Data or information

Dora Gelo Čolić[1]

[1]Affiliation not available

April 25, 2024

Abstract

The need for clarity in fundamental concepts within the discipline is constant. Still there is much confusion about these concepts. There are tendencies to take distinction between data and information as natural and obvious but in concrete research we can easily understand that the issue is far from resolved. The relationship between data and information and their differences are explored in this article. The concept of DIKW (Data-Information-Knowledge-Wisdom) hierarchy is analyzed empirically and theoretically. This paper will present an analysis of 6 sentences from media articles. The sentences will be analyzed as information and as data with the aim of discerning the difference that exists when the same set of words is considered information and is thus composed of data (analysis 1) and when it is considered a piece of data in relation to the article in which it is found (analysis 2). The paper thus attempts to show empirically what the real differences between the two entities are and in what relationship they exist. The obtained results are limited to the set research framework, however, they may serve as the initial setting for testing in a different framework as well.

Keywords: data, information, DIKW, mutual interference, empirical analysis

Introduction

The need for clarity in fundamental concepts within the discipline as Gray (2003) states is constant. Still there is much confusion about these concepts. "How they are distinguished depends on how they are individually defined, but it is almost universally agreed that "data" (whatever they may be) and "information" (whatever it is) are somehow different, i.e. that the terms have different referents." There are tendencies to take distinction between data and information as natural and obvious but concerned brief historical review shows that "distinction is neither natural nor obvious". Due to the lack of agreeement about the definitions, potential of language is one approach worth to explore "to ask to what problems or concerns the language responds" (Gray, 2003). Furthermore, Ridenour and Moulaison-Sandy (2022) analyze calls in the field to define data but emphasize that a fundamental agreement "on what data itself is lacks, and the distinctions between data and information is far from resolved".

The relationship between data and information and their differences are explored in this article.

The general definition of (Floridi 2011) information (GDI) is an operational standard that defines the information as an item that:

1. is composed of one piece of data or more data that
2. is well composed and
3. has meaning.

By well-composed data, we mean data that are composed in accordance with the syntax of a certain system, whereby syntax is seen as a term in a broader sense than linguistics, which includes the determination of form, construction, composition and structure (Floridi, 2011).

A primary characteristic of Floridi's semantic information, though not self-sufficient, is the meaning in a particular code, system or language implying compliance with the chosen code, system or language.

The fundamental nature of the data is where we see x different to y, where x and y are two uninterpreted variables, or in other words, the fundamental nature of data is exhibiting the anomaly.

The GDI endorses the following thesis: TaN) a datum is a relational entity. It implies taxonomic neutrality due to the fact that nothing seems to be a datum per se, classifying it as a relation. Although we have the slogan "data are relata", the GDI is neutral with respect to the identification of data with specific relata.

The second implication of the GDI analysed in this research is typological neutrality, which means that information can consist of different types of data as relata. Although the typology is not yet fixed and standard, it is quite common to have five classifications. Depending on the analysis conducted and the adopted level of abstraction, the same data can be understood as more than one type (Floridi, 2009)

Considering the importance of this concept and its immanence when it comes to discussion of nature of data and information, as well as of nature of knowledge and wisdom, DIKW (Data-Information-Knowledge-Wisdom) hierarchy is analyzed empirically and theoretically. Cornelius (2002) emphasizes that more of an issue here is relationship between information and knowledge and what concerns him is how information feeds and alters knowledge structure in a human recipient. This is certainly substantial question but still, distinction between data and information can't be avoided and considered as kind of convention if we observe DIKW as the concept in longlasting discussion. We still can argue that is even more substantial to try to create and recreate process of alteration and its implications on the journey from data to information.

The interpretation of the DIKW hierarchy serves to determine the relationship between fundamental concepts in the scientific discourse, and interpretations are carried out at different levels, of which the structural and functional ones are important for this research. The existence of different levels of interpretation opens up the possibility of different interpretations of the interrelationships of the elements of that hierarchy. Several researchers have dealt more seriously with this topic in the context of the field of knowledge management and the field of library and information science. The development of technology and science and the phenomena typical for this development are also linked to the DIKW hierarchy in such a way that: 1. the effect of big data analytics on the hierarchy is analyzed, 2. the concept is examined in relation to philosophical approaches to "nursing informatics", i.e. health systems supported by information technology, and 3. it is linked to risk management (Bosančić, 2017).

In the framework of the present research, it is particularly interesting to mention Donald Clark (2004), who imagines the DIKW hierarchy in a coordinate system, where the abscissa is understanding and the ordinate is context. In this system, the concept of data, which precedes the other concepts in time, is assigned the lowest value of understanding and context, and the concept of wisdom is assigned the highest. Data in this presentation are the least comprehensible and are outside of any context, while the other concepts of the hierarchy are "enveloped in increasing comprehensibility and contextuality during processing".

Unlike authors who consider the concept of data as the starting point of the model, there are those like Tuomi Ilkka (1999) who believe that knowledge must exist before any formulated information or data that can be measured with an appropriate scientific method. Knowledge exists at the beginning, and when it is assigned an unchanging representation and a standard interpretation – knowledge becomes data. In the framework of this research, it is possible to find space for such considerations.

Theoretical elements important for both analysis

1. Although there are different definitions of propositions, they all have several common denominators, and here it is particularly emphasized that they are the part of the meaning of a sentence that describes some kind of state of affairs (Šarić, 2006). When defining a proposition in a specific sentence, insight according to Kratzer (2010) has also been used, according to which propositions are sets of possible situations or characteristic functions of such sets.

2. The quantity of a sentence here is meant as the amount of content, and although the initial setting was to count all the words in a sentence including particles, in the second phase when the empirical analysis was done, it was decided that it was optimal to use the traditional approach according to which nouns, verbs and adjectives are substantive or auto-semantic expressions (Šarić, 2006); that is, very early on in the empirical analysis, although not without exception, it turned out to be most expedient to count verbs, nouns and adjectives as the bearers of content of a sentence.

3. The correctness of data is established when, at sentence level, there is no contradiction between the data of which the information consists (Floridi, 2005). The contradiction is examined at all levels – syntactic, semantic and grammatical. In addition to the context, a piece of data can also be examined in relation to external entities. A precise piece of data is the one whose repetition brings meaningful value (Floridi, 2005), while an imprecise piece of data is considered to be the one whose repetition, although possible, does not bring meaningful value either after the first, the second, or the tenth attempt. As well as when testing the correctness, the precision of a piece of data can also be tested in relation to external entities.

4. In the Shannon-Weaver model, information implies probability measurements, while Brookes (1980) conceptualizes information as a product of human thought embodied in a physical form, and this understanding carries the connotations of a measurable entity that can be coded and can be a cause for cognition. Ma (2012) argues that these two approaches contain fundamental errors, one of which boils down to the analogy of machine processing and human thought and communication, and the other is related to the causal relationship between data and information, and information and knowledge. However, despite their different meanings, the conceptualizations of information are very similar in terms of the empiricist epistemological determinations that underlie them. In the three mentioned approaches, within which the meaning of the concept of information is discussed, information is conceptualized as an entity or a unit that can be objectively measured, observed or processed (Ma, 2012).

