A well-founded ontological framework for modeling personal income tax

Isabella Distinto
Laboratory for Applied Ontology, ISTC-CNR
Trento, Italy
distinto@loa.istc.cnr.it

Nicola Guarino
Laboratory for Applied Ontology, ISTC-CNR
Trento, Italy
guarino@loa.istc.cnr.it

Claudio Masolo
Laboratory for Applied Ontology, ISTC-CNR
Trento, Italy
masolo@loa.istc.cnr.it

ABSTRACT

In this paper we present an ontological framework for modeling the core concepts of personal income taxes, based on the Italian law. The ontological analysis focuses on an high-level conceptualization of the main principles of tax legislation, and is largely based on contributions from legal doctrine and the most relevant Italian Constitutional Court's decisions in tax law. As such, the model may serve as a framework to be specialized by further ontological modules. In addition to the core ontological concepts of tax domain, an emphasis is given to the norms application process, which we believe helps to explain the complicated way taxes are imposed. In our approach, the final result of this process is a tax position, which accounts for the relationship between the taxpayer and the treasury with respect to a particular tribute.

Keywords

Tax systems, Personal Income Tax, Ontological analysis, Fiscal normative system

1. INTRODUCTION

One of the most influential economist of 20th century said that "the public finances are one of the best starting point for an investigation of society, especially though not exclusively of its political life" [21]. And this is certainly true. No other activity of modern States impacts and affects the everyday life of citizens like taxation. Taxes influence the political choices of people and their behaviors, their feelings toward social inequalities and each other's needs, their understanding of public services: in other words, their idea of *the social contract*. Therefore, investigating the ontological nature of taxation means investigating a relevant aspect of society and also a crucial relationship between Governments and citizens.

From another perspective, tax information systems represent a strikingly complex example of digital administration, since they typically work with an extraordinary amount of data of different nature: not only accounting data, but also personal data, cadastral information, enterprise data and so on. So, a shared ontology for this large amount of knowledge would be certainly beneficial for the improvement of data exchange and integration between partner administrations and, as a result, it would play a crucial role for checking credibility of tax returns, and for fighting tax evasion and fiscal frauds.

However, despite their obvious benefits, several methodological issues hinder the development of tax ontologies. The tax domain is usually recognized as a complex legal domain, which presents serious challenges from an ontological point of view.

First of all, it must be emphasized that the existence of taxes depends by law. The statement *no taxation without representation* is indeed one of the basic principle of taxation in modern democracies and means that tributes may be applied only if there is a pact agreed within a representative assembly of citizens, through law. Thus, since norms define tributes and their scope of application, all attempts of modeling an ontology of taxation should focus on an accurate examination of fiscal normative systems, which are typically encoded by large, very technical and detailed legal corpora.

In addition, fiscal normative systems often include a wide range of exceptions and exemptions, which constitute a not negligible part of the legal discipline. Furthermore, given the strong impact of fiscal regulatory norms on States' financial affairs, usually the matter is subject to many changes, especially during economic crisis periods. Hence, taking into account both the instability and the complexity of the matter, the modeler of fiscal ontologies should strive to uncover, identify and disclose the unchangeable essence at the very heart of the tax domain, finding a balance between current and future legislative frameworks and between simplicity in design requirements and complexity of underlying represented knowledge.

There is also another consideration that needs to be taken into account: understanding the principles of taxation touches the basic principles of public finance science and economics. Therefore, we may look at ontological analyses of such basic principles to develop an ontology of taxation. Unfortunately, however, as far as we know, the literature on ontological foundations of finance science and economics is very scarce. Indeed, ontology of economics has been mostly based on economic epistemology and methodology, rather than on the investigation and the analysis of the fundamental nature of economic reality. Ontology of economics is usually conceived as a tool for the verification of theories and mod-

ICAIL '13, June 10 - 14 2013, Rome, Italy

els and, as a consequence, the subject matter of ontological analysis is often confused with the method used for investigate it [20]. This is a pity, because since economic phenomena are similar in every country, a core tax ontology based on such phenomena could act as a bridge across country- and language-specific legal ontologies [14]. On the other hand, it is also true that the basic notions which inspire any tax system, such as the distinction between direct and indirect taxes, the progressively criterion or the ability-to-pay principle, are shared by many jurisdictions, and, therefore, any attempt to clarify the ontological foundations of such principles can serve to economics too, as an effective tool for understanding fiscal policies and strategies.

In this paper we propose an ontological approach for modeling a specific type of tribute: the personal income tax, as regulated by the Italian law. The model we propose is based on a foundational approach aimed at high-level conceptualization of the domain driven by an analysis of the Italian fiscal normative system, the most relevant decisions of the Italian Constitutional Court, and some relevant contribution of legal doctrine.

The rest of paper is structured as follows. Section 2 outlines previous work on modeling taxes. Section 3 introduces the core ontological model of personal income taxes while Section 4 introduces more details regarding the Italian case and discusses some possible extensions of our model. Finally, in Section 5 we draw the conclusions and a possible future work path.

2. PREVIOUS WORK

Despite the relevance of the matter, there are only few dated works concerning the modeling of tax knowledge. The majority of these works belong to the classical AI & Law field, and very few explicitly address ontological issues. Nevertheless, a review of previous works allows us to remark some methodological considerations underlying the ontological approach we have chosen to follow.

The most thorough attempt at modeling income tax is probably that of Sherman [24]. The author considered the Canadian Income Tax Act as ideally suitable to computerization because of its complete and explicit explanation of fact-situations which trigger tax effects. So, the research goal of this work was to construct a computer program which could analyze all of the facts of a case and determine their legal results as defined by the Income Tax Act. The research followed the trend pioneered by Sergot, Kowalsky and colleagues on representing legislation as logic programs [13, 23, 22]. Sherman was rather optimistic in considering the task manageable, despite his awareness of the typical problems characterizing the representation of legal knowledge, such as the "open textured" structure of legal concepts and the defeasibility of legal rules. Indeed, he was quite confident in counteract to McCarty's critique [17] to Hellawel's approach [10], which was quite similar to his own.

Those critiques were based on the experience gained in a project that is a classical piece of AI & Law literature, namely TAXMAN [16], whose goal was capturing legal reasoning in the corporate reorganisation taxation domain (as regulated by the 1954 US Internal Revenue Code). To achieve this goal, McCarty represented and made explicit the structure of relevant legal concepts by means of a "semantic network", defending the need of "deep conceptual models" [17] to properly address the issue of frequent changes in tax legislation. In a certain sense, TAXMAN can be considered therefore as an early type of legal ontology.

