Formalising the 'No Information without Data-representation' Principle

Patrick ALLO
Centre for Logic and Philosophy of Science, Vrije Universiteit Brussel, Belgium
IEG, Oxford University, UK
GPI, University of Hertfordshire, UK

Abstract. One of the basic principles of the general definition of information is its rejection of dataless information. In general, it is implied that "there can be no information without physical implementation" [1]. Though this is usually considered a commonsensical assumption, many questions arise with regard to its general application. In this paper, a combined logic for data and information is elaborated, and specifically used to investigate the consequences of restricted and unrestricted data-implementation principles.

Keywords. Data, semantic information, epistemic logic

1. Epistemic Logic and an `Information First' Approach to Epistemology

When we theorise about the relation between data and information, we automatically engage a debate whose ramifications are not confined to a single scientific or philosophical domain. This surely does not facilitate such an enterprise. For starters, in many domains a tight distinction between these obviously related but separate notions is rarely observed, and in practice the terms of data and information can be used interchangeably. By contrast, the distinction between data and information becomes much more urgent when looked at from the perspective of the physics of computation; especially when, as we find in Landauer, pieces of data are considered as a physical implementation of the information we compute with or reason about. Reverting to looser characterisations, pieces of data tend to be understood as syntactical entities, information rather as a semantical entity. Finally, the distinction between data and information is also crucial to the so-called *data-information-knowledge* hierarchy (henceforth, DIK), one of the central metaphors of information-science and knowledge management [2]. In view of this, the core aim of this paper is to give a sufficiently strong, yet not unrealistic interpretation and formalisation of the relation between data and information, and more precisely between the states of being informed and holding data.

Arguably, it is not a good idea to try to capture the relation between data and information on the basis of the wide-ranging uses of both terms. If feasible at all (which is rather doubtful since in most uses of these terms confusion is more common than an actual insight in the subject) this could only yield a very general theoretical basis. It is therefore preferable to properly constrain the issue by (i) providing general though sufficiently precise definitions of data and information, (ii) fixing the broader

context in which we envisage to use the notions of data and information, and (iii) choosing a suitable formal framework for the formulation of our theory.

As for the characterisation of both basic notions, the general definition of semantic information as *well-formed and meaningful data* is the most obvious starting point [1]. Although it is perhaps still too vague to be used to derive any substantial property of data and/or information, it already settles their distinctness. To a first approximation, any string of symbols should count as a piece of data, but it takes more to qualify as a genuine piece of information. Besides the demand that data need to be well-formed as well as meaningful to count as information, it is also commonly assumed that it is possible to turn data into information. Calling the process of turning data into knowledge the core aim of information-science is itself fairly uncontroversial. Yet, while the development of tools and techniques for doing so is valuable, for present purposes this is not as important as the insight that the adoption of a certain level of abstraction is crucial to obtain well-formed and meaningful pieces of data.

One of the most common methods we use to turn data into meaningful content is the use of a properly functioning language; in more formal terms: a language with a syntax and a semantics. Referring to a level of abstraction, then, is no more than a way of talking about the distinctions that can be effectively expressed in that language. In the case of natural languages such levels of abstraction are fairly hard to determine, but when we turn to formal languages this problem is easily dealt with.

While the received view does not include an explicit reference to the levels of abstraction that serve as an interface between data and information, it is also not opposed to it. Nevertheless, the above description is as far as the consensus about the relation between data and information goes. As soon as we start wondering whether any properly structured piece of data counts as information the opinions promptly diverge. One way to avoid the most common quibbles is to explicitly focus on *declarative, objective and semantic* information (henceforth, DOS). Intuitively, this is meaningful data that can be used to answer a question (data can be exploited by a system as input of adequate queries, see [1]) and is therefore cognitively valuable. As a more precise rendition of the general definition, this is sufficient to satisfy the needs of information-scientists and epistemologists alike. At this point an important question remains unanswered: is the process of turning data into meaningful and declarative content reliable in the sense that it should always yield genuine information?

