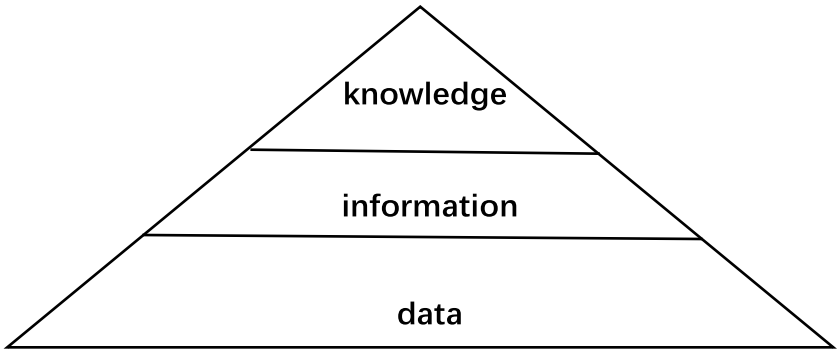


Figure 1 - Relationships of data, information, record and knowledge in data environment

4.2 Concept of structured data

4.2.1 Core characteristics of structured data

Based on the identification of concepts and characteristics in the structured data definitions from the existing sources in Table 1 (see Annex A), Figure 2 shows the formation of the concept of structured data, mapping with ISO 704:2009. From the analysis of characteristics of structured data in its definition, Table 2 and Figure 3 show the characteristics of structured data, which include: (1) use a pre-defined set of rules; (2) specified data model; (3) structured in an organized manner or identifiable way; (4) stored in distinct or fixed fields.

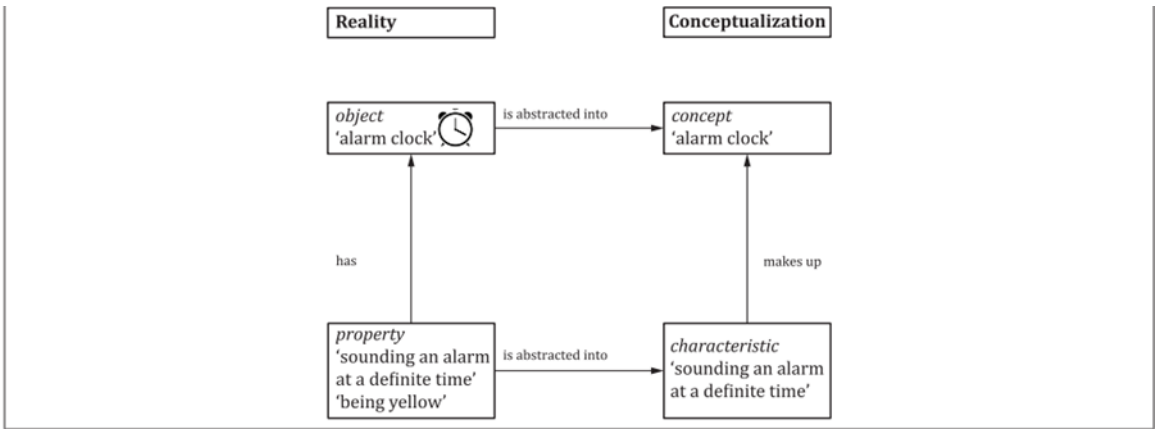


Figure 2—Example of formation of a concept (Source: ISO 704:2022,5.4.1)

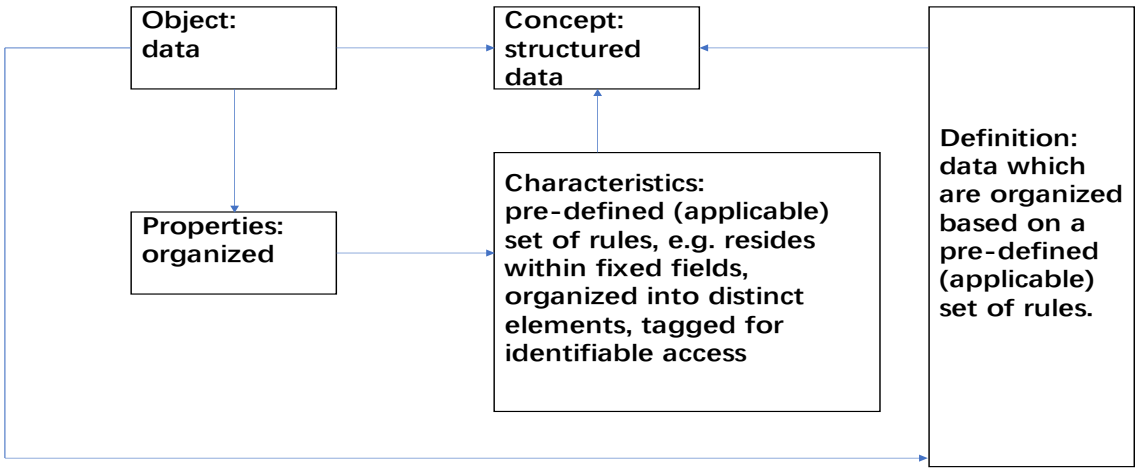


Figure 3—Formation of structured data concept

4.2.2 Relational database management system as a typical example of a structured data environment

4.2.2.1 Database management systems

A database management system (DBMS) is a software for establishing, using and maintaining databases. Most DBMS systems support [implement] data definition language and data manipulation language (DML) for users to define the database's structure and permission constraints, and perform operations such as querying, adding, updating and deleting data.

4.2.2.2 Relational database management systems

A relational database is the collection of all relationships corresponding to a relational model. DBMS require structures (such as tables) to be defined before storing data in a table, where each column (field) stores a specific type of information (data type). The collection of relationships constitute a relationship model.

4.2.2.3 XML Document type definition (DTD)

The XML standard describes how to create a DTD and how to associate it with the XML document written according to its rules. It also defines how the XML processor processes the DTD. A DTD can be used to detect whether the structure of the XML document is correct.

In a practical application, DTD's main functions include that it can:

- verify the validity of XML document data
- provide a unified format and the same structure for some kinds of XML documents
- ensure the exchange and sharing of XML document data within a certain range

4.3 Mapping between the concepts related to structured data and ISO 30300:2020

The Figure 4 illustrates the mapping between the specific characteristics of structured data and the characteristics of records defined in ISO 30300.

First set of characteristics (e.g. authenticity, usability, reliability and integrity, as defined in ISO 30300) is essential for the categorization of certain data as records.

The second set of characteristics is related to the management of data as records in compliance with the principles and requirements defined in ISO 15489-1 and ISO 30301. These requirements are applicable to all kinds of data (structured, semi-structured, or non-structured) that are managed as records.

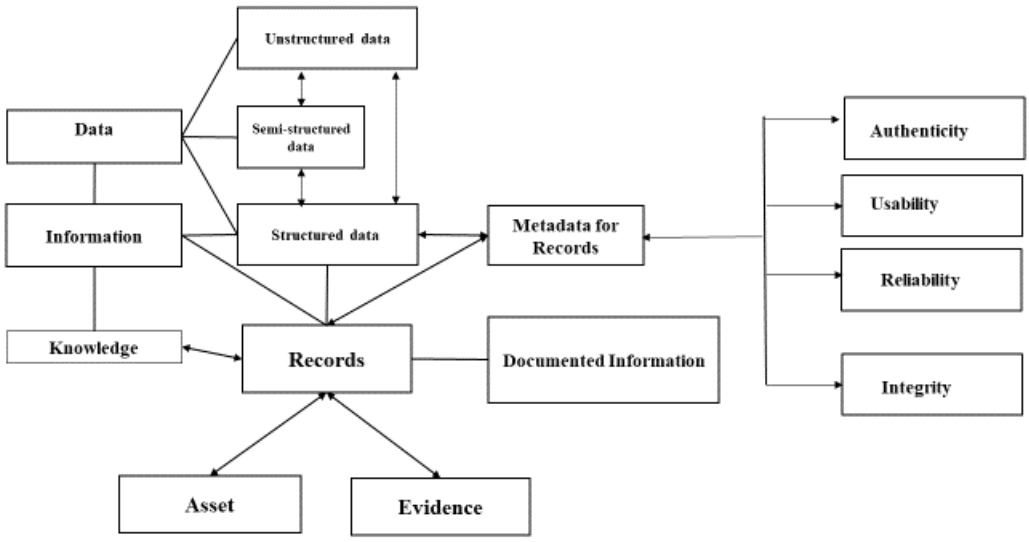


Figure 4- Mapping between the concepts related to structured data and ISO 30300 terms

4.4 Different interpretations of the concept of 'record'

4.4.1 Relationships between 'data' and 'record'

For database managers and administrators, the word "record" does not have the same meaning as in ISO 15489-1:2016 and ISO 30300:2020. From their perspective, a "record" is a 'database record' which is a tuple or row of data in a database table with data (values, information) in data fields. Tables are defined object types and table columns are predefined fields of attribute types. A database can contain multiple tables (object types) and every table can contain multiple rows or database records (instances of object types). So, for database managers and database administrators a record is primarily a data structure representing an information unit in a database.

4.4.2 Relationships among 'record', 'information' and 'data'

A record is defined as: "information created, received and maintained as evidence and as an asset by an organization or person, in pursuit of legal obligations or in the transaction of business" (ISO 30300:2020, 3.2.10).

Information is defined as "data in context with particular meaning" (ISO 30300:2020, 3.2.7).

Data is defined as "set of characteristics or symbols to which meaning is or could be assigned" (ISO 30300:2020, 3.2.4).

The relationships of record, information and data reflects a records continuum thinking in ISO 30300, in terms of trusted data, trusted record and trusted information regarding meanings to be given and given as evidence and asset in pursuit of legal obligations or in the transaction of business.

4.4.3 Issues of 'record' as both 'information asset' and 'evidence'

Records managers define a record as 'information created or received and maintained as evidence and as an asset by an organization, in pursuit of legal obligations or in the course of conducting business' (ISO 30300:2020, 3.2.10). That means that information is considered a record provided it has sufficient value (hence it is an asset). The value can be of different character, and the ISO 30300's definition specifically highlights the value of information for pursuing "legal obligations" (meaning compliance with legal and regulatory requirements, contractual obligations etc.) and the value for "conducting business" (that part covers legal, business and partially historical and societal value).

Note: It would be wrong to presume that the list of values provided in the ISO 30300 definition is exhaustive, since this definition was focused on common business-related records rather than on records in general. Importantly, in many cases an organization cannot determine the "recordness" of information at will, and the laws and regulations of the applicable jurisdictions, contractual obligations, organization's voluntary commitments, established business practices and standards (including codes of practice and ethics), societal expectations, the valid interests of various stakeholders, business context etc. also can play a role.

Looking further, one can read in the introduction of 15489-1:2016 that 'Records are both evidence of business activity and information assets. They can be distinguished from other information assets by their role as evidence in the transaction of business and by their reliance on metadata'.

Importantly, while a record is always associated with metadata, having metadata is not what defines the information as being a record. Information assets that are records can be distinguished from information assets that are not records by their role as evidence in the context of the transaction of business or legal obligations by an organization or person.

4.4.4 Management and technology perspectives on 'record' concepts

There are many definitions of 'record'. From a management perspective, records are seen both as instruments, as evidence and as reference material. Generally speaking, the definition of a record is based on the transaction of business and the value that if embodied.

Whereas from a technology perspective, they are identified as collection of related data (see Annex C for different viewpoints on records from different perspectives). Generally speaking, the definition focuses on the record's format and attributes, and emphasizes the reading, storage, and processing of records by computer software and hardware devices.

5 Challenges and considerations about records management in structured data environments

5.1 Identifying records in a relational database

5.1.1 The challenge in brief

The first challenge in managing records in relational databases is their identification. To this end, it has to be determined which records in the database was received and created as a result of business processes and is then required to be kept and maintained as evidence of those processes.

In addition, it has to be determined which of these records, once created and stored in the database, are to be reused in business processes that create other records, such as decisions, minutes, and reports containing copies of the original database data.

It is important to understand that meaningful records in a database often consist of quite complex combinations of data elements (also called items in relational database), where each data element is the content of a field in a row in a table of a database.

It is also important to be aware of the potential inability to dispose of records within the structured data environment because of its specific nature e.g. records/data in the blockchain or other distributed ledgers can’t be disposed of, so organizations need to think about the use of these environments beforehand.

5.1.2 Understanding data elements of a relational database and related records

Every database has a predefined name and the tables in a database also have predefined names. Every field in a row of a table is of a predefined field type and has a predefined field type name. Each column in that table is a field type. The content of every field in a row of a table is a data element.

In the example below, Bo Wang from Rotterdam owns a house in Gouda, Netherlands, that he wants to sell using John Johnson’s housing broker services.

A data element containing the string 'Gouda' may be the content of a data field of the field type named 'name_town' in a row about the entity Gouda in a table named 'towns'. This table may be part of a relational database named 'houses' owned and managed by John Johnson (or even an association of housing brokers). Such data about towns in a table of a relational database can be the result of a general process of gathering and storing data about towns.