Twenty years later, the same position was at the basis of the work by Melz and Valente [18]. They argued for the necessity of developing an ontology of the U.S. Internal Revenue Code (IRC), because such ontology (p. 653)"would enable the construction of a broad range of intelligent applications, including automatic auditing software, robust on-line help systems, and tax question-answering systems". The authors limited themselves to the ontological representation of a very specific normative provision of the IRC[1], suggesting that the same work should have been done with all the Code in order to have an ontology capable to support the development of a tax question answering system. Furthermore, they insisted that (p. 659) "not only would it be necessary to model *each of the roughly 50,000 rules in the IRC*, but it would be necessary to model a huge amount of world knowledge"[2]. For this task, they suggested to reuse an upper ontology such as Cyc[3] or a linguistic frame resource like FrameNet[4], but, for a high-quality rendering of the IRC, they claimed in addition the need of (p. 659) "a massive amount of human engineering", even with the contribution of experts, or through sophisticated NL learning techniques.

Looking at this history, we may conclude therefore that to solve the problems emphasized by McCarty, namely the frequency of changes in fiscal legislation and the intrinsic complication of the tax domain, a very large knowledge base resulting form some kind of brute force approach is mandatory. However, we believe that, although of course detailed legal knowledge cannot be ignored, we can attack the tax complexity problem by focusing on the fundamental categories inspiring the Tax Law, rather than formalizing each single rule. In this way, we may expect to characterize a relatively stable core ontological framework, which will be easily specialized by detailed modules, which are, in principle, more subject to changes.

In this foundational perspective, it is important to stress that the role of legal doctrine is definitely invaluable. The role of legal doctrine is indeed to show the overall vision of the fiscal normative system, by mediating between the frequent changes to the tax system motivated by contingent economic needs and the stable normative principles of taxation, such as those encoded in the Italian Constitution. For this reason, the analysis of the most relevant decisions of the Italian Constitutional Court is equally essential.

Such foundational approach may also be used, as general pattern, for comparing and integrating legislation among different jurisdictions. An attempt in this direction has been proposed [4] under the E-POWER project (European Program for an Ontology based Working Environment for Regulations and Legislation). The goals of the E-POWER project were the implementation of knowledge management solutions to improve the enforcement of law in administrative processes and to test the quality and impact of draft legisla-

[1]Precisely the IRC Title 26, Subtitle A, Chapter 1, Subchapter B, Part II, Sec. 109 that states: "Gross income does not include income (other than rent) derived by a lessor of real property on the termination of a lease, representing the value of such property attributable to buildings erected or other improvements made by the lessee."

[2]Our emphasis.

[3]www.cyc.com

[4]https://framenet.icsi.berkeley.edu/fndrupal/

tive proposals. The project was funded by the Dutch Tax and Customs Administration (DTCA), and one of its outcomes was an UML ontology of tax legislation. Since the project was not focused on the ontology in itself, but rather on the development of applications based on this ontology, there are no research papers or deliverables describing it.

More recently, another project, namely the POIROT project has been undertaken to develop a 'VAT topical ontology'[11]. The 'VAT topical ontology' aims at covering the VAT domain, as defined in the EU legislative framework, in order to support VAT applications, such as information extraction, document management or question answering systems. Because of these stated purposes, the 'VAT topical ontology' is more aimed at disclosing the relations between the VAT domain relevant terms, rather than at investigating the nature of VAT relevant facts or VAT taxpayers position.

3. THE CORE MODEL

3.1 Preliminaries

Taxes are typically divided into two categories: direct and indirect taxes. Even though we think that our core model can be applied to different kinds of taxes, in this paper we shall focus on the Italian personal income tax (namely the IRPEF[5]), which is an example of direct tax. Direct taxes are levies on incomes, on real estate possessions, or in general on wealth of natural or legal persons. Unlike indirect taxes, which are imposts on consumption or expenditures and other kinds of events that only indirectly reveal wealth, direct taxes are strictly based on the *ability-to-pay principle*.

This principle implies that direct taxes have a progressive rate structure based on the overall personal "situation" rather than on a single defined fact revealing wealth, such as the received income. For example, besides the earned income, which is the necessary condition for tax application, the IRPEF tax depends on a composition of diverse facts relevant for determining the ability to contribute, such as medical expenses, or personal and familiar states (such as the number of dependent children). These facts are taken into consideration by specific rules of favor, which determine the admissibility of deductions from the taxable income or from the gross tax. Finally, specific rules (often separated from the previous ones) determine the actual amount of such deductions.

Focusing on the *constitutive* aspects of tributes (leaving aside the *formal* aspects, concerning e.g. payment rules), we can distinguish three basic kinds of rules: (*i*) rules that define the necessary conditions for the tax to be applied; (*ii*) rules that define the conditions for exceptions, exemptions, deductions or credits; (*iii*) rules that collectively determine the tax position and, in particular, the *quantum*, i.e. the actual value of the imposition.

In the proposed core-model, illustrated in Figure 1, we focus our analysis on the *process* of applying a particular tax (IRPEF in our case), with the purpose of making explicit all the aspects that contribute to determine a taxpayer's status with respect to the fiscal law. Such process has two main parts: the *selection* of the facts relevant to the particular tax and the *work-out* of the taxpayer's position with respect to such tax, on the basis of the relevant facts. The selection process isolates the facts such that the necessary conditions for the tax application hold, and collects all the data necessary to determine the exceptions, exemptions, and deductions. These data are then worked out on the basis of the norms that apply to the specific case, and, as a result, the actual *tax position* is determined, i.e., the overall taxpayer's status relative to the considered tax. The application process and the tax position are analyzed in detail in Sections 3.1.1 and 3.1.2.