The standard definition of semantic information is commonly thought to imply that form and meaning alone suffice. Still, two objections to this claim are of particular interest. A first widespread objection suggests that *relevance* is of utmost importance; a second, still controversial objection points to the need for information to be truthful. In each case we should wonder whether 'irrelevant information' and 'false information' are kinds of information rather than no information at all. The question of relevance can quite easily be dismissed. Even if relevance is indeed a property of information we ought to value, we still obtain a more general and conceptually sound theory by considering relevance as a property we do not need to capture the conceptual nature of information itself. By contrast, and despite the debate surrounding this claim [3-6], it is not that obvious to make similar claims about false information. A pragmatic rather than a principled motivation for understanding declarative, objective and semantic information as *truthful, well-formed meaningful data* derives from the constitutive role of acquiring information as a means to attain knowledge. Provided knowledge gets its usual factive reading instead of the ultra-loose sense it gets in information-science, it surely makes sense to apply the same veridical standard to information. In short, while

false information might very well be presented as if it were a solid basis for knowledge, it can never be the stepping stone to knowledge that epistemological theorizing requires.

The above remark is especially relevant for epistemological theories in which the need for information is explicitly affirmed. This includes Drestke's knowledge as information-based belief [7], but it is also true for what I would call an *information first* approach to epistemology. Loosely speaking the latter could be thought of as taking the best of traditional epistemology and the previously mentioned DIK-hierarchy, but this is still a highly misleading description. Formulated as a broader programme, an information-based (formal) epistemology is more conveniently characterised as follows: where G is an epistemological theory, we say that G formulates an information-first approach iff (i) it takes the relation of *being informed* as one of the central cognitive attitudes; (ii) it attempts to explain knowledge as a way to be informed rather than as a peculiar way to believe; and (iii) it acknowledges the fundamental connection between being informed and being able to inform as a more basic connection between epistemic states and action than the necessary relation between knowledge and proper assertion.

Moving down the hierarchy, making the relation between data and information precise has a place within this broader programme. In the next section a basic logic for 'being informed' is introduced, and subsequently used to express both the basic features of an information-based epistemology and the problems we face when we include an explicit reference to data in this framework.

2. Logics for 'Being Informed' (and some of their rivals)

The formal language used to express the properties of being informed is derived from the standard approaches in modal epistemic logic. It is based on the language of propositional logic augmented with a set of modal operators. In addition to the standard connectives for conjunction (&), disjunction (Ú), negation (Ø), and implication (®), we need the following operators: IaA to express that a is informed that A and DaA to express that a holds data for A as well as the standard operators for knowledge and belief BaA to express that a believes that A and KaA to express that a knows that A. Finally, for the purpose of expressing logical relations we use G Þ A to express that A is a logical consequence of G and B Û A to express that A is logically equivalent to B. Using that language, we can elegantly express the basic properties of a purely information-based approach to knowledge, and also focus on the role of data. As a starting point, a fairly minimal characterisation is all we need. The following general properties are proposed:

non-doxasticism: knowledge is not defined as a kind of belief, and being informed does not have a belief-component. Hence we reject both the possibility of defining knowledge as $K_a\, p := B_a\, p \,\&\, p \,\&\, \ldots$as well as the inclusion of the $I_a\, p \rightarrow B_a\, p$ axiom. Crucially, this version of non-doxasticism does not deny that belief is necessary for knowledge, only that it cannot figure as a conjunct in a sufficient definition [8]. As a consequence, the $K_a\, p \rightarrow B_a\, p$ axiom which traditionally relates knowledge to belief might figure in a combined logic of knowledge and belief and still be consistent with the 'information first' approach we try to sketch.

