Aboutness: Towards Foundations for the Information Artifact Ontology

Barry Smith[1,*] and Werner Ceusters[2]

[1] Department of Philosophy, University at Buffalo, 126 Park Hall, Buffalo, USA

[2] Department of Biomedical Informatics, University at Buffalo, 921 Main Street, Buffalo, USA

ABSTRACT

The Information Artifact Ontology (IAO) was created to serve as a domain-neutral resource for the representation of types of information content entities (ICEs) such as documents, data-bases, and digital images. We identify a series of problems with the current version of the IAO and suggest solutions designed to advance our understanding of the relations between ICEs and associated cognitive representations in the minds of human subjects. This requires embedding IAO in a larger framework of ontologies, including most importantly the Mental Functioning Ontology (MFO). It also requires a careful treatment of the aboutness relations between ICEs and associated cognitive representations and their targets in reality.

1 INTRODUCTION

At the heart of the IAO is the term 'Information Content Entity' (ICE), which is currently defined as follows:

> INFORMATION CONTENT ENTITY =def. an ENTITY which is (1) GENERICALLY DEPENDENT on (2) some MATERIAL ENTITY and which (3) stands in a relation of ABOUTNESS to some ENTITY.

An ICE is thus conceived as an entity which is about something in reality and which can migrate or be transmitted (for example through copying) from one entity to another. In what follows we introduce and defend proposals to improve this definition and the IAO as a whole.

The relation of generic dependence was introduced into BFO 1.1 in order to capture the fact that some dependent entities – for example the dependent entity which is the pattern of ink marks in your copy of the novel *War and Peace* (a complex *quality* in BFO terms) – are able to migrate from one bearer to another (e.g. through use of a photocopier). Generic dependence can thus be defined as follows:

> *a generically depends on* *b* = def. *a* exists and *b* exists and: for some universal B, *b* *instance_of* B and necessarily (if *a* exists then some B exists)

In BFO 1.0 the migration of dependent entities from one bearer to another was excluded. Dependence was seen as amounting in every case to *specific* dependence, or in other words as a relation which obtains between one entity and another *specific* entity when the first is of its nature such that it cannot exist unless the second also exists. A smile is dependent in this sense on a certain specific face, a head-ache on a certain specific head, a charge on a certain specif-

ic conductor. Generic dependence, in contrast, obtains where the first entity is dependent, not on some *specific* second entity, but rather merely on there being *some* second entity of the appropriate type (Smith et al. 2015). A DNA sequence is generically dependent in this sense on *some* but not on any *specific* DNA molecule; a pdf file on *some* but not on any specific memory store; and so on.

A generically dependent entity is in each case *concretized* (see definition in section 5) in some specifically dependent entity (more specifically in some BFO:quality). For example, this DNA sequence is concretized in this specific ordering (pattern) of nucleotides in this particular molecule; this sentence is concretized in this pattern of ink marks on this piece of paper (or also in this pattern of neuronal connections in the brain of the subject who reads it). The term 'pattern' can thus be understood in two senses – as referring either (i) to what is shared or communicated (between original and copy, between sender and receiver), or (ii) to the specific pattern before you when you are reading from your copy of Tolstoy's novel.

We can now define:

> INFORMATION QUALITY ENTITY (IQE) =def. a QUALITY that is the concretization of some INFORMATION CONTENT ENTITY (ICE) (Smith et al., 2013),

noting that IQEs are called 'information carriers' in the current version of IAO.

All concretizations are qualities in the BFO framework. Such qualities can serve as the basis for dispositions. When we concretize a lab test order by reading the text of the order on our screen, then in addition to the mental quality that is formed in our mind as we read the text, there is also a disposition to be realized in our actions of carrying out the relevant test. This disposition may come into being simultaneously with the mental quality created through our understanding of the text, but it is still dependent on this quality, as is shown by the fact that the latter may exist even in the absence of any accompanying disposition.

We define 'artifact' and 'information artifact' as follows:

> ARTIFACT =def. a MATERIAL ENTITY created or modified or selected by some agent to realize a certain FUNCTION or ROLE (Examples: a key, a lock, a screwdriver)

> INFORMATION ARTIFACT =def. an ARTIFACT whose function is to bear an INFORMATION QUALITY ENTITY. (Ex-

---

* To whom correspondence should be addressed: phismith@buffalo.edu

amples: a hard drive, a traffic sign, a printed form, a passport, a currency note, an RFID chip, a SIM card)

As a matter of definition, therefore, all information artifacts are material entities. While every ICE is dependent upon some material entity that is its bearer ICEs themselves are not material entities.