The model in Figure 1 is an Entity-Relationship schema in Crow's Foot Notation.[6] To better characterise their intended nature, core entities and relationships appearing in the model are labeled with top ontological categories from the DOLCE ontology [15]. Conventionally, a label with the form '`A : B`' means that `A` subsumes `B`, i.e. $\texttt{B}(x) \rightarrow \texttt{A}(x)$. The basic DOLCE categories used in the model are summarised below (for more details and the formal characterisation see [15]):

- `ED`, *endurants* (also called *objects*) are *wholly* present at any time they are present, e.g., persons, houses, holes, laws, amounts of gold, etc.
- `PD`, *perdurants* (which here we shall also call *facts*[7] extend in time by accumulating different *temporal parts* so that, at any time, they are *partially present* (their past and future phases are not present), e.g., football matches, eatings, conferences, John possessing an house, etc.. Perdurants include *events* (which can be either *accomplishments* or *achievements*), *states*, and *processes* (see below).
- `APO`, *agentive physical objects* are endurants that have a direct localization in space and to which we ascribe *intentions*, *beliefs*, and *desires*, e.g., persons or animals.
- `ASO`, *agentive social objects* are endurants to which we ascribe *intentions*, *beliefs*, and *desires* but are only indirectly localized in space trough the physical endurants they depend on, and are created by a community of agents, e.g., institutions, companies, presidents, etc.
- `NASO`, *non-agentive social objects* are endurants with no direct spatial location, no intentionality, but, similarly to `ASO`s they are created by a community of agents, e.g., laws, economic systems, currencies, assets, etc.
- `ST`, *states* are perdurants that are both *cumulative* and *dissective*, i.e. the sum of two states of kind `K` is still a `K`-instance as well as all their parts, e.g., John sitting here, John possessing an house, etc.
- `ACC`, *accomplishments* are perdurants that are non-cumulative and non-atomic, e.g., conferences, ascents, performances, etc.

[5]Regulated by D.p.r. 22 december 1986, no. 917, Title I.

[6]In the following we shall adopt the extended ER model, which includes the *isa* relation.

[7]We are aware that there is a technical difficulty in this choice, since the ontological status of facts is still debated in analytic philosophy, and strictly speaking they should not be confused with perdurants. For instance, a precedence relationship between two perdurants is a fact, but does not correspond to a perdurants itself. However, in the everyday language the term "fact" is commonly used in the legal jargon to refer to things which happen in time, and we prefer to stick to the common practice. Context will certainly help to recognize the situations where the term has a slightly different meaning

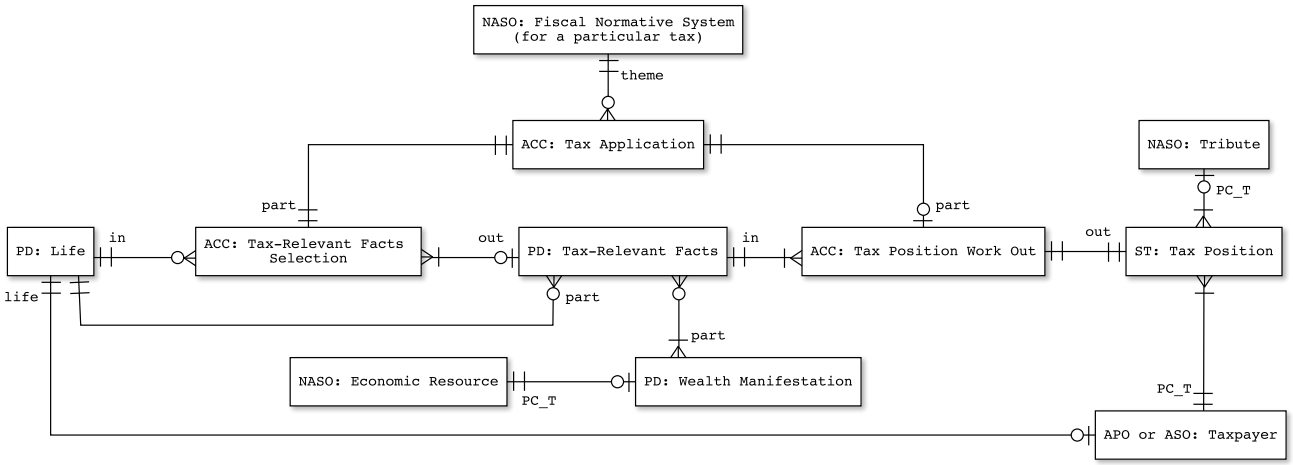

Figure 1: The core model.

- `ACH`, *achievements* are perdurants that are non-cumulative and atomic, e.g., departures, deaths, arrivals, etc.

3.1.1 The tax application process

A tax application process is modeled as an *accomplishment*, having as *theme* a system of fiscal norms, which may include all the different kinds of rules mentioned above (see Figure 1). A fiscal normative system is intended as the collection of all the norms that regulate a given tax (e.g., the Presidential Decrete 22 december 1986, no. 917, Title I), including the norms that are explicitly cited or implicitly assumed (e.g., the labour law or the family law), as well as the background systems (e.g., the Italian Constitution).

As we have seen, a tax application process can be split—at least conceptually—in two parts: a *selection* process, which isolates the facts relevant to the specific tax taken into account, including those needed to establish the ability-to-contribute, and a *work out* process, which determines the actual tax position.

The input of the selection process is the whole taxpayer's life during the fiscal year, and the output is the sum of all relevant facts. Among such facts, at least one wealth manifestation must occur. It must be emphasized that different taxes address different kinds of wealth manifestations. For instance, the value added tax (VAT) considers consumptions. One can argue that consumptions *indirectly* reveal the wealth of a person. However there are taxes that consider facts whose connection with the wealth of a person (or organization) is very weak or controversial. For example, consider the stamp duty (in Italy the so called "*imposta di bollo*") imposed for the validation of a document which certifies the identity and nationality of its holder, such as a passport: what wealth the request of a passport reveals?

3.1.2 The Tax Position

The final result of the tax application process is a specific tax position, which in DOLCE terms is a *state* involving at least two participants: the *taxpayer* and the *tribute*, i.e. a quantitative index that concisely represent the actual imposition for the considered fiscal year. The taxpayer is either an Agentive Physical Object (`APO`) (more specifically, a physical person), or an Agentive Social Object (`ASO`) (usually a company), while the tribute (typically an amount of money) is a Non Agentive Social Object (`NASO`) with a conventional nature and dependent on a community of agents that create and accept it.

Being a state, any tax position has a temporal extension. We assume that the tax position starts at the end of fiscal year and never ends. It records the taxpayer's fiscal situation related to a particular fiscal year. This assumption clarifies the fact that the tax position does not coincide with the taxpayer's *obligation* to pay some amount of money before a deadline: it just attests the way the taxpayer is taxed (relatively to a given tax) during a given fiscal year. For instance it represents the fact that John is subject to an IRPEF-imposition of 100 euro for the fiscal year 2012. Clearly a tax position can generate an obligation-state—e.g., the John's being obliged to, or John's having the obligation to, pay the tribute of 100 euro to the treasury before the 15th March 2013—but this is different from the tax position.

