The Internet of Things – Hype and reality: Crucial issues about Internet of Things technology and 'content' of human-machine communication

CHRISTIAN GALINSKI
Infoterm (International Information Centre for Terminology)

KEYWORDS: Internet of Things (IoT), 'things', eApplications, 'smart' applications, unstructured and structured content, microcontent, interhuman (H2H) communication, information object, content object vs. content meaning, codings, non-verbal concept representations, content interoperability, metadata, masterdata, controlled language, controlled vocabulary, persons with disabilities (PwD), augmentative and alternative communication (AAC)

1. THE RELATIONSHIP BETWEEN INTERNET OF THINGS AND 'TERMINOLOGY' OR 'STRUCTURED CONTENT'

This contribution aims at putting 'content' into perspective with Internet of Things (IoT) developments, achievements and challenges, its impacts on human society and of course on interhuman (H2H) communication which is increasingly supported by information and communication technologies (ICTs). Content can take many guises, such as texts, content in other modalities than spoken or written (e.g. other visual or acoustic signs and even information communicated in tactile or haptic form). With respect to content, the focus of this contribution is on structured content at the level of lexical semantics – also called 'microcontent'. Terminologies are the most important kind of microcontent in specialized communication whether in spoken or written form. However, increasingly non-verbal modalities are necessary in spoken or written specialized communication, and are appearing even in terminological data.

Still in recent past, major topics related to computer technology were information and knowledge societies, digitalization (e.g. digital library), office automation, mechatronics and the like. Today's buzzwords are Artificial Intelligence (AI), Semantic Web (SemWeb), Big Data, cloud computing, robotics (referring to cyber-physical systems), Industry 4.0 – not to mention many aspects of industry and society carrying the attribute ‘smart’ in front, such as smart cities, smart transport, smart homes, smart hospitals, smart buildings etc. Some of this ‘smartness’ is already implemented to quite an extent in the form of activities with an ‘e’ in front of it, such as eBusiness and eCommerce, eHealth, eGovernment, eLearning, and new ‘eApplications’, such as eAccessibility and eInclusion, eBanking, eProcurement, eInvoicing, etc. ***Can any of the above be useful without content?***

The IoT has the potential of combining all technologies and applications mentioned above in one large cyber-universe where data – of course also in the form of ‘content’ – are created not only by and for humans, but also by (and for) ‘things’ which are increasingly used for purposes hitherto not imagined. Therefore, terms frequently used relating to the ICTs and the IoT in particular, do not yet represent stable concepts.

So far, the focus of IoT developments is on technology, while the aspects of H2H communication – even, if ICT-supported – is largely neglected in pertinent literature.

2. DEFINITIONS

2.1. The Internet of Things (IoT)

IERC (http://www.internet-of-things-research.eu/about_iot.htm), the IoT European Research Cluster, defines IoT as “a dynamic global network infrastructure with self-configuring capabilities based on standard and interoperable communication protocols where physical and virtual ‘things’ have identities, physical attributes and virtual personalities, use intelligent interfaces and are seamlessly integrated into the information network” (see Vermesan e.a. 2011). As a vision (with a certain degree of hype), early IoT conceptions may have been purely technology-oriented futuristic visions, but over the last couple of years more and more developments reveal increasing market relevance. It looks as if the IoT will dramatically change human environment – worldwide – over the next years with respect to traffic and transport, housing and living, health technology and services, developments in agriculture and food industry, education, communication, etc. However, the aspect of ‘content’ more often than not is omitted from the discussions about these technical developments.

ITU (ITU-T Y2060:2012, 3.2.2) defines the IoT as:

> "A global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies.
>
> NOTE 1 – Through the exploitation of identification, data capture, processing and communication capabilities, the IoT makes full use of things to offer services to all kinds of applications, whilst ensuring that security and privacy requirements are fulfilled.
>
> NOTE 2 – From a broader perspective, the IoT can be perceived as a vision with technological and societal implications."

According to Dave Evans (2011 – CISCO Internet Business Solution Group, IBSG), the Internet doubles in size every 5.32 years. This somehow resembles Moore's Law forecasting the number of transistors in highly integrated circuits (semiconductors) approximately doubling every two years. Moore's Law ultimately resulted in the 'miniaturization' of electronic components – and thus the electronic devices built with these semiconductors became smaller while their functionalities were ever-increasing. The Internet's growth just follows the need to transport the exponentially increasing amounts of data.

According to D. Evans (ibidem) under the IBSG's technological and commercial perspective, the IoT simply is the point in time around 2008/2009 when more 'physical things' (such as electronic devices) were connected to the Internet than people. Today over 75% of world population is connected to the Internet in one way or other. Investigations show that the number of devices (especially electronic devices) connected to the Internet grew and continues to grow faster than that of humans connected to the Internet by means of computers or mobile phones. Thus, it could well be that already between 25bn and 50bn devices are connected to the Internet, today.

This could never have happened without the rapid development of mobile technologies not only for ICT-supported H2H communication, but also for data retrieval and other use of data. In addition, other 'physical things' – such as surveillance cameras, sensors for all kinds of purposes in industry etc. – entered the scene as 'originators' of data. Ericsson (2016) "conservatively" estimates the number of connected devices in the form of tangible facts at nearly 30bn in 2021, including only a bit more than 10bn devices for ICT-supported H2H communication. The latter number grows linearly while the number of devices for M2M communication grows much faster. D. Evans (ibidem) "optimistically" estimates 50bn devices connected to the Internet.

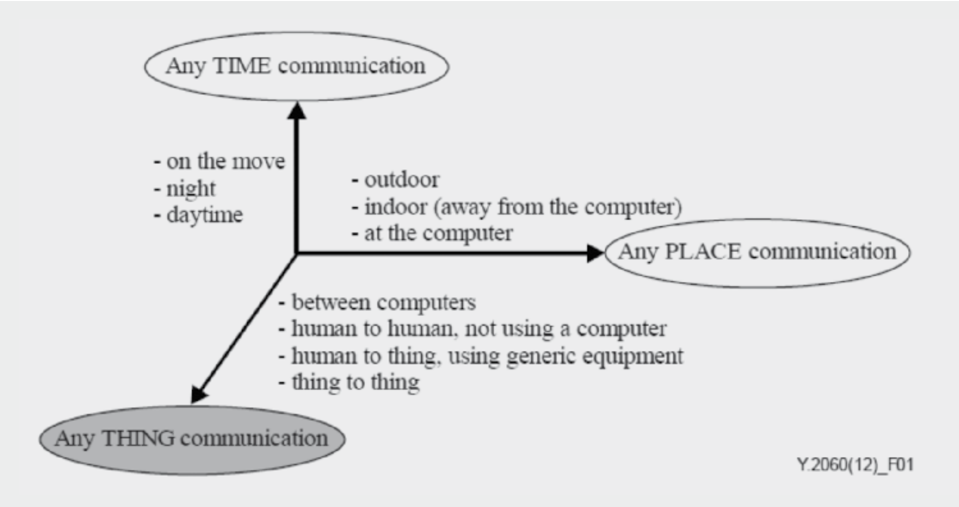
So far, the Internet arguably has not changed very much, fundamentally speaking. However, the IoT can be seen as the first real evolution of the Internet. It has already made the Internet sensory (temperature, pressure, vibration, light, moisture, stress), allowing us to become more proactive and less reactive.

ITU (ITU-T Y2060:2012, 6.1) states:

> "Through the exploitation of identification, data capture, processing and communication capabilities, the IoT makes full use of "things" to offer services to all kinds of applications, whilst ensuring that security and privacy requirements are fulfilled.
>
> NOTE – The IoT is expected to greatly integrate leading technologies, such as technologies related to advanced machine-to-machine communication, autonomic networking, data mining and decision-making, security and privacy protection and cloud computing, with technologies for advanced sensing and actuation.
>
> As shown in Figure 1, the IoT adds the dimension "Any THING communication" to the information and communication technologies (ICTs) which already provide "any TIME" and "any PLACE" communication."

