

A Typology of Temporal Data Imperfection

Nassira Achich¹, Fatma Ghorbel^{1,2}, Fayçal Hamdi², Elisabeth Metais² and Faiez Gargouri¹

¹MIRACL Laboratory, University of Sfax, Tunisia

²CEDRIC Laboratory, Conservatoire National des Arts et Métiers (CNAM), France

Keywords: Data Imperfections, Typology of Temporal Data Imperfection, Direct and Indirect Imperfections, Natural Language based Temporal Data, Context - Dependent Temporal Data.

Abstract: Temporal data may be subject to several types of imperfection (e.g., uncertainty, imprecision..). In this context, several typologies of data imperfections have been already proposed. However, these typologies cannot be applied to temporal data because of the complexity of this type of data and the specificity that it contains. Besides, to the best of our knowledge, there is no typology of temporal data imperfections. In this paper, we propose a typology of temporal data imperfections. Our typology is divided into direct imperfections of both numeric temporal data and natural language based temporal data, indirect imperfections that can be deduced from the direct ones and granularity (i.e., context - dependent temporal data) which is related to several factors that can interfere in specifying the imperfection type such as person's profile and multiculturalism. We finish by representing an example of imprecise temporal data in PersonLink ontology.

1 INTRODUCTION

Temporal data is an important aspect in representing ontologies such as PersonLink (Herradi et al., 2015) which is a multicultural and multilingual OWL 2 ontology for storing, modeling and reasoning on family relations. However, temporal data are subject to various kinds of imperfection. For instance, we find "I lived in Berlin before changing my job", "I traveled to Spain during 2005" and "My sister married before moving to Paris". Several typologies of data imperfections have been proposed. Some are generic (e.g., typology proposed by (Bouchon-Meunier, 1993)) and others are specific to some domains (e.g., typology of imperfections adapted to the context of archaeological data proposed by (Desjardin et al., 2012)). However, to the best of our knowledge, there is no typology of temporal data imperfections. Temporal data can have more imperfections compared to the ones proposed in the generic typologies. Thus, these latter are inadequate to be adapted to temporal data.

In this paper, we propose a typology of temporal data imperfections. It is classified into direct imperfections of both numeric temporal data and natural language based temporal data, indirect imperfections that can be deduced from the direct ones and granularity (i.e., context - dependent temporal data) which is related to several factors such as person's

profile and multiculturalism. The direct imperfections can be deduced directly from the given data: uncertainty, typing error, imprecision, missing and uselessness. The Indirect imperfections are those that can be generated from the direct ones. We distinguish three types of indirect imperfections: (i) The incoherence can be generated from the uncertainty and typing error. (ii) The incompleteness can be generated from the imprecision and the missing. (iii) The redundancy can be generated from the uselessness. Granularity is the general context of the given temporal data which can determine the type of imperfection.

The rest of the paper is organized as follows. Related work is reviewed in the next section. Then, we introduce our proposed typology of temporal data imperfections in Section 3. In Section 4, we represent an example of imprecise temporal data in PersonLink ontology. Section 5 summarizes our work and gives some future directions.

2 RELATED WORK: TYPOLOGIES OF DATA IMPERFECTIONS

Related work about the typologies of data imperfections is discussed in this section. We distinguish two

types of typologies: generic typologies of data imperfections and specific typologies of data imperfections in a given domain.

2.1 Generic Typologies of Data Imperfections

We identify four generic typologies of data imperfections:

- **(Niskanen, 1989)** proposes a typology of non-precision of the data. He defines four concepts which are uncertainty, imprecision, ambiguity and generality. Uncertainty is a concept associated with the accuracy and error definitions. Imprecision exists when objects are naturally imprecise or the verbal expression can have several meanings or when a human cannot identify an object exactly. Ambiguity appears when there are several points of view about the same subject of the real world. Generality is the multiple representation of reality according to the level of detail.
- **(Bouchon-Meunier, 1993)** in her first work distinguishes two types of imperfection which are uncertainty and imprecision. In her second work (Bouchon-Meunier, 1995), she distinguishes a third factor of imperfection, which is the incompleteness. She defines uncertainty by the validity of the data. The imprecision of the data is due to the vague or approximate nature of the used semantic. Incompleteness is related to incomplete data or lack of data.
- **(Klir and Yuan, 1995)** propose a typology of data uncertainty. The authors divide the uncertainty into two classes which are fuzziness and ambiguity. Ambiguity refers to conflict and non-specificity.
- **(Smets, 1997)** establishes a classification of the imperfection of the data. It is divided into three groups which are imprecision, inconsistency and uncertainty. The imprecision is relative to the content of the data that may or may not have errors. Inconsistency is related to conflicting. Uncertainty can be subjective (i.e. observer's opinion) or objective (i.e. data properties).

