

Reviewer #1

The authors are grateful to Oksana Tarasova for her very positive and helpful comments which definitely helped to sharpen the article and make the important messages more visible. Below, we respond to Oksana's comments point by point.

The paper "New directions in atmospheric research offered by research infrastructures combined with open and data-intensive science" provides useful view on the future of the atmospheric research. The paper contains a lot of information and in its current form it rather looks like a comprehensive review than an opinion. It will take a reader 25 pages of reading until the authors go to the point.

I would suggest that the author shorten the paper by at least 30% , which still would allow to make a point about future research approaches. In particular, this applies to the references and description of the infrastructures where the whole paragraphs are just copied from the referred publications.

Reply: To make the key messages more visible, we moved the entire former Section 3 on Research Infrastructures into an Appendix. However, we do not delete this section completely since from our discussions at multiple dissemination events in our research communities, we still recognized that not many scientists and in particular young researchers are aware of what research infrastructures are and how they can make use of them.

By rearranging the manuscript, the key message appears now much earlier, but the background information on Research Infrastructures is still available for those who are not familiar with this new tool.

I find the authors opinion interesting, though it would be useful to have authors opinion on two points which are currently not reflected in section 5. The first one: currently the performance of scientists is judged by a number of publications they produce. Within the envisioned new research environment, the research will be done as a collective one. This diminishes the role of the individual researcher and the current evaluation framework by publication number (as in the envisioned research environment the summaries can be easily generated by AI). The second question is related to the role of innovation (e.g. regarding analytical methods or measurement techniques and instruments) within the highly standardized research environment.

It would be useful if the authors reflect on these points in the next revision of the paper.

Reply 1/: The need to revise the reward system for researchers is being discussed as part of the open data and open source movements. In response to this discussion, the citation of datasets and their link to the responsible researchers has been improved by adding DOIs to datasets and referencing them in publications. This cultural change is taking place across the scientific community. It is also called for in the European Open Science Cloud Declaration of 2017 (https://eosc-portal.eu/sites/default/files/eosc_declaration.pdf). For this reason, we have decided not to focus on this issue as it is independent of the atmospheric sciences.

Reply 2/: The role of innovation, e.g., regarding analytical methods or measurement techniques and instruments, within the highly standardized research environment cannot be underestimated. Research infrastructures not only operate instruments following standardized operation procedures but improve existing or develop new techniques as part of their mission. This is an intrinsic activity of the research infrastructures. However, this activity is independent of the combination of data intensive science and research infrastructures, which is the topic of this opinion article. For this reason, we decided to leave the topic of technological innovation untouched in our opinion article.

General note on the use of language: the sentences are too long, and the idea is often get lost before one reads to the end. It would be magnificent if authors use more concise language.

Reply: The entire manuscript has been revised for a more concise language.

Additional comments:

1. Line 22: the term “variability” may be a bit better than “periodicity”

Reply: Thank you for this suggestion, we use now seasonality instead of periodicity.

2. Line 31: “which evolve from”; it is not clear from what opportunities evolve as you do not provide initial state of affairs. Maybe it would be better to use another works here, like “emerge”

Reply: Thank you, we use now “arise”.

3. Line 32: what do you mean with “emerging service ecosystem”

Reply: We used more concise language and write now: “and the developing collaboration platform ENVRI-Hub, hosted by the European Open Science Cloud.”

4. Line 44: “economic processes in the Earth system” – economic processes refer to human society rather than to the Earth system

Reply: We replaced “Earth system” by “on our planet” which includes both the Earth system and human activities.

5. Line 50: GCOS is a co-sponsored programme of WMO, IOC-UNESCO, UNEP and International Science Council

Reply: Thank you, this remark is included.

6. Line 55-56: do you consider vertical profile measurement as ground based or as in situ?

Reply: We specified as “ground-based and airborne in-situ observations”.

7. Line 62-66: lots of repetitions here

Reply: This paragraph has been significantly shortened.

8. Line 67: actually the threats are posed to humanity rather than to the planet

Reply: The threats of climate change target the human society but also biodiversity in all compartments of the Earth system. Therefore, we prefer to keep “our planet”.

9. Line 94-94: I would disagree with the comparison with astronomy infrastructure. Unlike astronomy, environmental infrastructure has an immediate value for multiple applications, including climate services for mitigation and adaptation, health and agriculture applications, hence this poses different requirements for timeliness of data availability.

Reply: The comparison with astronomy refers to the implementation of tools for data intensive science. However, the statement on the timeliness of data availability is highly valuable and added to the paragraph. The full paragraph reads now:

“Astronomy and high-energy physics can serve as role models for this kind of infrastructures (Bell et al., 2009), but with the essential difference that environmental observations always require the

operation of largely distributed measurement networks together with the handling of complex data streams from highly diverse sources and the need for short-term data availability.”

10. Line 120-126: it is not clear what point you are trying to make in this paragraph

Reply: Infrastructures and networks like ICOS contributed significantly to the data base behind the Nobel Prize 2021. Therefore, we want to keep this paragraph.