This can be exemplified by the following graph:

Figure 1: The extended dimensions of the IoT (ITU-T Y2060:2012, 6.1)



2.2. The Internet vs. ‘internet’

There are – even among ICT experts – misconceptions about ‘**internet**’ as made explicit by UNESCO (2017):

> “**Internet:** the worldwide public network of computer networks that provides access to a number of communication services including the World Wide Web (WWW) and carries e-mail, news, entertainment and data files.
>
> NOTE 1: The Internet (with capital ‘I’) refers to the huge global public network which also runs the World Wide Web. Other internets – also being networks of computer networks – are written with lower case ‘I’.”

Something similar also applies to the World Wide Web (WWW, also called Web or The Web) which can be described as (UNESCO ibidem):

> “the most popular of all Internet services and applications (often used interchangeably with the Internet) that provides users with the ability
> – to access information and services while connected to the Internet,
> – to publish information, and
> – to offer services that can be accessed by anybody else in the Internet.
>
> NOTE 1: The World Wide Web is one of the biggest services running on the Internet. The multitude of other services implemented over the Internet includes e-mail, file transfer, voice over IP (VOIP), digital TV, remote computer control, newsgroups, and online games. All of these services can be implemented on any internet, accessible to network users.”

Today the Internet as well as the WWW are largely based on standards, first of all the Internet protocol suite comprising:

> /a/ “computer networking model and communication protocols combining numerous technical standards used in the Internet and other networks based on the Internet Protocol Suite
>
> NOTE 1: The Internet Protocol Suite is the most important set of protocols of the Internet (collectively called Transmission Control Protocol and Internet Protocol Suite – TCP/IP) and allows large, geographically diverse networks of computers to communicate with each other quickly and economically over a variety of physical links. It is maintained and developed by the Internet Engineering Task Force (IETF).” (UNESCO ibidem)

The Internet Protocol (IP) can be described as:

> “main internetworking technical standard underlying the Internet that specifies how data is moved through it based on three principles: packetswitching, end-to-end networking, and robustness
>
> NOTE 1: The IP is implemented in two versions, IPv4 and IPv6 both based on technical standards of which different implementations exist. It is often

> used interchangeably to the Transmission Control Protocol and Internet Protocol Suite (TCP/IP), although the IP covers the first – but still most important – networking protocols defined in the TCP/IP." (UNESCO ibidem)

It is necessary to understand the differences between the *Internet* and the *WWW* – terms that are often used interchangeably. The Internet is the physical layer or physical network made up of switches, routers, and other equipment. Its primary function is to transport information from one point to another quickly, reliably, and securely. The WWW, on the other hand, is an application layer that operates on top of the Internet. Its primary role is to provide an interface that makes the information flowing across the Internet usable. Already quite early the WWW was called the "Web of Things".

2.3. What are 'things' in the IoT?

ITU (ITU-T Y.2060:2012, 3.2.3) defines 'things' as follows with regard to the IoT:

> "object of the physical world (physical things) or the information world (virtual things), which is capable of being identified and integrated into communication networks".

Things have associated information which can be static or dynamic. Equalizing 'things' with 'objects' conforms to traditional science theory, especially ontology. In terminology science (based on science theory) 'object of the physical world' is also called 'material object, (or physical object, real object, concrete object, and the like). In this connection ITU defines 'devices' as physical objects being

> /a/ "piece of equipment with the mandatory capabilities of communication and the optional capabilities of sensing, actuation, data capture, data storage and data processing".

In terminology science 'virtual thing' is called 'immaterial object' (or imagined object, abstract object, and the like). According to ITU virtual things

> "exist in the information world and are capable of being stored, processed and accessed. Examples of virtual things include multimedia content and application software." (ITU-T Y.2060:2012, 6.1)

This definition neglects the fact that 'content' – especially microcontent – is a universe of 'virtual things' in itself. Although software is needed to handle it, serious interoperability issues arise, if content is put into the same category as 'application software' – not least due to the fact that content and software have totally different life-cycles and requirements concerning project management and quality aspects just to mention a few.

Besides, the philosophical question here is, whether all 'things' are only "information objects", or whether they can be much more, if considered as (by humans) 'perceived objects' or individually 'conceived objects' in human reality. A radical reductionist view of everything being nothing else than 'information objects' from a technological point of view could turn out to be a big fallacy.

Content can also be regarded as 'things' (see also sub-chapter 2.5). Micro-content in terms of 'content objects' falls under 'physical things', the meanings of the content entities under 'virtual things'. Both necessitate metadata for their description and identification.

2.4. Interhuman communication

Ultimately any content has to be or must potentially be understood by humans.

Interhuman (H2H) communication is unique for its extensive use of 'language' – in all its meanings from 'langage' via 'langue' to 'parole' in French, according to the ideas of Ferdinand de Saussure. However, language and languages are not the only means of communication, neither is language only confined to spoken and written. The channels (or modalities) of interhuman communication can be visual, auditory, tactile (such as in Braille) and haptic, olfactory, electromagnetic, or biochemical. Besides, communication may take place at conscious and unconscious levels. Generally speaking, it can be defined as:

> "the act of conveying intended meanings from one entity or group to another through the use of mutually understood signs and semiotic rules". (Wikipedia – https://en.wikipedia.org/wiki/Communication)

The essential elements of this definition from the perspective of H2H communication are:

- intended meanings,
- conveyed among humans (whether individuals or groups),
- mutually understood signs/symbols and rules.

Others by comparison, emphasize the social interaction aspect:

> "Communication is the essence of human interaction and learning.
> The nature of communication is dependent on interaction between two or more individuals and understanding is constructed through that interaction". (ISAAC – https://www.isaac-online.org/english/what-is-aac/what-is-communication/)

ISAAC continues:

> "Communication is a basic human right and essential to our quality of life as a social species. As human beings, we use communication to: relate to others, socially connect, greet, call attention, share feelings, express an opinion, agree, disagree, explain, share information, question, answer, tease, bargain, negotiate, argue, manipulate, compliment, comment, protest, complain, describe, encourage, instruct, provide feedback, show humor, discuss interests, be polite, make friends, express interest or disinterest, etc."

For the above purposes and taking the social and interaction aspects of communication into account, non-verbal communication is well-known as major communicative means for supporting language in interhuman communication or even replacing it in certain communication acts. It is reckoned to represent two-thirds of all communication and involves conscious and unconscious processes of encoding and decoding (of meanings/messages conveyed). It can include visual cues, such as body language (e.g. postures), gestures and mimics, auditory cues of voice in speech (paralanguage), such as voice quality, rate, pitch, volume, speaking style, as well as prosodic features, such as rhythm, intonation and stress. Needless to mention that much of this can be highly culture and language dependent. (See Wikipedia – https://en.wikipedia.org/wiki/Communication)

Much of the study of nonverbal communication has focused on interaction between individuals, where it can be classified into three principal areas: environmental conditions where communication takes place, physical characteristics of the communicators, and behaviours of communicators during interaction. Written communication – especially under a 'localization' (L10N) perspective, such as in translation and technical writing/documentation – can also have non-verbal elements: (a) linguistic ones, such as personal style, and (b) non-linguistic ones, such as photos, images, drawings, graphs, diagrams, embedded acoustic signs (e.g. in eBooks), etc. Needless to mention, much of this may be highly culture and language dependent, and also even social group or language register dependent.

Taking the above-mentioned communication channels into consideration, the traditional definitions of H2H communication needs extension, as they largely focused on 'language' in a narrow sense. In the field of child raising and education, as well as in augmentative and alternative communication (AAC) for persons with communication disorders of different kind or origin, 'language' and H2H communication are used and must be used in a much broader sense. eLearning tools and assistive technologies are taking this into account – thus often developing apps or tools which are useful for everybody.

Increasingly all of the above-mentioned interhuman communication aspects can be supported by ICTs. Ultimately 'things' in the IoT will or – if need arises – may even have to be presented to human beings. Therefore, negligence of the complexity of the nature of H2H communication will not only curb efficiency and effectiveness of the technologies applied, but also lead to problems of quality of information, liability for wrongly presented information (to certain target groups or in general), corporate social responsibility (CSR), legal responsibility of all sorts, etc.