2.2 Specific Typologies of Data Imperfections

We identify seven typologies of data imperfections specific to a given domain.

- **(Gershon, 1998)** focuses on kinds of imperfection in information that might be provided to an

analyst or decision maker. He proposes a taxonomy of causes for imperfect knowledge of the information state. The first level of the taxonomy contains six inputs which are incomplete info, inconsistency, info too complicated, uncertainty, corrupt data/info and the quality of the presentation.

- **(Fisher, 2005)** propose a typology of uncertainty of geographic data. They classify the data into a well or a badly defined data. In the case where the data is well defined, it is subject to uncertainty. In other cases, the data is poorly defined and the imperfection of the data is due to imprecision, ambiguity and / or incompleteness. He focuses on ambiguity. It occurs when there is a doubt about how to define an object or phenomenon. Two types of ambiguities are recognized namely disagreement and lack of specificity.
- **(Olteanu, 2006)** propose a typology based on (Fisher, 2003) and (Fisher, 2005) to classify the imperfection of a set of textual data describing ethnographic objects. Four types of imperfections are distinguished: Imprecision, uncertainty, level of detail and incompleteness. Imprecision concerns the difficulty of expressing knowledge clearly and precisely. The uncertainty concerns a doubt about the validity of information. The level of detail is knowledge presented in several granularities. Incompleteness refers to the absence of information.
- **(Casta, 2009)** establishes a typology of the imperfection of the data resulting from the economic activity. It is divided into uncertainty, imprecision and error. Uncertainty refers to the state of an agent who does not know about the future because the set of possible events, or contingencies, includes more than one element and these elements do not obey strict determinism. Imprecision refers to a lack of rigor or an operational constraint that affects measurement or, most often in the social sciences, an ambiguity inherent in the formulation of concepts. The error is defined as the dispersion of probable gains and losses around their mean.
- **(Desjardin et al., 2012)** rely on the typology of (Fisher et al., 2005) to propose a typology of imperfection adapted to the context of archaeological data. They classify imperfections into uncertainty, imprecision, ambiguity and incompleteness. The uncertainty occurs when there is a doubt about the validity of knowledge. Imprecision is the difficulty in expressing knowledge clearly. The ambiguity is the difficulty in agreeing. Incompleteness is the fact that there is miss-

ing or partial knowledge.

- (M. Snoussi, 2014) propose a classification that concerns a large set of imperfections that was built specifically to characterize spatial data. This taxonomy distinguishes three types of imperfection: Imprecision occurs when the true value is located in a defined subset of values, inconsistency is the conflict or incoherence in the values and uncertainty is the partial knowledge about the true value of information.
- (Sta, 2016) proposes several types of imperfect data during the process of information retrieval and data integration in smart cities. This imperfection can have several forms: Uncertain information reflects the lack of knowledge, imprecision information translates the non-specificity, vague information reflects an ambiguous information and missing information reflects the not found or incomplete information.

2.3 Discussion

There is a big number of typologies of data imperfection. Authors such as (Niskanen, 1989) and (Bouchon-Meunier, 1995) start by proposing generic typologies. Then, to respond to several domains, authors such as (Fisher, 2005), (Desjardin et al., 2012) and (Sta, 2016) propose typologies for specific fields. However, some correspond better to a reality than others. Also, there is not a definitive terminology and definition of terms used to qualify imperfect data, such as uncertainty and imprecision. We also note that the majority of these typologies share three common concepts which are the imprecision, the uncertainty and the incompleteness. Imperfection types are interdependent. According to (Bouchon-Meunier, 1995), "Incompleteness leads to uncertainties, imprecision can also be associated with incompleteness and they generate uncertainties during their handling". According to (Smets, 1998), imprecision always refers to incompleteness. Imprecision may be the source of uncertainty, but not necessarily (Smets, 1998).