The choice of introducing explicitly this notion of tax position in our model has several motivations. *First*, our notion is perfectly compatible with tributes of value zero. This is the case of *exemptions* (see Section 4.2 for a detailed explanation). *Second*, while the temporal extension of the tax position is clear, there is a deep doctrinal debate concerning the time at which the obligation arises. For some scholars, the obligation arises automatically upon the occurrence of a fiscally relevant fact that is classified as taxable and it is not defeated by other circumstances manifesting an exempt status (declarative theory) [8, 27]; for other scholars, instead, the tax declaration is an essential part of the tax application process and, so, the obligation to pay the tribute arises only after the submission of the tax declaration (constitutive theory)[1, 25]. The cases of taxpayers that are employees are particularly tricky. The employer, indeed, is committed by law to pay the tribute on behalf of its employees, by deducting the amount from the taxpayer's salary. In this case, when does the employee's (taxpayer's) obligation arise? *Third*, it is easy to extend our core model by adding some participants to the tax position. For instance, in addition to the tribute, one can introduce the quantification of the ability-to-contribute, or a set of more detailed indexes

like the quantification of earnings or expenses, the deduced expenses and the taxed earnings, etc. In this way, the tax position becomes a report, a balance, which reflects all the relevant aspects of a taxpayer's position *vis-à-vis* a given tax system. Fourth, in this way we keep clearly distinct the tax position from the the tribute's payment implementation, which often involves a complex series of obligations and rights, like in the case of instalment payments, which require ad hoc models.

3.2 Axiomatization

After the general discussion above, let us now further characterize the intended semantics of the scheme reported in Figure 1. We shall first clarify the semantics of our graphical notation, and then introduce more specific constraints.

3.2.1 Semantics of the core model

Differently from classical ER modeling, we find it natural to allow in our schema for several occurrences of the same relation. For instance, in Figure 1, the `out` relation links both `Selection` with `RelevantFacts` and `WorkOut` with `TaxPosition`. This choice implies that, even if a relation only occurs once in the scheme (as is the case for `theme`), we still assume it may also hold somewhere else in a possible extension of the schema. Because of this fact, for a generic relation `r` from `A` to `B`, we cannot simply assume $\mathtt{r}(x,y) \rightarrow \mathtt{A}(x) \wedge \mathtt{B}(y)$ nor the weaker $\mathtt{A}(x) \wedge \mathtt{r}(x,y) \rightarrow \mathtt{B}(y)$.

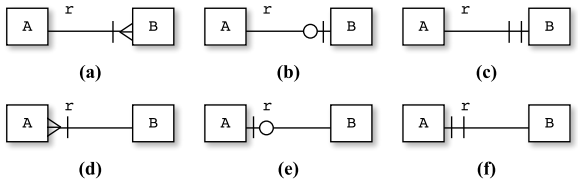

Figure 2: Different cases of cardinality constraints.

With reference to the different cases of cardinality constraints reported in Figure 2, we adopt therefore the semantics reported below[8]:

(a) $\mathtt{A}(x) \rightarrow \exists y(\mathtt{B}(y) \wedge \mathtt{r}(x,y))$
(b) $\mathtt{A}(x) \wedge \mathtt{r}(x,y) \wedge \mathtt{r}(x,z) \wedge \mathtt{B}(y) \wedge \mathtt{B}(z) \rightarrow y = z$
(c) $\mathtt{A}(x) \rightarrow \exists! y(\mathtt{B}(y) \wedge \mathtt{r}(x,y))$
(d) $\mathtt{B}(x) \rightarrow \exists y(\mathtt{A}(y) \wedge \mathtt{r}(y,x))$
(e) $\mathtt{B}(x) \wedge \mathtt{r}(y,x) \wedge \mathtt{r}(z,x) \wedge \mathtt{A}(y) \wedge \mathtt{A}(z) \rightarrow y = z$
(f) $\mathtt{B}(x) \rightarrow \exists! y(\mathtt{A}(y) \wedge \mathtt{r}(y,x))$

Let us start the analysis of the constraints expressed in the schema in Figure 1 by discussing the nature of the Tax Application process, which plays a key role in our core model. Note, first of all, that most of the entities appearing in Figure 1 are temporal entities, i.e. DOLCE perdurants. This is the result of a precise methodological choice: we are convinced that focusing on perdurants and their participants helps understanding what really *happens* in our domain. In particular, we believe it is useful to systematically reify temporally qualified relations as *states* that exist at the times the relations holds and have as participants all the entities that fulfill the arguments of these relations (see Section 4.2 for more details). States (and sums of states) are useful to describe both the tax position and the tax-relevant facts of the taxpayer's life.

The participation relationships reported in the schema are all DOLCE *total participations* ($\mathtt{PC_T}$) ([15], definition Dd65), holding between an endurant which participates to a perdurant throughout its duration. The relation between `Taxpayer` and `Life` is a refinement of the DOLCE *life* relation (`lf`) ([15], definition Dd68) that takes into account only the perdurants in which a person totally participates *during the fiscal year of interest*. This means that the life of a person (during a given year) is the mereological sum of all the perdurants to which the person totally participates (during that year). The `theme` relation between `FiscalNormativeSystem` and `Application` is also a total participation in this case, but this does not hold in general, i.e., `theme` does not always imply $\mathtt{PC_T}$ (see Section 4.1 for a counterexample). The relation `part` is the (non-temporary) *parthood* relation of DOLCE: it is defined between perdurants, it has no temporal qualification, and it obeys to the axioms of *classical extensional mereology* (in particular it is transitive, see [6] for more details).

Taking into account these constraints, the schema in Figure 1 makes explicit that not all selections[9] have an output; however, when a selection has an output, then it is unique. On the contrary, all work-out processes have an output, an unique tax position, that, in its turn, always requires a work-out process. In addition, all selections and work-out processes are part of a unique tax application. Vice versa, a tax application has always an unique selection part, but not always a work-out part (in the case the selection has no output). All tax applications have a theme—the normative system they apply—and the same system could be applied multiple times.

The result of the selection process is the sum of all tax-relevant facts in the taxpayer's life, relative to a particular tax. Such facts are defined and regulated by the normative system. In this respect, it is important to note that the Italian Constitution establishes that "*No obligation of a personal or financial nature may be imposed to any person except by law*" (Art. 23 It. Cost.)[10], therefore any tax is necessary imposed and regulated by a statutory law. Furthermore, the Italian Constitution assumes the ability-to-contribute (or ability-to-pay) principle: "Every person shall contribute to public expenditure in accordance with their capability." (Art. 53 It. Cost.)[11], therefore the taxpayer's life must actually manifest the ability-to-contribute. According to the Italian Constitutional Court, the ability-to-contribute should be understood as the recognition of a "*def-*

[8]We don't consider the case of a zero-or-more to zero-or-more relation `r` since it is not very informative (it just means that it is possible to have a relation `r` between the two classes). Note that in our schemas there are no constraints of this kind.