In the same way, the data element 'Bo Wang' can be the content of a data field of the data type named 'name_client' in a row about Bo Wang in a table named 'clients' in another relational database. This kind of data can be the result of John Johnson’s specific client intake process. Bo Wang's residential address details are in a table named 'addresses' with a reference from the table 'clients' to this table and another reference from the table 'addresses' to the table 'towns' which includes both 'Gouda' and 'Rotterdam'.

When Bo Wang's house is sold, the sale date and sale price can be added to another table called 'sales'. Thereafter, data already stored in the said tables and databases can be reused in a transaction document prepared by John Johnson and in another transaction document prepared by a notary public.

This example shows:

- how data are entered as data elements in several (or in more complex situations, multiple) data fields of several rows of several tables of one or more databases;
- how combinations of these data elements have meaning while a single data element like 'Gouda' or 'Wang Bo' has little or no meaning. This is the case even without added metadata, because the data fields have predefined names as a technical requirement for any operational relational database, and those names are usually interpretable;
- how combinations of data elements from a relational database can be reused in other processes and the documents that they produce;
- every data element in a relational database can be part of several meaningful combinations of data elements inside and outside this database, therefore, even the use of a relational database's data elements outside of the database may have implications for
    - the meaning of combinations of these data elements within the database
    - how these data elements are considered components of records
    - how records management rules such as for disposition and deletion are applied to these data elements.

Because of the mechanisms described here, it may be rather complex to identify which data elements and combinations of data elements in a relational database are considered as records.

5.1.3 How to deal with the challenge

Dealing with this challenge is considered to be understood and analyzed during the design of the relational database and the data model that predefines the structure and content of such a database. In addition, it is usually helpful to distinguish two types of tables: master tables and transaction tables.

Every master table and its columns (and constraints) contain data about one of the main entity types of a relational database such as client, town (or address) and house.

Every transaction table contains data about one type of transaction or events. Transactions and events are the results of processes or activities performed by entities as recorded in the master tables. Most transaction rows (or entries) have foreign keys that refer to master records.

5.2 Appraisal of records in structured data environment and its retention period

Whenever a structured data environment is used to support an organization's business or for compliance with legal, regulatory, and other stated requirements, such an environment creates and stores business records. These records can be subject to various legal, regulatory, business, and societal retention and disposition requirements of one or more jurisdictions.

In most cases, the retention periods for the identified records in the structured data environments are determined and the existing retention conflicts are resolved in compliance with common appraisal practices described in ISO 15489-1 and ISO/TR 21946. However, due to the high level of reuse of data elements and other records' components, the same data elements could be subject to numerous retention requirements. The situation is further complicated when the legal and regulatory requirements of several jurisdictions are applicable.

Technical feasibility of the disposition of the records and their components also could be an issue.

5.3 Disposition of records in the structured data environments

According to ISO/TS 7538, proper disposition of records generally offers a number of benefits including the following [Source: current draft of ISO/AWI TS 7538]:

a) From a business point of view:

- cost reduction, workload reduction and release of resources,
- enhancing operating efficiency (e.g. search efficiency),
- reduction of security risks,
- management of corporate data, information, records, and knowledge (including rearrangement, archiving),

b) From a societal point of view:

- preservation of memory and knowledge of permanent value,
- reducing the environmental impact of storage and management,

c) From a legal point of view:

- compliance (archival law, privacy laws, etc.),
- reducing the risks of disclosure of sensitive and confidential information resulting in investigations, court cases and fines.

Specific disposition problems in structured data environments include the following:

- **Identification and appraisal of the records** (see clauses 5) to be disposed of;
- **Technical feasibility** of destruction or transfer. In many cases, the records in the structured data environments can only be disposed of without the risk to the environment's integrity, when disposition of both data and corresponding metadata has been integrated by design into the environment's architecture and functionality. In some cases, the records can only be destroyed in the process of decommissioning the environment, especially when its architecture and design is intended to endure the immutability of data and records.
- **Need for disposition process automation.** Due to the high volumes and high creation rates, manual disposition is often problematic even in traditional records systems. Policy-driven, implemented by design automation is even more crucial for structured data environments where retention conflicts are common and, in the worst case, retention control has to be at the data element level. Automation also helps in conducting disposition regularly in the course of normal business rather than as *ad hoc* projects; which is important for the trustworthiness of records management operations and to avoid of legal risks in some jurisdictions.
- **Comprehensiveness.** From both legal and business standpoints, the records can only be considered as destroyed or transferred (without retaining copies) if all the controlled and as many as possible uncontrolled exemplars and copies held by the organization and its stakeholders are destroyed (backup copies, paper printouts and reports where appropriate, etc.). Additional challenges are also present in the case of third-party controlled (e.g. cloud) or distributed (e.g. blockchain and distributed ledger-based solutions) structured data environments, especially where they are not under the sole control of the organization.

- **Identification of proper archive(s) for transfer and isolation of organization's records**. Currently, it is not uncommon that the same structured data environment is used by several independent business entities or by organizations and bodies belonging to different (local, regional, or federal) governments and their branches, which could be obligated to transfer their records of permanent value to different archival institutions. It could also be difficult or impossible to clearly isolate an organization's records from those of other tenants of the shared environment. Archival and records management theory and practice underpinning records transfer in such circumstances is currently at an early stage of development.
- **Assurance of the timely, complete and irreversible destruction of data and records, and proper documentation of disposition**. In the context of structured data environments that are complex, distributed, shared and/or controlled by the third parties, it could be difficult to get trustworthy, verifiable assurances or attestations of the timely, complete, and irreversible destruction of the data and records. In some cases, this risk could be mitigated by internal governance and management measures or by inclusion of suitable clauses into contracts or service level agreements with relevant parties.
- **Making data inaccessible (e.g. by means of "cryptographic shredding", destruction of indexes and pointers) instead of destruction**. Making data inaccessible by whatever means (e.g. by destroying corresponding cryptographic keys) is not usually considered proper destruction as the information persists and could potentially be discovered with sufficient effort and expenditure. However, in certain situations, this could be the most feasible solution, though legal and long-term information security risks are thoroughly analyzed.

  Preserving long-term usability, accessibility and confidentially of the records after transfer. There is a clear lack of standards and tools supporting long-term usability, accessibility, and confidentially of data and records extracted from structured data environments. The structured data environments often rely on proprietary software, hardware, and data formats that are expensive and quickly become obsolete. Licensing policies could prevent the transfer of the corresponding software and hardware to archival institutions, as well as the re-engineering efforts in archives in order to ensure access to the obsolete formats.

5.4 Records management of current and variable data in relational databases

As stated in ISO 16175-1:2020, many business applications generate and store data that can be subject to constant updating (dynamic), are able to be transformed (manipulable), and only contain current data (non-redundant). While business requirements for dynamic, manipulable, and non-redundant data can be entirely legitimate, if records are to serve as reliable evidence of business functions and processes, they need to be fixed and inviolable. So, the capture of current data and the fixity of their content can be challenging.

If only current values have to be available, then securing these values against unauthorized changes through authentication and authorization measures may be sufficient. However, if historical current values, such as historical addresses of citizens, have to remain available as evidence, it may be challenging to create fixed-content records by capturing values that are initially current and not fixed and not linked to one date or point in time.

5.4.1 Dealing with the challenges

To be considered as authentic and reliable evidence, content has to be fixed to a point in time and unalterable. That is not the case for dynamic, current data that is regularly updated. Therefore, strategies have to be implemented to record the changes to these data by creating and maintaining records with fixed values.

ISO 16175-1:2020 mentions two strategies for ensuring the fixity of records in database applications that contain data that is frequently updated, manipulable, and non-redundant or current:

1. preventing the overwriting or deletion of specific data, enabling the updating of data and recording the previous values in a history status field. Records are formed by the combination of specified fields and the associated event history data;
2. bringing together the selected data elements (from within the same table or selected data from rows in relational tables) and creating a distinct digital object that is fixed and unalterable.

In practice, 1 is an obvious choice for storing captured records in the relational database containing current data, while 2 is a more obvious choice for storing captured records outside such a database.

Variants of strategy 1 are:

a. logging the changes of data resulting in entries that show the changes that are made to specified fields of the tables of a relational database by using the logging facilities of the database;
b. storing initial current data as fixed historical data (that can serve as evidence) for this purpose in designed pre-defined fields. Often, current data contain values that are valid for a certain period such as address data of citizens, price data of products, and roles or functions of employees. In that case, context data such as start and end dates that define a period in which specific values of current data were valid can be stored as fixed data in pre-defined fields together with the values that were current and valid but unaltered (till the end date) in such a period.

Implementation of these strategies are considered in the design of solutions, preferably together with designing the initial storage solutions for the mentioned current and variable data. The design of records management solutions for this data may include the design of data and metadata fields and application controls. Once implemented, these solutions enable the described methods for capturing current and variable data as records that can be used as evidence.

5.4.2 More types of current and variable data and associated challenges

Continuously changing measurement values in a technical or scientific environment provide a challenge with specific characteristics. Solutions may include:

- sampling the continuously changing values and capturing and storing the resulting sample values;
- determining averages of continuously changing values and capturing and storing the resulting average values;
- storing resulting values in a type of non-relational database that is more suitable for storing ranges of (measurement) values, for example a so-called time-series database.

Recording current output values of smart device sensors in communication networks is a similar challenge. This also includes smart devices connected by the Internet, resulting in a new digital

environment referred to as the Internet of Things (IoT).

5.4.3 Considerations

To ensure that data in a database can be captured as records, the requirements for their capture have to be incorporated into the database's design. This is because it is unlikely that they can be introduced after the database has been built without negatively impacting the established data relations. As such, it is critical for database designers and records managers collaborate closely during the design stage.

Records management by design is discussed in detail in Clause 6.

5.5 Potential conflict between disposition rules for records and the referential integrity rule for relational databases

If structured data in a relational database is to be considered records and/or if these data are stored as records in such a database, they will be subject to both disposition rules for records and the referential integrity rule for relational databases. While disposition rules may dictate the deletion of records, referential integrity rules for relational databases may dictate the retention of the same records—potentially creating a conflict.

This leads to the challenge of how to combine applying disposition rules for records in a relational database from the perspective of records management while maintaining the referential integrity of such a database from the perspective of proper database management.

5.5.1 The referential integrity rule for relational databases

A relational database satisfies the referential integrity rule for it to be able to operate properly. In a relational database, foreign keys (also referred to as secondary keys) are values in predefined foreign key fields of rows in the tables of the database. These foreign key values refer to values of primary keys of rows in other tables. The primary keys are the identification (ID) codes of rows. The references from secondary keys to primary keys are essential to a relational database's operation. Therefore, each secondary key has to reference a primary key value that exists and a row that exists and contains that primary key value. The deletion of a row referenced by a secondary key damages a relational database, impedes its proper operation, and prevents access to content that are retained. Such a loss of relational database integrity has to be prevented. In practice, it is common that the application controls of a relational database management system (the application for the management of such a database) prevent the execution of commands that would damage the database in this way.

5.5.2 Disposition rules

Disposition of records includes both the transfer of records and the responsibility for them to another organization, and the destruction of records at the end of their retention period. If records are stored in a relational database, the result of either process may be the removal of these records from the database. Migration of records from a relational database (system) to another digital system (application) in the same organization will also result in the removal of records from the database in question. So, there are several situations in which the management of records in a relational database and the rules for it can lead to the removal of records from the database. If these records are rows, or include rows to which secondary keys refer, this will violate the referential integrity rule for the database.

5.5.3 Recommendations for dealing with the challenges

Determining what is and is not are considered from the database perspective, including the predefined structure and the operation of a relational database. The structure of a relational database is laid down in the data model of the database. The way a relational database works is determined by the principles developed in 1970 by Edward F. Codd and further developed in 2006 by Hugh Darwen and Christopher Date.

If the challenge described involves an existing database, data model is to be known and analyzed to determine what content can be removed without damaging the database. If content has to be removed from the records management perspective but cannot be from the database perspective, alternative solutions may be sought such as masking content or shielding content from reading instead of removal. In such situations, one has to accept that such solutions have drawbacks and that sometimes one can only choose an approach of minimizing the risks by accepting the best possible (but not the perfect) solution.

A better way is to consider solutions before designing and implementing a relational database that will also store records. Even then, finding a perfect solution is not guaranteed, but this way offers more opportunities for design decisions at the data model and application logic level that prevent problems or make them easier to deal with.

The bottom line is that dealing with this challenge may be complex in both the design phase and the operational phase. Papers about the potential conflict described here, written by researcher Ahmed Ayaz Ataullah [Sources 2, 3, 4], confirm this conclusion.

6 Considerations for records management in structured data environments

The working group collected and analyzed four use cases （see Annex E,Annex F，Annex G，Annex H）about records management in the structured data environment, which were from the Commercial Aircraft Corporation of China, PetroChina, borui Database Company, and State Grid Tianjin Electric Power Company, of which one was from the database manufacturer and three were from the enterprise archives department. Through use case data analysis, the working group found that all the challenges mentioned in clause 5 were validated, and in addition, some macro management challenges were also identified, such as the lack of file management awareness of the leadership, the risk of technology, the shortage of human resources, the lack of funds, and the constantly updated industry rules, which had brought challenges to records management in the structured data environment.

