13. A procedural model for ontological analyses

Michael Rosemann, *Centre for Information Technology Innovation, Queensland University of Technology*
Peter Green, *UQ Business School, The University of Queensland*
Marta Indulska, *UQ Business School, The University of Queensland*

Abstract

In recent years, there has been a significant increase in the popularity of ontological analysis of conceptual modelling techniques. To date, related research explores the ontological deficiencies of classical techniques such as ER or UML modelling, as well as business process modelling techniques such as ARIS or even Web Services standards such as BPEL4WS, BPML, ebXML, BPSS and WSCI. While the ontologies that form the basis of these analyses are reasonably mature, it is the actual process of an ontological analysis that still lacks rigour. The current procedure is prone to individual interpretations and is one reason for criticism of the entire ontological analysis. This paper presents a procedural model for ontological analysis based on the use of meta models, multiple coders and metrics. The model is supported by examples from various ontological analyses.

Introduction

As techniques for conceptual modelling, enterprise modelling, and business process modelling have proliferated over the years (e.g. Olle et al., 1991), researchers and practitioners alike have attempted to determine objective bases on which to compare, evaluate, and determine when to use these different techniques (e.g. Karam and Casselman, 1993; Gorla et al., 1995) . However, throughout the 1980s, 1990s, and into the new millennium, it has become increasingly apparent to many researchers that without a theoretical foundation on which to base the specification for these various modelling techniques, incomplete evaluative frameworks of factors, features, and facets will continue to proliferate. Furthermore, without a theoretical foundation, one framework of factors, features, or facets is just as justifiable for use as another (e.g. Bansler and Bodker, 1993).

Wand and Weber (1989; 1990; 1993; 1995) have investigated the branch of philosophy known as ontology as a foundation for understanding the process of developing an information system. Ontology is a well-established theoretical domain within philosophy dealing with identifying and understanding elements of the real world. However, interest in, and the applicability of, ontologies today extends to areas far beyond philosophy. As Gruninger and Lee (2002, p. 13) point out, '…a Web search engine will return over 64 000 pages given 'ontology' as a keyword … the first few pages are phrases such as ''enabling virtual business'', ''gene ontology consortium'', and ''enterprise ontology''.'

The usefulness of ontology as a theoretical foundation for knowledge representation and natural language processing is a fervently debated topic at the present time in the artificial intelligence research community (Guarino and Welty, 2002). The use of ontologies as a basis for the analysis of techniques that purport to assist analysts to develop models that emulate portions of the real world has been growing steadily more popular. The Bunge-Wand-Weber (BWW) ontological models (Weber, 1997), for example, have been applied extensively in the context of the analysis of various modelling techniques. Wand and Weber (1989; 1990; 1993; 1995) and Weber (1997) have applied the BWW representation model to the 'classical' descriptions of entity-relationship (ER) modelling and logical data flow diagramming (LDFD). Weber and Zhang (1996) also examined the Nijssen Information Analysis Method (NIAM) using the ontology. Green (1997) extended the work of Weber and Zhang (1996) and Wand and Weber (1993; 1995) by analysing various modelling techniques as they have been extended and implemented in upper CASE tools. Furthermore, Parsons and Wand (1997) proposed a formal model of objects and they use the ontological models to identify representation-oriented characteristics of objects. Along similar lines, Opdahl and Henderson-Sellers (2001) have used the BWW representation model to examine the individual modelling constructs within the OPEN Modelling Language (OML) version 1.1, which is based on 'conventional' object-oriented constructs. Green and Rosemann (2000) have extended the analytical work into the area of integrated process modelling based on the techniques presented in Scheer (2000). Most recently, Green et al. (2003) have extended the use of this evaluative base into the area of enterprise systems interoperability using business process modelling languages like ebXML, BPML, BPEL4WS, and WSCI. Clearly, ontology is a fruitful theoretical basis on which to perform such analyses. However, while ontological analyses are frequently utilised, particularly in the area of conceptual modelling technique analysis, the actual process of performing the analysis remains problematic. The current process of ontological analysis is open to the individual interpretations of the researchers who undertake the analysis. Consequently, such analyses are criticised as being subjective, *ad hoc*, and lacking in relevance. There is a need, therefore, for the systematic identification of shortcomings of the current ontological analysis process. The identification of such weaknesses, and their subsequent mitigation, will lead to a more rigorous, objective, and replicable analytical process.