There is a series of standards which is based on a most comprehensive, though condensed combination of fundamental theories on interhuman communication and behaviour based on: ITU-T X.1081:2011 *The telebiometric multimodal model – A framework for the specification of security and safety Aspects of telebiometrics*. Experts of H2H communication probably may not find the link to ITU-T X.1081:2011 and related standards, because the titles of these standards and the names of the committees that work on them seem to have no connection to H2H communication. Experts of biometrics would hardly think of the H2H communication content and its semantic, social and cultural dimensions – unless these are in some way or other 'computable'. Today, however, nearly everything in terms of interhuman communication is or is going to be ICT-supported anyhow.

The ITU-T Recommendation was followed up by IEC 80000-14:2014 *Telebiometrics related to physiology* as a part of the ISO and IEC 80000 harmonized series and ITU-T Recommendation X.1082:2010 *Telebiometrics related to human physiology*. Standardizing work on the ISO and IEC 80000 harmonized series is shared by ISO/TC 12 and IEC/TC 25 both named *Quantities and units*, while related topics are dealt with by several other committees in ISO/IEC-JTC 1 *Information technology* and ISO/TC 68 *Financial services*, as well as several other standards developing organizations (SDOs), such as OASIS, ICAO, Interpol, etc. – not to mention numerous US agencies. ITU-T X.1081:2011 paved the way towards an interdisciplinary approach for standardizing also H2H communication aspects (including the respective methodologies, and – if necessary – also pertinent content repositories).

Under the new perspectives outlined above, parts of the ISO/IEC 80000 series of standards may become very useful with respect to ICT-supported H2H communication and content aspects in the future IoT.

2.5. Content

Even from a standardization perspective, more often than not the terms data, information and content are used interchangeably. Information processing theory argues that the physical world is made of information itself. Under this definition, 'data' is either made up of or synonymous with physical information. Data as an abstract concept can be viewed as the lowest level of abstraction from which information and then knowledge are derived. "Typically information is defined in terms of data, knowledge in terms of information, and wisdom in terms of knowledge". (Rowley 2007: 163)

Generally speaking, information and data have much in common and are often used as synonyms in pertinent literature. In ISO standards, 'data' is defined in relation to information and 'information' defined in relation to data – a typical case of circular definitions. Bob Boiko (2004: 5) concludes:

> "From the user's perspective, information is all content, while from the computer programmer's perspective, it is all data."

In the course of the development of the eApplications, 'content' has become quite a fuzzy term. In international standards, it is among others defined as follows:

- ISO/IEC 15938-5:2003 (3.3.2.9): *a representation of the information contained in or related to multimedia data in a formalized manner suitable for interpretation by human means. Content refers to the data and the metadata*;
- ISO/IEC 24800-3:2010 (3.1.5): *data and the associated metadata*;
- ISO 24531:2013 (4.11): <XML> *all data between the start tag and end tag of an element*.

The above shows that content and metadata are closely related to each other. The metadata applied or necessary to be applied are by no means confined to linguistic content, whether spoken or written. Given the increasing ubiquity and pervasiveness of ICTs, aspects of H2H communication beyond 'language' are gaining importance – also seen under requirements of 'content interoperability'.

ISO 9241-151:2008 defines 'content' (in the meaning of web content referring to the web user interface) as "set of content objects" (item 3.4) and 'content object' as "interactive or non-interactive object containing information represented by text, image, video, sound or other types of media" (item 3.5). Thus, from a technical point of view, content management takes content as 'content objects' depending on the media they require as:

- text (i.e. textual data, incl. all kinds of alpha-numeric data),
- image (graphical data),
- video (incl. multimedia data?),
- sound (audio data),
- other types of media.

"Other types of media" indicates that other 'modalities' (defined in ISO 5492:2008, 2.11, as "sensations mediated by any of the sensory systems, for example auditory, taste, olfaction, touch, somesthesis or visual modality") are not excluded. On the one hand, text, image and video (without sound) refer to the visual modality; on the other hand, different modalities today may occur imbedded in texts or documents, e.g. in eBooks. This reveals that *content* is not satisfactorily defined, if only from a technical point of view. From the point of view of semantics, the above definitions are insufficient.

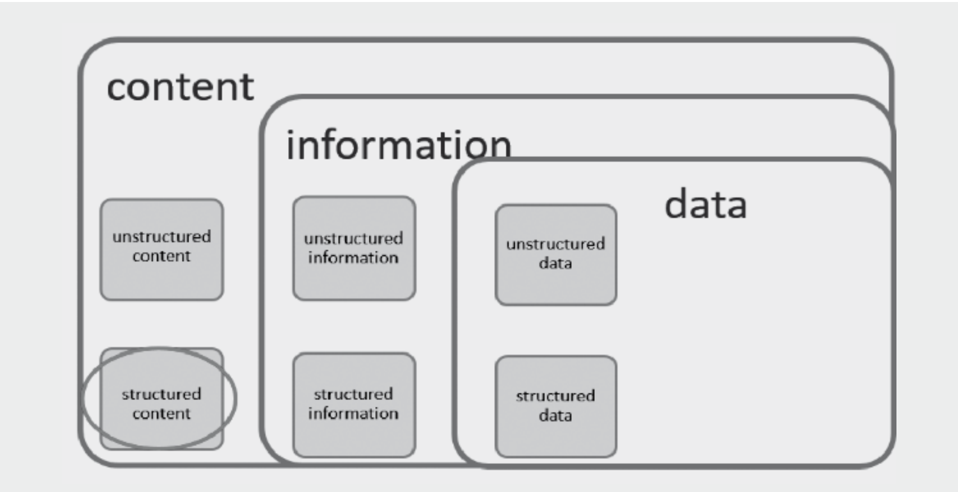
ISO/IEC 12785-1:2009 defines 'content' (item 3.7, in the meaning of 'LET content' from a learning technology and content packaging point of view) as "logical unit to represent usable (and reusable) information contained in or related to learning, education, and training (LET) data in a formalized manner suitable for interpretation by human means". It is further explained by the "EXAMPLE: In the instructional context, content can be web-based instructional materials". Thus, on the one hand a content entity is a logical unit representing *usable (and reusable) information* possibly contained in larger entities. On the other hand, "data in a formalized manner suitable for interpretation by human means" reveals that content "represents information" which must ultimately be "suitable for interpretation by human means".

From the definitions in current standards and discussions in pertinent literature it can be deduced that

- data need additional metadata to become information,
- information needs additional metadata to become content.

Summarizing, the relationship data – information – content can be exemplified as follows:

Figure 2: The relationship data – information – content



In this connection, it is usually neglected that the set of metadata needed for structured content is quite different from those for unstructured content, although the two sets overlap to some extent. Besides, metadata – even if skillfully applied and based on standards – do not yet guarantee interoperability. Here lies the reason for the need for 'content interoperability'. (See chapter 4)

Standardization of methods for managing different kinds of microcontent is a key for interoperability. This also applies to metadata, which formally speaking, are a kind of microcontent, too. However, it also needs more or less standardized 'good practices' with interoperability in mind in order to achieve 'content interoperability'.

From the above it becomes clear that 'data, information and content' convey different – though closely related and overlapping – meanings, although they are commonly used interchangeably. If "information is data that has been processed in such a way as to be meaningful to the person who receives it" (Riley 2012: s.p.), 'content' becomes the 'representation of information meaningful to the person who receives it'. The latter comprises also the communication aspect – both in terms of technical communication as well as of H2H communication – which is particularly important in eLearning and AAC. Jim Riley (ibidem) continues "Note the two words [...] processed and meaningful. It is not enough for data simply to be processed. It has to be of use to someone – otherwise why bother?!"

3. CONTENT FROM A TERMINOLOGICAL PERSPECTIVE

3.1. The multifaceted nature of content

"There is a proliferation of content – not only in the Internet /at large/, but also at organization level [...] Therefore, enterprises struggle to integrate content resources or at least make them interoperable. In line with this process, content should be analysed and low-quality content identified in order to be deleted or improved. This refers also to a large degree to structured content – here content entities at the level of lexical semantics, viz. microcontent, comprising linguistic and non-linguistic representations of concepts." (Galinski & Giraldo Perez 2014: 405)

'Content' has acquired new meanings and dimensions under the conception of 'eContent'. Originally derived from 'electronic content', eContent is defined as 'digital content' that can be transmitted over a computer network such as the Internet. (The Computer Language Company Inc. 2017-09-03) This definition lacks a 'purpose' (e.g. intended application), as the development of eContent is not a goal in itself (not even when applied for fine arts). In any case, the definition implies that eContent is developed in a computer-assisted way (which does not exclude conventional output) and, therefore, must conform to a minimum of standards, in order to be transmissible over the Web.