6.1 Pre-defined considerations for records management in structured data environments

Pre-defined considerations can be divided into those that are strategic and those that are operational.

6.1.1 Pre-defined strategic considerations

According to ISO 30301 and ISO 30302, the management system for records (including elements such as records policies and records objectives) should be established and compatible with the organization's strategic goals, and be imbedded in the organizational governance framework.

With organizations adapting to an increasingly digital society through digital transformation, this means that records management and IT professionals need to integrate their records and data management strategies as part of the organization's top-level digital transformation plan, to enable the design and development of data and information governance, infrastructure, architecture, and processes that incorporate records management requirements into the pre-defined database data model.

6.1.2 Pre-defined operational considerations

According to records management standards such as ISO15489 and ISO 16175, records process and controls should be controllable and embedded in the business system. Records management fundamentals such as methodologies, processes, controls, systems, and guidance for applying records system functionality should be considered in the initial planning stage of business system design as well as throughout an entire life cycle of the system. In addition, the conflict between records management rules and structured data rules (such as rules of relational databases) is to be considered, and a compromise solution at the data model design stage and application logic level design stage in to be chosen.

6.2 Records management in structured data environments by design and by default

6.2.1 Records management by design as a proactive top-down approach

The traditional records management approach starts with an existing environment in which information has already been created and received, and where there are already information systems comprising processes, people (typically employees), funding, and possibly a records management policy. The organization determines or has already determined through an appraisal process what information is to be captured and retained as records and for how long. The system for the retention of those records is often a dedicated system like an electronic documents and records management system (EDRMS) or a records management application (RMA).

The concept of records management by design is an approach that results in proactive top-down plans for records management. Records management principles are included at the initial design stage and throughout the lifecycle of products, process or service that involve managing records. Any initiative to change something in the processes also takes into account the requirements for records processes,

controls and systems. Changes include starting new processes, modifying existing processes, and changing information and information systems for these processes; such as creating or receiving new information types, developing, acquiring, and implementing information systems, and modifying information systems.

Records management by design establishes a pre-defined set of rules for records processes, controls and systems in a structured data environment because of the challenges of managing the records in such an environment. These challenges (described in Clause 5) are to be addressed in the database data models, the (predefined) data (or information) object types in those models, and their intended or expected use, which are determined in the database design phase.

6.2.2 By design for records management processes and controls in structured data environments

High-level process and control steps for developing by design records management solutions include but are not limited to:

- identifying which records are to be captured, retained and disposed of;
- identifying the requirements for the capture, classification, indexing, description, access control disposition, and preservation of these records;
- choosing where (in which information systems) to retain these records;
- designing the solution(s) for capturing, retaining and disposing of the selected records—including fundamental system connections;
- building or purchasing and implementing the solution(s) for capturing, retaining and disposing of these records.

6.2.3 By default as an enabler for configuring records management into structured data environments

Predefined records management processes and controls in structured data environments (such as those mentioned above) are default considerations of a by design system. This means that they are embedded in the system's design and operate automatically in the course of business transactions. Because the rules of consent are built into the system, seeking the consent of data creators, data managers and data users is no longer necessary.

As mentioned in clause 5, the retention periods of the records to be retained in the structured data environments, and the corresponding rules for records disposition based on ISO 15489-1 and ISO/TR 21946 are embedded at the beginning of the data model design so as to enable the machine to automatically understand the rules and undertake automatic processing.

A range of additional process are also needed to ensure that the records management rules:

- cover the whole life cycle of the structured data;
- are SMART (standard machine applicable, readable and transferable);
- are machine executable, interpretable, auditable, and adaptable to the applications of data-enabled technologies such as records management in AI (artificial intelligence) environment.

6.2.4 Special consideration of records management by design for structured data as records

Records management of structured data in a relational databases begins by analyzing the database's data model to identify the records to be captured and retained for reuse as evidence and/or which already-stored data can be considered and treated as records based on its use and meaning in processes and other information systems. If choosing to store records in the database itself, the next step may involve modifying the database's data model, for example, by adding rows and fields for historical and context data. After this, the needs and options for disposing of the records in the database are

determined, taking into account possible conflicts between applicable records management disposition rules and the referential integrity rule for relational databases (described in sub-clause 5.3). Suboptimal solutions are to be considered. Once this is completed, the functions necessary for managing the records in the database will be known and this will determine the application logic of the database system for their execution. Figure I-2 in Annex I shows the considerations of records management in structured data environment by design that call for collaboration among business managers, records managers and information system developers.

Annex A (informative) Identification of concepts and characteristics of structured data

No.	definitions of structured data	Object	Characteristics	Scenario
1	structured data - data which are organized based on a pre-defined (applicable) set of rules Note 1 to entry: The predefined set of rules governing the basis on which the data is structured needs to be clearly stated and made known. Note 2 to entry: A pre-defined data model is often used to govern the structuring of data. [SOURCE: ISO/IEC 20546:2019, 3.1.35; Note 3 to entry: Example of structured data are data contained in relational databases. [SOURCE: ISO/IEC 38505-3:3021, 3.15]	data	pre-defined (applicable) set of rules；	Use pre-defined data model; relational database
2	"Structured data is commonly defined as data that resides within fixed fields such as relational databases, as opposed to unstructured data which is not relational and does not fit into pre-defined data models." Source: John Isaza and Tom Reding, Defensible Disposition of Structured Data, Information Governance Solutions, 2015, see https://www.infogovsolutions.com/disposition-structured-data-1/	data	that resides within fixed fields	Fit into pre-defined Data model; relational database
3	structured data 1. Data organized into distinct elements established by a data model or standard. Note: Structured data usually refers to data that is atomic and stored in distinct fields according to the character or use of the type of data. Examples include relational databases and spreadsheets. By comparison, XML is a container that can provide a structure to a variety of data objects using tags to support access. [SOURCE: InterPARES Trust Glossary, see https://interparestrust.org/terminology/term/structured%20data]	data	Organized into distinct elements; stored in distinct fields	data model or standard; relational databases, spreadsheets
4	Structured data：Data that either has a structure according to a specified information model or is otherwise organized in a defined manner. [SOURCE: ITU-T Y.4500.1 (01/2018)]	data	has a structure; organized in a defined manner	a specified information model
5	Franks 2013 (†560 p.36): Structured data is organized in a way that makes it identifiable.	data	Organized in a way identifiable.	
6	Gingrich & Morris 2006 (†358 p. 31): Structured data is defined as data stored in fields and rows in tables of a relational database.... Examples would include databases containing accounting and financial data, customer data, and personnel data. (†348)	data	stored in fields and rows	in tables of a relational database; accounting and

				financial data, customer data, personal data

Annex B (informative) What is a database and what is a relational database?

Definitions

A database is a container for storing data. Software is used in the digital variant of a database. In practice, we see combinations of a digital database and software, so-called database management systems. A relational database is a type or a category of databases.

The standard ISO/IEC 2382 provides definitions for the concepts database and relational database and their system variants.

Database:

collection of data organized according to a conceptual structure describing the characteristics of these data and the relationships among their corresponding entities, supporting one or more application areas.

Database management system (abbreviation: DBMS):

system, based on hardware and software, for defining, creating, manipulating, controlling, managing, and using databases.

Note: The software for using a database may be part of the database management system or may be stand-alone.

Relational database:

database in which the data are organized according to a relational model.

Relational database management system (abbreviation: RDBMS):

database management system designed for relational databases.

In practice there are also other terms in use for the concept of a database management system such as 'database information system' and 'database application'.

Relational databases

A relational database organizes data into tables where each table contains rows (or tuples), columns and cells. Rows include cells that contain data of entities such as data about persons and products. Columns represent the predefined attributes for these entities such as IDs, first names, last names and product names. Every cell of a row contains the value of one attribute of the entity represented by that row.

The data in rows can refer (or relate) to data in other rows by using primary and foreign keys[1]. Primary keys are the IDs of rows. Secondary keys are pointers to primary keys of other rows.

The main idea behind the concept of a relational database and its design is to avoid data redundancy. Data redundancy is the term for storing the same data more than once in a considered environment. A relational database is designed to store the same data only once in such a database. This has both advantages and drawbacks.

The relationships between data in a relational database are organized by a set of predefined rules. One of those rules is the referential integrity rule. This rule states that for every secondary key there exists a row with a primary key of the same value so that every secondary key refers to something that exists. If this rule is not met, the database loses its referential integrity and can no longer function properly.

---

1 'Secondary key' is an alternative term for 'foreign key'.

This reduces the options for deleting data in a relational database if needed to meet records management requirements for the disposition of records. More about that in subclause relating to challenges of referential integrity.

SQL (Standard Query Language) is a query language for accessing and manipulating data stored in relational databases. Formally SQL is not a part of a relational database, but in practice it is important for working with these databases. An alternative name for a relational database is 'SQL database', especially in use when classifying databases.

Other types of databases

A common classification of databases is one that is based on the structure of a database. On a high level it divides databases into two categories:

- SQL databases for relational databases;
- NoSQL databases for non-relational databases, where NoSQL stands for Not only SQL.
- The NoSQL category is divided into several subtypes or subcategories.
  Examples of NoSQL database subcategories are:
- hierarchical databases containing parent-child relationships, like a family tree;
- network databases: hierarchical databases with non-hierarchical links;
- key-value databases containing two-column rows within every row a value and a key for that value;
- document databases or document-oriented databases, for example XML;
- graph databases with entities as nodes and connections with other nodes if these connections are meaningful;
- column databases with flexible columns;
- time-series databases for storing measured values that change over time.

There will not always be sharp boundaries between them. There are overlaps by databases with characteristics from more than one subcategory.

More insight in types of databases (optional)

The best way to get some more insight into the mentioned types of databases without being or becoming an expert is by looking at examples of their structure visualized by data model schema like in the images below.

