

# Technology to aid the analysis of large-volume multi-institute climate model output at a Central Analysis Facility (PRIMAVERA Data Management Tool V2.10)

August 31, 2023

We would like to thank the three reviewers for their thoughtful and constructive comments, which we believe have helped to improve the manuscript. We have implemented all of their suggestions and our responses to each comment are listed below.

## 1 RC1

- **The title** The manuscript's title has been changed slightly to emphasise that the paper talks about the technology rather than the analysis. The new title is "Technology to aid the analysis of large-volume multi-institute climate model output at a Central Analysis Facility (PRIMAVERA Data Management Tool V2.10)".
- **Page 2, lines 4-7** Hopefully, the flow has been improved.
- **Page 2, line 26** Additional information about the compute has been added.
- **Page 6, line 3** A clarification has been added to say that 440 TiB was the maximum available to the project.
- **Page 8, lines 3-4** The description of process has been elaborated on to hopefully allow future projects to work more efficiently. It has also been moved to the new Sect. 5.2 described below.

- **Page 9, first paragraph** The terminology was misleading. “unique” has been removed as the text already includes the phrase “... at least once...”, which means the same as unique.
- **Page 9, first paragraph** More information about purge policies, priorities and demand management have been added to a new paragraph in the new Sect. 5.2.
- **Page 9, second paragraph** Jupyter notebooks were used and these have been added to the list of tools that are mentioned.
- **Page 9, line 15** A response rate has been added.
- **Page 10, first paragraph** Use of JASMIN by our African partners is very much in its infancy and further work is required but a reference describing such work has been added.

## 2 RC2

- **Portability** A new Sect. 5.2 titled ”Design Considerations” has been added, which includes a discussion on how generalisable the current version of the DMT is to other institutions.
- **Design** The new Sect. 5.2 titled ”Design Considerations” discusses how the requirements for the DMT were developed and the design methodology used along with some pitfalls that should improved on in a future version. Several related paragraphs from elsewhere in the text have been moved to this new section.
- **Page 2, lines 5-7** The offending sentence has been moved to the second in the paragraph, which improves the flow.
- **Page 2, line 8** The DMP’s citation has been moved to here.
- **Page 2, line 10** The DMT’s citation has been moved to here.
- **Page 2, lines 9-10** A mention of the DMP data workflow steps implemented by the DMT has been added here, along with the steps assisted by it.
- **Page 2, lines 13-18** This paragraph has been moved to Sect. 10 Conclusions. The expansion of ESGF and its reference has been moved to the new first occurrence in Sect. 2.

- **Page 3, lines 23-24** “Sufficient resource” has been replaced by the number of person months dedicated to this work in the project proposal.
- **Page 5, lines 8-9** A sentence has been added to confirm that all data migrations are fully automated by the DMT to maintain the consistency of the catalogue.
- **Page 9, Sect. 7** This information was published in the “Lessons learnt” document and a reference has been added in Sect. 7.
- **Page 9, line 26** Changed to use acronym.
- **Page 11, lines 8-10** A reference to the AWS cost calculator and a date that these costs were obtained has been added. AWS S3 storage costs appear to be relatively static over time.
- **Page 11, lines 24-34** This paragraph was intended to be less about Pangeo and more about the concept of compute being closer to geographically distributed storage. The paragraph has been modified to reduce the emphasis on Pangeo (although it is still mentioned to introduce the concept) and to increase the emphasis on distributed storage and compute, orchestrated by a catalogue such as the DMT.
- **Pages 12-13** “...but continued expansions of...” has been changed to “...but continued expansion of the storage and compute of...” to elaborate our thinking.

### 3 RC3

- **Lessons learnt, problems encountered, and adjustments made** Reviewer RC2 made a similar comment and so a new Sect. 5.2 titled “Design Considerations” has been added, which includes a description of how the data management tool evolved as problems were encountered and lessons were learnt during the project.
- **Clarification question** Figure 1 has been amended to clarify that it is users internal to the project who have access to the data in stage 4 of the DMP. This has been clarified in the text too. The fourth paragraph of the Conclusions has been amended slightly to say that a wider range of users need access to CAFs.
- **Page 5, last line** Users were generally well behaved, with only occasional emails reminding them of good practice required when storage

was becoming limited. This has been clarified in Sect. 7 in response to another reviewer’s comment.

- **Page 8, lines 3-4** This explanation has been expanded.
- **Page 12, lines 17-18** Work has begun on a more generic version of the DMT at <https://github.com/MetOffice/primavera-dmt> and the community is invited to contribute there. A link to this new version’s repository has been added to Sect. 9 “Future Opportunities” and to the “Code availability” statement.