Besides, in the course of increasingly using eContent as a commodity, commercial and legal aspects (*inter alia* digital rights) are becoming increasingly important for the distribution and use of eContent. Pertinent technical standards as well as legal norms have been developed at international, regional and national levels – not to mention new methodology standards for accommodating commercial and legal requirements in the data models for eContent. As nearly any traditional content can be digitized with technical means and turned into eContent, 'content' today stands for traditional content as well as modern digital content and is used as such in this contribution. (See also Galinski & Giraldo Perez ibidem)

The above-mentioned proliferation of content – especially at organizations' level – can lead to very high costs if content is not integrated in terms of *system integration* as well as *content integration*. When content is integrated in large organizations without being fully interoperable, and then made accessible on the WWW, it adds to the emergence of 'big data', which is a great amount of unstructured and structured content. In this respect the WWW can be considered as the biggest resource of big data. Webopedia (s.a.) refers to 'big data' as "a buzzword, or catch-phrase, used to describe a massive volume of both structured and unstructured data that is so large that it is difficult to process using traditional database and software techniques". The IoT with its combination of eApplications and smart applications depending on highly reliable content cannot function on the basis of big data.

In any case, the IoT – being regarded as the first real evolution of the Internet – will create and use more content than any development before. New forms of content may arise and the amounts of content will grow exponentially. However, one should always remember J. Riley (2012: s.p.): "It is not enough for data simply to be processed. It has to be of use to someone – otherwise why bother?!"

3.2. Different types of content

As explained above, there are many facets of content. However, whether structured or unstructured, it can represent general knowledge (or world knowledge) or specialized knowledge (i.e. domain- or subject-specific knowledge), though there is no clear-cut borderline between the two. Some can be language-dependent, nearly all kinds of content are, or could be culture-dependent.

It can soon be recognized that there are:

- different kinds of unstructured content entities of all sizes from small to very large:
  - linguistic, or non-linguistic or combined linguistic-nonlinguistic ones,
  - entities of structured content explicitly comprised or embedded in unstructured content;
- different kinds of structured content – or better microcontent entities – of different comprehensiveness, depending on the number of metadata referring to 'meaning' (i.e. semantics) or the amount of 'context' (or rather 'co-text') added:
  - linguistic, or non-linguistic or combined linguistic-nonlinguistic ones,
  - systematically structured or nonsystematically structured ones,
  - being able to stand alone as intelligible 'knowledge units' or being unintelligible, if extracted out of 'context',
  - comprising elements of unstructured content,
  - fit or not fit for being used explicitly in unstructured content or for making unstructured content better usable and reusable, processable, manageable, etc.,
  - having different roles within a collection of structured content or in a combination of collected resources of structured content
    - for structuring collections of structured content,
    - to improve the accessibility to information in unstructured content;
- many ways – not least due to the fast-emerging language technologies – to combine different kinds of structured and unstructured content within themselves or among each other.

In the light of the importance of content to be used in and through the Internet, Infoterm launched the "Recommendation on software and content development principles 2010" that defines as basic requirements for the development of fundamental methodology standards concerning semantic interoperability the fitness for:

- multilinguality (covering also cultural diversity),
- multimodality and multimedia,
- eInclusion and eAccessibility,
- multi-channel presentations,

which have to be considered at the earliest stage of:

- the software design process, and
- mata modeling (including the definition of metadata),

and hereafter throughout all the iterative development cycles. (MoU/MG 2012)

> "The above Recommendation inevitably requires a higher degree of structural complexity, which has to be coped with by a higher degree of granularity of the data model. It may require additional URIs for the different language parts of the individual entries of structured content. Ultimately the time-honoured principle of term autonomy will have to be extended towards representation autonomy in the field of terminology management. This is anyhow necessary when re-purposing entities of structured content for instance for eLearning purposes. Because of the paramount phenomena of quasi-equivalence of concepts between languages links between parts of entries or even certain fields in a given entry to other entries (or parts thereof) will be necessary. As experienced in localization (such as in the field of technical documentation), the above even applies to non-linguistic representations due to cultural diversity factors." (Galinski & Giraldo Perez 2014: 415)

3.3. Metadata for content

There are different metadata standards for each field of application (e.g. museum collections, digital audio files, websites, etc.). Describing the content and context of data or data files increases their usefulness. But from a content perspective – especially regarding structured content – it must be borne in mind that the set of metadata needed for structured content entities and unstructured content entities may be quite different, although they overlap to some extent:

- In the case of unstructured content – considered to be a 'content resource' – the metadata comprise those related to:
  - the formal description and identification of each 'content object' (including source description and identification),
  - the data management of the 'content objects',
  - the knowledge-level of the 'content object', comprising
    - classification, typology, taxonomy, and similar means of categorization,
    - descriptors of documentation thesauri or other kinds of 'controlled vocabularies',
    - free indexing methods,

  (or a combination of the above);

- In the case of structured content – considered to be basic semantic entities – the metadata would comprise those related to:
  - the formal description and identification of each microcontent entity (including source description and identification),
  - the data management of the microcontent entity,
  - the knowledge-level of ‘microcontent objects’ similar to unstructured ones,
  - the ‘semantics’ of the microcontent entity, comprising
    - linguistic and non-linguistic information,
    - explanations or definitions or other kinds of denomination,
    - usage aspects of all sorts,
    - quality/reliability indications, etc.

In addition, the sets of metadata for different kinds of resources of structured or unstructured content differ, although they usually have some metadata in common (e.g. information on the source). Each kind of content has a specific minimum (but extendable) set of metadata which allows to process, use, reuse and re-purpose them. Even if only slight differences between the metadata sets are not respected, interoperability issues may arise.

In this connection, the complexity of a content entity in general may be determined in terms of:

- quantity of information covered,
- granularity of metadata applied,
- amount of explicit context and co-text provided,
- number of content types comprised,
- degree of cognitive processing required, etc.

The metadata applied – at least in the past – mostly refer to formal aspects, i.e. to syntactic structuring rather than ‘semantic structuring’. (See also: OASIS s.a.) The more information on ‘meaning’ is required, the more metadata related to semantics are necessary – especially in microcontent of all sorts with explanations, definitions, underlying conceptual structures, etc. – such as for terminological data. The above is also supported by Ben Steichen & Vincent Wade (2010: 5):

> “The heterogeneous support content that is available for software products needs to be transformed to a semantically richer form in order to allow reasoning, adaptation and personalisation across it. <...> Semantic Web technologies such as ontologies represent an opportunity to base such structuring and

> markup on. The different types of content can be broadly categorised by their amount of existing metadata and structure. Consequently, different types of usage can be drawn from each: whereas highly structured content (such as technical documentation) can be used to derive an ontology of the knowledge domain, unstructured content (such as forum posts) can be marked up in order to provide querying users with a larger range of problem solutions."

In this connection, a paradox has been pointed out by Sabine Zumpe & Warner Esswein (2002: 246) referring to the fact that a high degree of structural complexity in the form of higher granularity represented by more (incl. more different kinds of) metadata in fact reduces complexity from the point of view of information processing:

> "Highly structured knowledge bases permit a low degree of complexity to be managed by the information system. In contrast the degree of complexity is very high in weakly structured knowledge bases, whereby the user does only need a small amount of information about the meta-structure."

The appropriate application of metadata approaches alone does not yet guarantee content interoperability. Metadata came into use, when databases had to be designed. The original definition "data about data" (ISO 19115:2003, 4.8) can still be found in several standards. Later definitions are: "data [information] that provides information about data" (Merriam-Webster) or "information about a resource" (ISO 19115-1:2014, 4.10). Different definitions of metadata abound.