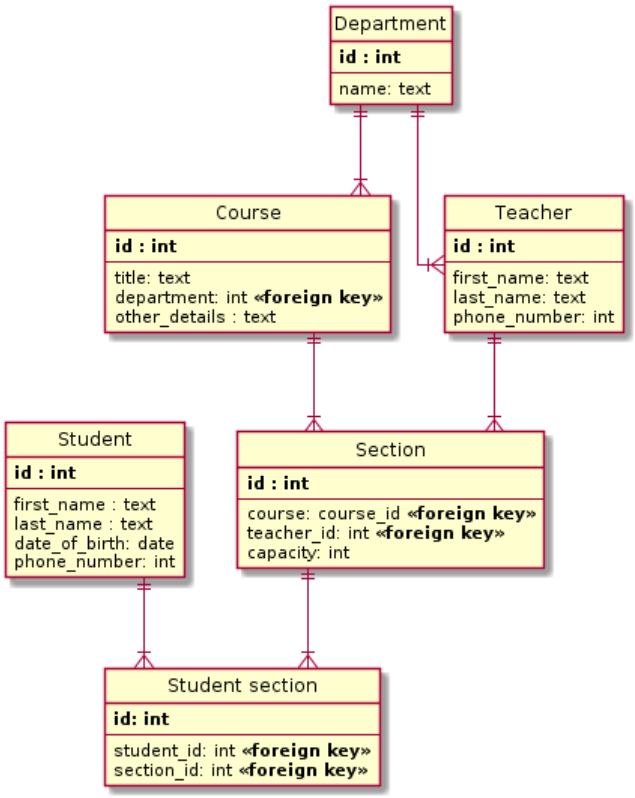


Figure B-1 Data model of a relational database

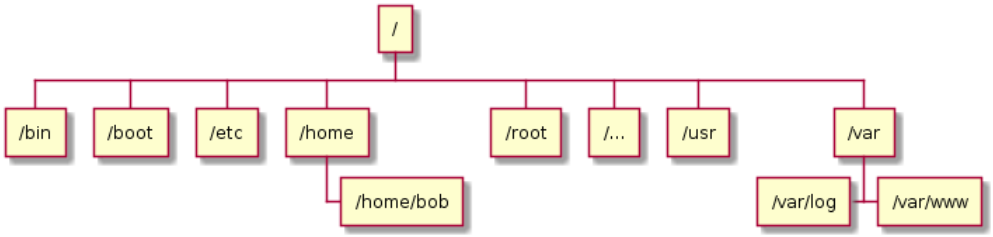


Figure B-2 Data model of a hierarchical database

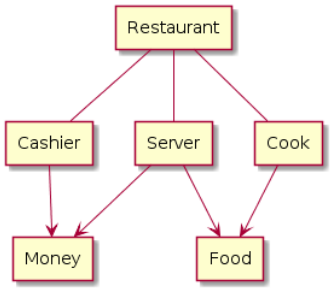


Figure B-3 Data model of a network database

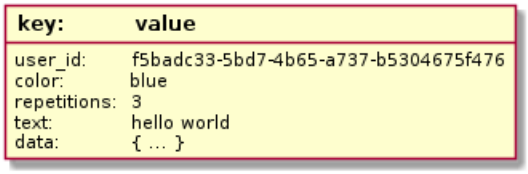


Figure B-4 Data model of a key-value database

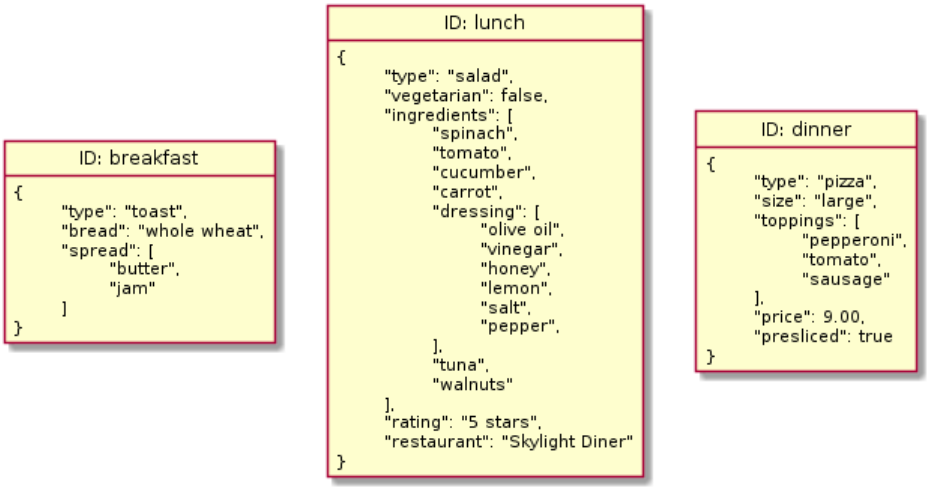


Figure B-5 Data model of a document database

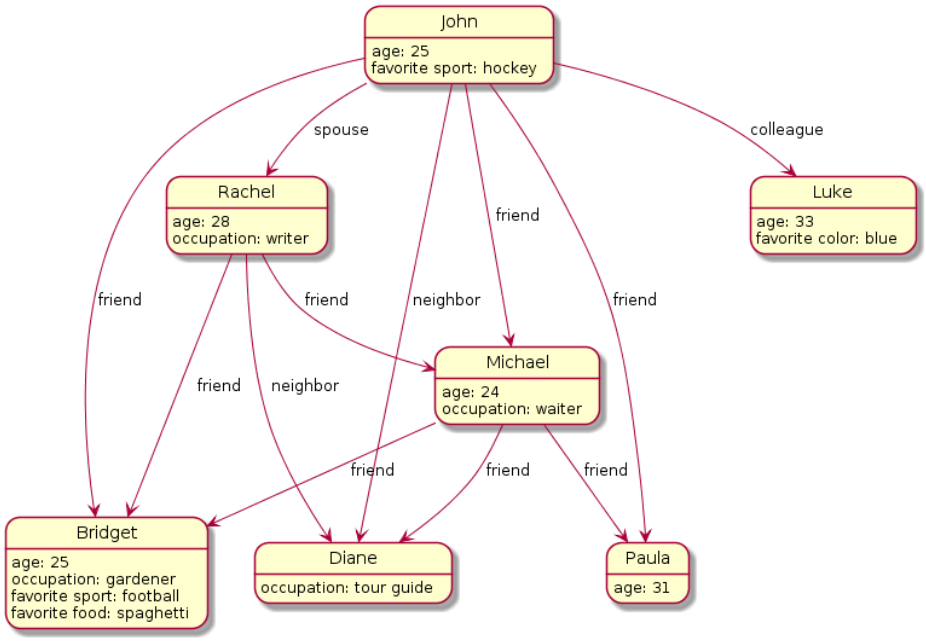


Figure B-6 Data model of a graph database

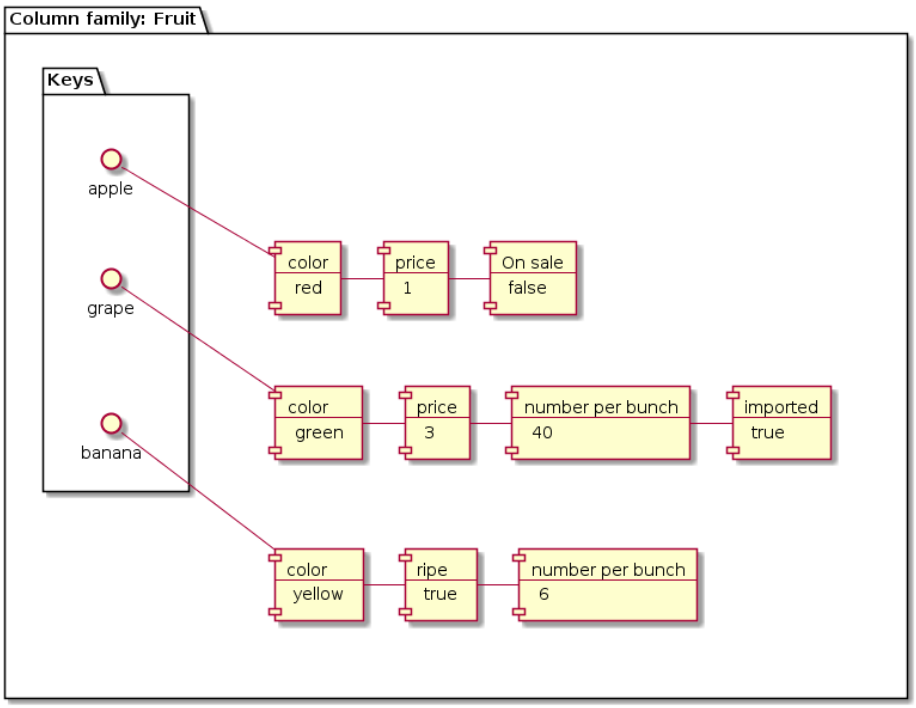


Figure B-7 Data model of a column database

Time	CPU Temp
2019-10-31T03:48:05+00:00	37
2019-10-31T03:48:10+00:00	42
2019-10-31T03:48:15+00:00	33
2019-10-31T03:48:20+00:00	34
2019-10-31T03:48:25+00:00	40
2019-10-31T03:48:30+00:00	42
2019-10-31T03:48:35+00:00	41

Figure B-8 Data model of a time-series database

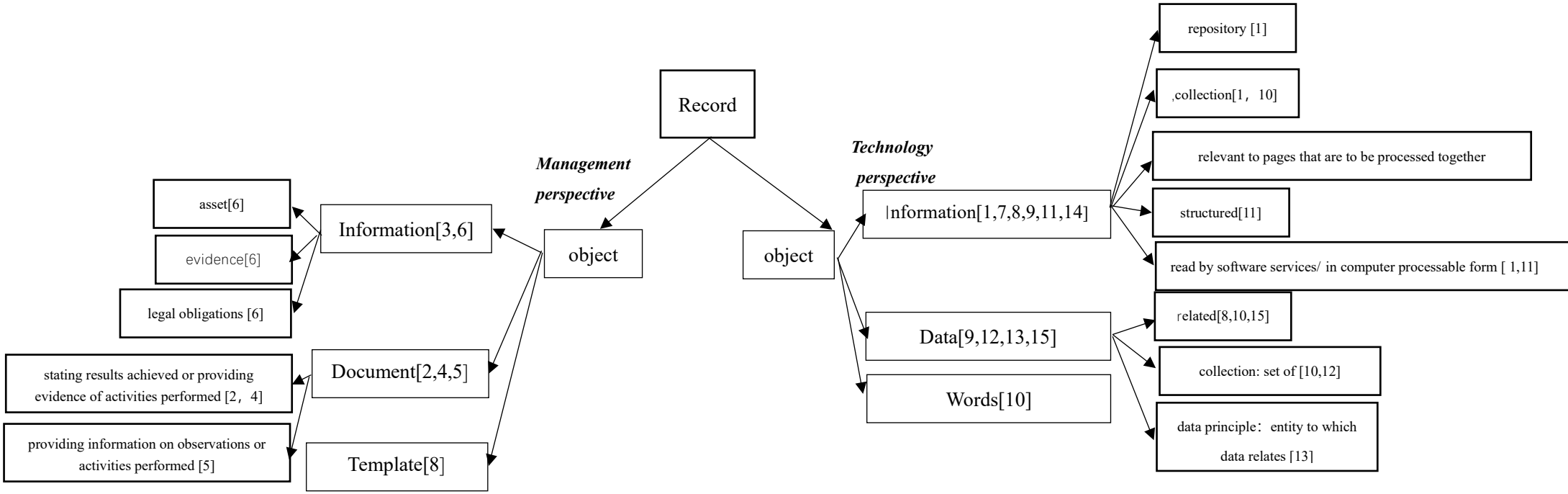
Some more about relational databases

A relational database is based on the relational model and principles as developed by Edward F. Codd in 1970 in a document called 'A Relational Model of Data for Large Shared Data Bank'. In this document Codd formulated 'twelve rules (0 to 12) for database systems'. Nowadays these rules are considered to be superseded by 'The Third Manifesto'-series of writings on the relational model by Hugh Darwen and C.J. Date.

Data (or information) in relational databases are considered as being structured data (or information) according to the definition of the concept 'structured data' as mentioned in clause 3. But note 3 to that definition ('Example of structured data are data contained in relational databases.') suggests that there are other digital environments that contain structured data too (maybe an XML file according to the specifications in an XML Document Type Definition?).

In the area of database administrators and database managers, rows in relational database tables are also known as 'tuples', 'database records' and in short 'records'. This does not mean that these 'records' are records as defined by TC46/SC11 (in the standard 15489-1). In the same area tables are also named 'relations'.

Annex C (informative) Concepts of record from management perspective and technology perspective



Annex D (informative) ISO/TC46/SC11/WG22: Template for use cases on records management in structured data environments, especially in relational databases

Version2022.11.25

Row ID	**Case content type**	**Case content**	**Actions and notes** (by WG22)
1.1	Use case code	to be completed by WG22	
1.2	Liaison officer national body → WG22		
1.3	Use case name The use case name needs to reflect the scenario of records management in the structured data environment. e.g. use case and requirements for relational database records scheduling.		
1.4	Additional notes		
About the organization(s)			
2.1	Name of the organization Name of author Country Email		
2.2	Brief description of the organisation's function, key roles, and tasks		

Row ID	Case content type	Case content	Actions and notes (by WG22)
2.3	Additional notes		
About records management of data in structured data environments in general			
3.1	Please provide a brief description of the use case and the associated relational databases. The use case brief introduction needs to reflect the following elements: what is the specific scenario of the structured data environment, what are the associated relational databases（name,type,function, predefined (applicable) set of rules?）? (Less than 200		
3.2	Please provides brief descriptions of types of actors involved in the management of data in relational databases and their roles and the relationships? **Actors who play the roles of managing records in structured data environment in the use case e.g.** (1) Business manager, system developer, data architect, database manager; （2）Data creator (records creation), data processor and controller (records retention, custody and disposition, authority control), data provider (records provider), data user (records user) Please draw a diagram to show relationships between various actors.		
3.3	Does the organization recognize that its relational databases contain data that should be subject to records management rules? If yes, why?		

Row ID	Case content type	Case content	Actions and notes (by WG22)
	Records management rules refer to records management processes and controls in structured data environment. E.g. Capturing, disposing, deleting records in relational database. (what are the rules of ISO 15489,ISO 16175? ISO 23081 that are implemented in relational database?) E.g. classification schema, metadata schema, description rule, access and permission rule, disposition authority		
3.4	Is the organization aware of issues and challenges related to applying records management rules to the data in relational databases? If yes, please describe the perceived or identified issues and challenges. (list of issues identified in WD)		
3.5	Does the organization work with other types of databases than relational databases? If yes: Please provide a brief description of these databases?		

Row ID	Case content type	Case content	Actions and notes (by WG22)
	Are these databases also considered to have challenges in applying records management rules to them? If so, please describe the perceived (or identified) issues and challenges.		
3.6	Additional notes		
About issues and challenges for records management of data in relational databases as already identified by WG22			
4.1	Issue/challenge 1: **Identifying records in relational databases (appraisal phase).** *Explanatory note:* In a relational database, data elements and combinations of these data elements can be used multiple times for different events and transactions, as well as in multiple information objects about these events and transactions. Examples include requests, decisions, permits, minutes, and reports that serve as evidence of business activities. This results in multiple many-to-many relationships between: • meaningful information elements in the real world; • data elements in the data fields of a relational database. As such, the challenge is to determine which data elements in a relational database are subject to records management rules. Does the organization recognize this as a problem or challenge that applies to the organization? If yes, how does the organization solve this problem or deal with it?		