Accordingly, this paper has several objectives. First, we aim to identify comprehensively the shortcomings in the current practice of ontological analysis. The identification of such shortcomings will provide a basis upon which the practice of ontological analysis can be improved. Second, we want to develop several propositions and methodology extensions that enhance the ontological analysis process by making it more objective and structured.

There are several contributions this paper aims to make. They are based on previous experiences with ontological analyses as well as observations derived from published analyses. First, the work presents a detailed analysis of the actual process of performing an ontological evaluation. We identify eight shortcomings of the current ontological analysis process, *viz*. lack of understandability, lack of comparability, lack of completeness, lack of guidance, lack of objectivity, lack of adequate result representation, lack of result classification, and lack of relevance. Each of the identified shortcomings is then classified as belonging to one of three phases of analysis, *viz.,* input, process, and output. Second, the paper presents recommendations on how each of the shortcomings in the three phases can be overcome. The recommendations, *inter alia*, include an extended

methodology for improving the objectivity of the analysis as well as a weighting model that aims to improve the classification of the results of any ontological analysis.

The remainder of the paper is structured as follows. The next section identifies eight current shortcomings of ontological analyses that are classified with respect to the three phases of analysis. The third section provides recommendations concerning how to overcome the identified shortcomings in each of the three phases. The final section provides a brief summary of the work and outlines possible future research in this area.

Shortcomings of current ontological analyses

An ontological analysis is, in principle, the evaluation of a selected modelling grammar from the viewpoint of a pre-defined and well-established ontology. The current focus of ontological analyses is on the bi-directional comparison of ontological constructs with the elements of the modelling grammar that is under analysis. Weber (1997) clarifies two major situations that may occur when a grammar is analysed according to an ontology. After a particular grammar has been analysed, an assessment of the modelling strengths and weaknesses of the grammar can be made according to whether some or any of the following situations arise out of the analysis.

1. *Ontological incompleteness (or construct deficit)* exists unless there is at least one grammatical construct for each ontological construct.
2. *Ontological clarity* is determined by the extent to which the grammar does not exhibit one or more of the following deficiencies:

a. *Construct overload* exists in a grammar if one grammatical construct represents more than one ontological construct.
b. *Construct redundancy* exists if more than one grammatical construct represents the same ontological construct.
c. *Construct excess* exists in a grammar when a grammatical construct is present that does not map to any ontological construct.

Though this type of ontological analysis is widely established, it still has a range of shortcomings. These shortcomings can be categorised into the three main phases of an ontological analysis: preparation of the input data, the process of conducting the analysis, and the evaluation and interpretation of the results. The first two identified shortcomings refer to the quality of the input data.

Lack of understandability

Most of the ontologies that are currently used for analysing modelling grammars have been specified in formal languages. While such formalisation is beneficial for a complete and precise specification of the ontology, it is not a very natural or intuitive specification. An ontology that is not clear and intuitive can lead to misinterpretations as the involved stakeholders have problems with the specifications. Furthermore, it forms a hurdle for the application of the ontology as it requires a deep understanding of the formal language in which it is specified.

Lack of comparability

The specification of an ontology typically requires a formal syntax, which allows the precise specification of the elements and relationships of the ontology. Such specifications are required, but are not necessarily intuitive. Consequently, textual descriptions of the ontology in 'plain English' often extend the formal specification. However, even if an ontology is specified in an intuitive and understandable language, the actual comparison

with the selected modelling grammar remains a problem. Unless the ontology and the grammar are specified in the *same* language, it will be up to the coder to 'mentally convert' the two specifications into each other for comparison purposes, which adds a subjective element to the analysis. Different languages can also lead to different levels of detail and further complicate the analysis. In any case, they make a more automated comparison practically impossible. This is the typical situation in nearly all previous analyses.

The three further shortcomings identified below are related to the process of the ontological analysis and refer to what should be analysed, how it should be analysed as well as who should conduct the analysis.

Lack of completeness

The first decision that has to be made in the process of an ontological analysis is on the scope and depth of the analysis. Even though most ontologies have been discussed for many decades, they still undergo modifications and extensions. It is up to the researcher to clearly specify the selected version of the ontology and the scope and level of detail of the analysis. In our work in the area of Web Services, for example, it was often not clear what constructs form the core of the standard and, in fact, two researchers who conducted independent analyses of the same Web Services standard selected a different number of constructs.