5. Shannon (1993) also supports insight on different meanings of various writers: "It is hardly to be expected that a single concept of information would satisfactorily account for the numerous possible applications of this general field."

6. Fricke (2009) provides arguments supporting the claim that DIKW is incorrect and methodologically undesirable. The same author, relying on Ackoff (1989), points out that the difference between data and information is functional, not structural, which will be examined separately in analysis 1 and 2. Since Fricke defines information as a subset of relevant data together with the results of inferences derived from that data, it clearly follows that the subset can be augmented by additional items inferred or calculated from the subset. The above removes the possibility of taking structure and structuring itself into account as a criterion for defining something as data or information.

7. Fricke (2009) concludes his approach to the nature of data, which he analyzes through truth and certainty, by suggesting that there is no certainty regarding the truth of a piece of data, but that it must be true regardless. Data are akin to facts and, despite a great number of theories, it is acceptable, in the theoretical framework that he provides in this article, to say that data are similar to facts, which he also calls logical atoms.

Methodological Approach

This paper is based on a review of the literature in the information sciences and empirical application of theoretical elements in two text analysis. As an example, this paper will present an analysis of 6 sentences from media articles. The sentences were analyzed as information and as data with the aim of discerning the difference that exists when the same set of words is considered information and is thus composed of data (analysis 1) and when it is considered a piece of data in relation to the article in which it is found (analysis 2). The paper thus attempts to show empirically what the real differences between the two entities are and in what relationship they exist.

In a research an iterative analysis method is used. Due to that it is possible to talk about a gradual reduction of the error after each step. Considering the complexity of the system being processed, it is to be expected that a method of this character and scope will be used in the following research, possibly in a more complex form by narrowing down and determining the number of steps of information/data analysis and/or by introducing the minimum allowed measure of error reduction in the results in the sense of limiting the legitimacy of the results.

It is important to state that parts of the methodological approach presented here e.g. final version of the selecting sentences from media articles, classification of data came as a result of empirical analysis of initial theoretical elements. It is expected as mentioned above more updates on results in future circles of analysis and possibly on methodological approach.

The whole process is meant like complex organism as a place of mutually interference of theoretical elements, methods and empirical outcomes.

The obtained results are limited to the set research framework, however, they may serve as the initial setting for testing in a different framework as well.

Paper discusses:

- distinction between data and information and its nature
- is there a border between them
- applicable approach in obtaning result
- possibity of standard process of the analysis

Free stream of analysis is intentionally left to be transparent in order to recognize all dilemmas encountered during the process. Also, it is important for the outcome of the research to enable its reproducibility, especially considering used methods and all theoretical inconsistencies and voids that this research tends to diminish. Additionally, transparency enables that another concept or definiton or approach could be applied e.g. another definiton of proposition in analysis 2, and potentially lead to different, more coherent result.

Selecting units of analysis

The process of selecting media articles was conditioned by the thematic framework in which the articles were selected according to participants, events and activities related to sensitive information in relation to the broader concept of national security in its non-military aspects, which relate to the vulnerability of political structures, institutions, social stability and cohesion (Buzan 1983).

Articles that were selected because of the characteristics that make them suitable for the set thematic framework are analyzed by first determining the context at the level of the entire article, then the zero event as the event that initiated writing or, where media reporting in a series is concerned, some other event that occurred after the zero event and followed from it. Methodological problem in this part could be due to fact that Croatian language is not internationally known. For that reason there is no possibility for reader to reconsider different context for particular media article but still this is implementation problem and not of

substantial nature. There is always opportunity to translate specific article in adequate language or try to analyse sentence and article in known language. Since this research originally deals with Croatian language and its rules, there might be different rules in the whole process of analysis due to the differences in language syntaxes and the rules in general.

Structural analysis of the text, syntax analysis, grammatical analysis and narrative analysis were used during the selection of units of analysis (sentences) from the media articles. It was done according to different criteria:

1. sentence that is related to the context of the article
2. sentence that doesn't include components of the context of the article
3. sentence that is interesting to research in regards to grammatical or syntactic elements

Extracting data in analysis 1

The next step will be to extract data from the information:

1. using grammatical analysis, semantic analysis and, where necessary, using syntax analysis and structural analysis what basically is language analysis
2. in relation to the context at the level of the entire article as a system in which a sentence exists.

Extracting will be done in this way to investigate difference between data and information in regards to function and structure. While functional difference is observed in relation to the context, structural difference is observed during grammatical, syntax and structural analysis.

The fact that extracting in relation to the context is not a rule without exception became evident when analyzing sentences which did not have a significant number of components of the context or any at all, and data were extracted on the basis of grammatical analysis/syntax analysis/structural analysis. This conclusion, however, does not mean that there is no relationship between data and context, because it is indisputable that a sentence is placed in the analyzed text, and that only in relation to context can it be concluded that data are not in context. Its functions accordingly needs to be examined.

Classification of data in analysis 1

Since the data/information in this case are a media portrayal of an event, they are as such parts of the system of the context in which they are presented, so an analogy is drawn that the context is a system/syntax in relation to which data are identified as primary data, secondary data, operational data, derived data and metadata (Floridi 2009).

During the analysis, a problem arose with the adjustment of the term "derived data", given that the type of data is determined in relation to the context. However, solutions during the analysis showed that data are extracted and that their type is determined in relation to the context precisely because it is only in relation to the context that it is possible to determine that data are not a part of it, i.e. that they originate from another source different from the system in which they are found. However, the fact is that they exist in information in a specific way and as such are derived from another source but have a specific function in the system. One of the results showed that their function is to support the context, but there are findings suggesting that they can also have the opposite effect. Further sentence research in this direction that has not been carried out in this analysis should focus on the function of such sets of words in a text as a whole. Therefore, by taking a fresh look at data after extraction, one can observe that regardless of the fact that each piece of data participates in the creation of meaning, it is not in itself information. A similar view can be found when Ma (2012) talks about information as an element of the DIKW model which is understood as organized and processed data. Since the GDI mentions that information consists of at least one piece of data, it remains to be discerned how to differentiate data from information that consists of one piece of data.

Other analysis settings

1. External entities are elements that are outside the sentence, but are connected with it. They can be persons, things, phenomena, processes, documents, etc.

2. The average reader as a recipient of media information is a person who is averagely informed and to whom valid information outside the public space is not available.

3. In contrast to the initial setting for determining the element of time as a nominal number of days, given the deficiency of the comparative approach due to large deviations in the calculations, it was concluded that it was constructive to adopt a scoring system for assessing the novelty of information, i.e. to determine a time index, so a scale was standardized according to which time (number of days) that was shorter than thirty days from the event to the publishing of the media article is worth a certain number of points when assessing the value of a sentence.