Row ID	Case content type	Case content	Actions and notes (by WG22)
4.2	Issue/challenge 2: **Conflicting retention periods (appraisal phase).** *Explanatory note:* Using structured data elements in a relational database multiple times in multiple cases and transactions, may result in multiple and conflicting retention periods for the same data elements/element combinations in the database. Does the organization recognize this as a problem or challenge for the organization? If yes, how does the organization solve this problem or deal with it?		
4.3	Issue/challenge 3: **Storing and keeping fixed-content records of current data (capture phase).** *Explanatory note:* In a relational database, if structured data elements need to be kept current by updating them, their contents are often not fixed. However, the content of records that serve as reliable evidence of business processes need to be fixed and inviolable. So, it may be challenging to create fixed-content records by capturing values that are initially current and not fixed (source: ISO 16175-1, clause 5.1). Does the organization recognize this as a problem or challenge? If yes, how does the organization solve this problem or deal with it?		
4.4	Issue/challenge 4:		

Row ID	Case content type	Case content	Actions and notes (by WG22)
	Potential conflict between records management disposition rules for data in a relational database (as records), and the referential integrity rule for a relational database (disposition phase). *Explanatory note:* If data elements in a relational database are to be considered records and/or if they have been stored as records in such a database, these records will be subject to both: • disposition rules for the records; • the relational databases' referential integrity rule (needed for proper database management and operation of a relational database). As such, a conflict can be created whereby disposition rules dictate the deletion of records, but the referential integrity rule for relational databases dictates their retention. Does the organization recognize this as a problem or challenge? If yes, how does the organization solve this problem or deal with it?		
4.5	Issue/challenge 5: **Collaboration with other professionals on records in a structured data environment.** *Explanatory note:* Record management specialists should collaborate with database managers, database developers, and information architects because of the need to analyze database data models, recognize and solve problems early on, and incorporate records management by design. This can be quite new and challenging if an organization is not yet		

Row ID	Case content type	Case content	Actions and notes (by WG22)
	accustomed to records management by design and a multidisciplinary approach to records management. Does the organization recognize this as a problem or challenge? If yes, how does the organization solve this problem or deal with it?		
4.6	Additional notes		
About more issues and challenges for records management of data in relational databases			
5.1	Has the organization experienced more specific issues/challenges around records management of data in relational databases (other than those above)? If yes, what are they and can you please describe what makes them an issue/ challenge?		
5.2	Do you have any suggestions for WG22 to address other problems or conduct further research?		
5.3	Additional notes		
In general			
6.1	Additional notes		

Annex E (informative) Use case-1 as an example

Row ID (by WG22)	**Case content type**	Content
	Date	2023-02-01
	Use case code	UC-1
	Liaison officer WG22 - National body	Xiaomi An… (CN)
	Use case name	Managing records in relational databases of telecommunication operator
	Additional notes	
	Name of the organization(s) Name of author Country	Beijing Borrui Data Technology Co., Ltd. Dowson Liu; Zhao Jinghua; Gao Yang; Piao Shenghong China
	Type of organization: role and tasks	Private Corporate Database Company
	Additional notes	

Row ID (by WG22)	**Case content type**	Content
	Brief description of the use case and the relational database or databases involved or in mind for this use case.	Relational databases are widely used in the telecommunication industry, supporting the data analysis of business data. This relational database accesses and processes data directly in memory. No waiting time of loading data from disk to memory is needed. Query processing request is broken up into many smaller tasks, which are distributed intelligently and executed in parallel across nodes for real-time processing and analysis. The distributed architecture of relational database allows a horizontal expansion of hardware. Additional nodes can be added to clusters to meet ever-growing business demand. The unique embedded federated connector system enables users to access various data sources with industry-standard SQL and JDBC interfaces, dispensing with the traditional Extract-Transform-Load (ETL) process and unnecessary data migration.
	Does the organization recognize that data in one or more of its relational databases should be subject to records management rules? If so, why?	Yes, it has been common sense in the industry. The query and analysis of the relevant information, such as account opening information, user call information, and location information for base station positioning are stored in a structured data environment. Organizations have realized that all information is valuable, and some information will be classified as data assets, but there is lack of awareness of records management rules and managing data as records is not adequate.

Row ID (by WG22)	**Case content type**	Content
	Does the organization have a presumption of the issues and challenges it has to face if applying records management rules to the data in its mentioned relational databases? If so, what are these supposed (or discovered) issues and challenges?	The development of industry and technology promotes the improvement of record management awareness, but the considerations of records management rules in the system is largely affected by traditional concepts, capital investment and technical limitations. 1. Leadership awareness: The quality of records management depends on the awareness of records management and the resources input from leadership. The more input, the better the records management. 2. Cost input: Cost control is the most important factor which can affect input to function and performance of system. Only when benefits outweigh costs, records management will be implemented smoothly. 3. Technical risk: Whenever records management are implemented, the new requirements of records management (e.g. immature technology or new technical solution) will bring challenges to the former system technical architecture. 4. Business conflict: The records management may challenges to the original business, retention period rules for example. 5. Other
	Additional notes	

Row ID (by WG22)	**Case content type**	Content
	Challenge 1: Identifying records in relational databases (appraisal phase). Explanatory note: The content of data elements and combinations of these data elements in a relational database can be used multiple times for different events and transactions and in multiple information objects about these events and transactions, such as in requests, decisions, permits, minutes and reports which has to serve as evidence of business activities. This leads to multiple n:n relationships between meaningful information elements in the real world and data elements in the data fields of a relational database. In this landscape, the challenge is to determine which data elements in a relational database are subject to records management rules. Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	Due to lack of identification rules for records, it is hard to identify records from numerous data for databases, which needs records/archives experts, business experts, technology experts, and other relevant departments to cooperate to set identification rules and guide the implementation of records management. 1. Matching relevant data elements with business (records/archives) requirements, and manage access rights, retention periods, and disposal of different data elements in layers. 2. In a structured environment, different management/storage strategies can be set for different types of data, and the system can be flexibly implemented when the business (file) management party puts forward clear requirements at the beginning. 3. The application of core and key data in different scenarios should be considered at the design stage. For example, in order to retain evidence information and avoid business risks, we store user basic account information, call record information location information of base station positioning, and contract information in a structured data environment based on the Archives Law and the Regulations of Telecommunications. The stored data is appraisal as a record according to the following rules: 1. It reflects the result or evidence of a business decision or action. 2. It has legal, financial, management or historical value. 3. It is required by regulations or standards.

Row ID (by WG22)	**Case content type**	Content
	Challenge/issue 2: Conflicting retention periods (appraisal phase). Explanatory note: If the content of structured data elements in a relational database is used multiple times in multiple cases and transactions, then this may result in multiple and conflicting retention periods for the same data elements or combinations of data elements in a relational database. Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	If the corresponding rules cannot be set at the system design stage, it will be difficult to solve the relevant problems during the process of system operation. Ensure that the system can retain the data according to the business scenario with the consideration of disposition rules and retention schedule at the very beginning of the rules design stage. For example, the call records of users may need to be kept for six months according to the telecommunications regulations, but only for three months according to the local provincial and municipal regulations. In this case, the longest storage period should be followed.

Row ID (by WG22)	**Case content type**	Content
	Challenge/issue 3: Storing and keeping fixed-content records of current data (capture phase). Explanatory note: If structured data elements in a relational database are current and need to be kept current by updating them, their content is not fixed, while the content of records that are to serve as reliable evidence of business processes, need to be fixed and inviolable. So, it may be challenging to create fixed-content records by capturing values that are initially current and not fixed (source: ISO 16175-1, clause 5.1) Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	Yes, it is a challenge. The rules should be set in advance and set the capture trigger/ node (including time, frequency, departure event, etc.) If there is a change in this type of data, the system will be triggered to automatically save the current fixed content and store it in the expected format. We use triggers, stored procedures, and other mechanisms to automatically take a snapshot of the data elements when they change and save it in another table as a record. Alternatively, we use specialized record management software or systems that work with relational databases and offer features for creating, capturing, and managing fixed-content records. For example, we have a table in a relational database that stores user information such as name, phone number, ID number, etc. We need to refresh this information periodically to keep it accurate and current. At the same time, we also need to make fixed-content records as reliable proof of our user information management process. We use the database snapshot function to take a snapshot of the database before or after each data update and keep it on the same server instance. This way we query the snapshot to see how the data looked at a certain point in time and compare it with other snapshots.

Row ID (by WG22)	**Case content type**	Content
	Challenge/issue 4: Potential conflict between disposition rules from the records management perspective for data in a relational database on the one hand and the referential integrity rule for a relational database on the other (disposition phase). Explanatory note: If data elements in a relational database are to be considered records and/or if these data elements have been stored as records in such a database, these records will be subject to both disposition rules for records and the referential integrity rule for relational databases as needed for proper database management and the operation of a relational database. While disposition rules may dictate the deletion of records, the referential integrity rule for relational databases may dictate the retention of the same records. That is the crux of a potential conflict. Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	Yes, it is a challenge. The rules of records management shall be incorporated into database processing rules. Once the database processing rules are set, it is difficult to change the rules, which the processing rules of the database must be assured at the design stage. It can meet the requirement of records management by modifying the relational database processing rules. We use the archive function to move a record from the main table to a special area for storing historical data or backup data instead of deleting it from the database when deleting it from the main table. This way, we can retain the original data and reference relationships, and facilitate future recovery or query operations. For example, we use the archive function to handle the relationship between user information and business information. In this case, the user information table and the business information table are linked by a foreign key through the phone number, and when creating the foreign key, specify that deletion is not allowed. In this way, when deleting a record of a canceled user from the user information table, it is not deleted from the database, but moved to a special area for storing historical data or backup data. This way, we retain the original data and reference relationships and facilitate future recovery or query operations.

Row ID (by WG22)	**Case content type**	Content
	Challenge/issue 5: Collaboration with other professionals on records in a structured data environment. Explanation: Record management specialists should collaborate with database managers, database developers and information architects because of the need for analyzing database data models, recognizing and solving problems early on and records management by design. This can be quite new and challenging if an organization is not yet accustomed to records management by design and a multidisciplinary approach to records management. Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	During the system design process, business experts, records/archives experts, and technical experts including data architects should be fully communicated，design a data management strategy that meets the requirements of records management should be formulated and implemented. However, currently, at the strategic level, both archival management personnel and business personnel lack of guidance of standards for records management to structured data environment. Therefore, it is difficult to formulate relevant rules for records management, which makes it difficult for IT technicians at the operational level to design practical standard requirements for records management, and system developers to jointly promote records management. We set up a cross-departmental working group or committee to coordinate communication and collaboration among different parties. This ensures that records management experts can timely understand and participate in the process of database design, development, and maintenance, and share records management standards, policies, and best practices with other professionals. We provide training and guidance to enhance the awareness and skills of all parties. We help records management experts and other professionals understand each other’s roles, responsibilities, and needs, and improve their ability to manage records in a structured data environment. At the same time, we also use software tools to support the implementation of records management by design. This enables us to leverage technology to automate, simplify and optimize records management processes, such as identifying, classifying, protecting, accessing, and destroying records, and improve data quality and security.
	Additional notes	

Row ID (by WG22)	**Case content type**	Content
	Has your organization experienced more specific challenges/problems in the area of records management of data in relational databases (other than those mentioned here)? If so, what are these challenges/problems and can you explain why they are a challenge/problem?	No

Row ID (by WG22)	**Case content type**	Content
	Does the organization work with other types of databases than the relational type? If so, what are these types of databases? And are they also in mind as a challenge when applying records management rules to them? If so, has the organization a presumption of the issues and challenges if doing so? If so, what are these supposed issues and challenges?	Yes, other databases are widely used, such as document databases and graph databases. Document databases mainly provide storage, retrieval, and analysis functions for documents. Compared with relational databases, document databases allow the creation of many different types of unstructured or arbitrary format fields. The main difference from relational databases is that they do not provide support for parameter integrity and distributed transactions, but they are similar to relational databases. They are not mutually exclusive, and they can exchange data with each other, thereby complementing and expanding each other. In this case, the document database is mainly used for storage and query of user credentials, various agreements, contracts, and other information. Graph databases are online database management system that handles create, read, update, and delete operations on graph data models. In this case, the graph database is mainly used to manage the network, access control, and customer relationship. The main challenge faced is the same as relational databases. The main challenge is the lack of awareness of record management rules rather than what types of databases are used.

Row ID (by WG22)	**Case content type**	Content
	Do you have any suggestions for WG22 to address other problems or conduct further research?	1. Due to the lack of awareness of functional requirements for records management during the development of information technology systems, it is hoped that standards for guidance and evaluation will be formulated to guide relevant parties to implement them. 2. Records management in digital environment, including all types of data not just structured data environment, and can not be separately considered from semi-structured or unstructured data environments. 3. Managing records in the SMART area, which are applicable, readable, and transferable by machine, deserves more attention from records management professionals 4. Managing records in AI and in SMART way needs the first priority considerations in future standards development.
	Additional notes	

Annex F (informative) Use case-2 as an example

Row ID (by WG22)	**Case content type**	Content
	Date	2022-12-1
	Use case code	UC-2
	Liaison officer WG22 - national body	Xiaomi An (CN)
	Use case name	Records management of structured data of aircraft products
	Additional notes	
	Name of the organization(s) Name of author Country	COMAC Shanghai Aircraft Design & Research Institute Wang Muliang China
	Type of organization: role and tasks	Manufacturing enterprise Aircraft design
	Additional notes	

Row ID (by WG22)	**Case content type**	Content
	Brief description of the use case and the relational database or databases involved or in mind for this use case.	In the process of aircraft product development, to record it authentically and completely, a lot of structural product data will be formed, such as Engineering Change Request, ECR and Engineering Change Proposal, ECP. These data which storage in relational databases of about thirty types and link to each other are indispensable records to prove that the process of aircraft product development has met regulatory requirements and that the product is of acceptable quality. The relational database to generate structural data is a part of the Product data Management (PDM) system. We rely on PDM to create, check, publish, transfer, storage, use and dispose data. The database volume reaches to 50 GB, Queries Per Second (QPS) is 265, and Transactions Per Second (TPS) is 1.15. We face the following challenges: 1. The formulation of data management rules involves multiple departments, which makes it difficult to coordinate among them; 2. Capital investment. Decision-makers may not be able to support this work because of large capital investment and relatively small benefits; 3. Human resources are insufficient. The enterprise may not know how to implement the records management rules due to the lack of professional personnel in the field of records management; 4.Data changes frequently. It may result in the misuse of historical version data. 5.Complex relationship between data. The processing of one type of data may make another type of data unavailable.