In reflection of the needs it was originally designed to address, the IAO is focused deliberately on ICEs associated with information *artifacts* – above all scientific publications and databases – thus with information entities which are continuants in BFO terms. No less important, however, is the occurrent side of the informational coin, which is made up of those processes – above all acts of thinking, speaking, hearing, writing and reading – through which ICEs are created, understood, and communicated. Given that thinking and speaking pre-dated writing, we know that acts of these sorts existed long before there were any information artifacts. They are of crucial importance to the ontological treatment of the phenomenon of aboutness because it is they which provide the relational tie between representations and their targets in reality.

If, therefore, we are to deal with these more fundamental aspects of the information pipeline, then we will need to embed the IAO into a wider framework of ontologies. This would include, on the one hand, all existing domain ontologies, which can be seen as representing the portions of reality about which we have information – they are ontologies of the various families of *targets* of aboutness. More importantly here, however, it would include on the other hand the Mental Functioning Ontology (MFO), which is designed to provide the resources to describe different types of cognitive acts, including those cognitive acts as a result of which ICEs are created (Ceusters & Smith, 2010).

2 ABOUTNESS AND PORTIONS OF REALITY

Aboutness corresponds to what is otherwise referred to by means of the expressions 'reference' or 'denotation,' (Yablo, 2014) but generalized to include not merely linguistic reference but also the relations of cognitive or intentional directedness that are involved, for instance, when a nurse is *measuring a patient's pulse rate* or a doctor is *observing a rash on a patient's thigh*. These processes are about, respectively, a *pulse* and a *rash*. When the nurse enters the string *72 beats per minute* in the medical chart of the patient, then there is an ICE that is concretized in the ink (or pixel) pattern exhibited on the chart, which inherits its aboutness from the aboutness of what we shall call the nurse's *direct cognitive representation* of the pulse. The latter is a (binary) relational quality; it links the nurse causally to the target of his observations. It is on this basis that, by entering data, he creates an ICE that is also tied relationally to its target in reality. Thus the ICE is not an abstract entity analogous to a 'proposition' in logical parlance. Rather it is a created, historical entity that is marked by the feature of indexicality: its

aboutness and its rootedness in time and context are analogous to those of an instruction issued by someone who points his index finger and says 'go *there now*.'

The current IAO definition of ICE can account for the aboutness involved in many examples of these sorts. However, we believe that it falls short when it comes to more complex cases. In (Ceusters, 2012) we proposed broadening the definition of ICE to require 'aboutness to some *portion of reality*' rather than just 'to some *entity*,' in order to allow the domain of the aboutness relation to include *inter alia*

- universals, for instance in the ICE concretized by the string *there are no instances of dinosaur which survive*,

- relations, for instance in the ICE concretized by the string *the part-whole relation is transitive*,

- other ICEs, for instance when someone asserts that what someone else just stated is true, and

- configurations, for instance in the ICE concretized by *Barack Obama is the current President of the USA*

– none of which is an entity in BFO terms.

The last example on this list is not only about Barack Obama but also about his *role* of being President of the USA and about the USA itself. But it is not only about these entities taken singly; in addition, it is about how the three entities are related to each other in a certain interval of time, and about the entire portion of reality – the configuration – made up by all of these together. This configuration is *asserted to exist* by a human subject using the corresponding sentence in a specific sort of context and with a specific sort of associated cognitive quality. But it can also be *referred to*, for instance when someone makes a second-order assertion using a nominalized expression, as in: *That Barack Obama is President of the USA is of epoch-making significance*.

3 INFORMATION AND MIS-INFORMATION

We can on this basis address another issue with IAO's current definition of ICE, which is that it does not give us a clear way of doing justice to the distinction between *information* on the one hand and what we might call *mis-information* on the other. Consider the ICE concretized in the sentence *Barack Obama was never President of the USA*, written on some piece of paper in 2015. This ICE is indeed about Barack Obama, the USA, and so forth. But what it communicates about these entities is something that is false. Our amended definition of ICE can allow us to accept that both information and mis-information exist, but also to recognize that the latter is not a special type of the former (that what some people might call 'false information' is not a special type of information, any more than a cancelled oophorectomy is a special type of oophorectomy). We achieve this by using our generalized definition of ICE to formulate a view according to which the relation of aboutness between a composite (for example sentential) ICE

and the associated portions of reality can obtain (or fail to obtain) simultaneously on two (or in principle more than two) levels: first, on the level of simple referring expressions such as 'Barack Obama' and 'USA'; and second, on the level of more complex expressions such as sentences and their nominalizations.