In order to apply ICTs for the handling and processing, as well as maintenance of any kind of microcontent, J. Riley (2017: 6) distinguishes four types of metadata:

- descriptive metadata (for indicating the role or function of the microcontent entity as well as search words/terms to facilitate the search for them),
- administrative metadata (incl. technical metadata for decoding and rendering files, preservation metadata for long-term management of files, rights metadata for intellectual property rights attached to content),
- structural metadata (relationships between microcontent entities, between elements of microcontent entities, different microcontent resources etc.),
- markup languages and harmonized data modelling approaches (to support a consistent methodology with content interoperability in mind).

These various categories of metadata support different target uses in information systems. As can be seen, semantic aspects are mentioned, but only marginally.

Metadata can be stored and managed in a database, often called a metadata registry or metadata repository. However, without context and a point of reference, it might be impossible to identify metadata just by looking at it. (Wikipedia – https://en.wikipedia.org/wiki/Metadata) As metadata description require metadata again, metadata as such are not different from general microcontent – only their distinct role or function for certain uses must be marked and respected.

3.4. Masterdata

The increasing need for information interchange via the Internet called for standards harmonizing minimum sets of metadata for various eApplications, such as Dublin Core (today ISO 15836:2009) for libraries, or data categories for terminological data (ISO 16642:2003).

In eBusiness and eCommerce such harmonized sets of metadata today are often called masterdata (e.g. referring to customer data, product data) which are also used to combine different kinds of data categories for certain transactions and additional metadata as well as factual data and related routines (e.g. price calculation), when needed.

Similar efforts are taking place in more and more standardization fields. Therefore, the number of new standards referring to content increases exponentially. These standards can refer to the standardization of:

- microcontent entities as such (e.g. standardized terminologies, quantities and units, symbols of all sorts, codings of countries, harbours, airports, containers, currencies, etc.),
- metadata pertaining to microcontent entities,
- masterdata pertaining to objects of the physical world (physical things) or to the information world (virtual things),
- links to respective metadata and master data registries or repositories.

It has become a common conviction of experts in the field that "standardization of content reduces the complexity of business processes".

However, there are many standards and standardized approaches concerning metadata and masterdata in various eApplications which are a hindrance to ‘content interoperability’. Although harmonization efforts are under way, there is an urgent need for intensifying harmonization and enforcing it. In order to help to improve this situation, Infoterm in co-operation with the Association for the Advancement of Assistive Technology in Europe (AAATE) launched the “Recommendation 2016 concerning standards on eAccessibility and eInclusion”. (See ANNEX)

4. THE ROLE OF STANDARDS

4.1. The nature of standards and standardization

There are enormous differences in the degree of acceptancy of standards in various scientific and technical communities. For example, engineers and other technical experts consider standardization as the basis of efficiency, effectiveness and innovation in industry and trade. (ETSI 2017-09-03: http://www.etsi.org/standards/why-we-need-standards)

“Standards provide

- **Safety and reliability**
- **Support of government policies and legislation**
- **Interoperability*** /see 4.3/
- **Business benefits,** such as:
  - o Open up market access;
  - o Provide economies of scale;
  - o Encourage innovation;
  - o Increase awareness of technical developments and initiatives.
- **Consumer choice**

Consider what the world would be without standards:

- Products might not work as expected;
- They may be of inferior quality;
- They may be incompatible with other equipment – in fact they may not even connect with them;
- In extreme cases, non-standardized products may be dangerous;
- Customers would be restricted to one manufacturer or supplier;
- Manufacturers would be obliged to invent their own individual solutions to even the simplest needs, with limited opportunity to compete with others.” (ETSI ibidem)

ETSI concludes: “Society needs standards”. From the point of view of engineers and technicians which is in fact true, however, not necessarily recognized nor acknowledged by other scientific quarters and people not familiar with standards.

Standardization takes place in a complex system:

- Official standardization bodies are developing ‘formal standards’ or ‘de jure standards’ at national, regional and international levels and there are established networking mechanisms among them;
- Other standards developing organizations (SDOs) – especially those from industry – are developing ‘de facto standards’ or ‘industry standards’.

Traditionally, the development of standards is closely connected to terminology standardization. Many larger SDOs have more than 100,000 standardized terminological entries in their databases, some several 100,000. Increasingly other kinds of microcontent is standardized, as already explained. In addition, there is quite a number of methodology standards about how to handle microcontent in its various forms.

“Recommendation 2016” (see ANNEX) was prepared given the fact that the IoT cannot function efficiently and effectively without content and that standards are of great importance for ‘content interoperability’ in all its facets.

4.2. Technical standards, legal norms and certification

> “While ’legal norms issued by the state (or other kind of legal) authority are generally binding rules of conduct’, technical standards are identifying the state-of-the-art of scientific, technical or methodological development. Only if they are referred to in a law or other legal regulation, they can become part of the respective legal norm.” (IN LIFE D9.8 2017: 48)

Whereas legal norms are issued by a state (or other kind of legal) authority, technical standards are issued by standardizing bodies (or standardizing organizations or standards developing organizations, SDOs) at international, regional or national level similar to the legal system. In some countries, standardizing bodies are national authorities whose technical standards are considered as legal norms. In European countries, standardizing bodies are operating formally in the private sector – usually based on a national law referring to technical standardization. To summarize:

- Standards are not in themselves [in the legal sense] regulatory in nature and their application normally is voluntary. Although standards are "only" [strong] recommendations, they are widely used due to the benefits they bring about and because they contain a concentration of qualified technical information.
- Standards only become [legally] mandatory, if they are referred to in private contracts or agreements, or in laws or regulations and their use is stated as a requirement. Standards can prevent legal disputes, because they set out unambiguous specifications.
- In reality, standards are taken – first of all in jurisdiction – as second to law.

Certification – especially if it is based on standards – can lead to a powerful enforcement of standards. Certification can be applied to:
- Products.
- Software/tools (e.g. with respect to interoperability).
- Processes.
- Services.
- Human resources:
  - personnel certification,
  - competences and skills.
- Training:
  - training organization,
  - trainers,
  - training material and tools.

Increasingly certification is also applied to content resources from the point of view of content quality and interoperability. This extends towards software and tools to manage content, pertinent services and other services reusing content resources, organizational requirements for managing content resources, and requirements concerning competences and skills of the human resources engaged in the management of content resources.

Given the fact that some most important and societally as well as politically sensitive transversal aspects are far from being solved, legal and technical regulations will play an important role in the further development of the IoT. These aspects comprise in particular: security, privacy, safety, integrity, trust, dependability, transparency, anonymity, ethics and under a wider perspective also interoperability, energy consumption and cybercrime. (Vermesan & Friess 2014: 31)

4.3. Standardization and interoperability

The following statement referring to micro learning objects rather than ‘content’ in general still holds true:

> “There is a proliferation of web-based content platforms that offer users one or multiple resources on the one hand, and there is a lack of theoretical-methodological foundation on the one side and a lack of orientation at best practices of content interoperability on the other hand. /The fact that more and more resources are also created and maintained with often deficient web-based cooperative / participatory and distributed methods will further add to this proliferation./ Therefore, a combination of means, such as standards, appropriate software /information and communication technologies/, certification schemes etc.) is necessary to assure the quality – i.e. first of all reliability – of structured content.” (Galinski & Giraldo Perez: 2011)

It has already been explained above that ‘technical interoperability’ may not guarantee ‘content interoperability’. Interoperability (IOp) in the quote from the ETSI Portal was obviously used in the narrow sense of ‘technical interoperability’. However, it is increasingly being recognized that semantic IOp has more dimensions than expected, and data quality strongly depends on content IOp. Data quality in turn heavily depends on the quality of the metadata, and content IOp largely depends on the interoperability of data models and ontologies (e.g. according to data quality monitoring project for open data ADEQUATe (2015)).

Comparing standardized terminology entries on ‘interoperability’ one can find (technical) interoperability in the field of the ICTs defined as follows:

> “4.17 interoperability
> capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units” (ISO/TS 19101-2:2008, 4.17).

Increasingly this definition is recognized as being insufficient with respect to the user’s role as shown by:

> “3.11 interoperability
> degree or extent to which diverse environments (hardware and software) are able to exchange information without loss of content and in a manner transparent to the user” (ISO/TS 22224:2009, 3.11).