Temporal data can have more imperfections compared to the ones proposed in the generic typologies because of the complexity of this type of data and the specificity that it contains (i.e., It can be numeric or natural language-based. It can also depend on the general context of the information that contains the temporal data and they may also be subject to several factors such as the multiculturalism). Thus, generic typologies cannot be adapted to temporal data (i.e, they are inadequate). For instance, if we have the information "I forget when last time I visited my uncle

who lives in Japan ", the temporal data indicates a "Missing" which is a type of imperfection that no one of the generic typologies contains. Another example is "The first day of the week, we will have a meeting". In this example the temporal data indicates a "circumlocution" which is another type of imperfection that no one of the generic typologies contains. Finally, to the best of our knowledge, there is no typologies dedicated to temporal data imperfection.

3 OUR TYPOLOGY OF TEMPORAL DATA IMPERFECTION

In this section, we introduce our typology of temporal data imperfections illustrated by Figure 1. Our typology is based on the studied mentioned typologies and collected real examples. We divide our typology into direct imperfections, indirect imperfections and granularity. The direct ones are those which can be deduced directly from the given data: uncertainty, typing error, imprecision, missing and uselessness. The Indirect imperfections are those that can be deduced from the direct ones. We distinguish three types of indirect imperfections: (i) The incoherence can be generated from the uncertainty and typing error. (ii) The incompleteness can be generated from the imprecision and the missing. (iii) The redundancy can be generated from the uselessness. Granularity is the general context of the given temporal data which can determine the type of imperfection.

3.1 Direct Imperfection Types

We distinguish seven direct imperfection types: uncertainty, typing error, imprecision, missing, circumlocution and uselessness which will be detailed one by one and illustrated by examples.

- **Uncertainty.** According to (Smets, 1999), each statement is whether true or false. However, the knowledge about the statement, if it is uncertain, does not allow you to decide either it is true or false. Uncertainty is partial knowledge of the true value of the data. It is the lack of information. In our typology, uncertainty has two categories. The first category includes uncertain data that respect the rules of common sense. It can be due to a doubt. As an example, "I'm not sure if it was the last summer or the one before when my aunt came from Paris to spend the holidays with us". In this example, the person is doubting about the given temporal data. However, the latter re-