4. The criterion of common knowledge is important for defining the limit for the possibility of defining which set of words can only be data, and not information.

5. The verifiability of data in analysis 1 and the characteristics of propositions in analysis 2 is an essential element in assessing the quality of information or data. The number of verifiable entities in a set of words is put in correlation with the precision and correctness of data in both analysis 1 and 2 in order to obtain a certain value that represents the degree of quality of the information or data.

Parallel empirical analysis 1 and 2

Free stream of analysis is intentionally left to be transparent in order to recognize all dilemmas encountered during the process. Also, it is important for the outcome of the research to enable its reproducibility, especially considering used methods and all theoretical inconsistencies and voids that this research tends to diminish.

1. Tportal, 4 November 2021 (accessed electronically in May 2022)

"Ministry of Defense reacts to Milanović's letter"

The Ministry of Defense (MOD) reacted to the letter of the President of the Republic (PRH) stating that the case of the commander of the Honorary Guard Battalion, Brigadier Elvis Burčul, does not concern the abuse of power and early retirement, but the termination of active military service with the right to a full old-age pension.

Context: the PRH and the MOD interpret the same event differently

Event 0: the retirement of Brigadier Burčul (20 October 2021)

Analysis 1

1.1 Data are extracted from information

The initial setting of data extraction was defined in such a way that certain contradictions in the information were examined during the extraction, so data from the first sentence were written in the following more structural meaning following language analysis and less functional way in order to examine Ackoff's (1989) theory that the difference between data and information is functional, not structural, but also that data are usually reduced when transformed into information).

Data 1 – the Ministry of Defense reacted to the letter of the PRH

Data 2 – the case does not concern the abuse of power

Data 3 – the case does not concern early retirement

Data 4 – the case concerns the termination of active military service

Data 5 – with the right to a full old-age pension

1.2 Verifiability of data

Nowadays, the right to access information can be exercised in accordance with the Act on the Right of Access to Information (2013; 2015; 2022), which makes more data verifiable. In this sense, the selected article is in a better position in terms of verifiability than some others that are also an integral part of the research and originate from the early 2000s when other criteria were valid. In this particular case, it should be taken into account that the article concerns an active military person and certain personal data, so the verifiability is limited in that part.

Data 1 – the Ministry of Defense reacted to the letter of the PRH

Data 2 – the case does not concern the abuse of power

Data 3 – the case does not concern early retirement

Data 4 – the case concerns the termination of active military service

Data 5 – with the right to a full old-age pension

Three verifiable data are established. Data 1 can be verified in the article itself as well as on the Ministry's official website. Data 2 and 3 could not be established at the time of the analysis, for several reasons. The first one is that the average reader is not familiar with the specific facts about Commander Burčul's years of service, and the second is that the responsible persons give contradicting statements about powers in relation to the case. It is possible to determine more reasons, but that would steer the analysis towards a deeper legal analysis, which is not important for this research.

Data 4 and data 5 can be verified on the Ministry's website.

1.3 Classification of data

Data 1 – the Ministry of Defense reacted to the letter of the PRH – primary

Data 2 – the case does not concern the abuse of power – primary

Data 3 – the case does not concern early retirement – primary

Data 4 – the case concerns the termination of active military service – derived

Data 5 – with the right to a full old-age pension – derived

1.4. Characterization of data

To assess the value of information, it is necessary to determine the characteristics of data in terms of precision and correctness.

Data analysis established that all data are correct and precise.

Analysis 2

The process of identifying a proposition which would be an entity ranked lower than a sentence, when viewed as a piece of data, produced the proposition

the termination of active military service

Such a result was the product of identifying a kind of center of the sentence – that which the sentence is about, that without which the sentence loses its meaning. This syntagm is something that is in the sentence as opposed to something that is not. It is possible to observe this proposition in the extended interpretation and in such a way that the given sentence is a minimal situation (Kratzer, 2010) of the identified proposition.Namely, although the sentence explicitly states that "the case does not concern the abuse of power" and that "the case does not concern early retirement", a possible situation, especially in

the defined context (the PRH and the MOD interpret the same event differently), is that somewhere it is true that the case concerns the abuse of power or that the case concerns early retirement, all in connection with the argument that more information about the context is needed to interpret expressions smaller than sentences (Šarić, 2006). It should be noted here that the set of words defined in analysis 2 as the center of the sentence, as a proposition and as a part without which the whole changes/loses its meaning, is identified in analysis 1 as derived data. Such a finding leads to further analysis of the function of the set of words in the text, which in its essence is derived from another source and not from the system to which it belongs (context).

2. Tportal, 4 November 2021 (accessed electronically in May 2022)

Likewise, the requirements of the service for active military personnel are not determined by the Chief of General Staff of the Armed Forces, but by the Minister of Defense in accordance with the law.

Context: the PRH and the MOD interpret the same event differently

Event 0: the retirement of Brigadier Burčul (20 October 2021)

Analysis 1

During data extraction, we examine the significance of the result that the data are extracted as follows:

a)

1 there are requirements of the service for active military personnel

2 there is a Chief of General Staff of the Armed Forces

3 the requirements of the service for active military personnel are determined by the Minister

4 there is a minister

5 there is a law

6 the requirements of the service for active military personnel are determined by the Minister of Defense in accordance with the law

or

b)

1 the requirements of the service for active military personnel are not determined by the Chief of General Staff of the Armed Forces 2 the requirements of the service for active military personnel are determined by the Minister of Defense

3 in accordance with the law

Several elements are examined here:

– if the difference between data and information is structural and not functional, what is the effect of that difference on the results of this research

– if data are relationships, in example b) is it necessary to add a part or even all of data 2 to data 3 in order for data 3 to even be a piece of data

– if the difference between data and information is functional and not structural, is data 1 from example b) information or data

– what would be secondary data in this kind of research framework

– is the extraction of data such as “there is a minister”, “there is a law” necessary

All data were assessed as correct.

Data 1 and 2 were assessed as precise.

Data 3 was assessed as imprecise because, in addition to not being clear as to the particular law concerned, it is consequently further not known which provision of a law it concerns, and this would remain unknown without further specification even if it were known which particular law is concerned, from which it follows that there are a total of two unknowns in data 3.

In addition to the finding that this data is identified as secondary because it refers to the primary data, but is omitted, such data is not precise.

Next, we analyze whether every secondary data is also imprecise, so in that sense, the characteristics of data will be differentiated, or not, from its identity.

If the analysis is in a defined context (the PRH and the MOD interpret the same event differently), the question arises of the need to extract the data that "there is a minister" and the data that "there is a law". On the one hand, those two data are essential components of the context, and one could conclude that the extraction is necessary and correct. One can agree that it is correct, however, if the data are well-known or self-intelligible in relation to the context, they as such do not represent an anomaly and with their correctness they can seemingly increase the value of the information since both are correct and precise, while they practically bring no value as data that are well-known. In this sense, we have begun to consider the criterion of common knowledge in relation to the context, and according to it data that are well-known or self-intelligible should not be extracted and included in the assessment of the value of information.