necessary data-component: it follows from its standard definition that information is a specific kind of data which therefore cannot exist without a data representation. Thus formulated there is nothing objectionable about the fact that we treat information as a kind of data, and that in order to be informed the possession of certain data is required. The problems that nevertheless arise from this principle are of two distinct kinds: one concerns the nature of the data themselves, the other concerns the amount of data required for being informed. Of these, the first issue should not surprise us, for it is one of the standard controversies in the physics of computation [9]. For now, we can easily sidestep the problems introduced by the requirement that information be backed up by a physical data-implementation as long as we ensure that the system we present does not substantially rely on any specific feature of the ontological status of data. The second problem is more intricate, for it only surfaces in connection to the relation of 'being informed'. Despite the fact that our treatment of information as a kind of data intuitively suggests that states of information somehow be supported by data, the explicitation of that support-relation is not really straightforward. Getting the relation between being informed and holding data is precisely the problem that needs to be tackled in the present paper.

primeness of being informed: holding a piece of well-formed meaningful data which incidentally happens to be true is not sufficient for being informed. This we see by considering Gettier-like cases in which true consequences can be derived from a set of (partially) false data. As a consequence, it is convenient to treat the state of being informed as a prime state, i.e. a state which satisfies a condition that cannot be decomposed into more basic (but still generally applicable) conditions [8]. Crucially, this also entails that, being informed cannot straightforwardly be identified with holding a piece of information, and that there is no exact match between information as a state and information as a commodity.

As for standard modal epistemic logics [10,11], the systems required to reason about data and information are obtained by adding to standard propositional logic axioms which specify the properties of the modal operators required to express the relations of being informed and holding data. As a logic for being informed, we adopt the proposal from [12] which takes I to be a **KTB**-modality satisfying all but the last of the axioms listed below:

(Nec) If $\vdash A$ then $\vdash I_a A$

(K) $I_a (A \rightarrow B) \rightarrow I_a A \rightarrow I_a B$

(T) $I_a A \rightarrow A$

(B) $A \rightarrow I_a \neg I_a A$

(4) $I_a A \rightarrow I_a I_a A$

Since this is not the place to repeat the reasons for using the resulting system as a proper formalisation for being informed, we should only remark that its main difference with standard epistemic logics is reflected in its exclusion of introspective principles.

Once we have a basic system for I, we can start to spell out how it relates to other notions. This is for instance what happens when the so-called entailment-thesis $K_a A \rightarrow B_a A$ is used as the primary principle for connecting the logics of knowledge and belief. Yet, while in such cases two existing logics are adopted from the beginning, this need not be so; we can equally well start from a logic for being informed, and derive the properties of data once we have settled on the proper connecting principles. Inspired by the standard principles used to connect knowledge to belief, three bridge axioms could in principle be considered:

(ID1) $I_a A \rightarrow D_a A$

(ID2) $I_a A \rightarrow I_a D_a A$

(ID3) $I_a A \rightarrow D_a I_a A$

As the state of being informed is characterised by a prime condition, such principles are the only type we need to consider – the relation of holding data is solely investigated as a necessary condition. To begin with, only the first of these is explicitly considered, the latter two are taken into account in due time.

So far the approach sketched above does not immediately yield a logic for data itself, all it implies is that the data one is actually informed of are indeed closed under logical consequence. Nothing is said on the properties of merely holding data. As long as we think about data in the same way as belief, this *logical inertia* of data should presumably count as a flaw for our theory. Intuitively, it surely makes sense to have an independent logic for data, for even if we ought only to reason on the basis of our information, it is often only feasible to reason on the basis of the data one holds. This intuition can only be respected if a rather strong interpretation of the relation of holding data is singled out. The fact is that a sufficiently strong interpretation is not necessarily the most obvious one.

The dilemma we face can be explained by distinguishing between (i) holding mere data, (ii) holding well-formed and meaningful data, and (iii) being informed. Clearly, what we ought to reason about falls under the third category, while what we mostly reason about belongs to the second. If, as often happens, the two latter categories are identified, one could easily conclude that simply holding data has no substantive logical property. Since that option is incompatible with the notion of DOS-information we privilege, another solution must be sought. By being attentive to a common ambiguity in the way we refer to data and information, a more robust solution can be discerned.