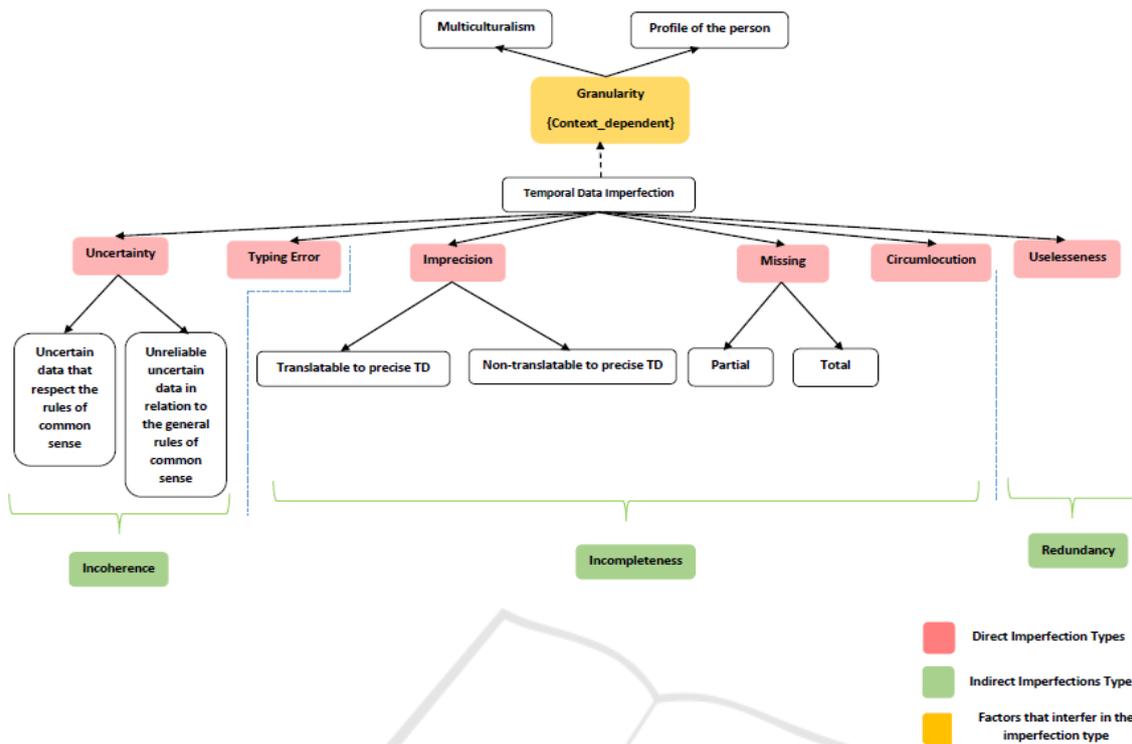


Figure 1: Our typology of temporal data imperfection.

spects the rules of common sense. The second category includes unreliable uncertain data in relation to the general rules of common sense. For example, "We swam in the sea during the new year holiday". This data does not respect the rules of common sense since the new year holidays are in winter and not in summer.

- **Typing Error.** In some cases like using a system, the entered data by the user may be subject to errors. For instance, if we have the information "The municipality employee typed my date of birth wrongly", the inserted date of birth is a typing error imperfection.
- **Imprecision.** Most of the temporal data are imprecise. Imprecision leads to the ambiguity about the information. Imprecise temporal data has two categories. The first one includes the ones that are translatable to perfect data. For example, "We moved to Tunis by 2008". "by 2008" is a numeric temporal data that can be translatable to a perfect interval that has two bounds (2007 .. 2009). Another example, if we have the information "During the summer vacation, they travelled to Spain". "During the summer vacation" presents a natural language-based temporal data which can be translatable to a perfect temporal interval [01/06 .. 31/08] since the bounds are precise. The

second category includes the ones that are non-translatable to perfect data. Several examples can be introduced, such as "I moved to my new job after my daughter's birth"; "After my daughter's birth" is imprecise since her daughter's birth is unknown. Another example is "He was sick before his death"; "Before his death" is also imprecise since we do not know how much time before his death he was sick. Another type of natural language based temporal data that is non-translatable to perfect data is the adverbs of time such as "Nowadays", "Currently", "Now" and "later".

- **Missing.** Temporal data may be subject to a missing. Missing can be either partial or total. For example, the following data "It was in March but I forget the year". In this example, the missing is partial. The person remembers the month however he forgets the year. Another example, if we have the information "I forget when last time I visited my aunt who lives in Paris". In this example, the missing is total. The person forgets totally the temporal data.
- **Circumlocution.** Temporal data can be the object of a circumlocution. For example, "The first day of the week, we will have a work meeting". The person used "the first day of the week" statement, however he means "Monday". Another ex-

ample, "The marriage is two days after the weekend". The person replaces "Tuesday" by the statement "two days after the weekend".

- **Uselessness.** This imperfection means that the given information is useless and there is not any added value to the given information. For example, if we have the information "During the Christmas in December, people celebrate all over the world", it is very enough to say "During the Christmas" since it is a general truth that the new year is obviously in December. Therefore, in this case, "in December" presents a useless temporal information.

3.2 Indirect Imperfection Types

We distinguish three indirect imperfection types: incoherence, incompleteness and redundancy. These imperfections are deduced from the direct imperfection types.

- **Incoherence.** To define the incoherence, we are based on (Gavignet et al., 2016). According to her, the incoherence refers to "the existence of contradictory data on the same object". Thus, uncertainty and typing error that can exist in a given temporal data may lead to the incoherence.
- **Incompleteness.** Missing and imprecise temporal data may lead to an incompleteness. For instance, if we have the information "My uncle will come next weekend"; "next weekend" is imprecise and, so incomplete. It could be "Saturday" or "Sunday".
- **Redundancy.** It can be concluded when uselessness interferes in the given temporal data. For instance, if we have the information "A long time ago, my grand-parents dead. It was several years ago", in this statement the person indicates two temporal data, "A long time ago" and "several years ago", that give the same meanings, which is useless and redundant.