Verifiability of data

All data were assessed as verifiable. When assessing verifiability, the difficulty of the verification process is not taken into account, but only the availability of verifiable sources and methods.

In this sense, even one piece of data that is treated here as secondary and imprecise can be verified. This means that the interested person can use appropriate methods, independently or with assistance, to limit the set of potentially applicable laws, and within that set search for the provision to which this piece of data refers.

Analysis 2

The proposition is identified further according to the criterion for identifying the part without which the whole either means nothing or means something significantly else. In this case, it was decided that the proposition is

the requirements of the service

When analyzing this sentence, we moved on to the characterization of the proposition, that is, we analyzed the intension of the concept. It was concluded that the proposition has two characteristics

they are not determined by the Chief of General Staff

they are determined by the Minister of Defense

The reason why the characteristics are separated in this way lies in the fact that in this way we can examine them equally and independently in relation to verifiability. Especially because in the defined context (the PRH and the MOD interpret the same event differently), there remains the possibility that the disagreement is not unambiguously resolved by a law or some other regulation, so the examination of one characteristic (yes – not that, but that) could be pointless and would not lead to a desired answer.

If the characteristic

they are not determined by the Chief of General Staff, but by the Minister of Defense

is found to be unverifiable/verifiable because one part of it is unverifiable/verifiable, we lose the possibility to get the result

that the characteristic

they are determined by the Minister of Defense

is verifiable

and the characteristic

they are not determined by the Chief of General Staff

is unverifiable

just as there would be the possibility that the characteristic

they are not determined by the Chief of General Staff, but by the Minister of Defense

is both true and false

which even more clearly confirms the justification of noting two characteristics.

3. Jutarnji list, 3 March 2022 (accessed electronically in May 2022)

MOD: "We will not reinstate the retired brigadier, the Administrative Court did not order us to do so."
Analysis 1

The Zagreb Administrative Court set aside the decision of the Ministry of Defense on the retirement of Brigadier Elvis Burčul, the commander of the Honorary Guard Battalion, but the Ministry of Defense said on Thursday that this was not a reason for Burčul's reinstatement.

0 Event: the judgment of the Zagreb Administrative Court concerning the retirement of Brigadier Burčul

Context: non-enforcement of the judgment of the Administrative Court concerning the retirement of Brigadier Burčul

Analysis 1

Data 1 – the Zagreb Administrative Court set aside the decision of the Ministry of Defense on the retirement of Brigadier Elvis Burčul

Data 2 – the Ministry of Defense claims that this is not a reason for reinstatement

When analyzing this sentence, the following is also examined:

– whether data 1 could also be information; whether this data, if it were to be viewed as information, can be divided into smaller units that are not well-known or self-intelligible in terms of context (analysis of the theory that the difference between data and information should be only functional, and not structural)

– whether it would be sufficient to replace the word "this" in data 2 with its content "the Administrative Court set aside the decision of the Ministry of Defense on the retirement of Brigadier Burčul" in order for that data to be identified as information

– whether it is necessary to extract, as the third piece of data, the part stating that the Ministry of Defense issued a decision on the retirement of Brigadier Elvis Burčul

In the conditions where a narrative is created in the public space through media articles or by the actions of one of the participants in the event referred to in the media article, which narrative leads to confusion regarding the interpretation of the applicable laws, the extraction of the third data would not be a bad choice. However, with regard to data 2, in which this very Ministry is making a statement, the extraction of the third data could be aimed at a more favorable/unfavorable assessment of the categories that depend on the number of data, their verifiability and quality.

Regardless of the fact that the sentence is analyzed as information, this is still done in context, so it is necessary, in order to increase the objectivity of the insight, not to extract well-known facts as data. In this

sense, there may be a problem with omitted predicates and subjects in the sentence, that is, potential nouns and verbs that, having been written in the previous sentence of the text, were not mentioned again in the sentence that is the subject of the analysis. Likewise, if the nature of data is boiled down to an anomaly, to difference, a question arises of whether a well-known fact or a fact that is clear from the context can or may contribute to the assessment of information, which further calls into question the relationship between data and information, but also the nature of data as such and information as such.

In order to assess the amount of semantic content in the third sentence, it turned out to be the most expedient to count verbs, nouns and adjectives as the bearers of content of the sentence (meaningful expressions). This is so because the writing style of an author of a text can be of high quality or of lower quality, so in this sense particles, necessary and unnecessary, as well as words superfluous as regards the content of the sentence may accumulate in the sentence. Although it is possible to examine the intention of using a certain style in order to achieve a certain goal, we decided not to engage in this type of interpretation of the text or characterization of the author, because that would conflict with the defaults of the entire research in connection with the solely thematic selection of articles, that is, the selection that does not include in the research the elements of a tacit or more visible ideological affiliation of the media, the following of certain authors, or the journalistic category or type, etc.

Verifiability of data

Here, both data were assessed as unverifiable. The first one is not verifiable because ambiguous comments are heard directly from the participants of the event (the Ministry, the President). In addition to the ambiguity of the comments, there is a lack of consistency in the statements, so that source (participants of the event) cannot be used for the purpose of data verification. Furthermore, probably due to the protection of personal data, access to the judgment through the Court's databases is not possible at the time of the research. An example of a possible change in the calculation of the value of information could be a change in the ability to access the judgment database.

The second data is not verifiable because the texts on the portals state that the Ministry replied as stated above at HINA's request. It remains to be believed that this is true, but it is impossible for any individual to find out whether it is true, on account of the legal restrictions in this case. However, the discussion about this remains open in the period during which such a statement is not denied. At the same time, it is necessary to establish a reasonable time-limit that is relevant for changing the verifiability assessment. One of the proposals is to use a time index scoring system where everything under 30 days is worth at least one point, which could analogously be applied to the relevant period in which the statement should be denied. However, such a method may be applicable *a posteriori*, and not *pro futuro*.

This raises the question of using the replies of the relevant institutions to the author's inquiries, which in a practical sense should not be problematic approach. The solution could be to attach the original statement, which can, however, then be attributed unoriginality, editing, partialness, various visual deficiencies, etc.

The time index is undetermined and undeterminable, which in this information is connected with the verifiability of the data of which the information consists. Namely, if the date of the judgment were known, there might be enough elements to verify the existence of the judgment. If it were not possible to determine the existence of the judgment in this way, the author would have presented the data more precisely and thus justifiably influenced the assessment of the value of the information.

On the other hand, the time index can remain an unknown whose possible value in the scoring system is not eliminated, but is marked as Y. The solution may also be to assign zero points due to the undetermined/undeterminable nature of time in which the zero event took place because, precisely because of that part of the data, the data cannot be characterized as precise.