[9]For brevity, we will call *selection* a selection process. Similarly for the other kinds of processes.

[10]"*Nessuna prestazione personale o patrimoniale può essere imposta se non in base alla legge.*", official translation source: http://www.quirinale.it/qrnw/statico/costituzione/pdf/costituzione_inglese_01.pdf.

[11]"*Tutti sono tenuti a concorrere alle spese pubbliche in ragione della loro capacità contributiva.*", official translation source: http://www.quirinale.it/qrnw/statico/costituzione/pdf/costituzione_inglese_01.pdf.

inite manifestation of wealth" (*"manifestazione determinata di ricchezza"*)[12], i.e., a wealth manifestation necessarily belonging to one of the kinds explicitly foreseen by the considered tax. Clearly, different taxes can address different kinds of wealth manifestations (see Section 4.1 for more details). Note that the choice of grounding our model on perdurants results useful here because, by means of participation relations, we are able to represent in a unified way all the data necessary for the tax position work-out. For instance, introducing an earning event implicitly tells us who earned, how much the person earned, when the earning occurred, etc. This also allows us to start from a quite abstract minimal model that considers the tax position work-out as a black box that still represents the core mechanism of a tax application, abstracting for instance from the various rules dealing with exemptions and detractions.

3.2.2 Further Constraints

Some constraints cannot be expressed by using the graphical notation. Here we add them as first order logic axioms.

(C1) A selection process with an output requires the existence of a work-out process (operating on that output) that is part of the same tax application:

$$\texttt{RelevantFacts}(f) \wedge \texttt{out}(f,s) \wedge \texttt{part}(s,a) \wedge \texttt{Application}(a) \rightarrow \exists w(\texttt{WorkOut}(w) \wedge \texttt{in}(f,w) \wedge \texttt{part}(w,a))$$

Vice versa, the input of a work-out process that is part of a certain norm application a needs to have been also selected by the same a:

$$\texttt{RelevantFacts}(f) \wedge \texttt{in}(f,w) \wedge \texttt{part}(w,a) \wedge \texttt{Application}(a) \rightarrow \exists s(\texttt{Selection}(s) \wedge \texttt{out}(f,s) \wedge \texttt{part}(s,a))$$

Note that there is no reference to *who* is actually applying the norms by executing first the selection and then, eventually, the work-out process[13]. Similarly, there is no reference to *how* the data about the life of a person are collected (for example through a self-declaration). The taxpayer's life represents what a person has done during the fiscal year. The tax application process acts as a filter, to select all and only the facts which are relevant from the perspective of a particular tax and finally produce (after having worked out the details) a tax position, which in a sense is a picture of the final state the taxpayer has with respect to the treasury. If the taxpayer's life does not satisfy the necessary conditions for the tax to be applied, then the tax application process just coincides with a simple check–a selection without output. If the necessary conditions are satisfied, then the selection has an output that implies a work-out process and a tax position.

Note that it is possible to have different work-out processes that consider the same facts as input, for instance when the same income is multi-taxed. In this case we need to exclude the possibility to have the same income taxed two times by the *same* system of norms [14]

$$\texttt{WealthManifestation}(m) \wedge \texttt{part}(m,l_1) \wedge \texttt{part}(m,l_2) \wedge \texttt{in}(l_1,w_1) \wedge \texttt{in}(l_2,w_2) \wedge \texttt{part}(w_1,a_1) \wedge \texttt{part}(w_2,a_2) \wedge \texttt{theme}(n_1,a_1) \wedge \texttt{theme}(n_2,a_2) \rightarrow n_1 \neq n_2$$

(C2) The tax position of a taxpayer p is the result of the tax application to the life of p:

$$\texttt{TaxPayer}(p) \wedge \texttt{life}(l,p) \wedge \texttt{in}(l,s) \wedge \texttt{Selection}(s) \wedge \texttt{out}(f,s) \wedge \texttt{in}(f,w) \wedge \texttt{WorkOut}(w) \wedge \texttt{out}(x,w) \rightarrow \texttt{PC}_\texttt{T}(p,x)$$

4. EXTENDING THE MODEL: THE ITALIAN PERSONAL INCOME TAX

The previous section focused on a core model that, in our opinion, is general enough to be applied to different tax-systems. In this section we focus our attention on the Italian personal income tax called IRPEF. The proposed modelling strategies are still quite general but the examples considered come from IRPEF.[15]

4.1 Selecting the relevant facts

In general, the sum of *relevant facts* selected by the application of a certain tax has parts of different kinds. We already said that, in principle, only wealth manifestations are taxed –this is why, in our core-model, at least one wealth manifestation must be part of the relevant facts. However, to establish the taxpayer' ability to contribute (or some specific allowances) and her/his tax position, other facts are usually necessary, e.g., *personal expenses*, *social circumstances*, or additional *personal data*. Many types of expenses are considered by the IRPEF law as eligible for gross tax deduction (in some cases, within specified limits of expenditure). Such fiscally relevant expenses are, for example, mandatory social security contributions paid to the social security authorities or to complementary pension funds, alimony payments to a spouse from whom the taxpayer is legally separated or divorced, medical expenses, but also charitable donations. In addition to these, social circumstances such as being a pensioner or a temporary employee, or having dependent children, are selected and regulated by norms of favor, as well.

Similarly to wealth manifestations, expenses are *perdurants* that necessarily have an economic resource as participant. In addition, they also have explicit *themes* (in the sense of the `theme` relation), the 'things' one has payed for, that can be goods (that, in general, are endurants) but also services (that are perdurants). For instance, medical expenses could regard purchases of drugs or medical examinations. Notice that, differently from other thematic roles, we believe that the `theme` relation is not a specialization of participation (which is defined only between endurants and perdurants): the theme of a event may be not a participant

[12]The concept of the "manifestazione determinata di ricchezza" is a jurisprudential result based on several decisions of the Italian Constitutional Court, namely the decisions n. 45/1964; n. 50/1965; n. 92/1972; n. 144/1972; n. 201/1975.

[13]Alternatively one can assume that it is a State's delegatee that applies the system of norms. Actually this is again an ideal perspective because it presupposes that the delegatee accesses all the data without mistakes.