Moreover, many ontological analyses focus solely on the constructs of the ontology and the constructs of the grammar but do not sufficiently consider the relationships between these constructs. The difficulty of clearly specifying the boundaries of the analysis, as well as the limited consideration of relationships between the ontological constructs, can lead to a lack of completeness.

Lack of guidance

After the scope and the level of detail of the analysis have been specified, it is typically up to the coder to decide on the procedure of the analysis, i.e. in what sequence will the ontological constructs and relationships be analysed? Currently, there are hardly any recommendations on where to start the analysis. This lack of procedural clarity underlies most analyses and has two consequences. First, a novice analyst lacks guidance in the process of conducting the ontological evaluation. Second, the procedure of the analysis can potentially have an impact on the results of the analysis. Thus, it is possible that two analyses of the same modelling grammar using the same ontological base, but that follow different processes, may lead to different outcomes.

Lack of objectivity

An ontological analysis of a grammar requires not only detailed knowledge of the selected ontology and grammar, but also a good understanding of the languages in which the ontology and the grammar are specified. This requirement explains why most analyses are carried out by single researchers as opposed to research teams. Consequently, these analyses are based on the individual interpretations of the involved researcher, which adds significant subjectivity to the results. This problem is further compounded by the fact that, unlike other qualitative research projects, ontological analyses typically do not include attempts to further validate the results.

The five shortcomings identified above have a common flavour in that they heavily depend on the researcher conducting the ontological evaluation. Three further shortcomings have been identified, namely lack of result representation, lack of result classification

and lack of relevance. These shortcomings are detailed below and refer to the outcomes of the analysis

Lack of adequate result representation

The results of a complete ontological analysis, i.e. representation mapping and interpretation mapping, are typically summarised in two tables. These tables list all the ontological constructs (first table) and all the grammatical constructs (second table) and the corresponding constructs of the other meta model. Such tables can become quite lengthy and are typically not sorted in any particular order. They don't provide any insights into the importance of identified deficiencies and they also don't cluster the findings.

Lack of result classification

As indicated above, it is common practice to derive ontological deficiencies based on a comparison of the constructs in the ontology and the grammar. Ontological weaknesses are identified when corresponding constructs are missing in the mapping obtained between the ontology and the grammar or one-to-many (or many-to-one or even many-many) relationships exist in the mapping between the two. Such identified deficiencies are the typical starting point for the derivation of propositions and then hypotheses. In general, the ontological analysis does not make any statements regarding the relative importance of these findings in comparison with each other. Though this seems to be the established practice, it lacks more detailed insights into the significance of the results. It is to be expected, however, that missing support for a core construct of an ontology should be rated of higher importance than missing a construct corresponding to a minor ontological construct or a relationship. This lack of a more detailed statement regarding the significance of a potential shortcoming makes it difficult to judge quickly the outcomes of the results of two different sets of analyses (e.g. an ontological analysis of ARIS compared to an ontological analysis of UML).

Lack of relevance

Finally, the results of an ontological analysis should be perceived as relevant by the related stakeholders. However, if an ontological analysis leads, for example, to the outcome that Entity Relationship Models do not support the description of behaviour then it would hardly be surprising if the IS community developed a rather critical opinion of the worth of the analysis since this is both obvious and well known. It seems that an ontological analysis has to consider the purpose of the grammar as well as the background of the modeller who is applying this grammar. The application of a high-level and generic ontology does not consider this individual context and there is a danger that the outcomes can be perceived as trivial.

Reference methodology for conducting ontological analyses

The shortcomings identified above have motivated the development of an enhanced methodology for ontological analysis. The main purpose of this methodology is to increase the rigour, the overall objectivity and the level of detail of the analysis. The proposed methodology for ontological analyses is structured in three phases: input, process and output.

Input

The formal specification of ontologies, together with the differences in the languages used to specify the ontologies and the grammars under analysis, have been classified as issues pertaining to the lack of understandability and comparability.

In order to overcome these shortcomings, it is proposed to convert the ontology as well as the selected modelling grammar to meta models using the same language (e.g. ER Models or UML Class Diagrams). This facilitates a pattern-matching approach towards the ontological analyses of completeness and clarity of a grammar. As a first step we converted, for example, the Bunge-Wand-Weber ontology into an ER-based meta model. This meta model includes 50 entity types and 92 relationship types. It has clusters such as system, property or class/kind. Such a meta model explains, in a language familiar to the information systems (IS) community, the core constructs of the ontology. It also highlights the underlying focus of the ontology. In the case of the BWW model, for example, it is obvious from a visual inspection of the meta model that the ontology is centred around the existence of a *thing*, which is the central entity type in the meta model.

