Where Ontology Affects Information Systems -Extended Abstract-

Roland Kaschek

Department of Information Systems
Massey University
Palmerston North, New Zealand
Roland.Kaschek@ieee.org

1 Introduction

Ontology has turned into a buzzword. Except from fashion there are reasons to deal with ontology from an information system (IS) perspective. This extended abstract is to explore some of these issues. In preparation of this first I explain how I understand IS and give a sketch of how I think about ontology.

1.1 A conceptual model of IS

The scandinavian school of IS (*Langefors* in particular) defined (the artefact part of an) IS as technically implemented media for recording, storing, disseminating linguistic expressions and derive linguistic expressions from them. The artefact is assumed to interact with users to whom in the best case it replies what they want. It is reasonable to understand the artefact as an intermediator between a user and an expert. Perceived in this way information systems are communication systems (see *Cragan and Shields*). They help exploiting knowledge or abilities of experts. This exploitation, to be economical, requires several users concurrently benefitting from the experts knowledge or abilities. Therefore these knowledge or abilities have to be somehow drawn away from the respective person. It must be presented in a form in which it easily allows users concurrently sharing it. Often this can be achieved by modelling a particular universe of discourse (UoD). This model both conserves and helps operationalizing the experts knowledge and abilities. (In general most likely only a rather small ratio of it I guess.)

I use the information space metaphor (derived from *Mayr* et al.) to support reasoning about IS. According to this metaphor IS offer their users an information space, i.e., a set of locations at which functionality is available. Users of the IS then are able and invited by the artefact to enter the space, to position at a certain location within it, to re-position within it, to use the functionality allocated to the location positioned at, and to leave the space.

Functionality offered is understood as data, operations on data, and import as well as export of data. According to the *van Neumann* principle the distinction between data and operations is not an ontological, but a pragmatic one. Consequently data may be perceived as operation and vice versa. Data as well as operations may be of a primary concern to the user because these need to know or apply them. They additionally might be of a secondary concern for users in that they respectively index data or functionality.

The information space metaphor implies a usage model for IS in which humans interact with an artefact to ship, process and receive linguistic expressions. The artefact processes a linguistic expression shipped to it by a user by:

1. translating the expression into a processable form, and
2. obtaining a result matching the translation in an optimum way, and
3. translating the result in a user friendly form and ship it to the user.

The optimality of the result obtained in step 2 implies that the artefact has incorporated a representation of a user model. This model might neither be explicitly specified nor be differentiating users from each other.

Looking at IS as systems mediating a user-expert communication from an economical point of view suggests the metaphor of resource management system to characterize an aspect of IS. Resources except from the linguistic expressions can be classified according to their capacity to:

- store,
- generate, delete, and modify, or
- transport linguistic expressions.

However, concerning these resources IS users in general only have the choice to use them or not but not to modify them. Rather than all of them in the sequel I thus focus on the linguistic expressions. These here are understood as composite signs. They are generated out of atomic signs according to a set of composition rules, i.e. a grammar. At least some of the linguistic expressions thus can be expected to be used to refer to something.

Designing information systems often is approached in a manner which focuses on the use of the IS. This leads to the core areas of ontology affecting IS development as discussed below.

1.2 A naive view of ontology

Originally ontology is a branch of philosophy stating that what exists. However humans never can be sure about that what exists independently and outside their mind. They only can generate ideas about it. So ontology is about theories on constitutions of possible worlds. They are models of these possible worlds.

Let I be a human having a particular purpose P. Let W be a possible world, and O, an ontology. Then I suppose $M(W, O, I)$ a modelling relationship (*Stachowiak*). Several quality aspects, i.e., aspects of the consistency of O making it useful with respect to purpose P (*ISO* definition of quality) might be used to assess O. Such as:

- **coverage**, referring to the part of W that can be addressed in terms of O.
- **essentiality**, referring to the part of O that states assertions about W
- **consistence**, referring to whether O contains conflicting assertions about W.
- **correctness**, referring to whether all assertions in O as far as they concern states of affair in W are true.
- **precision**, referring to the degree to which the true statements about W help the human I meet the purpose P.

- **size**, referring to the number of assertions in O.
- **complexity**, referring to the assertions in O referring to other assertions in O.
- **language**, referring to the language used to formulate O.

I do not consider language **formality** as an essential characteristic of ontology. For a classification of languages based on *Gethmann's* schema see the table in figure 1. In the table a '+' and a '-' respectively indicate independency or dependency of speaker or topic.

		Speaker	
		+	-
Topic	+	private language	technical language
	-	natural language	formal language

Fig. 1. Languages classified based on *Gethmann*

Understanding ontologies as models suggests comparing their mode of reference to their subject. Reference modes of models known in IS development are **descriptive mode** and **prescriptive mode**. The difference between them can be paraphrased as the difference between the utterances 'it is so!' and 'so shall it be!'. In contrast to this ontologies might be ascribed a **constitutive mode** of reference. This reference mode would be different from the prescriptive mode in that it introduces terminology needed to talk about the possible world while the terminology in the prescriptive mode already exists but the possible world is required to change.

2 Possible worlds ontology support is asked for

Ontologies appear to me as models of possible worlds. They thus should help conceptualizing and getting information concerning these possible worlds. In the sequel I briefly discuss the possible worlds that according to my knowledge of it play a role in IS development. The point here is not to give latest or deepest knowledge about the respective world but just to outline it such that folks aware of IS issues can understand it. What I hope that ontology folks can / will do, is pointing out results, methods, and references that can be used to improve understanding of the six possible worlds listed below.