Row ID (by WG22)	**Case content type**	Content
	Does the organization recognize that data in one or more of its relational databases should be subject to records management rules? If so, why?	Yes, 1.Implementing the records management rules is conductive to support high-quality information resources for product development, customer service and other businesses, contributing to stability and continuity, and improving the efficiency of operation. 2.To meet the needs of governmental supervision and third-party review, and provide authentic and complete business records. 3.Provide information resources support for the development of new business areas.
	Does the organization have a presumption of the issues and challenges it has to face if applying records management rules to the data in its mentioned relational databases? If so, what are these supposed (or discovered) issues and challenges?	The challenges include: 1. Capital investment. Decision-makers may not be able to support this work because of large capital investment and relatively small benefits; 2. Human resources are insufficient. The enterprise may not know how to implement the record management rules due to the lack of professional personnel in the field of records management;
	Additional notes	

Row ID (by WG22)	**Case content type**	Content
	Challenge 1: Identifying records in relational databases (appraisal phase). Explanatory note: The content of data elements and combinations of these data elements in a relational database can be used multiple times for different events and transactions and in multiple information objects about these events and transactions, such as in requests, decisions, permits, minutes and reports which has to serve as evidence of business activities. This leads to multiple n:n relationships between meaningful information elements in the real world and data elements in the data fields of a relational database. In this landscape, the challenge is to determine which data elements in a relational database are subject to records management rules. Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	Yes. 1.Determine whether it is within the scope of management according to *the Archive Scope and Retention Period Table*; 2. Hold an appraisal committee meeting to discuss the importance of data and determine whether the records management rules are fully or partially implemented. 3. Formulate relevant institutional document to clarify the record management rules to be followed within the enterprise and their scope of application.

Row ID (by WG22)	**Case content type**	Content
	Challenge/issue 2: Conflicting retention periods (appraisal phase). Explanatory note: If the content of structured data elements in a relational database is used multiple times in multiple cases and transactions, then this may result in multiple and conflicting retention periods for the same data elements or combinations of data elements in a relational database. Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	Yes, For example, the ECR is a prerequisite for the ECP. The ECP cannot be created without the ECR, the ECR is referenced when the ECP is established, the two belong to the reference relationship, and the relationship is stored in the database. When establishing the retention period, we identified the correlation between the two types of data, so as to set the same storage period.

Row ID (by WG22)	**Case content type**	Content
	Challenge/issue 3: Storing and keeping fixed-content records of current data (capture phase). Explanatory note: If structured data elements in a relational database are current and need to be kept current by updating them, their content is not fixed, while the content of records that are to serve as reliable evidence of business processes, need to be fixed and inviolable. So, it may be challenging to create fixed-content records by capturing values that are initially current and not fixed (source: ISO 16175-1, clause 5.1) Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	Yes, for example, ECP may change after it was published. We take the approach of version control to face this challenge. Each release of a version requires the data of that version to be fixed and at the same time captured and archived.

Row ID (by WG22)	**Case content type**	Content
	Challenge/issue 4: Potential conflict between disposition rules from the records management perspective for data in a relational database on the one hand and the referential integrity rule for a relational database on the other (disposition phase). Explanatory note: If data elements in a relational database are to be considered records and/or if these data elements have been stored as records in such a database, these records will be subject to both disposition rules for records and the referential integrity rule for relational databases as needed for proper database management and the operation of a relational database. While disposition rules may dictate the deletion of records, the referential integrity rule for relational databases may dictate the retention of the same records. That is the crux of a potential conflict. Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	Yes, two rules sometimes conflict. The method we take to deal with it is, When making the deletion decision, the potential impact of the referential integrity rules of the relational database should be fully considered and included in the evaluation scope as the decision input.

Row ID (by WG22)	**Case content type**	Content
	Challenge/issue 5: Collaboration with other professionals on records in a structured data environment. Explanation: Record management specialists should collaborate with database managers, database developers and information architects because of the need for analyzing database data models, recognizing and solving problems early on and records management by design. This can be quite new and challenging if an organization is not yet accustomed to records management by design and a multidisciplinary approach to records management. Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	Yes, collaboration between different professionals is greatly challenging. To cope with the challenges of product data management between different professionals, we established a favorable collaboration mechanism, including: At the beginning of database/software design, seek the opinions of records management experts and incorporate relevant opinions into database/software scheme design and detailed design records; Listen to the opinions of records management experts when dealing with database data; Convene coordination meetings regularly; Incorporate relevant activities into the internal management system.
	Additional notes 补充说明	

Row ID (by WG22)	**Case content type**	Content
	Has your organization experienced more specific challenges/problems in the area of records management of data in relational databases (other than those mentioned here)? If so, what are these challenges/problems and can you explain why they are a challenge/problem?	Is the long-term preservation of data records in relational databases dependent on its native database/software or a special electronic records management system? The method adopted is to archive structured data into the electronic records management system with long-term storage function for safekeeping and disposition.
	Does the organization work with other types of databases than the relational type? If so, what are these types of databases? And are they also in mind as a challenge when applying records management rules to them? If so, has the organization a presumption of the issues and challenges if doing so? If so, what are these supposed issues and challenges?	No
	Do you have any suggestions for WG22 to address other problems or conduct further research?	Implementing records management rules in a structured data environment is costly, sometimes even huge. Therefore, the degree of implementing records management rules in structured data environment should be distinguished. Special important data shall be fully implemented; For data of general importance, it can be partially implemented; For unimportant data, it can not be implemented.
	Additional notes	

Annex G (informative) Use case-3 as an example

Row ID (by WG22)	**Case content type**	Content
	Date	2022-12-1
	Use case code	UC-3
	Liaison officer WG22 - national body	Xiaomi An （CN）
	Use case name	Use cases and requirements of relational database in the field of petroleum engineering technology
	Additional notes	
	Name of the organization(s) Name of author Country	China National Petroleum Corporation Jingsheng Fu，Xinghua Su，Zhenghe Song，Guodong Su China
	Type of organization: role and tasks	Chinese central State-owned enterprises； It is a comprehensive international energy company with oil and gas business, engineering technology services, petroleum engineering construction, petroleum equipment manufacturing, financial services, new energy development and other main businesses, and is one of the major oil and gas producers and suppliers in China
	Additional notes	

Row ID (by WG22)	**Case content type**	Content
	Brief description of the use case and the relational database or databases involved or in mind for this use case.	This use case shall be used for data collection, storage and management of relational database in various petroleum engineering technology professional systems. This use case can significantly improve scientific decision-making and economic benefits by storing data generated in production management processes such as automatic collection, manual input, scientific calculation, research results, production management, etc. into relational databases for scientific research, production, management and other applications. The data mainly includes design data of oil drilling, drilling process data, and drilling results data. The relational database used in this use case is Oracle database. Oracle database system is a popular relational database management system in the world at present. The system has good portability, convenient use, and strong functions, and is applicable to various large, medium, small, and microcomputer environments. It is a database scheme with high efficiency, good reliability and high throughput.

Row ID (by WG22)	**Case content type**	Content
	Does the organization recognize that data in one or more of its relational databases should be subject to records management rules? If so, why?	In the petroleum engineering technology industry, there is a strong sense of management of construction achievements data (drilling well history, cementing summary, etc.) as records. However, although the construction process data (caliper data, daily data, etc.) are stored in the structured data environment for query and analysis, the awareness of management of data as records is relatively unconsciousness. However, the petroleum engineering technology industry has realized that all data are valuable and is gradually strengthening the records management requirements and the specification for all data.

Row ID (by WG22)	**Case content type**	Content
	Does the organization have a presumption of the issues and challenges it has to face if applying records management rules to the data in its mentioned relational databases? If so, what are these supposed (or discovered) issues and challenges?	With the increasing emphasis on the development of information technology in the petroleum engineering technology industry, some traditional concepts and capital constraints have improved, but still face the impact of industry specifications, technical risks and safety requirements. 1. Industry specifications: due to the possibility that the industry specifications used for some data are outdated or obsolete, the resulting records may face usability and security risks. 2. Technical risk: adding records management to the original old system may cause technical incompatibility or adaptation conflicts and other risks, and will face a lot of compatibility challenges. 3. Security requirements: records management of original data may cause information security risks, such as the record being obtained by unauthorized institutions or individuals. 4. Challenges to traditional concepts and regulations: The traditional thinking and regulations for arching paper records need to be updated.
	Additional notes	

Row ID (by WG22)	**Case content type**	Content
	Challenge 1: Identifying records in relational databases (appraisal phase). Explanatory note: The content of data elements and combinations of these data elements in a relational database can be used multiple times for different events and transactions and in multiple information objects about these events and transactions, such as in requests, decisions, permits, minutes and reports which has to serve as evidence of business activities. This leads to multiple n:n relationships between meaningful information elements in the real world and data elements in the data fields of a relational database. In this landscape, the challenge is to determine which data elements in a relational database are subject to records management rules. Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	The petroleum engineering technology industry lacks identification rules for data. At present, through the cooperation of archives, business, technology and other departments, corresponding identification rules have been formulated for some business data according to the actual production and management needs and relevant information systems; The method of rulemaking is mainly based on business logic, but it is difficult to archive process data, such as large amounts of data, high collection frequency, large space consumption, and storage cost performance.

Row ID (by WG22)	**Case content type**	Content
	Challenge/issue 2: Conflicting retention periods (appraisal phase). Explanatory note: If the content of structured data elements in a relational database is used multiple times in multiple cases and transactions, then this may result in multiple and conflicting retention periods for the same data elements or combinations of data elements in a relational database. Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	For records with retention period requirements, the data records shall be saved and managed according to the retention period of the business with consideration of the longest retention period in each business system.

Row ID (by WG22)	**Case content type**	Content
	Challenge/issue 3: Storing and keeping fixed-content records of current data (capture phase). Explanatory note: If structured data elements in a relational database are current and need to be kept current by updating them, their content is not fixed, while the content of records that are to serve as reliable evidence of business processes, need to be fixed and inviolable. So, it may be challenging to create fixed-content records by capturing values that are initially current and not fixed (source: ISO 16175-1, clause 5.1) Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	At present, the industry captures current data by setting records saving nodes in fixed business processes. For example, we take a certain period of time for well completion as the deadline for archiving, which is to capture and archive records based on the completion of records processing. If such data elements are changed, the new version of the record should be saved again to capture the latest data.

Row ID (by WG22)	**Case content type**	Content
	Challenge/issue 4: Potential conflict between disposition rules from the records management perspective for data in a relational database on the one hand and the referential integrity rule for a relational database on the other (disposition phase). Explanatory note: If data elements in a relational database are to be considered records and/or if these data elements have been stored as records in such a database, these records will be subject to both disposition rules for records and the referential integrity rule for relational databases as needed for proper database management and the operation of a relational database. While disposition rules may dictate the deletion of records, the referential integrity rule for relational databases may dictate the retention of the same records. That is the crux of a potential conflict. Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	Modify the length of the relevant field type of the relational database or reassign its content to make it meet designed records management requirements.