The obtained meta model can now be used for a variety of ontological analyses. Moreover, it allows a critical review of the BWW model by a wider community. The approach, however, is not without its limitations. Commonly used modelling techniques such as ER or UML are often widely accepted but they have not been designed for the purposes of meta modelling. Thus, they occasionally lack the required expressiveness. Figure 13.1 provides an impression of the size and complexity of the meta model for the BWW ontology.

Figure 13.1. The BWW meta model.
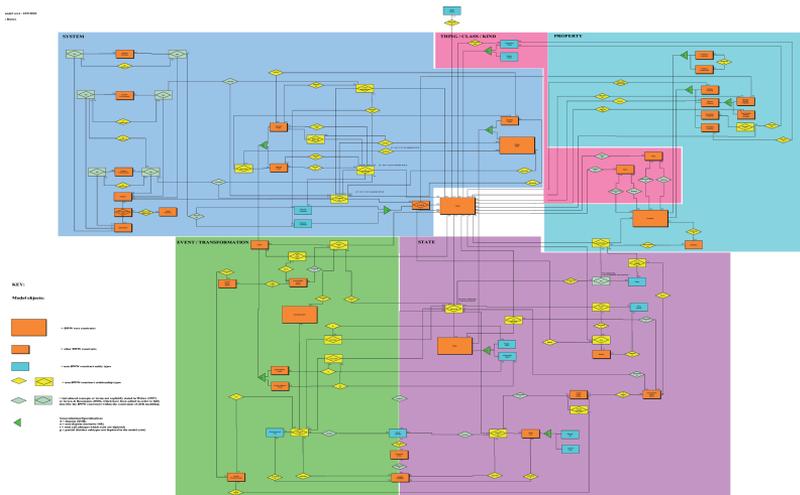
While an ER-based meta model helps to overcome issues related to the understandability of an ontology, a corresponding meta model of the analysed grammar is required to deal with the lack of comparability issue. Many popular modelling techniques (e.g. ARIS or UML, and also interoperability standards such as ebXML) are already specified in meta models using ER-notations or UML Class Diagrams. If the meta models for the ontology and the modelling technique are specified in the same language, the ontological analysis turns into a comparison of two conceptual models. As part of the analysis, it will be required to identify corresponding entity types and relationship types in both models. It also becomes immediately obvious if the paradigm of the analysed grammar differs from the ontology. In the case of ARIS or many Web Services standards, for example, the

meta models are centred around *functions* or *activities* instead of being centred around *things*.

Process

The issues related to the process of conducting an ontological analysis have been described as lack of completeness, lack of guidance and lack of objectivity.

Based on the assumption that corresponding meta models for the ontology and the analysed grammar are available, it is possible to clearly specify the scope of an analysis using those meta models. A selection of clusters, entity types and relationship types would define all elements that are perceived of relevance for the analysis. An analysis of an ER-based notation, for example, could be focused on the BWW clusters *thing*, *system* and *property* and could exclude the more behavioural-oriented clusters *event* and *state*. Such boundaries of an analysis could be easily visualised in the meta model and would provide a clear description of the comprehensiveness of the analysis.

The existence of two corresponding meta models and a clear definition of the scope of the analysis are necessary but not sufficient criteria for a well-guided process. Further guidelines are required regarding the starting point of such a process and the actual sequence of activities. Based on our experiences, we recommend starting with the representation mapping; that is, selecting the meta model of the ontology and subsequently identifying corresponding elements in the modelling grammar. The first construct to be analysed should be the most central entity type. For example, in the case of the BWW model, the entity type *thing* is the appropriate starting point. Our previous work provides a strong argument that this analysis should follow a cluster-by-cluster approach. Starting with the core constructs in a cluster allows a more structured and focused analysis of the completeness of a modelling grammar. The analysis of the entity types is followed by the relationships and the cardinalities. Constructs in the meta model that have only been introduced for reasons of correctness of the meta model, but that do not reflect ontological constructs, are excluded from the analysis. The representation mapping is followed by an analysis of the clarity of the target grammar, i.e. the interpretation mapping. In this case the meta model of the grammar under analysis is the starting point. The general procedure is similar. A primary advantage of a cluster-based analysis is that the structure of the two meta models provides valuable input for the ontological analysis. An example is the analysis of generalisation-specialisation relationships in the meta model of the grammar. We propose to ontologically classify the super-type first and then to inherit this ontological classification to all sub-types. This streamlines the process of the analysis and increases consistency.