A true sentence on the upper level is about a corresponding configuration (where the term 'configuration' is to be understood in a way similar to the way 'fact' or 'obtaining state of affairs' are understood by some philosophers (Wittgenstein, 1961)). We can now capture the fact that a given compound expression may inherit aboutness from some or all of its constituent simpler referring expressions but fail in its claim to aboutness (and thus to convey information) when taken as a whole.

If someone writes on a piece of paper the sentence *Barack Obama is President of Russia*, then there is an ICE – concretized by this written string and by any copies made thereof – which is generically dependent on the piece of paper and which is about (on the aforementioned lower level) Barack Obama, his being president, and Russia. But this ICE is not about any corresponding configuration, simply because there is no corresponding configuration. It is for this reason that the given sentence, while it is *about* certain entities in reality, is nonetheless not *true of* those entities. This strategy can be used also to explain how a fictional sentence such as *Sherlock Holmes was a user of cocaine*, can concretize an ICE – by inheriting aboutness from one or more of its components (here for example the string *cocaine*, which is about a corresponding universal) – even though the sentence as a whole is not about anything in reality.

A related problem with the current IAO is that it does not provide us with the resources to do justice to what happens with certain types of ICE when what they are about changes over time. The problem here is that the ICE concretized by the sentence *Barack Obama was never President of the USA* written on a piece of paper in 2007 was true when it was written; yet it appears that this very same sentence, when read by some observer in 2009, would be false.

This appearance is misleading, however, for it is not the case that the ICE in question changes in the intervening period. Rather, what has changed is the first-order reality that this ICE claims to be about. Certainly as a result of these changes in first-order reality there came into existence many new ICEs relevant to Obama, the presidency and the USA, with many new concretizations. But the original ICE, with its original concretization born with its original act of creation, must nonetheless still be evaluated as true. This is because, as in the case of the nurse's data entry above, the ICE in question has its time of origin baked into it through the indexicality of the *was* in *was never President*.

We shall presuppose in what follows that information artifacts do not bear information in and of themselves, but only because cognitive subjects associate representations of certain sorts with the patterns which they manifest. We thus view the aboutness that is manifested by information content entities in accordance with the doctrine of the 'primacy of the intentional' (Chisholm, 1984), according to which the aboutness of those of our representations formulated in speech or writing (or in their printed or digital counterparts) is to be understood by reference to the cognitive acts with which they are or can in principle be associated. The entry *72 beats per minute* is about what it is about because of what the nurse himself directly observed when he measured the patient's pulse (or, in the case where the ICE is created by sensor devices automatically adding data to the chart, it is about what the nurse would have observed in the given circumstances).

At higher levels we may have ungrounded representations, as illustrated for example in the letter published by Urbain Le Verrier in 1859 (Le Verrier, 1859) in which there appears an intended reference to a planet that is asserted to be intermediate between Mercury and the Sun, a planet which in 1860 Le Verrier baptised 'Vulcan'. This intended reference depended on a certain belief on Le Verrier's part in the existence of an intra-Mercurial planet. When we understand Le Verrier's text today, however, then we have a different sort of cognitive representation – involving what we refer to below as a *recognized non-referring representational unit* (RNRU) – in which this intended reference to a planet has been cancelled.

Such changes in our understanding of the reference of terms are of course a common phenomenon in the world of ontology, and specifically in the world of ontology versioning. Paying careful attention to these changes forms the basis for the strategy for ontology evaluation we have outlined in (Ceusters & Smith, 2006).

4 REPRESENTATION AND REFERENCE

We build on the notions of representation and representational unit informally introduced in (Smith et al., 2006). A representation is there described as *an idea, image, record, or description which refers to (is of or about), or is intended to refer to, some entity or entities external to the representation*. Note that 'representation' is thus more comprehensive in scope than 'ICE,' even on our proposed more inclusive definition of the latter, since an ICE must in every case be *about* some portion of reality, where the aboutness in question must always be veridical, so that 'being about' is a success verb. A representation, in contrast, is required merely to intend to be about something, and this intention might fail (as when a child draws what she thinks of as a unicorn).

We provided a formal definition of 'representation' along these lines in (Ceusters & Smith, 2010):

> **REPRESENTATION** =def. a QUALITY which *is_about* or is intended to *be about* a PORTION OF REALITY (POR).