However, it was also recognized that technical interoperability may not be really interoperable without ‘organizational interoperability’ – and further ‘semantic interoperability’ (semIOp) which is defined as:

"2.34 semantic interoperability
ability for data shared by systems to be understood at the level of fully defined domain concepts" (ISO 11354-2:2015, 2.34).

In this connection, syntactic interoperability – which refers to the packaging and transmission mechanisms for data – is considered a prerequisite for semantic interoperability. (Wikipedia – https://en.wikipedia.org/wiki/Semantic_interoperability) Programming practice, however, commonly considers only monolingual and mono-modal applications, which is certainly insufficient for coping with the various facets of interhuman communication.

Linguists would differentiate 'semantic interoperability' – still largely at lexical or vocabulary level – into:

- lexical IOp (covering a certain degree of syntactic IOp in a linguistic sense),
- conceptual IOp,
- pragmatic IOp.

In all the above, the interoperability of content/communication items in different modalities across systems and applications has not been sufficiently considered. 'Content IOp' is going a step further than 'semantic IOp' by covering (in the meaning of interhuman communication) lexical, conceptual and pragmatic interoperability. ISO/TC 37 "Language and terminology" is focusing on 'structured content at the level of lexical semantics' which comprises terminologies and other language resources as well as non-linguistic/non-verbal resources of – what is called 'microcontent'. The Committee recognizes non-linguistic/non-verbal resources of microcontent as important 'other content resources' (in the meaning of 'other kinds of communication' <at the level of lexical semantics>). Therefore, it uses the term 'content IOp' which goes beyond the 'content interoperability' as used in the EPUB world (namely for the same layout as well as look and feel of content used through various reading devices, such as eBooks, readers, iPad etc.). ISO/TC 37 is becoming aware of the fact that it may have to embark on standardization activities concerning issues related to this broader concept of content IOp.

The EU project universAAL (2012: 16-17) refers to further facets of interoperability including:

- "Protocol Interoperability: The ability to share, bits and bytes over a network;

- Service Interoperability: The ability to exchange messages in well-defined format;
- Application Interoperability: The ability to interpret the exchanged data uniformity;
- User Perceived Interoperability: The components of a system to communicate effectively, accurately and provide the services expected by the user."

> "The four different classes are based on each other. Interoperability from the perspective of a human user will only be possible if the application systems and the applications used by him are interoperable. This in turn requires that the network services used by these applications to exchange messages, commands, files, images and sounds are interoperable. In addition, this requires that the actual exchange of raw data ("bits and bytes") is interoperable, which is only possible if the electrical and electronic components (connectors and cables) or wireless components are."

In conjunction with the three forms of cooperation of systems: compatibility, interactions with de-facto standards and interoperability through standards, universAAL (2012: 15) pointed out: "For this, usually it is necessary that common norms and standards are observed." The same applies also to content interoperability which – in terminology management – requires methodology standards related to

- **Data categories** (not quite identical with metadata) used in the conceptual design of the entries of structured content;
- **Data models** and **data modeling methods**;
- **Metamodels** to make competing data models interoperable.

These standards should be applied consistently and stringently – which requires that the methods they stipulate are interoperable. One can refer to this as 'methodology IOp' and even 'standards IOp'. Only then content IOp can be achieved – which also facilitates assuring the quality of microcontent. This means that we still need more standards – and "better" ones from the point of view of interoperability which is one of the prerequisites of sustainability in this field.

The above also applies in particular to 'controlled communication'. Today's computers in terms of full-fledged semantic interoperability still cannot cope with the whole range of variations of language (viz. pragmatics – including 'spontaneous variations'), even less, if interoperability is required in the non-verbal communication among and with PwD. Therefore, the unlimited potential for variations may need to be constrained in some way or other in the form of one or the other kind of controlled communication – an extension of commonly known 'controlled language' of all sorts. Controlled communication, too, is composed of or comprises microcontent entities. In the eApplications, standardized or even harmonized codings (of content items) abound – they can be regarded as specific kind of controlled vocabulary.

The "Recommendation on software and content development principles 2010" (MoU/MG 2012) in fact aims at 'content interoperability' – in particular with respect to microcontent. Besides, it looks as if true content interoperability is more feasible for microcontent than for unstructured content, because its metadata largely refer to 'meaning', not only to the content 'information object' as a whole.

5. CONTRIBUTION OF TERMINOLOGY SCIENCE AND ITS APPLICATIONS

Christian Galinski & Blanca Stella Giraldo Perez (2012: 8) stated:

> "[...] there may be also acoustic/audible symbols, haptic/tactile symbols, and others, which, in terminology management could occur as **designations**, or even **concept descriptions** (such as non-verbal representations [ISO 10241-1:2011]). There may be further information – and the respective data categories – required for a systematic approach in managing structured content at large."

Numerous examples of non-linguistic concept representations or mixed linguistic–non-linguistic representations were presented to prove that the concept of 'microcontent' comprises many kinds of structured content at the level of lexical semantics which hitherto were considered as totally different kinds of data requiring different methodologies, such as:

- Lexicographical data, such as:
  - o word entities (including compounds etc.),
  - o morphemes (morphology),
  - o collocations, metaphors, and other word combinations;
- Terminologies and similar kinds of language resources and other content resources, such as:
  - o nomenclatures, taxonomies, typologies, glossaries, vocabularies etc.,
  - o terminological morphemes (morphology),
  - o terminological phrasemes (phraseology),

  - proper names of all sorts as used for instance as items in different kinds of directories,
  - graphical symbols and other non-verbal designations,
  - (product) properties, characteristics, attributes etc.;
- Thesauri, classification schemes (see also ISO/DIS 22274:2011), keywords and other kinds of documentation languages (or controlled vocabularies);
- Encyclopedic (concise) lexical entries, such as in the form of:
  - knowledge-enriched terminological entries,
  - (explained) proper names and other kinds of data closely related to proper names,
  - nomenclatures (if they are not covered under "terminology");
- Ontologies, topic maps and other kinds of knowledge-structuring systems;
- Metadata and data categories (called masterdata or core components in other applications).

Among the above kinds of microcontent, lexicographical data, terminological data, controlled vocabularies, ontology entities and metadata, among others have specific roles or functions each which must be respected in datamodeling. Depending on the complexity and length of text used, encyclopedic (concise) lexical entries may be unstructured content rather than microcontent depending on the length of the text. Like in terminology, any kind of microcontent listed above needs a designative representation for the purpose of communication, such as:

- (one or more) linguistic representation spoken or written or in sign language,
- (one or more) graphical and other non-verbal designative representations such as graphical symbols of all sorts, acoustic symbols, or symbols in other modalities.

More often than not they need descriptive representations (or at least some sort of 'context'), such as

- (one or more) verbal descriptive representations, e.g. a definition, a description, an explanation or the like (which also can be rendered spoken or in sign language),
- (one or more) non-verbal descriptive representations (some in addition to a verbal representation, others created as non-verbal representations independently from verbal ones).

Non-verbal kinds of structured content are particularly meaningful in applications like eLearning and of vital importance in the H2H communication:

- with and among PwD (directly or supported by ICT devices functioning as assistive technologies),
- between PwD (and their carers) and the devices they use, and
- among these devices.

As the terminological approach is 'language-independent' (i.e. unlimited multilingual) and 'amodal' (i.e. in principle fit for including any modality) from the outset, it can be applied to any kind of microcontent – duly respecting the different roles or functions which different kinds of microcontent may have in H2H communication.

As was explained already, this applies to the metadata themselves, too.

6. CONCLUSIONS

By the very nature of the IoT, humans are part of it – and so is their interhuman (H2H) communication (whether in written or oral form, in whatever language, or other communication modalities). Ultimately any 'information object' in the huge amount of data originated by 'things' in the IoT needs or may need interpretation to become information and knowledge. Interpretation cannot do without 'meaning' – and, if it is about specialized information and communication, the most important basic building blocks of 'meaning' (i.e. semantics) are concepts represented by terms or other kinds of concept representations.

In this connection, the concept of 'terminology' as commonly understood by experts with linguistic background, needs extension: in terminology standards, technical documentation (especially, if new media are involved), eBooks etc. non-verbal representations of concepts are quite common. For H2H communication in connection with eLearning and PwD these other modalities are indispensable.

