A formal ontology is about “being”. It gives classes of individual items, properties and logical rules constraining their combinations that approximate at a categorical level how we perceive that certain things and phenomena of reality are and behave, including our descriptions of it. It describes “possible states of affairs”[1].  We require that these concepts are not only conventions between humans, but also sufficiently close to reality so that valid deductions about reality can be drawn from the ontology and instances of it, obtained under theoretical, perfect conditions of observation. The deviations in precision and coverage (i.e. wrt exceptions) of the ontology from reality as an idealized, logical approximation should be understood and tolerable for the purpose of the respective research. Only things and phenomena of reality that behave close enough to the logical form of a formal ontology can be usefully described by it.

We regard knowledge as justified belief of proposition X of a form that makes sense in “I know that X holds”. Besides formulating the proposition X as an expression of information, a human stating “I know that X holds” must be able to relate all classes, properties and identifiers (names) in such an expression with situations and individual things of the real world as a correct characterization. Therefore only humans have knowledge.

A knowledge base in the sense of the CRM is an information object that instantiates the formal ontology with statements of propositions that the maintainer of the knowledge base believes, i.e., regards it as “the best of my knowledge”. There are subtle, but substantial differences between registered knowledge and reality, because it includes contradictions, alternatives and uncertainties. The maintainer of the knowledge base may be an individual person or a team, trusting each other and sharing the same contextual knowledge of the world (see Doerr, Meghini & Spyratos 20. I agree. As Levesque says, at the knowledge level a KB is an agent, meaning that the KB is not only a set of statements, but it also has reasoning abilities that can be used through the interface of the agent. This interface provides two methods: Ask (to ask questions) and Tell (to tell statements).

The maintainer of the knowledge base is its ultimate provenance, providing (or not) trust in the care and honesty of the described propositions. The maintainer should not appear in the knowledge base as propositions of provenance, but be described as metadata about the knowledge base as a whole, exactly as we do not repeat the author of a book in each phrase. The two things are compatible, in the sense that the provenance info is part of the metadata.

The knowledge may be direct or indirect:

Direct knowledge is that believed out of good, explicable reasons of observation or inferencing by the maintainers themselves.

Indirect knowledge is that that the maintainer adopts or refers to from other sources. In that case, the formal knowledge of the maintainer is restricted to the information as a formal expression and its provenance. The maintainer may or may not believe this information. Therefore, the knowledge base should contain the adequate propositions about its provenance (believed by the maintainer). The maintainer may express doubts about the correctness of this information, if indicated. I agree Martin, but there is a little contradiction here: if knowledge is justified belief, then also indirect knowledge must be believed by the maintainer, otherwise it’s not knowledge. One possible solution is to say that the KB contains knowledge (justified belief of the maintainer for which no provenance is required) and *some other statements* that the maintainer needs to record for some reason but has no justification for believing. For this latter kind of statements provenance information is necessary.

It should be possible to communicate with the maintainer and discuss justifications and possible corrections of errors.

Ideally, the source of indirect knowledge should contain further provenance statements about indirect knowledge its author has used. The ideal would be, to link all those provenance statements together until they direct us to all direct knowledge used. This is, of course, impossible, but nevertheless we have the means to document, increase and link our provenance knowledge to larger and larger chains, which will be extremely useful for validating and improving our overall knowledge. I do not see why it should not be possible to link statements together marking some of those as “believed by X”

The maintainer of a knowledge base may decide to document provenance of provenance, if there is no reliable digital resource to link to next statement in the provenance chain, or if a local copy of parts of the provenance chain appears to be useful. There has to be a limit to this regression of provenance. And I believe the limit is when you reach someone’s beliefs. These statements do not have any provenance, they can be considered as the primitives of knowledge (for some agent).