The question remains whether such imprecise data can be identified as secondary. If the type of data is determined in relation to the event as it was initially conceived, then the event is the system, and the primary data are the basic data of that system, and then although this piece of data is imprecise, it is

nevertheless primary.

According to the characterization of the piece of data in relation to the defined context (non-enforcement of the judgment), both data are operational.

In conclusion, the analysis of this sentence did not establish the necessary connection between impreciseness and the secondary nature of data, moreover, it was found that there is none.

Both data were found to be correct.

Analysis 2

a) If the proposition were

the decision of the Ministry of Defense on the retirement of Brigadier Burčul

the characteristics would be

is set aside

is not a reason for reinstatement

b) If we identified the proposition

setting aside the decision of the Ministry of Defense on the retirement of Brigadier Burčul is not a reason for the reinstatement of Brigadier Burčul

and if, by extensive interpretation entailing that, if this sentence negates the possibility that this decision is a reason for the brigadier's reinstatement, somewhere there is a possible world in which the decision could be a reason for his reinstatement, we considered this sentence to be a minimal situation of a proposition identified in this manner, then this manner would connect two principles of proposition identification – the principle of situational semantics and the principle of extensive interpretation of the sentence and all its parts, that is, the interpretation of the deep structure of the sentence and its connection with external entities.

In example b), however, we are left without any characteristics of the proposition, which leads to the inability to determine its verifiability through the verifiability of the characteristics. The finding in analysis 1 of the same sentence as regards the verifiability of the data where it was determined that the data are not verifiable, and the fact that here it is impossible to determine verifiability, given that there are no characteristics, lead to the conclusion that the set of words has a low value both as information and as data.

On the other hand, by identifying the proposition in example a), we would get one characteristic which is verifiable and remains for further iterations and comparison of the results through the analyses of other sentences to reach a solution regarding the optimized manner of conducting the both analyses, both in relation to the correctness of the extraction and in relation to the practical value and focus on achieving optimal analysis of the text in order to find a model, which is the goal of the research in its entirety.

4. Jutarnji list, 3 March 2022 (accessed electronically in May 2022)

MOD: "We will not reinstate the retired brigadier, the Administrative Court did not order us to do so."

Banožić issued the decision on the retirement contrary to the proposal of Chief of General Staff Robert Hranj, who requested that the brigadier's service be extended.

Context: non-enforcement of the judgment

0 Event: the judgment of the Zagreb Administrative Court concerning the retirement of Brigadier Burčul

Analysis 1

1	Banožić issued the decision contrary to the proposal of the Chief of		
GS		2	Chief
requested that the brigadier's service be extended			

Both data are operational because they refer to the operationalization of the system in which they are found. Since that system is the defined context (non-enforcement of the judgment), the data refer to how the non-enforcement of the judgment occurs.

Both data are correct and precise.

Analysis 2

The proposition is

the decision on the retirement

in order to examine the situation where the object, in terms of the function of words in a sentence, would be a proposition as one of the iterations in the research.

The criterion that this must be a part of the sentence without which the sentence loses its meaning or its meaning is significantly changed is not met here. Even without that syntagm, the sentence still concerns a disputed issue between the minister and the chief and the fact that the Chief of GS requested the opposite. In relation to the context "non-enforcement of the judgment", this kind of proposition does not correlate syntactically, and the semantic connection is recognizable only to the extent that it is clear that the decision on the retirement preceded the judgment, and the judgment says something about the decision, but the decision does not say anything about the judgment.

After identifying the propositions in the sentence, we isolate its characteristics

contrary to the proposal of the Chief of GS

Banožić issued it

We examine whether it is justified to isolate a characteristic that meets the criterion of common knowledge in contrast to the conclusion in analysis 1 that the extraction of such data was unjustified, especially in the context that the number of characteristics, and therefore the number of verifiable characteristics that, like data, affect the assessment of the value of information, directly affects the assessment of the value of a piece of data.

Here we get a hint of the first essential difference in the usefulness of a sentence as information and a sentence as data. Namely, in both cases, both categories increase with the number of verifiable entities (data or characteristics), but unlike well-known or self-intelligible data in relation to a specific context or in relation to an external entity, all characteristics of a proposition are isolated. The reason lies in the description of the concepts of data and characteristic. Data is given as that which is an anomaly, a difference from the previous which is theoretically and practically a black dot on white paper. However, as Lyon criticized (2016) Floridi's interperetation: "the diaphoric interpretation of data lets too many things count as data" adapting the content of the concept to the framework, we tested concerned criticism and have concluded that not every "black dot on white paper" can be considered data because, within a specific context, this is not the case, as elaborated in analysis 1. Characteristics due to the property of notoriety cannot be excluded from the process of the intension of a proposition. No matter how well it meets the criterion of common knowledge, a characteristic must be extracted, so it may happen that the same set of words is not extracted as data in analysis 1, but is isolated as a characteristic of a proposition in analysis 2.

The verifiability of a characteristic is subject to the same criteria as the verifiability of data. It is crucial to define the criteria for the verification of characteristics. In doing so, it is important to keep the list of criteria open and constantly update this list. As in the case of data verification, first verification possibilities can be found in transparent logs, records, public reports, publicly verified documents, public registers and transparent actions of the actors involved, such as direct television interviews or any other similar direct source. For texts created in earlier periods method of five open sources that confirm the information can be used. Nowadays, due to lots of copying the last one is not applicable.

Comparing the two data that were extracted in analysis 1

1. Banožić issued the decision on the retirement contrary to the proposal of the Chief of General Staff

2. Chief requested that the brigadier's service be extended

and the identified proposition "the decision on the retirement", which has two characteristics, we observe that in analysis 2 an important part is lost – the description of the content of the characteristic.

We find that either the proposition is incorrectly extracted or the difference between data and information is structural and not just functional. The finding is also conditioned by the type of sentence, which is a complex sentence, so it might be good to extract two propositions.

If a proposition is incorrectly identified, and the newly attempted proposition is

the decision on the retirement is issued contrary to the proposal of the Chief of GS

so its characteristic is

Banožić issued it

the content of the second clause "who requested that the brigadier's service be extended" is lost.

However, the sentence does not lose its meaning or significantly change its meaning, but rather only the content of that which is contrary is not specified.

Or if we take into account the fact that this is a complex sentence, and not a simple sentence, and we identify two propositions

the decision on the retirement

the extension of service

and we have the characteristic of the first proposition –

contrary to the proposal of the Chief of GS

and the characteristic of the second proposition –

the Chief of GS requested

we can calculate the values of both clauses separately but show them together as a sum.

5. Dnevnik.hr, 14 March 2022 (accessed electronically in May 2022)

"Don't look up and don't look down!": A complete communication failure concerning the aircraft crash in Zagreb"

Because an official military source in its latest official appearance says that an aerial bomb was dropped on Zagreb, while an unofficial military source close to the official military source says that the bomb weighed as much as 120 kilograms and that, fortunately, it exploded underground.