The aspect of different kinds of 'terminological units' also need to be revisited. In the eApplications there are many more kinds of terminological units than just the commonly considered terms and abbreviations (and maybe symbols in seemingly linguistic form). In any case, terminological units are most important for specialized communication – in whatever modality.

The terminological approach which is concept-oriented can be applied to any kind of microcontent – in particular, if multilinguality and multimodality is (even only potentially) required. In principle, it is language-independent (and therefore multilingual) and amodal (and therefore multimodal) from the outset. This fact is still not sufficiently considered in textbooks and standards about terminological principles and methods.

In order to capture the 'meaning' of microcontent objects, metadata – and the more constrictive use of metadata in the form of core metadata or masterdata or the like – is indispensable. Especially with respect to metadata and masterdata there are many standardized approaches even in ISO and IEC. This urgently needs harmonization.

'Competing' standards are very detrimental to 'interoperability' – violating one of the main objectives of standardization. Concerning different kinds of microcontent, harmonization of the content of pertinent methodology standards is most important. This would have a beneficial impact on the development of all kinds of information management systems for the sake of content integration and content interoperability.

One of the ways to improve this situation – which in fact leads to very high costs in organizations everywhere – may be initiatives to improve the coordination between standardizing committees in order to make content across standards and even different standardization areas more coherent, such as initiated by the "Recommendation 2016 concerning standards on eAccessibility and eInclusion". (See ANNEX)

This contribution tries to bring an array of seemingly unrelated aspects together to create awareness for important issues to be considered, and thus pave the way for improved system development and more effective content integration and content interoperability.

REFERENCES

ADEQUATe-Qualität von Datenportalen. Forschungdatenbank Fachhochschule Krems. 2015 (http://www.donau-uni.ac.at/opt/fdb/projects/view/4294968866).

Boiko B. 2004: Defining Data, Information, and Content. – B. Boiko, *Content Management Bible*, 2nd Edition, Indianapolis: Wiley Publishing Company, 3-12.

The Computer Language Company Inc. (1981-2017). – *Computer Desktop Encyclopedia* (http://www.computerlanguage.com/).

Evans D. 2011: *The Internet of Things. How the next evolution of the Internet is changing everything*, IBSG White Paper (CISCO Internet Business Solution Group) (file:///C:/DATEN/daten/Infoterm%20Members/LA/COL%20Workshops/Medellin%20_2014&2016/CISCO-2011_IoT_IBSG_0411FINAL.pdf).

Ericsson 2016: Ericsson Mobility Report. On the pulse of the networked society, Stockholm (https://www.ericsson.com/res/docs/2016/ericsson-mobility-report-2016.pdf).

ETSI Portal ETSI 2017-09-03. (http://www.etsi.org/standards/why-we-need-standards).
Galinski Ch. and Giraldo Perez B. S. 2011: The work of today's TC/37 in the light of 'content interoperability': Standardization as an issue of corpus planning taking eLearning as an example. – Proceedings of the IITF Colloquium on Language Policy and Language Planning, Perm, Russia, 22-23 August 2011. – Journal *Terminology Science and Research* (TSR), vol. 22, 5-19. ISSN 1017-382X (file:///C:/Users/chris/AppData/Local/Microsoft/Windows/INetCache/IE/X8YVWQ2G/Vol22.pdf).
Galinski Ch. and Giraldo Perez B. S. 2012: Content Interoperability: Re-purposing structured content for eLearning seen under a standardization perspective". International Conference on Terminology: State-of-the-art and the future of terminology methodology and application (TMA 2011), Vilnius, Lithuania, 12-13 October 2011. – *Terminologija* 19, 6-32.
Galinski Ch. and Giraldo Perez B. S. 2014: Typology of structured content in eApplications. Under a content interoperability, quality and standardization perspective. – Budin G. and Lušicky V. (eds.), *Languages for Special Purposes in a Multilingual, Transcultural World*. Proceedings of the 19th European Symposium on Languages for Special Purposes, 8-10 July 2013, Vienna, 405-417. ISBN: 978-3-200-03674-1. (Available with Creative Commons Non-Commercial 4.0 International License at: https://lsp2013.univie.ac.at/proceedings).
IERC – http://www.internet-of-things-research.eu/about_iot.htm (IoT European Research Cluster)
IN LIFE. D9.8 – Dissemination & Standardization Plan. 2017 (EU Project: **IN**dependent **LI**ving support **F**unctions for the **E**lderly).
ISAAC – https://www.isaac-online.org/english/what-is-aac/what-is-communication/ (International Society for Augmentative and Alternative Communication).
Merriam-Webster. – https://www.merriam-webster.com/dictionary/metadata.
MoU/MG (Management Group of the ITU-ISO-IEC-UN/ECE Memorandum of Understanding concerning eBusiness standardization) 2012: Recommendation on software and content development principles 2010. (MoU/MG/12 N 476 Rev.1, 2012) Available at: http://isotc.iso.org/livelink/livelink/fetch/2000/2489/Ittf_Home/MoU-MG/Moumg500.html.
OASIS (Advancing Open Standards for the Information Society) (s.a.). Retrieved 2013-07-01, from OASIS Unstructured Information Management Architecture (UIMA) TC: https://www.oasis-open.org/committees/uima/charter.php.
Riley J. 2017: *Understanding metadata. What is metadata and what is it good for? A Primer Publication of the National Information Standards Organization* (*NISO*). Baltimore: NISO (http://www.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf).
Riley J. 2012: ICT: The difference between data and information. Study notes: https://www.tutor2u.net/business/reference/the-difference-between-data-and-information retrieved 2017-09-03.
Rowley J. 2007: The wisdom hierarchy: representations of the DIKW hierarchy. – *Journal of Information Science* 33 (2), 163–180.
Steichen B. and Wade V. 2010: Adaptive Retrieval and Composition of Socio-Semantic Content for Personalised Customer Care. – Federica Cena et al (eds.). *Proceedings of the International Workshop on Adaptation in Social and Semantic Web* (SASweb – 2010), Hawaii: EEUU, 1-12 (http://ceur-ws.org/Vol-590/).
UNESCO 2017: Infoterm, ICANN. *Glossary of the Internet governance terms* (IGG), Paris: UNESCO (http://en.unesco.org/sites/default/files/igg2017_en.pdf).
universAAL Consortium. D8.3 2012: *Standardization usage plan and contributions*.
Vermesan O., Friess P., Guillemin P., Gusmeroli S. et al. (eds.) 2011: *Internet of Things. Strategic Research Agenda*. Chapter 2: Internet of Things Strategic Research Roadmap, Aalborg (DA): River Publishers. ISBN 978-87-92329-67-7 (http://www.internet-of-things-research.eu/pdf/IoT_Cluster_Strategic_Research_Agenda_2011.pdf).
Vermesan O. and Friess P. 2014: *Internet of Things. From research and innovation to market deployment*. Aalborg (DA): River Publishers, ISBN 978-87-93102-94-1 / ISBN 978-87-93102-945-8.
Webopedia – http://www.webopedia.com/.
Wikipedia – https://en.wikipedia.org/ all quotations retrieved 2017-09-03.
Zumpe S., Esswein W. 2002: Simplification of Knowledge Discovery using "Structure Classification". – Gaul W., Ritter G. *Classification, Automation, and New Media: Proceedings of the 24th Annual Conference of the Gesellschaft für Klassifikation e.V., University of Passau*, March 15-17, 2000, Dresden, Germany: Springer: 245–252. Available from: http://link.springer.com/chapter/10.1007/978-3-642-55991-4_26.

