Dear Dr Kuhn,

We would like to thank you for providing us with the opportunity to revise our manuscript, and we would like to thank the reviewers for their dedicated comments that helped to improve our manuscript significantly.  We have now revised our manuscript, and provided a detailed response to each reviewer comments and requests below.

**Editor comments:**
The most important shortcoming of your current paper are the missing position paper aspects as pointed by the second reviewer:
missing conclusion / general landscape description
missing road-map / vision / challenges / opportunities
a missing bold look into the future
Additionally, I think you should consider including how human factors and cognitive computing relate to the discussed issues, as pointed out by reviewer 4.

**Response:** We would like to thank the editor for highlighting the main points that should be considered under revision. We have now significantly revised our manuscript, and we have restructured and rewritten it to: give a clear general overview of the landscape; to state more clearly our vision of the challenges and opportunities and a roadmap to tackle them; and in the last section a bold look into the future of data science, where the role of human factors and cognitive computing have been included in the discussion of limits of AI algorithms that data science can be key to push forward.

**Editor comments:**
In addition to the various minor comments by the reviewers, I have the following comment:
"The OBO Flatfile Format and many biomedical ontologies utilize this definition pattern for edges": Is this really restricted to this format? Isn't it a property of OBO in general?

**Response:** We have restructured and rewritten the entire manuscript according to the reviewers suggestions and comments and the discussion in which this sentence was embedded has been removed from the text.

**Reviewer 1:**
The 'position' of the paper, though topical, is also not a very surprising one. As the authors themselves already note, there exist several efforts in this direction already.

**Response:** We have now changed our manuscript and made the relevance for Life Sciences more explicit, which we hope improves the focus of the manuscript and provides several novel aspects.  We also put a challenge for data science at the end of our manuscript; research on this challenge is, to the best of our knowledge, not a major active research area (on the contrary, major research trends seem to move towards more focus on data-driven research), and the challenge has also not been stated in the context of data science; we hope that this is somewhat surprising.

**Reviewer 1:**
The paper is overly repetitive and detailed in explaining and introducing symbolic approaches, while statistical approaches are only mentioned by name. Technologies such as RDF, OWL/DL and Knowledge Graphs are introduced in a way that they seem to be very distinct,

whereas they are strongly interconnected. I am sure this is merely a matter of presentation, but it makes the paper a rather dense read, and distracts from the message, the position, that the authors want to get across.

**Response:** We would like to thank the reviewer for these comments. We have now significantly revised our manuscript, removed the repetitions, and strengthened the manuscript's focus. We have improved the manuscript in order to address the reviewer's points, and now we think our messages are more clear. In particular, we have reduced the verbosity by removing redundant explanations such as for symbolic approaches. We have expanded and balanced the introduction and description of statistical approaches with respect to the symbolic ones. We also revised our introduction of knowledge representation technologies such as RDF, OWL or knowledge graphs, and connected them more explicitly. We hope that with these changes the reading of the manuscript is substantially better.

**Reviewer 1:**
The discussion of the "grand challenge" for AI to solve is interesting (discovering the principle of inertia), but not detailed enough to be convincing. What concrete evidence is there that "an addition creative step" (sic) needs to be taken to meet that challenge?

**Response:** We have revised this aspect of section 4, and now provide more details as well as the relation to other efforts that may prove to be useful in solving this challenge (and provide an argument for continued human involvement in such challenges, at least for now).

**Reviewer 2:**
However, as a vision/summary paper, it lacks, in my opinion, more clear discussions on how these two areas can actually intersect, and on the actual challenges and opportunities that this combination can bring in.

**Response:** We have now added a new section called "Data and knowledge in Life Science research" in which we discuss the opportunities that the intersection of both areas can bring within the Life Sciences. We also significantly revised our manuscript to remove duplication, highlight particular challenges, and areas in which we believe (or hope) data science and AI will move.

**Reviewer 2:**
However, reading this paper does not bring in clearly a very clear landscape of where we are (just to name an example, I have not seen clear references to ontology learning techniques i section 3), but mostly a sequence of descriptions of existing efforts that are in between both areas, which do not flow very clearly and do not tell a clear story. Which are your main conclusions from the analysis that you have done? Which are the actual challenges and opportunities that you refer to in the abstract and title? You refer in section 4 to the limits of data science and the relationship to some theories in Science, but these are just examples, and it is not clear how symbolic AI can help in those specific cases, where it seems that mathematical formulations are potentially more useful.

**Response:** We would like to thank the reviewer for these comments. We have rearranged and modified the content of sections 2, 3 and 4 in order to clearly discuss our analysis and vision on the current challenges and opportunities that the combination of symbolic and connectionist AI approaches can bring in for the development of methods and technologies used in data science. As suggested, we have included in section 3 references to ontology learning techniques. In section 5, we have made a bold statement on the limits of data science, and we have placed a key role of data science on the future landscape of knowledge discovery in Science. In our analysis, we describe the complex interactions between symbolic and connectionist AI approaches, human factors and cognitive computing will play in surpassing this limitation.

**Reviewer 2:**
But I am missing a clear picture, a clear roadmap, for such a vision paper, where the low hanging fruits are identified, where the longer-term opportunities are discussed, and where the main challenges and potential limitations are also identified and discusssed, so that the paper can provide a good analysis of how such intersection of areas may happen, even if we may be totally wrong when looking back at this paper in a couple of years' time.

**Response:** We have aimed to add such a statement of challenges for the next years throughout our manuscript. We have also restructured our manuscript significantly to clearly distinguish between the current state of the field and future directions (as we see them).