Zero event: drone crash on 10 March 2022

Context: the vagueness and inconsistency of the statements of the responsible persons

Analysis 1

In this sentence, data are extracted in a way that follows their functionality

1 an official military source says that an aerial bomb was dropped on Zagreb

2 an unofficial military source is close to the official military source

3 an unofficial military source says that the bomb weighed 120 kg

4 an unofficial military source says that the bomb exploded underground

When we try to divide data 1 into smaller units that are not well-known or self-intelligible in terms of context, for the purpose of analyzing the theory that the difference between data and information should be only functional and not structural, the extraction looks as follows:

1a an official military source says

1b an aerial bomb was dropped on Zagreb

One can observe that 1a can be a piece of data in itself in the context of its basic definition as difference from the existing. However, the question still remains whether this is truly different from the situation where this would not have happened as different from the previously existing. If the General Definition of Information is an operational standard that defines information as an entity that:

1. consists of one or more data which

2. is/are well-formed and

3. is/are meaningful,

and well-formed data means data formed in accordance with the syntax of a certain system, where syntax is understood in a broader sense than the narrow linguistic understanding, which includes the determination of form, construction, composition and structuring (17), the set of words "an official military source says" can be considered as a piece of data that is well-formed with the piece of data "an aerial bomb was dropped on Zagreb". From this it would follow that data 1 (1a+1b) is information.

Parts of the sentence such as "in its latest official appearance" and "fortunately" are not extracted because no function for those parts has been observed. They are treated as redundant words, in accordance with the theory that the difference between data and information is functional, not structural. While it is evident that the part of the sentence "fortunately" does not have a semantic content of significant value, the part "in its latest official appearance", when observed as a structural unit (it has two adjectives), is not unequivocally a part that should not be calculated in the assessment of value. However, in the functional sense, it makes no semantic contribution in this sentence in the defined context (the vagueness and inconsistency of the statements of the responsible persons).

Data 2 cannot be extracted differently because it fully and directly describes a relationship. Potential data extraction such as

2a there is an official source

2b there is an unofficial source

is not shown to be purposeful. By simple observation, one can get the impression that such extraction is useful in terms of indicators of duality and difference of sources. However, in this sentence, there is no difference in the statements given by those two sources, and there is inconsistency between the statements of the responsible persons (who, by their function, are participants in the event), which is the context of this sentence. From this finding, it is evident that the participants of the event within the framework of this research do not necessarily have to participate in the event in terms of time and space, but rather the criteria for their identification are broader, and are most often related to the function they perform. In this sense, they can be parameters for placing a specific text in a thematic framework.

In conclusion, this kind of extraction does not contribute to a more accurate result either in relation to the context or in an alternative manner, in relation to an external entity.

In order to test the options, we try to analyze a combination of extractions. Data 2a confirms data 1 and, in relation to 1a and 1b, it confirms data 1a. Confirmation of data 1a that there is an official source leads to the conclusion that there is a high probability that there is also an unofficial source. Taking into account the positive dimension related to the mutual confirmation of data, which is a significantly better relative situation within the information than contradiction, such a superanalysis does not significantly affect the

results of this research. Moreover, there is potentially the possibility of an adverse effect. For example, in connection with this sentence, we would repeatedly confirm something that, in the defined context and in the light of the zero event, could be considered a well-known or self-intelligible fact, but could also lead to an increase in the number of unverifiable and imprecise data, which affects the assessment of value, but has neither a theoretical basis nor practical value.

Analysis 2

If we identify a proposition

a bomb was dropped

so the characteristics are

it weighs 120 kg

it exploded underground

it is evident that the characteristics defined the content of the concept of this specific bomb that was dropped.

If the proposition is identified in relation to the information we have about the context, and here we have a defined context (the vagueness and inconsistency of the statements of the responsible persons), we examine whether it is justified not to isolate the evident characteristic of the proposition – that it was dropped on Zagreb. The argument that the interpretation of expressions smaller than sentences requires more information about the context (13) could support the decision not to isolate the mentioned third characteristic, because there is no vagueness and inconsistency in the statements of the responsible persons regarding that part, so this characteristic as such does not support the context. However, such a finding also has its value in terms of sentence analysis and cannot cease to be a characteristic of the proposition just because it does not support the context.

If the characteristics are incorrectly isolated because it is considered pointless to isolate the characteristics of the bomb that are not related to the context, and then if the proposition is

a bomb was dropped

and we isolate the characteristic

an official military source in its latest appearance says so

we lose the content of the concept of this specific bomb that was dropped.

If the proposition

a bomb was dropped

is an incorrect choice in relation to the context

so, in the second attempt, we identify the proposition of the first of the two clauses of the analyzed compound sentence:

Because an official military source in its latest official appearance says that an aerial bomb was dropped on Zagreb

an aerial bomb was dropped

so the characteristics are

an official military source says so

it was dropped on Zagreb

and the proposition of the second clause

an unofficial military source close to the official military source says that the bomb weighed as much as 120 kilograms and that, fortunately, it exploded underground

bomb

and the characteristics

an unofficial military source close to the official military source says that it weighed 120 kg

an unofficial military source close to the official military source says that it exploded underground

we can calculate the values of both clauses, but show these values separately, taking into account the fact that these are compound sentences, as opposed to complex sentences when we can consider the possibility of addition of the obtained values as in the analysis of one of the previous sentences.

6. Dnevnik.hr, 14 March 2022 (accessed electronically in May 2022)

Don't look up and don't look down!: A complete communication failure concerning the aircraft crash in Zagreb

Not even two years of the pandemic were enough to establish a normal communication channel.

Zero event: drone crash on 10 March 2022

Context: the vagueness and inconsistency of the statements of the responsible persons

Analysis 1

Data 1 the pandemic lasted two years

Data 2 no normal communication channel has been established

The reason why data 3 "two years were not enough to establish a normal communication channel" is not extracted here lies in the examination of the relation which the author applied in this unit of research. If the intention was to emphasize the two years in which a normal communication channel was not established, it is unclear why the author emphasizes that precisely those two years are relevant for establishing that channel. If the intention is to draw a parallel between the pandemic and the establishment of a normal communication channel, that is, between the pandemic and the communication channel as such, it is not clear where the connection between the event and the pandemic is established. Given that the analysis could not establish a unit that would have significance in the research, the data were extracted in the manner shown.

If the relationship between these two data is observed, the quality of being well-formed and the meaning cannot be established, so it would be more accurate to observe data 1 as information that is not epistemologically relevant (Floridi, 2008), but is present, and data 2 as data that leans towards becoming information. However, in the defined context and in relation to the event, it has been decided that data 1, which perhaps may be observed as information, is data in this framework, given that its structure, cannot bear the features of relevance in this framework and cannot be broken down into smaller parts that. Furthermore, data 1 is observed as a piece of data that only together with another piece of data in the information belongs to the defined context. It remains to examine the function of that part of the sentence and all future equivalent parts. Later in the analysis, there will be examples of sentences as information that, regardless of the fact that they are not directly related to the context, in the case of a sufficient amount of such sentences in a complete text, can create another context that has a potential role in the characterization of the specific context.