[14]The so called *double taxation* of IRPEF is explicitly excluded by the Italian law (art. 163 d.p.r. no. 917/1986). The double taxation cases are particularly interesting from an international tax law perspective, where two different systems of norms belonging to two different jurisdictions could tax the same income of a person. However, it must be emphasized that many agreements between States regulate these cases in order to avoid double taxation.

[15]When there is no risk of ambiguity, we shall use IRPEF to refer both to the tax and the system of norms defining it.

to that event. So both goods and services can be considered as themes of an expense event, even if services are perdurants.

Social circumstances and personal data are more tricky to be explicitly taken into account in the model. Our choice is to represent them as (mereological sums of) *states* (in the sense of DOLCE) that are part of the taxpayer's life. To explain the rationale behind this choice, let us consider, for example, the case where John has two dependent children, Luc and Mary, during the fiscal year $\mathtt{t}$. In first order logic, this situation is usually represented by the formula $\mathtt{DepChild}(\mathsf{luc}, \mathsf{john}, \mathtt{t}) \wedge \mathtt{DepChild}(\mathsf{mary}, \mathsf{john}, \mathtt{t})$. To be able to refer to such situation, we need to *reify* this formula.

To do so, we start from the two atomic formulas, above, concerning each one child, and represent them as two different *states* in which both the child and the parent participate (clearly with different roles) and both temporally located at the time $\mathtt{t}$. The reification of the complex formula can then be obtained by summing up the two atomic states[16]. This strategy allows us to represent wealth manifestations, expenses, and social circumstances in an homogenous way: they are all facts (i.e., perdurants).

Let us remark that this is not a new idea. In [12], Kim introduced what he calls *events*. An event is the exemplification of a property P (a relation R) by a 'substance' s (several substances $s_1, \ldots, s_n$) at a given time t and it is noted $[s, P, t]$. For instance in "the collision of the Titanic with the iceberg", s_1 = Titanic, s_2 = the iceberg, R = colliding with. An event $[s, P, t]$ exists if and only if 's has P at t', i.e. in a logical perspective, if and only if $P(s, t)$. The events $[x, P, t]$ and $[y, Q, t']$ are identical if and only if $x = y$, $P = Q$, and $t = t'$.[17] Our *states* are very similar to the *events* of Kim, but (i) we assume other kinds of perdurants, like changes, that cannot be reduced to states[18]; and (ii) we do not reify the properties in the domain of quantification.

All the facts that result from the selection process are relevant by definition (of course with respect to a particular normative system). This means that our core model just makes explicit that the selection 'filters' what the taxpayer has done during the fiscal year according to what established in a given normative system, without stating what the *specific kinds* of facts considered to be relevant are.

To address this aspect, we have to consider several issues. *First of all.* we should remember that the result of the selection process is the simple collection (mereological sum) of all the facts relevant for the given system. Because of that, the 'filter' implemented by the selection can act only on the *intrinsic* properties of the facts it receives in input[19]. It sequentially analyses single facts without considering their interrelations. These interrelations are taken into account only by the work-out process. *Second*, the filter relies on notions introduced by laws and regulations *external* to the tax regulation as, for instance, the labour law, the house law or the family law –that however, according to our definition, are part of the fiscal normative system and can then be used by the selection process. *Third*, the selection does a first *classification* of facts. For instance, IRPEF distinguishes different kinds of incomes, expenses, personal data, etc. *Fourth*, taxes not only apply to specific kinds of wealth manifestations but also to specific kinds of taxpayers. For instance, IRPEF is a *personal* income tax, it applies to persons not to companies. The model must address also this constraint.

In the following the case of IRPEF will be used to motivate some axioms representing specific cases of the previous constraints in our framework. Although these axioms illustrate the general mechanisms underlying IRPEF, they are not to be intended as a complete representation of IRPEF. On the other hand we think that these constraints can be easily adapted to other taxes, therefore they also have a conceptual value.

In our framework, there exist two ways for representing the fact that IRPEF is a personal income: (i) assume the existence of a pre-check, a check done before the application of the normative system, or (ii) assume that the check is done by the selection process. In both cases, the following constraint holds (where irpef is the constant that represents the IRPEF *system* of norms not only the IRPEF regulation):

$$\mathtt{Relevant}(x, \mathsf{irpef}) \wedge \mathtt{part}(x, l) \wedge \mathtt{life}(l, y) \rightarrow \mathtt{Person}(y)$$

but in the case of the first assumption, one needs to introduce

$$\mathtt{Selection}(s) \wedge \mathtt{Application}(a) \wedge \mathtt{theme}(\mathsf{irpef}, a) \wedge \mathtt{part}(s, a) \wedge \mathtt{in}(l, s) \wedge \mathtt{life}(l, y) \rightarrow \mathtt{Person}(y)$$

IRPEF considers the following (general) *kinds* of wealth manifestations: *employment earnings*, *self-employment earnings*, *capital earnings*, *business earnings*, *miscellaneous earnings*, and *real estate possessions. First of all*, note that most of these wealth manifestations do not include the *possessions* (as opposed to the earnings) of economic resources. IRPEF focuses on *incomes*, i.e. on the various forms of earning economic resources. For instance, the possession of a given capital is not taxable, only the interests and dividends coming from this capital are taxable. Relying on the notion of *possession* of an economic resource (at a given time), one could assume that a wealth manifestation of a resource r during a fiscal year implies the non-possession of r during the previous year. However, there are possessions, namely *real estate possessions*, that are taxable according to IRPEF even though they do not produce income –for instance, when the possessed house is not let. In this case, IRPEF relies on a mechanism that, on the basis of some characteristics of the real estate, *conventionally* assigns a kind of 'virtual income' to the possession [20]. *Second*, note that the previous kinds of wealth manifestations are actually defined independently from the tax regulation (the IRPEF in this case). Indeed, these concepts are established by the norms that regulate work's contracts, estate possessions, etc. The tax system must necessarily refer to these external notions to *ground* the criteria for selecting the relevant facts.

By assuming a set of predicates that represent all the kinds of wealth manifestations taken into account by IRPEF, it is

[16] In this way logical disjunction is not easy to represent. This is not a real limiting factor in our case, however.

[17] From the logical perspective this implies that we need to specify the identity condition of predicates. In general one assumes logical equivalence, but more intensional identity criteria could also been considered.

[18] This was also one argument against the approach of Kim that considered only events.

[19] Note that the properties of the participants to these facts are also included in their intrinsic properties.