3.3 Context - Dependent Temporal Data

Granularity presents the scale or level of detail in a set of data. The context of a given temporal data determines the type of imperfection. For instance, if we have "We can meet next weekend"; "next weekend" is imprecise. It could be "Saturday" or "Sunday" or both. However, if we have "We will spend next weekend in Berlin", it is clear from the context that "next weekend" means the whole weekend (i.e., both of Saturday and Sunday) and thus the data is precise. Besides, several factors may interfere in determining the

type of imperfection such as the profile of person and multiculturalism.

- **The Profile of the Person.** From the person themselves we can decide the type of imperfection. For instance, the temporal data given by an Alzheimer's patient is probably uncertain. For example "I am not sure it was in the morning or at night". Another example concerns the children who frequently get confused about "yesterday", "tomorrow" and "today".
- **Multiculturalism.** Multiculturalism aspect reflects the cultural differences where ideas are focused on the beliefs ways of societies. For instance, "the weekend" in Europe is "Saturday and Sunday". It differs from the one in the Arab World which is "Friday and Saturday".

4 REPRESENTING EXAMPLES OF IMPRECISE TEMPORAL DATA IN PERSONLINK ONTOLOGY

In this section, we represent some examples of imprecise temporal data in PersonLink¹ ontology based on our previous work (Ghorbel et al., 2018) in which we extend the 4D-fluents approach (Welly, 2006) to represent imprecise time intervals and crisp temporal interval relations.

Let's have the following example: "Peter was married to Stephanie just after he was graduated with a PhD and it lasts 15 years. Peter was graduated with a PhD in 1960". This example contains a natural language-based temporal data which is "just after he was graduated". It is an imprecise temporal data. Let $I = [I^-, I^+]$ be an imprecise time interval representing the duration of the marriage of Peter and Stephanie. Assume that $I^- = \{1960...1963\}$ presents the interval of the beginning bound and $I^+ = \{1975...1978\}$ the interval of the ending bound. Figure 2 illustrates a part of PersonLink ontology which represents this example.

Another example, we consider the following information: "John was married to Béatrice since about 10 years. John was married to Maria just after he was graduated with a PhD and it lasts 15 years. John was graduated with a PhD in 1980". Let $I = [I^-, I^+]$ and $J = [J^-, J^+]$ be two imprecise time intervals representing, respectively, the duration of the marriage of John and Béatrice and the one of John and Maria.

¹<http://cedric.cnam.fr/hamdif/ontologies/files/PersonLink.html>

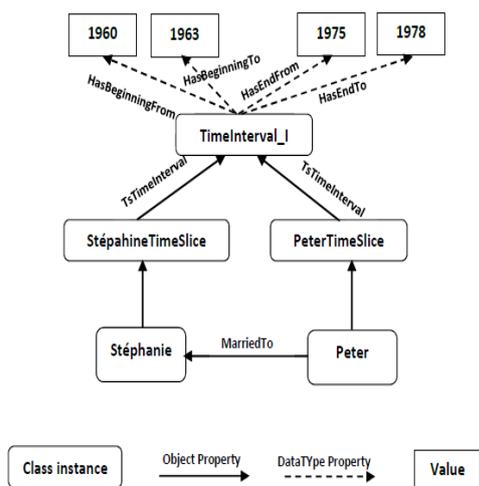


Figure 2: Representation of the first example of imprecise temporal data in PersonLink.

Assume that $I^- = \{2007...2009\}$, $I^+ = 2018$, $J^- = \{1980...1983\}$ and $J^+ = \{1995...1998\}$. Figure 3 illustrates this example.

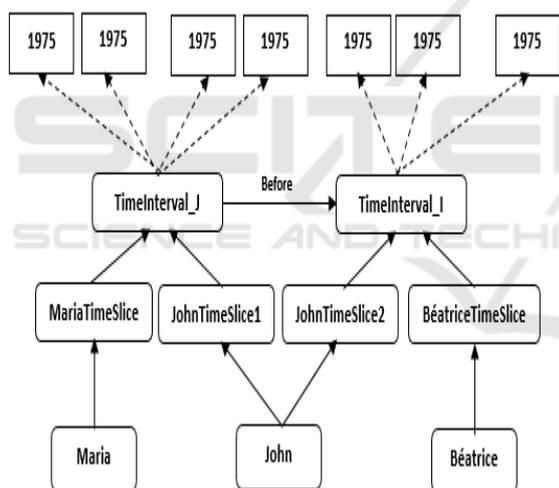


Figure 3: Representation of the second example of imprecise temporal data in PersonLink.