We can now single out cognitive representations (representations of the sorts instantiated in the brains of beings like ourselves) by means of the terms:

MENTAL QUALITY =def. a QUALITY which *specifically depends on* an ANATOMICAL STRUCTURE in the cognitive system of an ORGANISM.

COGNITIVE REPRESENTATION =def. a REPRESENTATION which is a MENTAL QUALITY.

defined in the Mental Functioning Ontology. We are here attempting to remain neutral as concerns the precise nature of cognitive representations; thus it does not follow from the definitions that such representations involve something like images; nor does it follow that they must all be *conscious* representations.

As concerns occurrents in the realm of cognition, it is clear that mental processes, too, for example processes of thinking or imagining or remembering, may be about or be intended to be about some portion of reality. We hypothesize, however, that such occurrent representations are always such as to inherit their intended aboutness from some underlying continuant representation. When the doctor sees, and recognizes, for example, that there is a rash on her patient's leg, then her act of recognition coincides temporally with the beginning to exist of a correspondingly targeted (relational) mental quality on her part (Smith, 1987).

As we saw above, cognitive representations may be more or less complex. When analyzed into their constituent parts, however, then we arrive at what we called 'representational units' (RUs), defined as *the smallest constituent sub-representations*, including *icons, names, simple word forms, or the sorts of alphanumeric identifiers we might find in patient records*. (Smith et al., 2006)

Subtypes of representational unit can then be defined as follows (Ceusters & Smith, 2010):

1. *Referring representational unit* (RRU)*:* an RU which is both intended to be about something and does indeed succeed in this intent.

2. *Non-referring representational unit* (NRU)*:* an RU which, for whatever reason, fails to be about anything.

3. *Unrecognized non-referring representational unit* (UNRU): an NRU which, although non-referring, is intended and believed to be about something;

4. *Recognized non-referring representational unit* (RNRU): an NRU which was once intended and believed to be about something, but which, as a result of advances in knowledge, is no longer believed to be so;

5. *Representational unit component* (RUC): a component of a representation that is not intended by the artifact's authors to refer in isolation;

RU	'Paris'
NRU	'Atlantis'
UNRU	'Vulcan' (as used by Le Verrier in 1860)
RNRU	'Vulcan' (as used now when referreing to Le Verrier's error)
RUC	'Le' (as it appears in the third row of this table)

Table 1: Examples of types of representational unit

Note that, as the 'Vulcan' case makes clear, classifications of representations under headings 1. to 5. may change with time. Note, too that, while items 2. to 5. on this list signify one or other kind of shortfall from aboutness, representations under item 1. include the fundamental (grounding, target-securing) cases of *direct cognitive representation* referred to in the case of the nurse taking someone's pulse as in our example above.

5 PROPOSAL

5.1 Primitives and elucidations

To do justice formally to the foregoing we propose the following primitive relational expressions. These cannot be defined, but only elucidated by means of examples and informal specifications of their meanings.

x is_about y means**:**

x refers to or is cognitively directed towards y. **Domain:** representations; **Range:** portions of reality. **Axiom:** if *x is_about y* then *y* exists (veridicality).

x concretizes y at t means:

x is a QUALITY & *y* is a GENERICALLY DEPENDENT CONTINUANT

& for some material entity *z*, *x specifically_depends_on z* at *t* & *y generically_depends_on z* at *t*

& if *y* migrates from bearer *z* to another bearer *w* then a copy of *x* will be created in *w*.

x is_a_direct_cognitive_representation_of y means:

x is a COGNITIVE REPRESENTATION in some subject *s*

& *x is_about y* & *x* comes into existence, as a result of a causal process initiated by *y* and in a way appropriate to *y*, in the cognitive system of *s*. **Example:** a causal process of visual perception initiated by an object presented visually to *s*.

5.2 Definitions

x is_a_representation_of y =def. *x* is a REPRESENTATION & *x is_about y* (where *y* is a portion of reality). Note that not all representations are about something.

x is_conformant_to y =def. *x* is an INFORMATION QUALITY ENTITY & *y* is a COGNITIVE REPRESENTATION & there is some GDC *g* such that *x concretizes g* and *y concretizes g*. **Example:** *x* is a sentence on a piece of paper, *y* is the

belief of the author of the sentence who wrote the sentence as an expression of her belief, and *g* is the ICE (the content) that belief and sentence share.