There is no such finding in this case.

Analysis 2

The proposition is

a communication channel

and the characteristic is

in two years of the pandemic it did not become normal

In analysis 1, when this sentence was observed as information, we elaborated the reason why here we did not extract data 3 “two years were not enough to establish a normal communication channel”, but instead

Data 1 the pandemic lasted two years

Data 2 no normal communication channel has been established

We examine the relations applied by the author in this unit of research. If the intention was to emphasize the two years in which a normal communication channel was not established, it is unclear why the author emphasizes that precisely those two years are relevant for establishing that channel. If the intention is to draw a parallel between the pandemic and the establishment of a normal communication channel, that is, between the pandemic and the communication channel as such, it is not clear where the connection between the event and the pandemic is established. Given that the analysis could not establish a unit that would have significance in the research, the data were extracted in the manner shown.

Further to the elaboration in analysis 1, it now happens that the same set of words, which in analysis 1 was assessed as an entity of a vague character, is a characteristic of a proposition in analysis 2, regardless of the value of the sentence in the semantic sense.

Results of empirical analysis 1 and 2

In relation to the initial setting that was examined, which is that the difference between data and information is functional, not structural, in this framework it cannot be unambiguously concluded whether a functional or structural difference is what makes up the difference between data and information. By combining functional principles and structural principles, one sentence can be analyzed both as information and as data, and individual findings can be obtained where one part of the sentence can no longer be broken down into smaller units by applying the criterion of common knowledge, so in this framework, that is the reason why something is considered data, and not information. There are no findings in the opposite direction, because that would concern the boundary beyond which information turns into knowledge, and this was not examined.

In addition to the initial definition i.e. the general definition of information in this research, after analysis 2, the definition that information can be a subset of the data or a subset of the data augmented by additional items inferred or calculated or refined from that subset (Fricke, 2009) was deemed applicable with the exception of the part of the definition where Fricke is refering to relevant data. Unlike him, in this framework there could be only relevant information that is calculated from data that by themselves are not examined in this category.

The established correctness that there is some form of hierarchical relationship between data and information in a general sense, in this research, is more like the presentation of the DIKW hierarchy according to Clark, who imagines it in a coordinate system in which understanding is represented by the abscissa and context by the ordinate. In this presentation, data are the least comprehensible and are out of context, and through processing the other concepts of the DIKW hierarchy become more comprehensible and have a stronger contextuality (Clark, 2004). At first glance, this is in contradiction with the default framework in which the context has been examined as an important determinant when extracting and classifying data, however,

· if we look at the setting of the process in which the concept increases through the interference of different forces,

· if we take in account results when data were out of the context but placed in the information that is part of the media article that makes them, despite unknown function, indisputable part of concerned article

then the link between this framework and Clark's presentation is clearer.

The aforementioned similarity with Clark's coordinate system is supported by the findings where the data are extracted and identified by grammatical and structural analysis and their alienation from the context is proven precisely in relation to the context, when the function of alienation is examined with regard to the indisputable belonging to a specific set of words. However, even in that case, the link exists if the processing goes in both directions, that is, that the piece of data determines the meaning and scope of the information and that the information determines the meaning and scope of the piece of data. The similarity between Clark's view and the hierarchical relationship in this research is supported by the result of the empirical analysis, which is to some extent different from the initial settings according to which the context is an important determinant for data extraction, and consists in the fact that the context is consistent with the data to the extent that the sentence has a direct connection with the context. This is visible in the analysis of the sentences that have only an indirect connection with the context, which consists of their placement in the text of the article, where the extraction was performed by combining functional and structural factors using grammatical and syntactic analyses and where the relationship between the data and the context was established only *post festum*, in connection with their classification.

There are differences between the results of analysis 1 and analysis 2, that is, differences between a sentence as a piece of data and a sentence as information. Regardless of the fact that, in the view accepted by this research as well, in the hierarchy of entities, a piece of data is a lower entity than information, analyses have shown that the same set of words can be categorized as both information and data until the fulfilment of the criterion of common knowledge, due to which a set of words can only be characterized as data, and not information. We can define that set of words as data when possibility of splitting a set of words into units that can stand independently in relation to the context ends, while therefore, they do not have the character of notoriety or self-intelligibility both in relation to the context and in relation to external entities.

After such a finding, the question remains as to what the hierarchical relationship between data and information consists of if the same set of words can be characterized as both data and information under different conditions. Its importance lies in the fact that we have a standard of concepts which, although not precisely defined, is still limited by parameters, so we indisputably call the lower entities data, and the higher ones information, which enables further research into the concrete sentence meaning.

The theory about the existence of knowledge before data (Ilkka, 1999) is not insignificant because it automatically comes to mind. In this sense, we can conclude that the hierarchy can be correct, especially in this framework where the sentences that are analyzed are selected according to a certain criteria related to thematic framework that surely entails some type and level of knowledge about the particular selected sentence. The hint at the possible correctness of that theory also lies in the need for the above-mentioned additional condition for the correspondence of Clark's presentation of the DIKW hierarchy, i.e. the need to condition the processing of the concepts of the hierarchy also in the opposite direction as well as the need for parallel analyses, analysis 1 and analysis 2. However, it must be emphasized that knowledge before sentence selection and before data extraction is not the same as after (what Ilkka calls data), especially with regard to the specific goal of the research. It should thus be emphasized that the theory about the existence of knowledge before data should distinguish the type and level of knowledge before and after, that is, it should clarify whether the disagreement/difference is only terminological or substantive.

The set of words as a piece of data had certain differences in relation to the set of words observed as an information. One of those differences regards the principle determination of the characteristics of the proposition which, analogously to data as a lower entity in relation to information, is extracted as an entity lower than data as well as the examination of their verifiability. Since in analysis 2 the direction of the analysis was oriented towards the depth of the concept, and not so much towards its scope, verifiability was put in correlation in such a way that it speaks of quality when its share is detected in the sum of the shares

of precision and correctness in a piece of data as opposed to the assessment of the quality of information when the verifiability of data is the coefficient of the sum of precise and correct data in the information. The share of correctness in a piece of data is assessed in such a way as to calculate the share of correct data in the total number of data in the information in analysis 1, and analogously, the share of precision in a piece of data is assessed in such a way as to calculate the share of precise data in the total number of data. The difference that can be observed in relation to the category of average data quality in the information is that in the mentioned category the number of verifiable data has the function of a coefficient.