5 CONCLUSION AND FUTURE WORK

In this paper, we introduce our typology of temporal data imperfection. It is classified into direct imperfections of both numeric temporal data and natural language based temporal data, indirect imperfections that can be deduced from the direct ones and granularity (i.e., context - dependent temporal data) which is related to several factors such as person’s profile and multiculturalism.

Direct imperfections consist of uncertainty, typing error, imprecision, missing, circumlocution and uselessness. Indirect imperfections are generated from the direct ones. Uncertainty and typing error lead to the Incoherence. Imprecision and missing lead to incompleteness. Uselessness leads to redundancy.

We finish by representing an example of imprecise temporal data in PersonLink ontology.

As for continuity of our work, we plan to pursue on treating the other proposed types of temporal data imperfections in our typology starting by the uncertainty.

REFERENCES

Bouchon-Meunier, B. (1993). La logique floue.

Bouchon-Meunier, B. (1995). La logique floue et ses applications.

Casta, J. F. (2009). Incertitude et comptabilité.

Desjardin, E., Nocent, O., and De Runz, C. (2012). Prise en compte de l'imperfection des connaissances depuis la saisie des données jusqu'à la restitution 3d. *established by: Mauro Cristofani and Riccardo Francovich*, (supplémento 3):385–396.

Fisher, P., C. A. . R. R. W. (2005). Nature de l'incertitude pour les données spatiales. pages 49–64. *Qualité de l'information géographique*, Paris : Hermès et Lavoisier.

Fisher, P. (2003). Models of uncertainty in spatial data. pages 191–203. *Geographical information systems*.

Gavignet, É., Leclercq, É., Cullot, N., and Savonnet, M. (2016). Raisonner en logique modale sur l'incertitude liée aux données-application en archéologie. *Revue Internationale de Géomatique*, 26(4):467–490.

Gershon, N. (1998). Visualization of an imperfect world. pages 43–45. *IEEE Computer Graphics and Applications*, 18(4).

Ghorbel, F., Métails, E., and Hamdi, F. (2018). A crisp-based approach for representing and reasoning on imprecise time intervals in owl 2. In *International Conference on Intelligent Systems Design and Applications*, pages 640–649. Springer.

Herradi, N., Hamdi, F., Métails, E., Ghorbel, F., and Soukane, A. (2015). Personlink: an ontology representing family relationships for the captain memo memory prosthesis. In *International Conference on Conceptual Modeling*, pages 3–13. Springer.

Klir, G. J. and Yuan, B. (1995). *Fuzzy sets and fuzzy logic: theory and applications*, volume 574. Prentice Hall PTR New Jersey.

M. Snoussi, P.-A. D. (2014). Methodological proposals to handle imperfect spatial and temporal information in the context of natural hazard studies. pages 495–517. *International Journal of Geomatics and Spatial Analysis* 3-4, 23.

- Niskanen, V. A. (1989). Introduction to imprecise reasoning, uncertainty, decision making and knowledge engineering.
- Olteanu, A. M., M. S. R. A. . D. E. (2006). L'apport de données spatiales pour une base de données ethnographique. Actes de Conférence SAGEO.
- Smets, P. (1997). Imperfect information: Imprecision-uncertainty en uncertainty management in information systems: from needs to solutions. pages 225–254.
- Smets, P. (1999). Imperfect information: Imprecision-uncertainty en uncertainty management in information systems: from needs to solutions. amuhai motro and phillip smets.
- Sta, H. (2016). Quality and the efficiency of data in “smart-cities. Future Generation Computer Systems.
- Welty, C., . F. R. (2006). A reusable ontology for fluents in owl. pages 226—336. 4th Int’l Conf. on Formal Ontology in Information Systems (FOIS 2006).

