Comment on gmd-2022-73
David Huard (Referee)

Referee comment on "25 years IPCC Data Distribution Centre at DKRZ and the Reference Data Archive for CMIP data" by Martina Stockhause and Michael Lautenschlager, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2022-73-RC3, 2022

General comments

The paper is a useful contribution as it discusses the important role played by data curators in the climate science ecosystem. I would like to see the DDC history expanded a little, and more context given for some of the discussion points of the paper.

Specific comments

6 Review the abstract, I don't think it quite captures the essence of the paper, and the syntax is awkward at times.
13 Not clear how visibility and exposure pose "a challenge and opportunity to operationalize ...".
14 There is not much talk in the paper about this long term DDC strategy
15 It's not clear which part of the paper discusses the political and societal realm of climate change.
16 The history section starts with the current composition. Wouldn't it be clearer to describe the DDC history in chronological order?
20 Shouldn't CEDA be part of that list, as they're still actively engaged in the final data archival and curate final datasets from AR6.
34 Would be interesting to know what situation spurred the creation of the DDC.
42 Here and elsewhere, there is a lot left unsaid on the reasons for decisions taken. I understand getting into the historical details is not the objective, but it leaves the reader without a storyline.
52 Explain why long-term archives are useful. What actual problems do they solve? Is it to provide original model data to future researchers? If so, can you cite papers using those data?
58 Do these downloads eventually translate into publications? Is the DDC cited often as a data provider for CMIP data from previous cycles?
102 No "NO" branch on "Data QC L3 passed"? "Replicated" bubble should be tied with arrow to diagram, no? Why is data replication a necessary condition for DDC archival?
123 The IPCC is not a research activity, but an assessment of the science. So I'm not sure the sentence "The FAIR data principles describe requirements for datasets to become an integral part of the research environment." really captures the IPCC motivation for this. Also, in my mind the IPCC FAIR Guidelines is a document mostly meant for IPCC authors, so it's not clear what the TRUST principles have to do with this, as it pertains to the archive curator.
144 An important outcome of the FAIR guidelines is also to give due credit to IPCC "chapter scientists", who are often heavily involved in IPCC graphics preparation.
173 How many authors ended up making active use of these virtual workspaces?
198 I think its worth adding a sentence describing the actual CMIP6 licensing issue.
267 "The importance of data has reached the policymakers." The example here seems a bit anecdotal? Can we really say this then follow-up with the fact that DDCs face funding uncertainties?
271 I think it would be useful to mention how DDC activities were funded up to now, what challenges does DDC face now, and discuss the rationale for changing the funding model.

Technical corrections

10 IPCC Sixth Assessment Report
11 Suggest to refer to the "IPCC FAIR Guidelines", not just FAIR guidelines, to avoid confusion.
13 has become more exposed -> has gained exposure and visibility
20 "DKRZ is the only remaining founding partner, who has been operating the DDC for 25 years." I would keep that comment for later, after discussing the formation of the DDC. Maybe rephrase since there were only two founding members if I understood correctly.
21 Sentence on MoU is unconnected to rest of text
27 Review sentence "The core role..." Not clear what "support in data" means.
35 The events are not in chronological order. It makes the presentation confusing.
41 syntax (again).
54 "downstream users without specific knowledge of climate model applications." What kind of users would fit this description?
55 via ca?
65 proprietary
72 syntax
76 Not sure what the value of this table is.
87 distributed
90 Use white background
115 Use expression "COVID-19 pandemic"
116 lack -> delayed start
127 I think it would be clearer if the history of the IPCC FAIR Guidelines was concentrated in one paragraph.
133 Confusing that FAIR Guidelines were approved on the first TG-Data meeting, yet the official version was only adopted in 2022. I think TG-Data approved an early draft version in 2019, giving it the mandate to discuss it subsequently with IPCC authors at the various Lead Author Meetings. Version 1.0 got approved at the last TG-Data teleconference in 2022.
137 the analysis script applied to produce the figures -> the analysis scripts generating the figures
137 I feel the word "collected" undervalues the considerable work of IPCC authors to prepare clean, documented versions of figure generation scripts.
145 I don't think DFG is a familiar acronym outside of Germany. Clarify reference.
153 What is "use metadata"?
156 Orphan sentence.
157 "Use metadata"? Should there be a step about the creation of entries in the DDC catalog?
161 Sixths ... -> AR6
178 Not convinced the list of partners is really relevant here, it is not accompanied by text describing the actual collaborations.
236 ... evolving requirements ... 
236 ... have adapted to developments ...
237 AR6 stands for ... review syntax
240 On the one hand ... on the other hand is used to express contrasting points of view. This is not the case here.
242 und - -> and
262 Link with "Good Scientific Practice" unclear.