3. States and Commodities

While it goes without saying that, as a theory, the DIK-hierarchy is hardly helpful; it remains a valuable source of common intuitions. In general it makes sense to define our attitude towards the practice of information-science and knowledge management in terms of an inheritance of the problems it identifies, rather than in terms of the solutions it proposes. In that perspective, two concerns should draw our attention: the

status of the hierarchy itself, and the gradual shift from properties of things to properties of people.

The first of these concerns relates to different ways in which the hierarchy can be read: top-down by identifying necessary relationships like "no knowledge without information" or as "no information without data-implementation"; or bottom-up either by identifying what, say, information lacks to count as knowledge, or by describing the procedures required to "turn data into knowledge." Interestingly, as is the case for epistemologists, the identification of necessary relations (top-down) is less tedious than the formulation of sufficient conditions (bottom-up). However, as the notions of knowledge and information presumed by the hierarchy do not correspond with ours, we should not extensively rely on this analogy. A more interesting concern centers on the tension between treating information as a commodity, a thing that can be stored, passed on, sold etc., and information as something that qualifies the cognitive state of an agent. In the literature, this tension arises in at least two ways. In the view that the notions higher up in the hierarchy define properties of humans whereas those lower in the hierarchy define properties of things, and in the presumably more problematic differentiation between implicit and explicit knowledge.

> [T]he distinction between explicit knowledge and information is even less defensible. If knowledge is a property of people, and embodies prior understanding, experience and learning [45, pp. 9--10], it is difficult to argue that explicit knowledge, recorded in documents and information systems, is any more or less than information. [13]

Yet, while the implicit/explicit distinction is rightly criticised by Rowley, it is therefore not implausible that for each level in the hierarchy we sometimes need to refer to properties of things and sometimes to properties of cognitive agents. For that purpose we might – instead of using the confusing implicit/explicit terminology – want to introduce two distinct levels at which data, information and knowledge are assessed: as states and as commodities.

If we want to reformulate the claim that knowledge is primarily a property of people in the standard philosophical terminology, we should probably say that knowledge is a state of mind. More precisely, when we focus on "knowing that" knowledge can be characterised as a propositional attitude: a relation of subjects to propositions. The contrast with our understanding of data as a thing is obvious, for its reformulation into the philosophical vocabulary does not need to refer to minds, subjects or even propositions. By treating data as constraining affordances representable as strings of symbols, the relation of holding data can conveniently be modelled as a relation between an agent and a particular syntactical object. As such it should probably not be understood as a state of mind, and not be modelled as a relation to a proposition. Finally, if we recall the tension between the relation of being informed as it is modelled by the modal operator I and the relation of merely holding a piece of information, we might conclude that being informed can be conceived as a relation to a proposition *and* as a relation to a particular syntactical object.

The best method for illustrating the consequences of modelling data and information in two distinct ways exploits the contrast between explicit syntactical models and mainstream possible worlds models known from epistemic logic [14]. Roughly, if being informed is modelled on the mainstream approach we have that

$$I_a A \ll I_a \mathrm{B} \text{ iff } A \hat{\mathrm{U}} B$$

That is, if being informed is modelled as a relation towards propositions, for every sentence A out of a set of logically equivalent ones, any $I_a A$ expresses one and the same relation. Nevertheless, this should not be confused with claiming that holding a piece of information A is really the same as holding a logically equivalent piece of information B. For by making such claims, we automatically revert to a way of reasoning that treats information as a particular commodity. This, in its turn, calls for a different, more refined, model where, given a new relation $\mathcal{I}$, $\mathcal{I}_a A$ and $\mathcal{I}_a B$ only express the same relation iff A and B consist of one and the 'same' string of symbols. On the face of it, each of these approaches partially captures our intuitive understanding of being informed; while the former largely agrees with the standard approach for knowledge, the latter acknowledges the insight that being informed requires us to hold a particular piece of information. Pieces of information being a kind of data, this way of reasoning about being informed presumably inherits most constraints we pose on our reasoning about data.