Row ID (by WG22)	**Case content type**	Content
	Challenge/issue 5: Collaboration with other professionals on records in a structured data environment. Explanation: Record management specialists should collaborate with database managers, database developers and information architects because of the need for analyzing database data models, recognizing and solving problems early on and records management by design. This can be quite new and challenging if an organization is not yet accustomed to records management by design and a multidisciplinary approach to records management. Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	In the records management design stage, according to the design requirements, multiple departments such as archives, business and technology are required to jointly discuss and implement. When IT designers design, they should form a project team with business and archives related personnel to fully communicate with each other, so as to design a strategy that meets the records management requirements of each department. We standardize relevant personnel in the data chain through regulations, such as training for data entry personnel, strengthening the implementation of verification mechanisms for auditors, and participating in the review of major milestone events during the system construction phase.
	Additional notes	

Row ID (by WG22)	**Case content type**	Content
	Has your organization experienced more specific challenges/problems in the area of records management of data in relational databases (other than those mentioned here)? If so, what are these challenges/problems and can you explain why they are a challenge/problem?	None
	Does the organization work with other types of databases than the relational type? If so, what are these types of databases? And are they also in mind as a challenge when applying records management rules to them? If so, has the organization a presumption of the issues and challenges if doing so? If so, what are these supposed issues and challenges?	Yes, PI real-time database has been used to store real-time data of production equipment. If records management is needed, data timeliness and space requirements are the main challenges.
	Do you have any suggestions for WG22 to address other problems or conduct further research?	It is hoped that the working group can formulate relevant standards or specifications on records structure in structured data records management to guide the records management design process.
	Additional notes	

Annex H (informative) Use case-4 as an example

Row ID (by WG22)	**Case content type**	Content
	Date	2023-3-8
	Use case code	UC-4
	Liaison officer WG22 - national body	Xiaomi An (CN)
	Use case name	Records Management of Relational Databases in Business Systems
	Additional notes	
	Name of the organization(s) Name of author Country	State Grid Tianjin Electric Power Archives Tingting Liu China
	Type of organization: role and tasks	Central enterprises State Grid Tianjin Electric Power Company is a subsidiary of State Grid Corporation of China. It is responsible for the planning, construction, and operation of the Tianjin power grid, and is committed to providing clean, low-carbon, safe, and efficient electricity and energy supply for the economic and social development of Tianjin.
	Additional notes	

Row ID (by WG22)	**Case content type**	Content
	Brief description of the use case and the relational database or databases involved or in mind for this use case.	The relational database mentioned in this use case exists in the course of various business activities of the company. The company has established corresponding work platforms and relational databases for various businesses, in which relevant functions such as business processing, data query, and information storage are implemented.
	Does the organization recognize that data in one or more of its relational databases should be subject to records management rules? If so, why?	With the approaching of the digital age , the company's various businesses have established a corresponding working platform, on which to implement business handling , data query, information storage and other related functions. But the awareness of considering these data as records to manage and archiving is still relatively weak.

Row ID (by WG22)	**Case content type**	Content
	Does the organization have a presumption of the issues and challenges it has to face if applying records management rules to the data in its mentioned relational databases? If so, what are these supposed (or discovered) issues and challenges?	Yes. There are still various problems and challenges in applying records management to data management in the operation system. It mainly includes the following aspects: 1. Data in partial operation system has strong timeliness and needs to be invoked, processed and analyzed at any time. The operating department believes that managing this part of data in accordance with records not only cannot meet the requirements of timeliness but will increase the cost of construction and communication. 2. Data in partial operation system can meet partly requirements of daily work of this business , while managing it as records will face the problem of archiving in the future as well as need to meet the requirements of archive standards, which including the storage format , metadata, electronic signature, encapsulation and backup requirements involved in electronic records management. For professional departments, it is difficult to implement. 3.Data in partial operation system can be preserved for a long time and called at any time. The operating department considers that managing and archiving it as records belongs to repetitive construction, and unfavorable to later utilization.
	Additional notes	

Row ID (by WG22)	**Case content type**	Content
	Challenge 1: Identifying records in relational databases (appraisal phase). Explanatory note: The content of data elements and combinations of these data elements in a relational database can be used multiple times for different events and transactions and in multiple information objects about these events and transactions, such as in requests, decisions, permits, minutes and reports which has to serve as evidence of business activities. This leads to multiple n:n relationships between meaningful information elements in the real world and data elements in the data fields of a relational database. In this landscape, the challenge is to determine which data elements in a relational database are subject to records management rules. Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	Yes. The challenge lies in how business department and archives department work together to formulate the rules. That is, the business department must first identify which data is necessary to comply with the rules of record management to manage, and the archives department puts forward the corresponding management requirements , then the business department makes corresponding adjustment in their own operating business according to this requirements . However, because record system is the tail end in actual business while the operating system is the department that generates the front-end metadata, it is greatly difficult to modify the data rules in the operating system with the record management rules under the premise that the system can meet routine work.

Row ID (by WG22)	**Case content type**	Content
	Challenge/issue 2: Conflicting retention periods (appraisal phase). Explanatory note: If the content of structured data elements in a relational database is used multiple times in multiple cases and transactions, then this may result in multiple and conflicting retention periods for the same data elements or combinations of data elements in a relational database. Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	Yes. Data elements can be used multiple times in many matters and processes, which is different from traditional carrier archiving. Therefore, it is suggested to keep structured data according to the longest retention period to solve this kind of problem.

Row ID (by WG22)	**Case content type**	Content
	Challenge/issue 3: Storing and keeping fixed-content records of current data (capture phase). Explanatory note: If structured data elements in a relational database are current and need to be kept current by updating them, their content is not fixed, while the content of records that are to serve as reliable evidence of business processes, need to be fixed and inviolable. So, it may be challenging to create fixed-content records by capturing values that are initially current and not fixed (source: ISO 16175-1, clause 5.1) Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	Yes. This problem involves how the business system interfaces with the record system. At the beginning of docking, time nodes or event nodes should be set up according to the actual situation of business departments. When data is pushed, it should be pushed and stored based on a certain time node or event node. If subsequent data elements change, it should be overwritten or stored separately according to the actual situation.

Row ID (by WG22)	**Case content type**	Content
	Challenge/issue 4: Potential conflict between disposition rules from the records management perspective for data in a relational database on the one hand and the referential integrity rule for a relational database on the other (disposition phase). Explanatory note: If data elements in a relational database are to be considered records and/or if these data elements have been stored as records in such a database, these records will be subject to both disposition rules for records and the referential integrity rule for relational databases as needed for proper database management and the operation of a relational database. While disposition rules may dictate the deletion of records, the referential integrity rule for relational databases may dictate the retention of the same records. That is the crux of a potential conflict. Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	It's a challenge. Some data is necessary but has no archival value to the business department. The unqualified data can be filtered when the data elements are pushed from the business system to the record system.

Row ID (by WG22)	**Case content type**	Content
	Challenge/issue 5: Collaboration with other professionals on records in a structured data environment. Explanation: Record management specialists should collaborate with database managers, database developers and information architects because of the need for analyzing database data models, recognizing and solving problems early on and records management by design. This can be quite new and challenging if an organization is not yet accustomed to records management by design and a multidisciplinary approach to records management. Does the organization recognize this as a problem or challenge? If so, how does the organization solve this problem or deal with it?	It's a challenge. First, it is necessary to improve the awareness of archives of all staff, especially the archiving awareness of business personnel. At the beginning of business system design, it is necessary to embed the rules of record management. The second is the need to formulate corresponding standards and simplify complex problems, so that business departments have the incentive to make changes. Third, improving the applicability of the record system to make it can be timely and rapidly response in the business department of data mobilization.
	Additional notes	
	Has your organization experienced more specific challenges/problems in the area of records management of data in relational databases (other than those mentioned here)? If so, what are these challenges/problems and can you explain why they are a challenge/problem?	无 None

Row ID (by WG22)	**Case content type**	Content
	Does the organization work with other types of databases than the relational type? If so, what are these types of databases? And are they also in mind as a challenge when applying records management rules to them? If so, has the organization a presumption of the issues and challenges if doing so? If so, what are these supposed issues and challenges?	Yes. In addition to relational databases, we also use web pages, news and other databases with a large number of photos and audio-visual , as well as unstructured databases in some operation systems. The challenge lies in the need to conduct a large number of selection and identification of such data to determine whether the data is managed according to the record management rules and whether it has archival value. If so, how to formulate corresponding rules to screen it.
	Do you have any suggestions for WG22 to address other problems or conduct further research?	No
	Additional notes	

Annex I (informative) By design approach

1 By design and its core concepts from relating designations and the definitions

Following table shows core concepts relating to by design from the existing definitions. By design as approach or as measure and its application scenarios can be divided into four types:(1)processing; (2)goods quality control;(3)information system;(4)software and hardware development.

Table E-1- By design and its core concepts from relating designations and the definitions

code	definition about by design	objects	characteristics	application scenarios	sources
D1	Privacy by design: approach in which privacy is considered at the initial design stage and throughout the complete lifecycle of products, processes or services that involve processing personally identifiable information	approach (to privacy protection)	consider at the initial design stage; throughout the complete lifecycle of products, processes or services	personally identifiable information processing	ISO/IEC TS 27570:2021(en), 3.21
D2	Dataprotection by design: technical and organizational measures designed to implement data protection principles.	measures (to data protection)	technical and organizational measures;	data protection processing	BS EN 17529:2022
D3	Quality by design: systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, employing statistical, analytical and risk-management methodology in the design, development and manufacture of goods.	approach (to goods quality control)	predefined objectives; systematic approach; process control; statistical methodology, analytical methodology, risk-management	goods quality control	ISO/TS 23565:2021(en), 3.16

			methodology		
D4	Archiving by design: during the design, purchase, construction, furnishing and/or modification of information systems , the appropriate measures (among other aspects) are also determined and implemented so that the information in the system becomes and remains permanently accessible .	measures (to ensure information permanently accessible)	appropriate measures to enable the information in the system becomes and remains permanently accessible	information system	The national archives of Netherland, Handbook archiving by design
D5	Archiving by design: during the design or adjustment of information systems, the appropriate measures are taken to ensure that the information becomes, and stays, sustainably accessible	measures (to ensure information sustainably accessible)	appropriate measures to ensure the information becomes and stays sustainably accessible	information system	ISO/TC46/SC11/AHG8 N9
D6	Recordkeeping by design: an approach to recordkeeping that enables it to be built into the design and architecture of information systems, business processes and network infrastructure.	approach (to recordkeeping)	approach enables it to be built into the design and architecture of information systems, business processes and network infrastructure	information system	Tasmanian Archive & Heritage Office, Information Management Advice 17: Recordkeeping by Design - Requirements for Managing State Records in New Business Systems

D7	Security by design: an approach to software and hardware development that seeks to minimise systems vulnerabilities and reduce the attack surface through designing and building security in every phase of the SDLC(system development lifecycle).	approach to (cyber security)	approach to minimise systems vulnerabilities and reduce the attack surface；in every phase of the system development lifecycle	software and hardware development	Cyber Security Agency of Singapore, 2017

2 **Characteristics of by design mapping with ISO 30300:2020**

Following figures shows the general characteristics of by design approach, which include:

- Characteristics of objective: it is considered to have predefined objectives.
- Characteristics of process: it is considered at the initial design stage, throughout the complete lifecycle of products, processes or services, enables it to be built into the design and architecture of information systems, business processes and network infrastructure; in every phase of the system development lifecycle.
- Characteristics of ways: it could be technical and organizational measures; systematic approach; Statistical/analytical/risk-management methodology.
- Characteristics of quality: enable the information in the system becomes and remains permanently accessible; ensure the information becomes and stays sustainably accessible; minimise systems vulnerabilities and reduce the attack surface.

It also shows special considerations of by design for records management in terms of records management process and controls and for structured data environment in terms of predefined set of rules with clear objectives.