[20] This virtual aspect complicates the analysis of the ontological nature of the wealth manifestation, which is only partially taken into account in this work.

easy to formalise the previous constraint (to individuate the wealth manifestations relevant for IRPEF):

$$\texttt{WealthManifestation}(w) \wedge \texttt{Relevant}(w, \mathsf{irpef}) \rightarrow (\texttt{EmploymentEarning}(w) \vee \texttt{SelfEmploymentEarning}(w) \vee \texttt{CapitalEarning}(w) \vee \texttt{BusinessEarning}(w) \vee \texttt{MiscellaneousEarning}(w) \vee \texttt{RealEstatePossess}(w))$$

It is trivial to adapt this formula to account for the kinds of expenses and personal or social facts relevant for IRPEF or for another tax systems.

One could also introduce sufficient conditions like:

$$\texttt{part}(w, l) \wedge \texttt{life}(l, p) \wedge \texttt{Person}(p) \wedge (\texttt{EmploymentEarning}(w) \vee \texttt{SelfEmploymentEarning}(w) \vee \texttt{CapitalEarning}(w) \vee \texttt{BusinessEarning}(w) \vee \texttt{MiscellaneousEarning}(w) \vee \texttt{RealEstatePossess}(w)) \rightarrow \texttt{Relevant}(w, \mathsf{irpef})$$

that makes explicit which kinds of (personal) tax manifestations have to be necessarily taken into account by IRPEF. Even though it is debatable, we do not exclude the case that the same wealth manifestation results to be relevant for more than one tax system.

4.2 Working out the tax position

Relevant facts could be further filtered by taking into account the interrelations they have with other selected facts. This refinement is necessary to individuate the *taxable* wealth manifestations and the *deducible* expenses.

For instance, suppose John is the owner of an house that he let out. In this case, John has two wealth manifestations: the house possession –as we have seen, the Italian tax system arbitrarily assigns a 'virtual income' (the so-called *rendita catastale*) to the house possession alone– and the rental earning. However, IRPEF considers as taxable only one of these wealth manifestations, the one that generates the greater income[21]. So, once a real estate possession has been selected as relevant for IRPEF, to understand if it is taxable, one needs to check if other relevant wealth manifestations have been selected (earnings from location in this example), and take them into account in the proper way (compare the economic resources in this example). This constraint can be represented in the following way, where $\texttt{Relevant}(x, p, t)$ stands for "the fact x *regarding the payer* p is relevant for the tax system t" and $x \geq_I y$ stands for "the economic resource that participates in x is greater than the one that participates in y" [22]:

$$\texttt{Relevant}(x, p, t) \triangleq \texttt{Relevant}(x, t) \wedge \exists l(\texttt{life}(l, p) \wedge \texttt{part}(x, l))$$

$$\texttt{TaxableRealEstatePossess}(x, \mathsf{irpef}) \triangleq \exists p(\texttt{Relevant}(x, p, \mathsf{irpef}) \wedge \forall y(\texttt{Relevant}(y, p, \mathsf{irpef}) \wedge \texttt{RentalEarning}(y) \rightarrow x \geq_I y))$$

Similarly in the case one finds a rental earning.

At the end of this refinement subprocess, we have all the *non-excluded* wealth manifestations (with the corresponding economic resources) that, according to IRPEF, are classified under one (and only one) type, of taxable incomes or deductible expenses. At this point, some *calculation* subprocess is needed to individuate the gross incomes and expenses of all the different types, the tax- and deduction-rates (specific of the previous types of incomes or expenses), and to finally calculate the net total tax. Typically, to establish the tax rate, IRPEF does not consider single incomes but considers the sums of incomes of the same type, i.e., the tax-rate too can be established only by taking into account the global situation of the taxpayer. For instance, one can assume that the rate for incomes from employment is 15% if the total employment income is less than 10.000 euros.

According to the legal doctrine [7, 26], while *exclusions* are incomes that are not classified (by IRPEF-types, in our case), i.e., they are excluded *a priori*, *exemptions* are incomes that are classified as taxable but not taxed, i.e., they are excluded *a posteriori*. Exclusions are not taxable, exemptions are taxable but not taxed. For instance, if the sum of all incomes coming from wealth manifestations of a given type is lower than a given threshold then these wealth manifestations, in principle taxable, are not taxed.

Similarly, in addition to excluded expenses, there are expenses that are classified as deducible but are not deducted. For instance, IRPEF establishes that the medical expenses that are deducted are those which exceed a given threshold (with the deduction-rate of 20%). In this case, all the expenses for purchases of drugs or medical examinations are classified as deducible, while only one part of them are *deducted* (from the gross income or the gross tax) by applying to them the corresponding deduction-rate.

5. CONCLUSIONS AND FUTURE WORK

In this paper we have presented an ontological framework for modeling the core concepts of personal income taxes, based on the Italian law. We have given a lot of emphasis to the norm application process, since we believe it helps to explain the complicated way taxes are imposed, by distinguishing two main steps: selection of the relevant facts, and work out of the actual tax. In our approach, the final result of this process is a tax position, which accounts for the relationship between the taxpayer and the treasury with respect to a particular tribute.

The tax position has been modelled as a *perdurant*, together with with the various events and states of fiscal relevance which constitute what we have collectively called the *relevant facts*. Indeed, the presence of so many perdurants in our model is the result of an explicit choice, which helps to understand the nature of our domain by distinguishing between events and states on one hand, and their participants on the other hand.

For instance, we have considered incomes as economic resources participating in wealth manifestations, and we have considered deductions (such as expenses and social circumstances), exclusions and exemptions as events as well.

The ontological analysis focuses on an high-level concep-

[21] While we were writing this paper, the IRPEF law has been modified by the introduction of another tax, the so called IMU (*Imposta municipale propria*), a local tax on real estate which has substituted IRPEF for the case of mere real estate possession, while IRPEF still applies only to rental earning. We have decided to show this previous case anyway because, in our opinion, it is particularly interesting and also because it is a real example of the way in which fiscal knowledge changes continuously. Note also that the calculation in the example shown (related IRPEF before the IMU introduction) is more complex, but the present simplification is enough for our illustrative goal.

[22] Here we assume that, for a given house, during a fiscal year it is possible to have just one rental earning. To manage the possibility of getting multiple rental earnings from one single real estate, one needs to introduce a sum between economic resources.

tualization of the main principles of tax legislation, and is largely based on the contributions of legal doctrine and on some relevant decisions of the Italian Constitutional Court. As such, the model may serve as a framework to be specialised by further ontological modules.