If, as suggested above, we leave room for theories and logics of 'being informed' that treat it as a relation to propositions, we face a problem when we try to simultaneously retain the strictly syntactical reading of `holding data'. Yet, since for data the syntactical reading is by far the most plausible one, it is well worth trying to save it. The problem arises roughly in the following way. Let I_a be a **KTB**-modality, and assume that the relation between being informed and holding data is fully determined by

$$I_a A \mathrel{®} Ð_a A$$

Assume now that we try to enforce an otherwise unobjectionable syntactical reading of holding data by, for instance, specifying that $Ð_a A$ is only closed under a highly limited set of syntactical manipulations. Thus we get $Ð_a [A;B]$ iff $Ð_a A$ and $Ð_a B$ as a minimal *logical* constraint on holding data (where square braces and semicolons are used to represent the complex syntactical objects obtained by concatenating simpler syntactical objects). Yet, merely constraining $Ð_a$ is not sufficient to ensure the failure of an argument of this form.

(1) $A \mathrel{Û} B$
(2) $I_a A$
(3) $I_a A \mathrel{«} I_a B$
(4) $I_a B$
$\backslash\ Ð_a A \mathbin{\&} Ð_a B$

Viz. if one holds data in virtue of being informed, one's data is thereby also logically closed and $Ð_a$ does not after all succeed to capture a reading of holding data that is systematically weaker than the propositional one. The culprit is nevertheless easily identified, for by formalising the connection between information and data as a straightforward implication, we effectively align them on to the same (propositional) reading. This fact calls for a refinement of the basic data-entailment thesis. An important consideration, in that respect, is the correct interpretation of the view that there is no information without data. That is, if one is informed that A this only means there ought to be some piece of data in virtue of which one is so informed, *not* that there must be a piece of data that somehow perfectly matches one's information that A. As a first refinement, the following revised principle is proposed:

(ID*) $I_a\,A$ only if there is an A' where $I_a\,A' \Subset I_a\,A$ such that $Đ_a\,A'$

where, in virtue of the failure of the above argument, $Đ_a$ effectively captures a non-propositional reading of holding data. Even then a different but equally objectionable argument is not yet avoidable:

(1) $A \vdash B$
(2) $I_a\,A$
(3) $I_a\,A \rightarrow I_a\,B$
(4) $I_a\,B$
$\setminus\ Đ_a\,A \,\&\, Đ_a\,B$

for some A' and B' such that $I_a\,A' \Subset I_a\,A$ and $I_a\,B' \Subset I_a\,B$

As before, since the move from A to B need not be warranted by the weak syntactical manipulations we allow the relation of holding data to be closed under (let A for instance stand for $(\varnothing\,p \cup q) \,\&\, p$, and B for q) $Đ_a\,B'$ is itself not directly implied by $Đ_a\,A'$ but merely inherited from the closure conditions at the propositional level of being informed. Hence, a further refinement of our connecting principle must be sought.

(ID)** $I_a\,A$ only if there is an A' where $I_a\,A' \vdash I_a\,A$ such that $Đ_a\,A'$

Using this revised principle, the strictly syntactical reading of holding data is duly ensured. For the purposes of the arguments that shall be presented further on, one can uphold that this version basically gets the relation between data and information as well as the syntactical nature of data right. The final refinement we still might want to propose is therefore of less immediate importance, for all it does is integrating the syntactical operations on data within the connecting principle itself.

(ID*)** $I_a\,A$ only if there is a complex string of data [A', ...] such that (i) A' can be extracted from [A', ...] ;(ii $I_a\,A' \vdash I_a\,A$; and (iii) $Đ_a$ [A', ...]