This is mostly due to the fact that, with regard to the isolation of the characteristics of a proposition as the here selected concept lower than data, all characteristics are registered in the finding, while with regard to the extraction of data as a concept inferior to information during analysis 1, some data were not extracted because they were found to be well-known in relation to the context or an external entity, so in this sense, in this empirical framework, they do not represent the true nature of a piece of data, that is, an anomaly in relation to the previous state. Furthermore, another contribution of analysis 2 in analysis 1 is that a piece of data observed in-depth that, when concentrated, gives different results contributes to the analysis in which the same set of words is observed as information and broken down into several parts that themselves again, solely because of the conditions of the observation, become data, but given that they are the same set of words, they necessarily become data with a lower concentration in terms of the amount of content. This also supports the finding that the nature of the difference between data and information cannot be unequivocally established.

As regards the verifiability of characteristics, a difference was made in such a way that in the set of words that was previously viewed as a higher entity, and now as a lower one, the analysis was directed more towards the depth of the concept and less towards its scope. We examine the proposition as a lower entity than the data entity, and in order to identify it we search for the core of the set of words. We examine what the difference is in assessing the verifiability of the characteristics of a proposition and the verifiability of data. In this sense, a difference was observed in the observation of a sentence as information and a sentence as data. The observed difference contributes to solving any uncertainties in analysis 1, and above all during the extraction of data from information and their characterization.

The correctness and precision of an expression is unchanging regardless of the change in form, that is, the change in the observation of the set of words. What happens is that, in order to establish quality of data correctness and precision are expressed differently, that is, in the form of a share in one entity, and not as the sum of several entities in one higher umbrella entity as it was done in order to establish quality of information.

Conclusion

a)

1. There are problems in making a definitive and unequivocal conclusion about the difference between data and information, even within a certain framework as this one. The difference between data and information in this framework of research is not reduced to one dimension, and in this sense it is not possible to distinguish whether it is a functional or a structural dimension. One set of words (sentence) was examined both as information and as data and there are differences between the results of analysis 1 and analysis 2, that is, differences between a sentence as data and a sentence as information. More important than defining the difference between data and information as structural or functional is determining line that defines what the data is. When possibility of splitting a set of words into smaller parts according to the structural principle and according to the functional principle, while applying the criterion of common knowledge, ends we define that final result as entity that can be just data and not information. In practice that looks like the possibility of splitting into units that can stand independently in relation to the context, while therefore, they do not have the character of notoriety or self-intelligibility both in relation to the context and in relation to external

entities. Beside the fact that in this research the approach to the hierarchy of entities in which data is a lower entity than information is accepted and considered confirmed, the analyzes showed that the same set of words can be categorized both as information and as data until the point when the criterion of common knowledge is met when, consequentially, a set of words can only be characterized as data, and not as information.

2. For a detailed finding in one sentence, it is good to do both analysis 1 and analysis 2 and extract the combined result as an interference of the individual ones.

3. Following on from conclusion 2, it is concluded that when the information contains elements of the context, it also contains data that is in relation to the context. However, this does not lead to the conclusion that context is a decisive parameter essential for the process of data extraction. If it were the opposite, then in the sentence that is not part of the context, while doing analysis 1, the data could not be extracted. On the contrary, it was done through a combination of functional and structural factors. After extracting data from a sentence that does not contain elements of context, the relationship between data and context can be established. The possibility should be left here that perhaps some data after the extraction process will have some component of the context core and that the positioning of the whole sentence related to the context will change to that extent.

b)

1. The typological property of secondary data is not related to the characteristic of data preciseness.

2. Verifiability of data as a coefficient most strongly affects the calculation of the value of information (verifiability coefficient), and in a practical sense, if the total sum of correctness and preciseness is greater than zero, can rehabilitate the lack of preciseness or correctness of the data that the information consists of.

3. The average quality of the data that makes up the information can be increased by increasing the number of data if the information contains at least one verifiable data (verifiability coefficient greater than zero) and if it has at least one correct or at least one precise data.

References

1. Ackoff R. L. From data to wisdom. Journal of Applied Systems Analysis 1989; 16: 3-9.

2. Bosančić B. DIKW hijerarhija: za i protiv (DIKW hierarchy: pros and cons). Vjesnik bibliotekara Hrvatske 2017; 60 (2-3):1-24.

3. Brookes B. C. The foundations of information science: Part I philosophical aspects. Journal of Information Sciences 1980; 2 (3-4): 128-133.

4. Buzan B. People, States, and Fear The National Security Problem in International Relations. Brighton, Sussex: Wheatsheaf Books, 1983.

5. Clark D. The continuum of understanding. http://www.nwlink.com/~donclark/performance/understanding.html (2004, accessed 10 December 2023)

6. Cornelius, I. 2002. Theorizing information for information science. Annual Review of Information Science and Technology, 36(1): 392–425.

7. Gray R. L. A brief historical review of the development of the distinction between data and information in the information systems literature. In: AMCIS (American Conference on Information Systems) Proceedings, 2003. 372, pp. 2843-2849

8. Floridi L. Philosophical Conceptions of Information. In: Sommaruga G (eds) Formal Theories of Information: From Shannon to Semantic Information Theory and General Concepts of Information (Lecture Notes in Computer Science, 5363). New York: Springer, 2009, pp. 19-20.

9. Floridi L. The Philosophy of Information. Oxford: Oxford University press, 2011, p. 83, 84

10. Floridi L. Is Semantic Information Meaningful Data? Philosophy and Phenomenological Research 2005; 70: 351-370.

11. Floridi L. Understanding epistemic relevance. *Erkenntnis* 2008; 69 (1): 69-92.

12. Fricke M. The knowledge pyramid: a critique of the DIKW hiearchy. Journal of Information Science 2009; 35 (2): 131-142.

13. Ilkka, T. Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. Journal of Management Information Systems 1999; 16 (3): 103-117.

14. Kratzer A. Situations in natural language semantics. Zalta. E.N.(ur.), The Stanford Encyclopedia of Philosophy, https://plato.stanford.edu/entries/situations-semantics (2010, accessed 10 December 2023)

15. Lyon A. Data. In: Paul Humphreys (ed) The Oxford Handbook of Philosophy of Science. Oxford: Oxford University Press, 2016, pp. 738-758.

16. Ma L. Meanings of Information. The Assumptions and Research Consequences of Three Foundational LIS Theories. Journal of the American Society for Information Science And Technology 2012; 63(4): 716–723.

17. Ridenour L and Moulaison-Sandy H. Give or take?: Conceptualizations of Data and its creation in Informaion Science. In: 85th Annual Meeting of the Association for Information Science and Technology, Pittsburgh, 29 October-1 November 2022, pp. 782-784.

18. Shannon C.E. Collected Papers. In: Sloane, N.J.A., Wyner, A.D. (eds.). IEEE Press,

New York, 1993, p. 180

19. Šarić, Lj. Formalna analiza značenja u prirodnim jezicima (Formal analysis of meaning in natural languages). Prolegomena: Časopis za filozofiju 2006; 5 (1): 65-88.

20. Act on the right of access to information, Official Gazette of the Republic of Croatia no. 25/13, 85/15 and 69/22)