In this regard, a useful future scenario may be the recognition of links between the ontological classes and the norms which define them, by using legal XML formats such as CEN Metalex [3] or Akoma Ntoso [5, 2] for the specification of legislative documents or a rule interchange language for the legal domain, such as LKIF-rule [9] or LegalRuleML [19] for the representation of legal norms' structure.

Finally, it may be important to investigate the epistemological aspects concerning tax declaration, in order to recognize discrepancies between the actual facts and the declared facts. This investigation will be certainly beneficial to several applications in tax evasion fighting.

5.1 Acknowledgements

Part of this work has been done in the framework of a cooperation with SOGEI SpA, the society in charge of the Italian fiscal information system, with the purpose of developing a methodology for common conceptual modelling for the fiscal context.

6. REFERENCES

[1] E. Allorio. *Diritto processuale tributario*. Milano, 1942.

[2] G. Barabucci, L. Cervone, M. Palmirani, S. Peroni, and F. Vitali. Multi-layer markup and ontological structures in akoma ntoso. In Casanovas et al. [5], pages 133–149.

[3] A. Boer. Metalex naming conventions and the semantic web. In *Proceedings of the 2009 conference on Legal Knowledge and Information Systems: JURIX 2009: The Twenty-Second Annual Conference*, pages 31–36, Amsterdam, The Netherlands, The Netherlands, 2009. IOS Press.

[4] A. Boer, T. van Engers, and R. Winkels. Using ontologies for comparing and harmonizing legislation. In *Proceedings of the International Conference on Artificial Intelligence and Law (ICAIL'2003)*, pages 60–69. ACM Press, 2003.

[5] P. Casanovas, U. Pagallo, G. Sartor, and G. Ajani, editors. *AI Approaches to the Complexity of Legal Systems. Complex Systems, the Semantic Web, Ontologies, Argumentation, and Dialogue - International Workshops AICOL-I/IVR-XXIV Beijing, China, September 19, 2009 and AICOL-II/JURIX 2009, Rotterdam,The Netherlands, December 16, 2009 Revised Selected Papers*, volume 6237 of *Lecture Notes in Computer Science*. Springer, 2010.

[6] R. Casati and A. Varzi. *Parts and Places. The Structure of Spatial Representation*. MIT Press, Cambridge, MA, 1999.

[7] F. Fichera. *Le Agevolazioni Fiscali*. CEDAM, Padova, Italy, 1992.

[8] A. D. Giannini. *Il rapporto giuridico d'imposta*. Milano, 1937.

[9] T. F. Gordon. Constructing legal arguments with rules in the legal knowledge interchange format (lkif). In P. Casanovas, G. Sartor, N. Casellas, and R. Rubino, editors, *Computable Models of the Law*, volume 4884 of *Lecture Notes in Computer Science*, pages 162–184. Springer Berlin Heidelberg, 2008.

[10] R. Hellawell. A computer program for legal planning and analysis taxation of stock redemptions. *Columbia Law Review*, 80(7):1363–1398, 1980.

[11] K. Kerremans and G. Zhao. Topical ontology of vat: Ff poirot deliverable 2.3. Technical report, CVC, VUB, 2005.

[12] J. Kim. Events as property exemplifications. In M. Brand and D. Walton, editors, *Action Theory*, pages 159–177. Reidel, Dordrecht, 1976. reprinted in Casati and Varzi, 1996.

[13] R. A. Kowalski. Legislation as logic programs. In G. Comyn, N. E. Fuchs, and M. Ratcliffe, editors, *Proceedings of Logic Programming in Action, Second International Logic Programming Summer School, LPSS '92*, volume 636 of *Lecture Notes in Computer Science*, pages 203–230. Springer Verlag, 1992.

[14] D. Liebwald. Semantic spaces and multilingualism in the law: The challenge of legal knowledge management. In P. Pompeu Casanovas, M. A. Biasiotti, E. Francesconi, and S. M. T., editors, *Proceedings of the 2nd Workshop on Legal Ontologies and Artificial Intelligence Techniques*, volume 321 of *CEUR Workshop Proceedings*, pages 131–148. CEUR-WS.org, 2007.

[15] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, and A. Oltramari. Wonderweb deliverable d18. Technical report, CNR, 2003.

[16] L. T. McCarty. Reflections on taxman: an experiment in artificial intelligence and legal reasoning. *Harward Law Review*, 90:837–893, 1977.

[17] L. T. McCarty. Intelligent legal information systems: Problems and prospects. *Rutgers Computer and Technology Law Journal*, 9(2):265–94, 1983.

[18] E. Melz and A. Valente. Modeling the tax code. In R. Meersman, Z. Tari, and A. Corsaro, editors, *On the Move to Meaningful Internet Systems 2004: OTM 2004 Workshops*, volume 3292 of *Lecture Notes in Computer Science*, pages 652–661. Springer Berlin Heidelberg, 2004.

[19] M. Palmirani, G. Governatori, A. Rotolo, S. Tabet, H. Boley, and A. Paschke. Legalruleml: Xml-based rules and norms. In F. Olken, M. Palmirani, and D. Sottara, editors, *Rule - Based Modeling and Computing on the Semantic Web*, volume 7018 of *Lecture Notes in Computer Science*, pages 298–312. Springer Berlin Heidelberg, 2011.

[20] J. Potts. Ontology in economics. In R. Poli and J. Seibt, editors, *Theory and Applications of Ontology: Philosophical Perspectives*, pages 277–285. Springer Netherlands, 2010.

[21] J. A. Schumpeter. *The Economics and Sociology of Capitalism*. Princeton University Press, 1991.

[22] M. J. Sergot. Representing legislation as logic programs. In J. E. Hayes, D. Michie, and J. Richards, editors, *Machine intelligence 11*, pages 209–260. Oxford University Press, Inc., New York, NY, USA, 1988.

[23] M. J. Sergot, F. Sadri, R. A. Kowalski, F. Kriwaczek, P. Hammond, and H. T. Cory. The british nationality

act as a logic program. *Communication of the ACM*, 29(5):370–386, 1986.

[24] D. M. Sherman. A prolog model of the income tax act of canada. In *Proceedings of the 1st international conference on Artificial intelligence and law*, ICAIL '87, pages 127–136, New York, NY, USA, 1987. ACM.

[25] F. Tesauro. *Il rimborso dell'imposta*. Torino, 1975.

[26] F. Tesauro. *Istituzioni di Diritto Tributario - Parte generale*. UTET, x edition, 2009.

[27] G. Tesoro. *Principi di diritto tributario*. Bari, 1938.