Generally speaking, the adoption of this final version has the benefit that no specific assumptions need to be made with regard to the syntactical manipulations that are part of the closure conditions of holding data. This, in its turn, has the advantage that – except for its syntactical representation – the nature of data, and especially its being well-formed, is not fixed by the underlying logic for $Đ_a$. Finally, the separation of mere syntactical manipulations and logical relations results in an agnostic attitude towards the ultimate division of labour between mere symbol manipulation and semantically sensitive inferences that play a role in the process of turning data into information.

Despite the complication the latter two revisions introduce, they are crucial for our ability to reason simultaneously about data as concrete or particular entities, and about information as more general cognitive states. The two levels that are introduced certainly do not rest on *ad hoc* mechanisms. Instead, they serve to make the levels of abstraction at which data and information are evaluated explicit. This results in a more precise reading of the intuitively valid conditions that (i) information needs to be supported by some amount of data, and (ii) being informed involves having precisely that data at one's disposal. This way of limiting the amount of data that is required is particularly valuable in contexts where, because data consist of (or are stored as)

physical entities, dealing with the problem of logical and deductive omniscience becomes even more pressing.

Before we start using these new principles for the design of a combined system of data and information, a few general considerations regarding the method of abstraction should be included. For our purpose, a level of abstraction is best considered as a level of (logical) discrimination; that is, a specified way of, respectively, telling cognitive states (in a formal model) or formulae (in a formal language) apart. Evidently, any account of logical equivalence provides such a level, and the propositional and syntactical accounts of being informed and holding data are obvious examples of this. Besides making the levels of abstraction precise, we also need tools for drawing conclusions at one level of abstraction on the sole basis of our knowledge of what happens at a different level of abstraction. In a simplified formulation, we only have this ability to make inferences that go from one level to another in virtue of a certain degree of inter-LoA coherence. In the present context, being able to do so facilitates the hitherto avoided modelling of data at the propositional level without the loss of the connection with the more fine-grained syntactical approach on which most of our intuitions depend. Its relation to the syntactical approach follows immediately from the above description of the relation between data and information.

(D)** $D_a\, A$ only if there is an A' where $D_a\, A' \vdash D_a\, A$ such that $Ð_a\, A'$

Consequently, while this reintroduces the simple relation between data and information

$I_a\, A \mathrel{®} D_a\, A$

it also raises the question of the correct interpretation of this new, less discriminating way of assessing data. A quick and rough answer is this: $D_a\, A$ expresses the semantic relation of holding data which carries the content that A as opposed to the syntactic relation of holding a piece of data A. As such, it does not necessarily refer to a particular object, but only to a state that is warranted by a particular object. The functioning of this notion which can serve as an interface between being informed and holding bare data is best understood in line with the following two principles:

△ If every state which satisfies $Ð_a\, A$ also satisfies $Ð_a\, B$, then every state which satisfies $D_a\, A$ also satisfies $D_a\, B$.

▽ If some state which satisfies $D_a\, A$ does not satisfy $D_a\, B$, then some state which satisfies $Ð_a\, A$ does not satisfy $Ð_a\, B$.

In particular, this entails that if at the syntactical level two sets of data are the same, they remain the same on the propositional level; inversely, if they are distinct at the propositional level, they stay so at the syntactical level. Using only these two principles, several problems can be solved in a more satisfactory manner.

4. The Trouble with Meta-data

A mostly convenient context to exploit the method we sketched above relates to the problems posed by meta-data. Basically, meta-data play a role that is similar to that of reflective states in standard epistemic logic. When B expresses the same proposition as $D_a\, A$, then B is meta-data for A: meta-data about the data one holds. Trivially, $D_a\, B$ then expresses that one holds such meta-data; in short: $D_a\, D_a\, A$. With this in mind, one should then ask how $D_a\, D_a\, A$ relates to $D_a\, A$. This relation splits up in two fragments. First we should inspect the prospects and consequences of the principle stating (at the propositional level) that there is no meta-data without data.

$$D_a\, D_a\, A \mathrel{®} D_a\, A$$

Call the principle in question the 'non-corrupted meta-data' thesis, and note that it is harder to reject than either $D_a\, A \mathrel{®} A$ or $D_a\, (D_a\, A \mathrel{®} A)$. The converse of these principles states that one cannot hold data without also having meta-data for it

$$D_a\, A \mathrel{®} D_a\, D_a\, A$$

As before, we cannot conclude the falsity of this principle, from the sheer fact that we do not hold data for all truths $A \mathrel{®} D_a\, A$. Rather, the question becomes whether a should be able to obtain the meta-data for all the data it holds by purely logical means. Call the principle which supports this the "free meta-data" thesis.