The lack of objectivity issue, on the other hand, frequently stems from the analysis being performed by a single researcher. This situation results in an analysis that is almost certainly biased by the researcher's background as well as their interpretation of the specification of the grammar. In order to improve the validity of the analysis, a research method can be adopted that involves individual analyses of a particular grammar by at least two members of a research team, followed by discussion and hopefully consensus as to the final analysis by the entire team of researchers. The method consists of three steps:

1. *Step 1:* Using the specification of the grammar in question, at least two researchers separately read the specification and interpret, select and map the ontological constructs to candidate grammatical constructs to create individual first drafts of the analysis.

2. *Step 2:* The researchers involved in Step 1 of the methodology, meet to discuss and defend their interpretations of the representation modelling analysis. This meeting should lead to an agreed second draft version of the analysis that incorporates elements of each of the researchers' first draft analyses. The overlap in the selection of the grammatical constructs and in the actual ontological analysis can be quantified by various figures that are used in content analysis and other more qualitative research.

3. *Step 3:* The second draft version of the analysis for each of the interoperability candidate standards is used as a basis for defence and discussion in a meeting involving the entire research team. The outcome of this meeting forms the final analysis of the grammar in question.

Just such a method was employed in a project that sought to apply the BWW representation model analysis to a number of the leading potential Web Services standards: ebXML, BPML, BPEL4WS and WSCI. The project team was composed of four researchers and the standards were analysed in the order: ebXML à BPML à BPEL4WS à WSCI. Two researchers were involved in Steps 1 and 2 of the method (the individual analysis of a standard followed by a meeting of the two researchers in order to obtain an agreed mapping). This was followed by a meeting of the entire team in order to discuss the mapping and arrive at the final analysis. The process was performed for each of the four standards. Table 13.1 shows the recorded agreement statistics at the second step of the applied method while Table 13.2 shows the recorded agreement statistics at the third step of the method.

Table 13.1. Summary of Step 2 mapping agreement between both researchers

Web Service Language	Construct Mapping agreed upon by both researchers	Total number of specification constructs identified	Mapping conference
ebXML	43	51	84%
BPML	36	46	78%
BPEL4WS	30	47	63%
WSCI	39	49	79%

Table 13.2. Summary of Step 3 mapping agreement

Web Service Language	Construct Mapping agreed upon by the team	Total number of specification constructs identified	Mapping conference
ebXML	49	51	96%
BPML	41	46	89%
BPEL4WS	42	47	89%
WSCI	46	49	94%

The adoption of such a method can be seen to have greatly improved the objectiveness of the carried-out analyses.

Output

The three main shortcomings related to the outcome of an ontological analysis have been characterised as the lack of adequate result representation, lack of result classification and the lack of relevance.

The meta models, which have been used as input for the ontological analyses, are an appropriate medium to visualise the outcomes of the entire analysis process. In our work on the analysis of ARIS, we derived a meta model of the BWW model that highlighted all constructs of the ontology that did not have a corresponding construct in the grammar under analysis. That is, we visualised incompleteness in the model using simple colour

coding. In a similar way, we derived three ARIS meta models that highlighted excess, overload and redundancy in ARIS. Such models form a very intuitive way of representing the identified ontological shortcomings. The underlying clustering of the models also helps to quickly comprehend the main areas in which there are shortcomings.

At the present time, the process of an ontological analysis results in the identification of ontological incompleteness and ontological clarity through the identification of missing, overloaded or redundant grammatical constructs. While the end result identifies such problems, it fails to account for their relative importance. For example, *thing* is one of the fundamental constructs of the BWW model. Therefore, a lack of mapping to a modelling grammar for this construct should be considered a more important shortcoming than the lack of mapping for, say, the *well-defined event* construct. There is a need for the development of a scoring model that enables the calculation of the 'goodness' of a grammar with respect to the ontology. In such a scoring model, each of the ontological constructs has a value assigned to it that reflects the relative importance of the construct in the ontology. Core constructs would therefore have high weightings whereas less important constructs would attract lower weightings. Following an ontological analysis of a particular grammar, the weighting of all missing constructs would be calculated to arrive at one value that generally reflects the outcome of the analysis.