11. Line 146-151: not clear what the relevance of this paragraph to this paper

Reply: This paragraph was removed for the sake of conciseness.

12. Line 154: air quality is a part of atmospheric composition

Reply: This is correct, but we wanted to mention it explicitly. So, we added “atmospheric composition in general”.

13. Line 170-176: another paragraph of repetitions

Reply: This paragraph was removed for the sake of conciseness.

14. Line 187-190: the name of infrastructures should be spelled out here, rather than later in the text (lines 205-210)

Reply: This paragraph was removed for the sake of conciseness. Details refer only to the atmosphere-centric infrastructures.

15. Line 189: I was under impression that ACTRIS also recently became ERIC

Reply: That is correct, the reference is updated.

16. Line 203: it is not necessary to spell out the title of the book, reference would be sufficient

Reply: The title of the book was removed.

17. Line 214: “in providing” (“in” is missing)

Reply: Corrected.

18. Figure one: the name of the phenomena is “sand and dust storms” not “desert storms”

Reply: Corrected.

19. Table one: “aerosol particles” – use either of the words, not both

Reply: From a scientific perspective, the term is correct since it refers to the particle phase of the aerosol which is composed of both the particle and the gas phase.

20. Comma is missing after O3 in IAGOS section (Table 1)

Reply: Corrected.

21. Line 237: what do you mean with “research infrastructure process”?

Reply: We added “implementation process” for clarification.

22. Line 240-241: “Before the consolidation phase, data have been locked in silos and were poorly standardised over decades, hampering scientific progress severely” – I disagree with this statement as it diminishes the role of the programmes like the Global Atmosphere Watch

Reply: We modified the statement and included the role of GAW. The sentence is now: “On a global scale, first harmonisation efforts were undertaken by the Global Atmosphere Watch programme of WMO. During the consolidation process, the ATMO-RIs have developed cutting-edge standardisation for their observations and implemented highly efficient data life cycle management systems, leading to timely, open, and FAIR data access by scientists all over the world.”

23. Line 251: “periodicity in the presence of” is better to reformulate as “variability in atmospheric levels/burdens of”

Reply: as suggested by Reviewer #2, we rephrased the sentence as follows: “The detection of trends and seasonality in the presence of greenhouse gases ... “.

24. Line 259-296: what are you trying to demonstrate with the presented three examples?

Reply: Among the most important achievements of Ris is the securing of long time series collected in projects which would have been lost otherwise after the finalization of these projects. This is what we want to demonstrate with the presented examples.

25. Line 277: “Data which were collected before ICOS measurement protocols were put in place, have been secured physically”. Could you please explain this statement? Most of ICOS atmospheric data are part of GAW and those data were and are achieved in the World Data Center for Greenhouse Gases supported by Japan

Reply: This statement refers to the work performed by the RIs for lifting the historic greenhouse gas data to the level of the ICOS data protocol. In this respect, there is an additional achievement of ICOS beyond archiving station data in the GAW data base.

26. Line 286: the referred “latest assessment” is 5 years old. Are there updates on this publication?

Reply: The assessments are published every 5 years. So, a new assessment is in preparation but not yet published.

27. Line 299-300: could you please clarify the statement “Trends are detectable in the data, possible only due to collection of data with consistent quality and operating procedures over long time periods.”

Reply: Trend detection requires long time series of consistent data quality. We believe that the Ris make a significant contribution to the provision of time series which can be used for trend analyses.

28. Line 319: ICOS is a regional network

Reply: Of course, the ICOS observation network is densest in Europe, and in that sense it is of continental scale. But still, ICOS is also the densest greenhouse gas observation network worldwide. Therefore, we suggest to keep the formulation.

29. Line 322: “which investigates the processes” – who investigate the processes?

Reply: We refer to ACTRIS as an infrastructure which is investigating the atmospheric processes.

30. Line 323: what do you meant with “controlled atmospheres”?

Reply: The term “controlled atmosphere” refers to the concept, that we can fill atmospheric simulation chambers with an atmosphere of known composition and investigate chemical processes under these specific conditions.

31. Figure 4: I disagree with the proposed adjustment of the figure as it completely changes its sense. The original figure was designed to present a limited set of key climate indicators and each box includes only one variable, while the added box includes tens of variables disturbing the intended meaning. I would suggest to remove this figure from the paper.

Reply: In accordance with the response to a comment by Reviewer #2, we decided to use the original figure here. The sentence describing this figure has been adapted.

To make the difference between GCIs and ECV – which include short-lived climate forcers – clearer, we rephrased this paragraph to:

“To transform the challenging task of integrated Earth observation into a concept towards a global climate observation system, WMO/GCOS has defined a set of global climate indicators (WMO-GCI), which stretches beyond the boundaries of traditional scientific disciplines like, e.g., atmospheric sciences, ocean sciences, or biology; see **Fehler! Verweisquelle konnte nicht gefunden werden.** This set of GCIs defines the indicators which point to long-term changes in the Earth system. Consequently, WMO requests the continuous observation of these indicators for monitoring the state of Earth’s climate. However, this undertaking requires a largely interdisciplinary approach.