2.1 Possible world 1: UoDs

The IS artefact takes linguistic expressions from users and ships such expressions to them. Consequently a language being the base of the expression exchange is needed that is understood by the users. This language may be a technical language. Its vocabulary as well as its syntactical rules up to some degree can be derived from the UoD model. It thus for IS development would be very convenient to have some kind of tool set for suitable UoD constitution. What at the time being often is used is the method of *Abbot*, i.e., somehow making experts contribute the generation of texts specifying the required UoDs and then read the text against a semantic model. Here I understand a **semantic model** as a set of modelling notions and abstraction concepts. Reading a text against a semantic model means firstly,

classifying text parts, in particular words, as instances of modelling notions or abstraction concepts, and secondly to apply the definitions of modelling notions and abstraction concepts to better understand the text or enhance it by answers to questions asked to the experts. Popular semantic models for this purpose are the ER-model (*Chen*), the relational model (*Codd*), semantic nets, and object models.

2.2 Possible world 2: Resources

The linguistic expressions managed by the IS were said to be the resource to be focused on in this extended abstract. Identification of resources affects IS development. Some resources even might give IS completely new characteristics as is the case with inter space transitions. Dealt with them properly introduces Web IS as opposed to IS. Furthermore system development shall support the user behavior that fits to the system purpose and shall make it difficult for users to use the system in a way contradicting the system's purpose. I have identified three dimensions of resource classification. The respective dimensions are:

- **focus**, i.e., the user refers to the UoD in its totality or only to a part of it, giving rise to the values 'global' and 'local' of the focus.
- **modus**, i.e., the user refers to the UoD with respect to a particular state of affairs or with respect to transitions of such states. This gives rise to the values 'static' and 'dynamic' of the modus.
- **kind**, i.e., the user refers to something because of his interest in it or because of its relation to something else. This gives raise to the values 'self contained' and 'referential'.

Identifying these dimensions of references and their scale values leads to resources as classified according to the table in figure 2

Resource	**Focus**	**Modus**	**Kind**
Data	global	static	self contained
View	global	dynamic	referential
Operation	local	static	self contained
Dialogue	local	dynamic	self contained
Inter space transition	global	dynamic	referential
Data model	global	static	self contained
Data subset	local	static	self contained
Intra space transition	local	dynamic	referential

Fig. 2. Resources of information systems classified

The idea to classify IS resources like this was derived from *Thalheim's* co-design approach and extended by the 'kind' dimension. However, *Thalheim* appears not to mention the idea of resource. In his abstraction layer model he mentiones 'data', 'view', 'operation', and 'dialogue' as aspects of IS that have to be observed throughout IS development. In my terminology he proposes to have in mind all the time throughout IS development all the resources and design the IS such that an optimum resource management can take place.

2.3 Possible world 3: Information space use

Users can enter and leave the information space, they can import and export data, as well as relocate in space and invoke offered functionality on available data. This gives raise to the following areas of user activity needing developers care during IS development to enable the IS supporting users in performing the respective activities:

- **information**, i.e., identifying data and operations being required to perform the actual task as well as the locations in space at which they are accessible to the user. This might involve using a help functionality, or using personal data or functionality introduced into the IS as additional usage aid (personalization).
- **navigation**, i.e., identifying a procedure for proceeding from a given location to a target location. This might include using space exploration means as well as asking for proceeding related suggestions or simply following a more or less likely path through the space to explore it and get familiar with it and the way to use it.
- **processing**, i.e., identifying how to effectively use the functionality of the IS to meet the actual purpose.
- **handling**, i.e., using the IS efficiently to meet the respective goals. This might include using the help functionality or documentation or similar.

2.4 Possible world 4: Users

The IS artefact needs to have incorporated a representation of a user model. It thus is reasonable to care for what users actually exist. **User profiling** is a technique used to support the task of personalization of the IS interface or the whole application. User profiling consists in identifying user types and allocating to each type a profile, i.e. a tuple of dimensions. To each dimension there is then associated a scale, i.e., toset. User types are proposed (*Thalheim and Schewe*) to be constructed such that in the space of all user dimensions equipped with the scales give rise to a convex region. User types can be identified by means of the method applied to a text describing the users of the system as anticipated by the system inventors or by the marketing, sales or human resources people of the company running the IS. These here would function as experts of the UoD of system users.

2.5 Possible world 5: Human inference making

I like to conceptualize human inference making by presupposing that it is essentially determined by main questions as follows:

1. **What exists?**, i.e., what makes up the actual problem and its context?
2. **What should I do?**, i.e., what is a solution to the problem at hand? In particular, what is a good solution to the problem. Connected to this, once a solution has been chosen, which solution plan should be implemented? The latter question then asks for the individual solution steps and their relative order.
3. **What can I know?**, i.e., which transitions from already obtained statements about W to new ones can be justified?

In an attempt to use traditional philosophical terminology I refer to the theories respectively giving answers to the questions listed above as: **ontology**, **ethics**, and **epistemology**. The interesting point now is that one is free to chose a triplet (O, T, E), i.e. a particular ontology O, ethics T, and epistemology E. Since one is free in this choice one can try to make a reasonable choice. It might be the case that with respect to a purpose P_1 the triplet (O_1, T_1, E_1) is the first choice. However, with respect to a purpose P_2 the triplet (O_1, T_1, E_1) might not be the first choice. Clearly, to answer these questions one could analyze problem solution plans and specify the most important quality characteristics. Then one could study how choices of ontology, ethics, and epistemology impact identification of problem solutions as well as identification and realization of problem solution plans.

3 Resume

This extended abstract started with the presentation of a simple conceptual model of IS. It then proceeded with a naive view on ontology. Finally five possible worlds were discussed concerning which the IS community might wish to benefit from the ontology community. These possible worlds were introduced. Except concerning the last one it was roughly explained how IS folks try to obtain the knowledge they require with respect to the possible worlds.