An example for such a classification could have the following structure. All core constructs of an ontology (and the modelling grammar) would get the value one. All other constructs represented as an entity type in the meta model of the ontology would receive the value 0.7, and all other constructs get the value 0.3. Such a weighting would then be applied to the outcomes of the ontological analysis. The scores would be aggregated across the ontology and modelling grammar. They could also be calculated separately for completeness, excess, overload and redundancy. Furthermore, they could be aggregated per cluster, which allows a more differentiated view of the particular strengths or weaknesses of a modelling grammar. Though the consolidated score of such an evaluation should not be overrated, it provides better insights into the characteristics of the ontological deficiencies and provides a first rating of the significance and importance of the identified shortcomings.

Apart from the lack of result classification that is addressed by the scoring model, another problem with the outcome of the analyses has been the perceived lack of relevance of the results. Since most modelling grammars focus on modelling a subset of the phenomena that occur in the real world, it would follow that not *all* constructs of an ontology are necessary in order to analyse such a grammar. If the full ontology is used in the analysis, the result may identify potential problems that would not, in reality, occur, because the modelling grammar is not used to model any phenomena described by the missing constructs. Further, there may also be a need for specialisation of some of the ontological constructs in order to enhance analysis of a grammar pertaining to a particular domain.

Figure 13.2. An extension of ontological analysis through the use of focused ontologies.
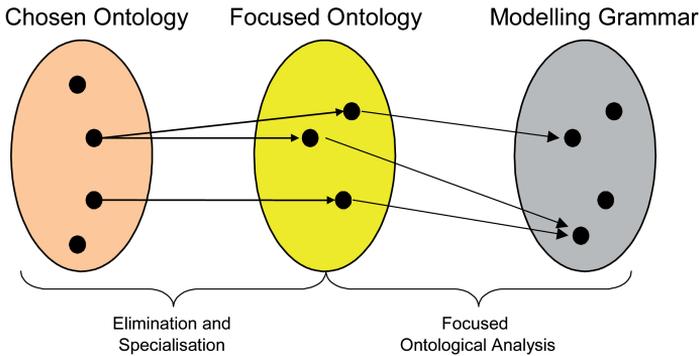
Indeed, the outcomes of the ontological analyses of different modelling grammars to date appear to support the need for a focused ontology, which consists of different subsets of the ontological constructs for different domains. The analyses of the examined grammars consistently show that the constructs *conceivable state space, conceivable event space* and *lawful event space*, for example,have no representative constructs in the grammars. Such missing constructs, if identified as unnecessary for the particular domain, can be ignored, leading to a simpler analysis that does not consider phenomena that are deemed to be outside of the scope of the target grammar.

Summary and future work

There has been a marked increase in the application of ontologies for the purposes of analysing modelling grammars. For example, a literature review identified more than 25 papers that applied the Bunge-Wand-Weber ontology for the analysis of modelling grammars such as ER (e.g. Wand and Weber, 1989; Wand and Weber, 1993; Wand and Weber, 1995), OMT, UML (e.g. Burton-Jones and Meso, 2002; Opdahl and Henderson-Sellers, 2002; Shanks et al., 2002), Petri-Nets, ARIS (e.g. Green and Rosemann, 2000; Green and Rosemann, 2002; Rosemann and Green, 2002) and Web Services standards such as ebXML, BPEL4WS, BPML or WSCI (e.g. van der Aalst et al., 2002; Wohed et al., 2002; Green et al., 2003). In general, selected ontologies and their interpretations, from an information systems viewpoint, are reasonably advanced. However, the actual process of conducting an ontological analysis is still rather immature. At this stage, the process is focused on the identification of the cardinality of the relationships between corresponding elements in the ontology and the modelling grammar under analysis. In our analysis, eight shortcomings of the current ontological analysis process have been identified and categorised into issues related to the input, process and output of the analysis.

This paper has proposed to further enhance the current process of ontological analysis. The objectives of such a method are:

1. to provide guidance for researchers who are interested in conducting ontological analyses;
2. to add rigour to the entire process and reduce the dependence on subjective interpretations of the involved researchers, and
3. to increase the credibility of the ontological analysis and its results.

Examples from our ontological analyses of ARIS and various Web Services standards have been used to exemplify this method. As a consequence, we hope the more rigorous process that has been presented here will increase the overall acceptance of using onto-logies for the analysis, comparison and engineering of various modelling grammars.