In view of the $\triangle$- and $\triangledown$-principles presented above, it should be obvious that the distinctness of $D_a\, A$ and $D_a\, D_a\, A$ cannot be argued for by merely referring to the actual distinctness of states wherein one holds data, and states wherein one holds meta-data. That is, the presumed invalidity of $Đ_a\, A \mathrel{®} Đ_a\, Đ_a\, A$ and $Đ_a\, Đ_a\, A \mathrel{®} Đ_a\, A$ cannot be used to infer the falsity of the corresponding principles we are actually interested in. Yet, using the refined data-principles we can approach the problem more carefully along the following lines.

$D_a\, A$ only if there is an A' where $D_a\, A' \mathrel{⊢} D_a\, A$ such that $Đ_a\, A'$

$D_a\, D_a\, A$ only if there is an A'' where $D_a\, A'' \mathrel{⊢} D_a\, D_a\, A$ such that $Đ_a\, A''$

Quite naturally, for holding data to imply holding meta-data, or conversely, for holding meta-data to imply holding data, the above instances of our general principle at least require that they be supported by a single syntactical entity or datum. This at least requires that A' and A'' could be the same datum. Two separate cases arise: either that datum can be represented by a purely factual (i.e. propositional) formula, or it can be represented by a data (i.e. modal) formula. To check the former option, we replace A' and A'' by the propositional formula B; to check the latter, we replace A' and A'' by the modal formula $D_a\, B$. Since A' can easily be substituted for a propositional formula, and A'' for a modal one, the following reasoning should work out. If the replacement by B works out, the "free meta-data" principle turns out to be unproblematic; if the replacement by $D_a\, B$ works out, the "non-corrupted meta-data" thesis is unproblematic. Finally, this requires that for all A there exists some propositional formula B such that, respectively, $D_a\, B \mathrel{⊢} D_aD_aA$, and $D_a\, B \mathrel{⊢} D_aD_a\, A$. Since this approximately brings us

back where we started, this means that given the relation between data-particulars and propositional data we defined, natural constraints on particulars do not fully determine the constraints on propositional data.

This conclusion should, however, not pose a problem for the present enterprise. Rather, it shows that while modeling data, information, and their relation at the propositional level, we enjoy a certain freedom that goes beyond the mere abstraction from actual syntactical specificities. We can also abstract from the difference between data and meta-data, and this is just one of the several choices we face when constructing a formal model. This apparent freedom does not deny that many external considerations can serve as a guide. One such kind of consideration derives from our interpretation of D_a. For instance, when D_a is a relation to propositions, we suggest it should be read as "holding data *for* ..." instead of "holding a datum ..." In the same vein, if iterations of D_a intendedly refer to different levels of meta-data, it is somehow incoherent for both $D_a A \rightarrow D_a D_a A$ and $D_a D_a A \rightarrow D_a A$ to be valid. This is especially true if one thinks that even the very basic kind of meta-data an iteration of D_a refers to has a value that exceeds what is already available at the primary level. For if that is the case, then at least $D_a A \rightarrow D_a D_a A$ has to go. By contrast, if one thinks that when the particular datum which initially supported $D_a A$ is destroyed, the propositional content of $D_a A$ cannot infallibly be recovered from that of $D_a D_a A$, then $D_a D_a A \rightarrow D_a A$ should perhaps also go.