To further develop this interdisciplinary approach towards a global climate observation system, WMO has introduced the concept of essential climate variables supplementary to the GCIs. Essential climate variables for atmospheric composition are defined as a physical, or chemical variable or a group of linked variables that critically contributes to the characterization of the atmospheric composition. This set of variables includes the long-living greenhouse gas CO₂, but also many of the short-lived climate forcers.”

31. Line 346: “this approach” – which one?

Reply: We refer here to the largely interdisciplinary approach needed to build such a global climate observation system. This approach is mentioned in the previous sentence, but for clarification we added the term “interdisciplinary”. See also reply to item 31.

33. Line 348: could you please provide the source for the definition of “essential variable”. Is it introduced by the authors of the paper?

Reply: We used “essential variables” as synonym for essential climate variables. For clarification, we rephrased the sentence. See the reply to item 31 for the modified paragraph.

34. Line 353: where does this number of 12 application areas come from. There are 10 application areas under atmospheric domain as one can see in the OSCAR database
<https://space.oscar.wmo.int/applicationareas>

Reply: Thank you for pointing at this error. We updated the number.

35. Line 379: OSCAR is a database, not a table.

Reply: To stay consistent with the WMO nomenclature, we replaced “table” by “tool”.

36. Table 2a: is the word “distribution” missing in the “aerosol number and size”? In GAW most of the ozone depleting substances are included under greenhouse gases

Reply: Thank you, the term “distribution” was added.

37. Table 2b: GCOS does none use volume mixing ratio for ozone variables. The correct variable is mole fraction

Reply: Thank you, the unit was adjusted.

38. Line 405: please use either upper air or upper atmosphere, not both

Reply: This sentence was removed during the reorganization of the text.

39. Figure 5: typo at the end of the first line (“to;” is not needed)

Reply: Figure caption was corrected.

40. Line 419: I guess “hundreds of stations” applies to temperature observations as this is not the case for greenhouse gases

Reply: The word is just used as a synonym for “many”. We replaced the term by “numerous”.

41. Line 430: the term “validation” is not used in the Paris agreement

Reply: The term “validated” was replaced by “checked”.

42. Line 444: what do you mean with “organised the implementation of the FAIR principles”. The principles can be applied to something, rather implemented

Reply: For clarification we replaced “implementation of the FAIR principles” by “implementation of the respective FAIR enabling resources”.

43. Line 448: “FAIR enabling resources need to be implemented” – resources can be used, rather than implemented

Reply: In this context the phrase “implementation of FAIR enabling resources” is frequently used. So, we prefer to keep this sentence.

44. Line 459-468: is this paragraph needed? Line 468: “composed of” or use another word instead of “composed”

Reply: The content of this paragraph is needed because composable workflows are a key tool of this new scientific approach. The paragraph was shortened and significantly rephrased.

45. Line 487: how satellite and modelling data re integrated in this cloud?

Reply: The integration of data and services into the EOSC is driven by the providers of these data. Thus, we can speak only for our involved Research Infrastructures.

46. Line 496: could you please address the risks of the data misuse and misinterpretation by citizens without appropriate background? Training platforms cannot substitute years of professional education.

Reply: The issue of misuse or misinterpretation of data in citizen science is a huge one, and certainly beyond the scope of this article. As citizens are only one of many users of environmental data, we prefer to refrain from commenting on this topic. This would require a separate article.

47. Line 497: please spell out SMEs

Reply: “SMEs” is replaced by the general term “industry”.

48. Line 501: could you please use another word for “uses”

Reply: “uses” was replaced by “applications”.

49. Line 505: please spell out abbreviations here, it reads like a lot of slang

Reply: The terms GIT repository, EGI and ZENODO are not acronyms, but are used to refer to open source programming tools (GIT), a large European e-Infrastructure (EGI Foundation) and the Open Access repository operated by OpenAIRE.

50. Line 515: “metadata catalogue to access information from the RI of interest”

Reply: WE clarified the phrase as: “browses then the metadata catalogue for information about the access to the RI of interest”

51. Line 517: “composability” -please clarify

Reply: The composability of workflows is a standing term in the context of workflows. For clarity, we rephrased it as “for re-use and combination with other workflows”.

52. Line 533: “on the use of license on data”

Reply: The sentence was rephrased as : “Also, an atmospheric community standard has evolved on the license for data use”.

53. Line 578-579: “for machines, knowledge recorded in scientific articles is not accessible” – this is incorrect and the text below referring to the AI had been used for this purpose. Later in the text (line 599) the authors refer to the publication of the scientific papers, which according to the initial statement would be a waste of resources

Reply: The point we want to make here is that knowledge recorded in scientific articles on a narrative basis is indeed not accessible to machines without the help of humans who identify relevant knowledge in publications and transform it into a machine-readable format. This cannot be done by AI. However, the publication of scientific papers will always be one way of generating and disseminating scientific knowledge, but there will develop other ways, too. Therefore, we prefer to keep this paragraph as it is.