6 DISCUSSION

Although it is a requirement that the target of aboutness be a portion of reality (POR), there is no requirement that the relevant POR exists at the time when the associated cognitive representation exists. Thus a patient can contemplate a past disorder, for instance by regretting his not having accepted the advice of some clinician. His thoughts are then about that very disorder, and not for example about his memories thereof. This is so independently of whether the nature of the disorder is known to him or not.

There is also no requirement that the agent of a veridical representation knows what the portion of reality is that his representation is about: even a baby, or a cat, may see a flow cytometer. We can directly represent an object even though we are ignorant of or mistaken about what universal it instantiates.

There is also – as is illustrated by the case of believers in the Higgs boson before there was evidence for its existence – no requirement that aboutness must imply that the subject *knows* that what he is representing exists – he must merely *believe* that it exists.

Although neuroscience, to our best understanding, is not yet sufficiently advanced to provide answers to the question what the precise physical basis of a mental quality exactly is – for example whether it is certain spatial configurations of one or more molecules in one or more brain cells – we believe that the following hypothesis is correct: that an anatomical structure in which there *can* inhere a mental quality need not always *have* a mental quality inhering in it (in this respect *having a mental quality* is comparable to having the quality of *being pregnant* and is to be contrasted with qualities such as height and mass, given that something in which there can inhere a height or a mass must always have a height or mass of some determinate sort). From this, it is then just a short step to the question of whether there can be unconscious representations, a question which, however, we must here leave aside for reasons of space.

7 CONCLUSION

IAO was designed to deal with information artifacts, which is to say with continuants such as the information stored in hard drives or formulated in written sentences or in printed texts – thus with information that is shareable between multiple bearers, including bearers existing at different times. As will by now be clear, the IAO must be embedded in a broader framework of ontologies, including the Mental Functioning Ontology (Hastings *et al., 2012)*. In the future we must address for example how an agent can use sight (or, in the case of Braille, touch) to process concretization in such a way as to generate mental representations that are conformant to the associated ICEs. For this we will require a Language Ontology – extending the Ontology of Document Acts proposed in (Almeida, et al. 2012) – that will allow us to do justice to the ways in which sentences can be not merely believed and thought but also asserted, heard, seen (for example in the case of sign language), understood, and formulated in written or printed texts.

ACKNOWLEDGEMENTS

REFERENCES

Almeida, M.B., Slaughter, L., & Brochhausen, M. (2012). Towards an ontology of document acts: Introducing a document act template for healthcare. Lecture Notes in Computer Science, 7567, 420-425.

Ceusters, W. (2012). An information artifact ontology perspective on data collections and associated representational artifacts. Stud Health Technol Inform, 180, 68-72.

Ceusters, W., & Smith, B. (2006). A realism-based approach to the evolution of biomedical ontologies. AMIA Annu Symp Proc, 121-125.

Ceusters, W., & Smith, B. (2010). Foundations for a realist ontology of mental disease. Journal of Biomedical Semantics, 1(10), 1-23. doi: doi:10.1186/2041-1480-1-10

Chisholm, R. M. (1984). The primacy of the intentional. Synthese, 61(1), 89-109. doi: Doi 10.1007/Bf00485490

Hastings, J., Ceusters, W., Jensen, M., Mulligan, K., & Smith, B. (2012). Representing mental functioning: Ontologies for mental health and disease Towards an ontology of mental functioning (icbo workshop), proceeedings of the third international conference on biomedical ontology.

Le Verrier, U. (1859). Lettre de m. Le verrier à m. Faye sur la théorie de mercure et sur le mouvement du périhélie de cette planète. Comptes rendus hebdomadaires des séances de l'Académie des sciences (Paris), 49, 379-383.

Smith, B. (1987). On the cognition of states of affairs. In K. Mulligan (Ed.), Speech act and sachverhalt (Vol. 1, pp. 189-225): Springer Netherlands.

Smith, B et al. (2015). Basic formal ontology 2.0 draft specification and user manual. from http://bfo.googlecode.com/svn/trunk/docs/bfo2-reference/BFO-Reference.docx

Smith, B., Kusnierczyk, W., Schober, D., & Ceusters, W. (2006). Towards a reference terminology for ontology research and development in the biomedical domain Kr-med 2006, biomedical ontology in action. Baltimore MD, USA

Smith, B., Malyuta, T., Rudnicki, R., *et al*. (2013). Iao-intel: An ontology of information artifacts in the intelligence domain. Paper presented at the STIDS.

Wittgenstein, L. (1961). Tractatus logico-philosophicus. London: Routledge and Kegan Paul.

Yablo, S. (2014). Aboutness, Princeton, NJ:: Princeton University Press.