**Reviewer 3:**
Uncarefull writing, with typos, unclear notations and sometimes sketchy structure.

**Response:** We have rewritten and restructured the manuscript to enhance its clarity.

**Reviewer 3:**
Page 1: world in which we „are living” or in which „we live”?

**Response:** We have rewritten the sentence.

**Reviewer 3:**
Page 2, second paragraph: the topic of this paragraph is unclear. Is this mostly about scientific experiments, or about data life cycle? Please choose one main topic per paragraph.

**Response:** We have restructured our Introduction section and have split this paragraph.

**Reviewer 3:**
It also reads sketchy at some parts. It would be better to provide a complete list of steps of a *typical* data life cycle instead of giving only examples (maybe the sentence could be rewritten to avoid „mainly consists” and „among others”).

**Response:** We have rewritten our manuscript according to the suggestions.

**Reviewer 3:**
What is „archival”?

**Response:** We have removed the word from the manuscript.

**Reviewer 3:**
Page 3: something wrong with this sentence:

„In a physical symbol systems [35], physical entities (tokens, symbols) stand for, or denote, entities, are combined with other symbols to form complex symbol structures, and are manipulated by processes”

**Response:** We split the sentence and rewritten this part of the manuscript to make our intentions more clear.

**Reviewer 3:**
Page 4:
Fig.1: It is unclear to what refer the processes of „creating”, „processing” etc. To „data” coming from social, technology and science disciplines?
Can A.I. be classified under both „science” and „technology”? Why e.g. „Computing” is not under „Science” and „Meteorology” under „technology”. Why „Energy” is under Social? What is exactly meant by „energy”? There is a typo in the word „philosophy”.
Overall, the figure makes an impression of a bit too arbitrary. It could probably be more tidied up.

**Response:** An explanation of the kinds of data that we consider in the focus of data science has now been included in the caption of the figure. We have improved the figure, selecting carefully the disciplines included in each category. We drop disciplines such as AI as we agree that their classification is ambiguous.

**Reviewer 3:**
The definition of „ontology” is given two times, once in Section 2 (Page 4) and then again (very similar definition) in Section 3. Should I suspect that specific paragraphs were written by different authors but without consolidating and proof-reading the final result?

**Response:** We apologize for this oversight, we have now restructured our manuscript to remove many duplications, and "ontology" is now only defined once.

**Reviewer 3:**
Last paragraph of Page 4: tenses are mixed (once there is „have been developed” and the other times „were developed”).

**Response:** Fixed.

**Reviewer 3:**
The citied references dealing with knowledge graphs seem to be picked a bit ad-hoc. Could there be more structure in referencing them?

**Response:** We have restructured this part of the manuscript according to the reviewer's suggestions.

**Reviewer 3:**
Page 5:
Fig. 2: What kind of a relation/process is meant by two big black arrows? Maybe this can be explained in the caption.
„if”->”whether”?

**Response:** We have explained the meaning of the big black arrows in the caption of the figure.

**Reviewer 3:**

I do not agree that only some OWL axioms may give rise to a graph structure. In principle, all OWL axioms can be serialized as RDF, i.e. as labelled graphs.

**Response:** We intended to distinguish between TBox and ABox axioms; whereas ABox axioms can be "naturally" represented as graphs (at least in their corresponding model structures), representing TBox axioms as graphs is not as "natural". However, we agree with the reviewer that our statements were confusing and have removed this sentence from our manuscript.

**Reviewer 3:**
Fourth sentence from the end: the symbol of logical „models” or „entails” is corrupted.

**Response:** We removed the formula from the manuscript.

**Reviewer 3:**
Section 2 would benefit from more structured discussion on what are the particular problems of treating knowledge as data. The current one is a bit too sketchy.

**Response:** We have now restructured section 4 (and the other sections) and hope that our argumentation follows a clearer structure.

**Reviewer 3:**
Page 7:
What is „almost infinite”? Is it finite or not or just huge?

**Response:** We have rewritten this statement (it's just "huge").

**Reviewer 3:**
References:
Reference [1] lacks the name of the author.
There are missing letters in the names of some authors, e.g. [9, 20]. Some authors are referenced using their first name and some not.
Some typos:
-grpah
-Page 7: an addition -> in addition?

**Response:** All fixed.

**Reviewer 4:** I found it is a bit unfortunate that the authors totally missed the consideration of human factors (such as human guided intelligence or human-machine interactions [1]) in this discussion. From the data science's point of view, it is very important to understand what humans are good at (putting in the era of big data) and how such human-factors can be utilised in or used to guide AI algorithms, such as [2]. From the societal point of view, with the increasing concerns from the public about Ai taking over our jobs, this topic also seems unavoidable. Just as Stephen Hawking pointed out "AI will be 'either best or worst thing' for humanity".

**Response:** We have now added a new discussion (to the "Limits" section) to discuss human factors and the role humans will likely play in the future of scientific research, as well as to the disruptive effect AI and data science have on our society.

**Reviewer 4:** Following this direction, I think it is also worth touching the topic of cognitive computing [3], which was proposed as a interdiscipline between computing science, neuroscience and nanotechnology and is now more about technologies to mimic the functioning of the human brain.

**Response:** We have added the discussion of cognitive computing, and research on computational creativity, to the manuscript (in the "Limits" section as a way to overcome current limitations of data-driven scientific research).

**Reviewer 4:** In addition to the related work introduced in the paper, there is another line of work of integrating reasoning with machine learning at the workflow level that is "an algebraic operation in a space of (machine learning) models"[4]. Furthermore, the paper mentioned "knowledge graph" in various sections without a definition. It would be helpful to provide a brief introduction or references, such as [5].

**Response:** Thank you for the references. We have added a better description and reference about knowledge graphs to the manuscript, as requested.

**Reviewer 4:** minor suggestions

**Response:** All fixed.