REFERENCED STANDARDS

IEC 80000-14:2014 *Telebiometrics related to physiology as a part of the ISO and IEC 80000 harmonized series*.
ISO/IEC 15938-5:2003 *Information technology – Multimedia content description interface*. – Part 5: Multimedia description schemes.
ISO/IEC 12785-1:2009 *Information technology – Learning, education, and training – Content packaging*. – Part 1: Information model.
ISO/IEC 24800-3:2010 *Information technology – JPSearch*. – Part 3: Query format.
ISO 5492:2008 *Sensory analysis – Vocabulary*.
ISO 9241-151:2008 *Ergonomics of human-system interaction*. – Part 151: Guidance on World Wide Web user interfaces.
ISO 10241-1:2011 *Terminological entries in standards*. – Part 1: General requirements and examples of presentation.
ISO 11354-2:2015 *Advanced automation technologies and their applications – Requirements for establishing manufacturing enterprise process interoperability*. – Part 2: Maturity model for assessing enterprise interoperability (https://www.iso.org/obp/ui/#iso:std:iso:11354:-2:ed-1:v1:en).
ISO 15836:2009 *Information and documentation – The Dublin Core metadata element set*.
ISO 16642:2003 *Computer applications in terminology – Terminological markup framework (TMF)*.
ISO/TS 19101-2:2008 *Geographic information – Reference model*. – Part 2: Imagery (https://www.iso.org/obp/ui/#iso:std:iso:ts:19101:-2:ed-1:v1:en).
ISO 19115:2003 (series) *Geographic information – Metadata*.
ISO 19115-1:2014 *Geographic information – Metadata*. – Part 1: Fundamentals.
ISO 22274:2013 *Systems to manage terminology, knowledge and content – Concept-related aspects for developing and internationalizing classification systems*.
ISO/TS 22224:2009 *Health informatics – Electronic reporting of adverse drug reactions* (https://www.iso.org/obp/ui/#iso:std:iso:ts:22224:ed-1:v1:en).
ISO 24531:2013 (4.11) *<XML> all data between the start tag and end tag of an element*.
ITU-T Y.2060:2012 Series Y: *Global information infrastructure, Internet Protocol aspects and next-generation networks. Next Generation Networks – Frameworks and functional architecture models*.
ITU-T X.1081:2011 *The telebiometric multimodal model – A framework for the specification of security and safety aspects of telebiometrics*.
ITU-T X.1082:2010 *Telebiometrics related to human physiology*.

ANNEX

Recommendation 2016 concerning standards on eAccessibility and eInclusion

(drafted at the IN LIFE Workshop "Strategic Standardization Issues Concerning eAccessibility & eInclusion" at the 15th International Conference on Computers Helping People with Special Needs – ICCHP 2016, in Linz, Austria, in July 2016)

Purpose:

Increasingly R&D projects and the software industry – especially for mobile technologies – consider the "Recommendation on software and content development principles 2010" whereby

> "decision makers in public as well as private frameworks, software developers, the content industry and developers of pertinent standards /should be/ aware that multilinguality, multimodality, eInclusion and eAccessibility need to be considered from the outset in software and content development. These considerations are required in order to avoid the need for additional or remedial engineering or redesign at the time of adaptation, which tend to be very costly and often prove to be impossible".

Since 2010, hundreds of standards about eAccessibility and eInclusion were developed or revised by technical committees in standards organisations at international, European, or national level – not to mention many industry standards developed by other standards developing organisations (SDO). In addition, there are possibly thousands of standards that have a bearing on persons with disabilities (PwD). In particular, 'accessibility' in a broad sense rarely occurs in the title or in the body of these standards.

"Recommendation 2016" addresses critical issues identified in recent conferences, initiatives and projects dealing with eAccessibility and eInclusion and related topics concerning the difficulties faced by system developers, their customers, health care providers and end-users when trying to find and apply pertinent standards.

The organizations endorsing Recommendation 2016 call upon stakeholders of eAccessibility and eInclusion, in particular standards developing organizations (SDOs) to:

- develop a more refined classification or keywording approach to identify content in standards with a bearing on eAccessibility and eInclusion
- register the potential relevance for eAccessibility & eInclusion of an emerging standard right from the beginning of a standardization activity
- cross-reference standards having a bearing on eAccessibility and eInclusion
- encourage the formulation and use of consistent vocabulary / terminology
- implement search functionalities that ease the use of standards
- facilitate the active involvement of PwD as end-users in standardizing activities among others by providing standards documents in an 'accessible' format

Implementing the above measures would enhance interoperability of eAccessibility&eInclusion related products and services and thus benefit users of standards and standardization at large.

Recommendation:

Standards development processes and monitoring in conjunction with standards about eAccessibility and eInclusion and related aspects, should allow the coordination of standardizing activities across technical committees and SDOs, leading to content coherence among standards about similar themes. This would help industry and other organizations to comply with standards' requirements referring to corporate social responsibility and risk management, as well as with the latest legal regulations on accessibility in eProcurement and public websites.

Supportive measures may be worthwhile pursuing, to (a) promote certification schemes based on standards about eAccessibility and eInclusion, (b) encourage education and training activities regarding such standards, and (c) enhance the positive role that media (both institutional and social) and civil society can play here.

Online endorsement: aaate-endorsed-the-recommendation-2016-concerning-standards-on-eaccessibility-and-einclusion/

DAIKTŲ INTERNETAS – LŪKESČIAI IR REALYBĖ: SVARBIAUSIOS DAIKTŲ INTERNETO TECHNOLOGIJŲ PROBLEMOS IR ŽMOGAUS BEI KOMPIUTERIO SĄVEIKOS TURINYS

Daiktų internetui skirtų mokslinių tyrimų ir technologinės plėtros progresas labai spartus. Daiktų internetas gali technologiškai sujungti daugmaž visas elektroninių paslaugų teikimo sritis, įskaitant e. sveikatos, e. verslo ir prekybos, e. valdymo, net e. švietimo ir kitas paslaugas, o kur dar junginiais su žodžiu „išmanus“ ar „pažangus“ įvardijami projektai (tokie kaip išmanieji miestai, išmanusis transportas, išmaniosios ligoninės). Turint omenyje šiuos dalykus, vis labiau ryškėja žiūrėjimo į *sąveikumą* pirmiausia per technologijų prizmę (techninis sąveikumas), taip pat tam tikru mastu ir per semantikos prizmę (semantinis sąveikumas) trūkumai. Internetinių turinio platformų, vartotojams siūlančių vieną ar daugelį išteklių, gausa iš pirmo žvilgsnio atrodo pribloškianti. Atidžiau pažiūrėjus į šių išteklių turinio sąveikumą ir kokybę matyti, kad jiems dažnai trūksta teorinio-metodologinio pagrindo. Situacija tampa dar sudėtingesnė, kai turinys – ypač įvairių rūšių mikroturinys – yra kuriamas, prižiūrimas ir siejamas su kitu turiniu interneto naudojimu pagrįstais metodais. Kai daiktų interneto *daiktų* „sukurti“ duomenys turi būti interpretuojami, kad taptų informacija ir žiniomis, nebepakanka paprasto *objektų* (fizinio pasaulio objektų ar virtualių dalykų) identifikavimo ir žodinio ar nežodinio *vardo* priskyrimo, neatsižvelgiant į atitinkamas sąvokas. Kai *daiktus* turi atspindėti skirtingų kalbų žodžiai ar nežodiniai ženklai esant kitoms jutiminio suvokimo formoms, *semantika* turi remtis sąvokomis. Kaip viena pamatinė idėja ar sąvoka, pasiekiama per atskirą universalųjį adresą ar pastoviąją nuorodą, apibrėžtas mikroturinys apima daugelį įvairių turinio rūšių – galiausiai net joms apibūdinti naudojamus metaduomenis. Kyla klausimas, ar šios įvairios mikroturinio rūšys, įskaitant leksikografinius ir terminologinius duomenis (taip pat ir mokslines nomenklatūras), valdomuosius žodynus (tezaurus ir kitas žinių tvarkybos sistemas), žinių struktūrinimo sistemas (ontologijas, temų žemėlapius ir kt.) ir atitinkamus metaduomenis, gali būti plėtojamos ir valdomos, remiantis viena plačios aprėpties metodologija – terminologijos metodologija. Žinoma, reikia atsižvelgti į skirtingų mikroturinio rūšių skirtingus *vaidmenis*.

Gauta 2017-09-07

Christian Galinski
International Information Centre for Terminology (Infoterm)
c/o CO-SPACE
Gumpendorferstrasse 65/1, 1060 Vienna, Austria
E-mail cgalinski@infoterm.org