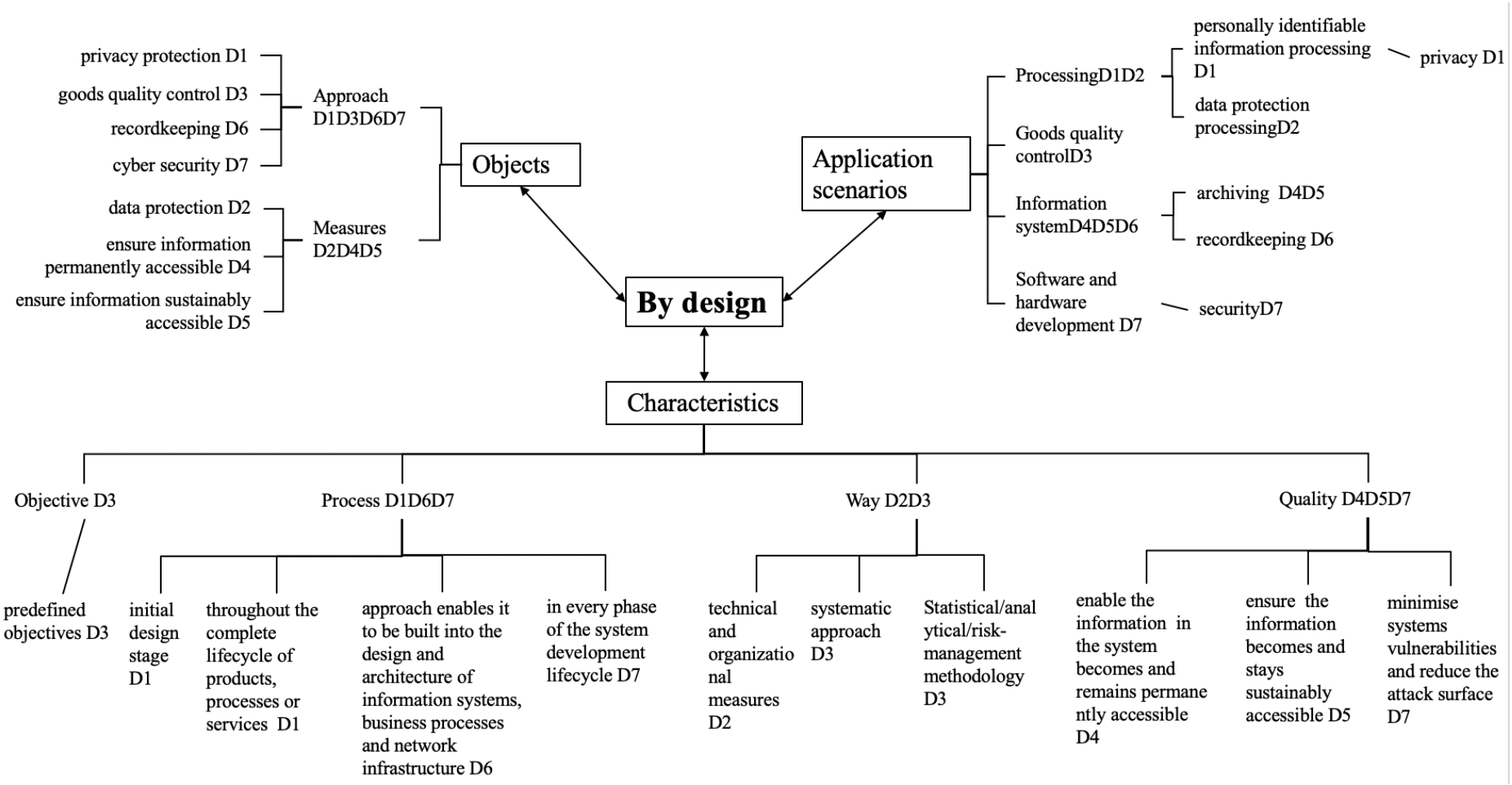


Figure I-1 Core concepts relating to by design in D1-D7

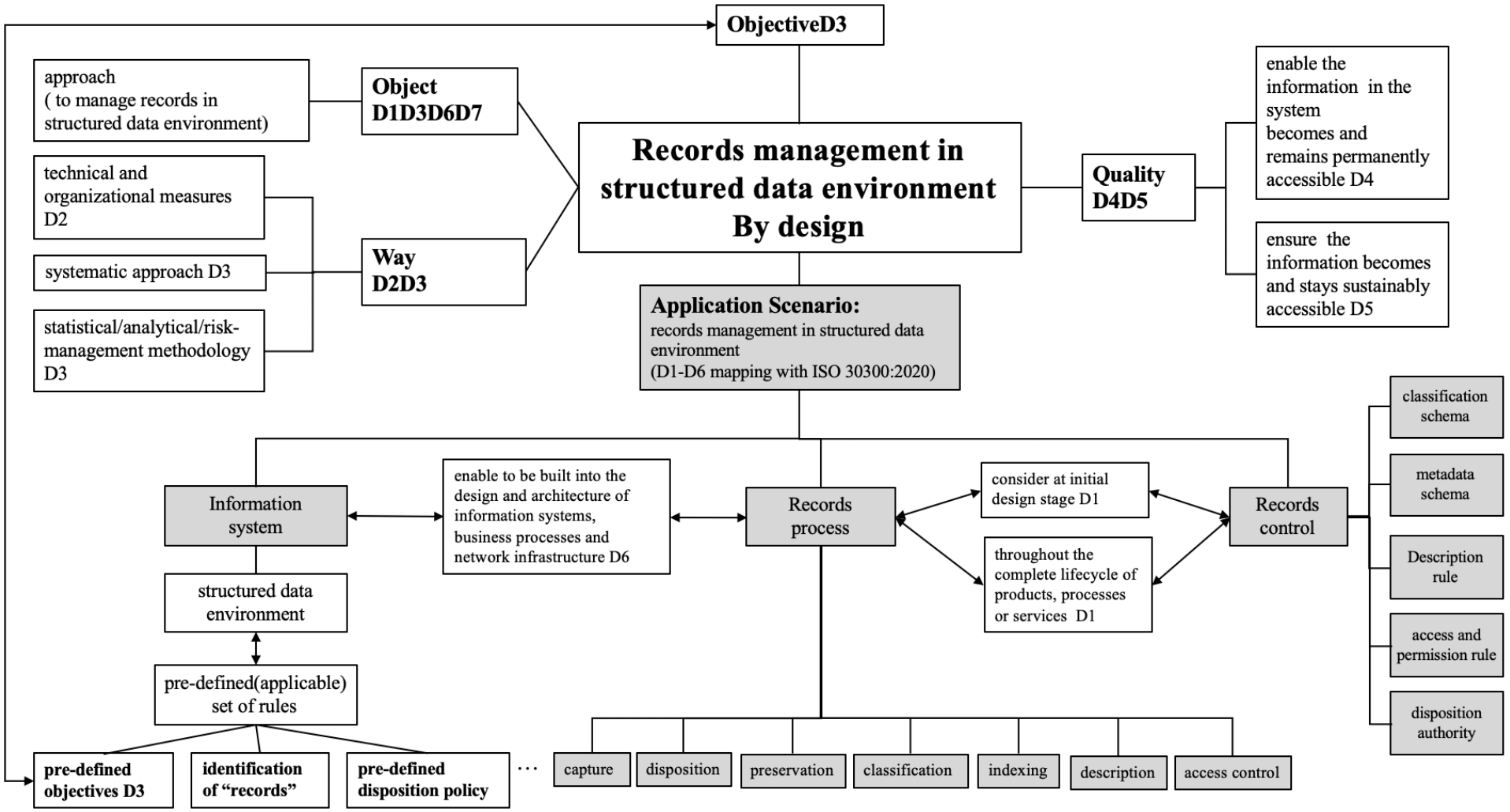


Figure I-2 Core concepts relating to records management in structured data environment by design

3 By design approach means for information management

The by design approach is not only well known in records management, but is also available and used to design solutions for other aspects and associated challenges within the broader field of information management. Therefore, this introduction begins with what the by design approach means for information management.

In the context of this document we distinguish three levels of the by design approach:

a. the by design approach in information management;
b. records management by design;
c. records management by design in structured data environments (and more specifically in relational databases).

4 The by design approach information management

In information management, the by design approach means embedding a specific goal in the requirements and the design of processes, information and information systems. In full, 'by design' means 'by means of design'.

The goal of a by design approach may be to meet the requirements for multiple aspects of information management. These aspects may include the security of information, the public access to information, the privacy of people, the quality and usability of information and also archiving information or records management. From an overall information management perspective, the best and most comprehensive way is to take into account all these goals and requirements together. At that level records management by design only pertains to one of the multiple aspects that need to be taken into account in the field of information management.

5 Records management by design: a definition

Proposal for a definition of *Records management by design*

> Designing and realizing the needed solutions for managing records simultaneously and in conjunction with designing and realizing the environment that will create and receive the records to be managed.

6 Process steps for records management by design

High-level process steps for developing records management solutions by design might be:

1. identify the business process(es) to be considered;
2. identify which records are captured, stored, retained and disposed of;
3. identify the requirements for the capture, storage, retention and disposal of these records;
4. choose where (in which information systems) to store and retain these records;
5. design the solution(s) for capturing, storing, retaining and disposing of the identified records including fundamental system connections;
6. build or purchase the solution(s) for capturing, storing, retaining and disposing of these records and implement them.

7 Records management by design for structured data

Records management of structured data in a relational database begins by analyzing the database's data model to identify which records are captured, stored and retained for reuse as evidence and/or which already stored data are considered and treated as records based on its use and meaning in processes and other information systems. If the choice is made to store records in the database itself, the next step may involve modifying the data model of the database, for example by adding rows and fields for historical and context data. After that the needs and possibilities for disposing of the records in the database are determined, taking into account possible conflicts between records management rules for disposing of these records and the referential integrity rule for relational databases (as described in subclause @@). This may require opting for suboptimal solutions. After that, the needed functions for managing the records in the database will be known and this will determine the application logic of the database system for executing these functions.

8 Multiple axes or dimensions of integrality

An important characteristic of the by design concept is that its core is a comprehensive and integral approach with multiple axes or dimensions such as:

- the time dimension: from policy and project-based design to implementation and daily operation;
- the involvement and collaboration of different disciplines and professionals such as executives, policy staff, information managers, records managers, process and information architects, software developers, database managers and application managers. Collaborating with other disciplines can be new and uncomfortable for records managers, but it is an important prerequisite for a successful by design approach;
- information systems including dedicated document management systems and records management applications and their functionality;
- information systems and their place in the application landscape of an organization or even a wider environment such as a branch or sector and the system connections between these systems. A design that is created in isolation will be rarely a good design and often the result will not even work in practice;

- all elements that are important for the proper operation of information systems, such as policy (purposeful and consistent), government, leadership, finance, knowledge, regulations, standards, primary processes, records management processes, supporting (secondary) processes (application management, incident management, problem management, change management and quality management), software, hardware, the network, instructions, professionals from different fields, and last but not least all digitally working employees and the way they execute their records management tasks. It is important to organize this properly and as a whole.

9 Records management functionality of information systems

If records are digital and are captured, stored, retained and disposed of digitally, then application logic needs to provide the functions to support these processes. If an application (information system) receives or creates information to be used as evidence and captured and stored as records, then this functionality are available for this information. But this functionality needs not to be provided by the application in which the information is received, created and/or stored. Nor does the information has to be stored in the same application used to receive or create it. The bottom line is that the functionality for capturing, storing and managing records is available in the application in which those processes take place. The availability of this functionality can be delivered by the same application or as a service by another application. More on possible configurations in subclause 8.6 Assessing options for deploying functionality in one or more software applications of the ISO standard 16175-1:2020.

Another perspective of how an organization has organized its digital records management functionality is the extent to which records management functions have been automated, thus relieving the burden on digital workers. Maturity levels of this aspect can be:

- all basic functions for managing digital records are provided by a central dedicated system such as a DMS or RMA. Filling this system with records is done by manually performed tasks for all digital workers with support from the central system's digital functions. This is the most basic configuration for how these digital functions can be organized;
- the basic digital records management functions are available for the information received and created by all (business) information systems within them. To this end, these functions are built into the (business) systems or are available as services from a dedicated records management system or other business systems with built-in records management functions. If functions are available as services delivered by other systems, system connections are included to make the functions available;
- the records management functions are more advanced, but still need to be executed or started manually. This may be combined with system connections. An example of this 'level' is an email system with a button to select an email and select or open a file in a dedicated records management system to include the selected email in that file;
- there are sophisticated records management functions including artificial-intelligence and self-learning logic. These functions are performed in the background as much as possible. Digital workers are relieved of their records management duties to the maximum extent possible (without relieving them from their responsibility for proper records management in their own processes).

10 Out-of-the-box thinking

Another perspective of records management by design is out-of-the-box thinking. Laws, regulations and standards are and always will be important for records management, but good design also requires an attitude with an open mind.

Avoid automating 1:1 the way records management has been done in the past for non-digital records, without analyzing what is the best way to do it in the digital world.

Also, avoid automating the way records management has always been automated. Instead, look at the primary processes, records management requirements, new techniques and the behavior of the people who have records management tasks. Analyze this as a whole and consider how records management can be done more effectively and efficiently. In doing so, also look beyond the current situation and consider future developments.

Sometimes the conclusion is that the software meets the right requirements, but the style and behavior of managers and employees are a problem. If employees are unclear about how to properly perform their records management tasks, better instructions and documentation may be a solution. If they do know how to perform their tasks properly, but do not have enough time to do so, more realistic planning of capacity and tasks are considered. If executives are not convinced of the importance of records management and do not direct and monitor the execution of it, leadership is the problem. Records management is more than providing rules and tools.

11 General preconditions for a successful by design approach

For successful implementation of the by design approach it is embedded in the policy and governance of both records management and information and data management.

Another important precondition is focusing on process and information architecture and the availability of expertise in this field. Knowledge and experience from this field can contribute to the integrated approach and the overview of the information and information technology landscape required for a successful by design approach. The ISO document 'NPR-ISO/TR 21965: 2019 Information and documentation - Records management in enterprise architecture' can be helpful when introducing the discipline process and information architecture in the records management field.

Bibliography

[1] ISO 704:2022, Terminology work — Principles and methods

[2] ISO 16175-1:2020, Information and documentation — Processes and functional requirements for software for managing records — Part 1: Functional requirements and associated guidance for any applications that manage digital records

[3] ISO/TS 16175-2:2020, Information and documentation — Process and functional requirements for software for managing records — Part 2: Guidance for selecting，designing，implementing and maintaining software for managing records

[4] ISO 19626-1:2020, Processes, data elements and documents in commerce, industry and administration - Trusted communication platforms for electronic documents — Part 1:

[5] ISO 23081-1:2017, Information and documentation — Records management processes — Metadata for records — Part 1: Principles

[6] ISO 23081-2:2021, Information and documentation — Metadata for managing records — Part 2: Conceptual and implementation issues

[7] ISO 23952:2020, Automation systems and integration — Quality information management (QIF) -An integrated model for manufacturing quality information

[8] ISO 30300:2020, Information and documentation — Records management — Core concepts and vocabulary

[9] ISO/IEC TR 38505-2:2018, Information technology — Governance of IT — Governance of data — Part 2: Implications of ISO/IEC 38505-1 for data management

[10] Information Management Advice 17: Recordkeeping by Design-Requirements for Managing State Records in New Business Systems (2015)[2022-01-01] https://www.informationstrategy.tas.gov.au/Records-Management-Principles/Document%20Library%20%20Tools/Advice%2017%20Recordkeeping%20by%20Design.pdf.

[11] Kevin Williams. A Design Science Approach to Deletion in Transactional Processing Relational Databases. (researchgate.net).2016