With the minimal relation between 'being informed that' and 'holding data for' expressed by **DD**** and a plea for a principled distinction between data and meta-data, a sufficiently large set of constraints for a combined logic for data and information has been obtained. Thus, a system can be fully described on this basis, and the value of some additional connecting principles can also be assessed. Starting with I_a as a **KTB**-modality, we add $I_a \mathrm{A} \rightarrow D_a \mathrm{A}$ as a by now well-motivated expression of the necessary relation between being informed and holding data. Finally, we also specify that meta-data should not come for free, hence $D_a A \rightarrow D_a D_a A$ is not allowed to come out as valid. This is the main guide for the evaluation of the acceptability of ID2 and ID3 – the two remaining connecting principles whose status we did not yet settle. As a matter of fact, they can both be dismissed in an entirely uncontroversial way. Namely, since $I_a D_a \mathrm{A} \rightarrow D_a D_a \mathrm{A}$ and $D_a I_a \mathrm{A} \rightarrow D_a D_a \mathrm{A}$ are already valid, the combination of $D_a \mathrm{A} \rightarrow I_a D_a \mathrm{A}$ with the former or $D_a \mathrm{A} \rightarrow D_a I_a \mathrm{A}$ with the latter yields the validity of $D_a A \rightarrow D_a D_a A$, which is exactly what had to be avoided. It is therefore easy to conclude that both axioms are at odds with the idea that meta-data actually do add something not yet present in the primary data itself.

5. Concluding Remarks

To conclude this article, two specific virtues of its general methodology deserve to be highlighted. Both concern the meaning and function of iterated *D*-modalities. The first benefit derives from the specific argument that was presented for the invalidity of $D_a A \rightarrow D_a D_a A$. When compared to standard rejections of the intuitively related KK-thesis, it should be noted that no appeal was made to either computational concerns or the higher standards required for *reflective* states. Instead, only the assumption of the added value of meta-data had to be introduced. In short: if meta-data were free in the sense of being obtainable by logical means only, it would be of no value at all. This

relies on a typical informational-theoretical concern which connects cost or value to informational content. If content, value, and logical consequence are so related, free-meta data are in fact informationally empty meta-data, and therefore only meta-data by name.

The second benefit derives from the interaction of the rejection of free meta-data and the failure of introspection for being informed. For starters, the rejection of $D_a A \rightarrow D_a D_a A$ is motivated independently from the rejection of $I_a A \rightarrow I_a I_a A$ in a system where I is a **KTB**-modality. This is important in view of the fact that the free meta-data principle need not to be rejected in order to invalidate introspection for I. Instead, a logic for data which does not validate $D_a A \rightarrow D_a D_a A$ combines more elegantly with a logic for being informed which does not validate $I_a A \rightarrow I_a I_a A$. To wit, combined with the unobjectionable $I_a A \rightarrow D_a A$, $I_a A \rightarrow I_a I_a A$ enforces the clearly unacceptable $I_a A \rightarrow D_a A \,\&\, D_a D_a A \,\&\, D_a D_a D_a A \,\&\, \ldots$). This suggests that the rejection of free meta-data is in a sense more basic than the rejection of introspection for being informed: introspection trivializes the value and hierarchy of meta-data. Free meta-data, to the contrary, does not yield introspection for being informed.

References

[1] Floridi, L., 2005, Semantic Conceptions of Information, *Stanford Encyclopedia of Information*, Zalta, E. N., ed., Stanford.

[2] Ackoff, R. L., 1989, From datc (t) Tj-0.024 Tc ( ) Tj0k4[lckoff, R. L., 1989, FroJo Tc (ckoff, R.u0.05 Tc (o)rn) Tj0.15 Tc ( ) T0.05 Tc (o) a rd
[1ec (t) Tj-0.024 Tc ( ) Tj0kzer, J.[Tj0 Tc (2004, ITj-0.024 Tc (t) Tj0 Tc989, Fr) Tj0.05 T (f8(da) Tj-0Tc ( ) Tj0 Tc (Concep) TTj-0.024 Tc () Tj0.15 Tc (