

CMIP6 Citation Service Survey Results

M. Stockhause (v0.1; 2021-09-28)

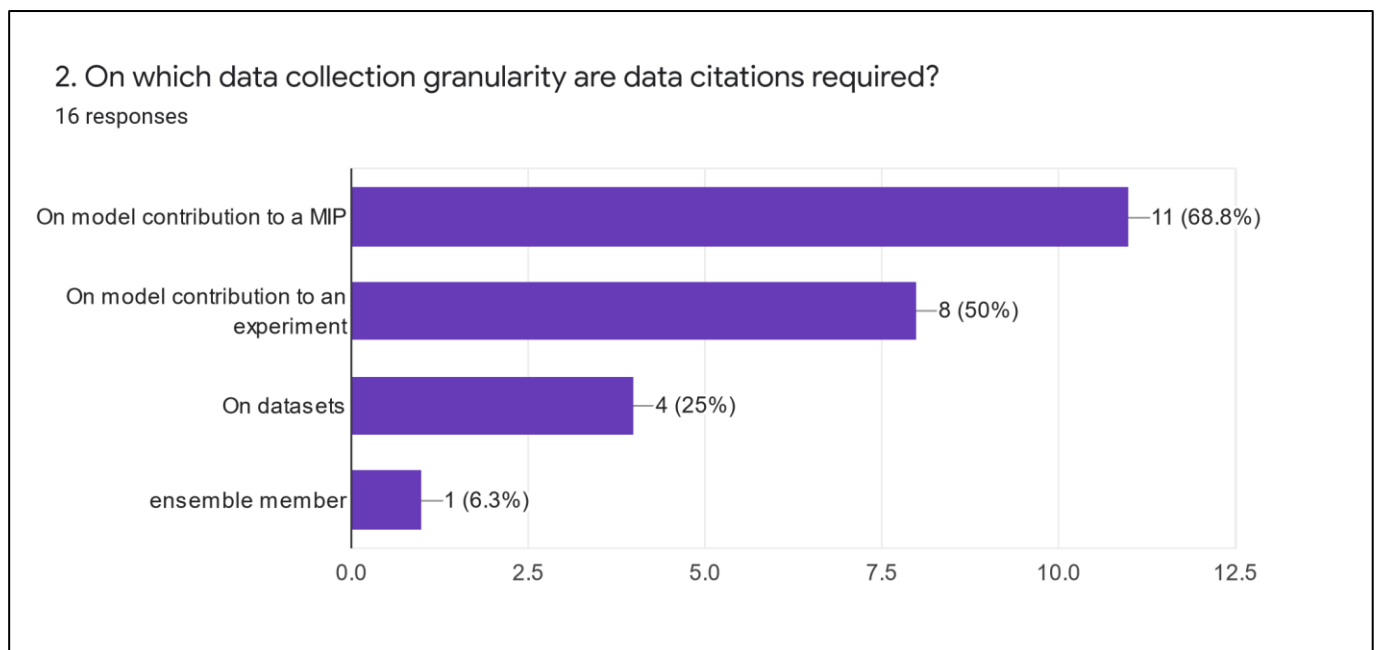
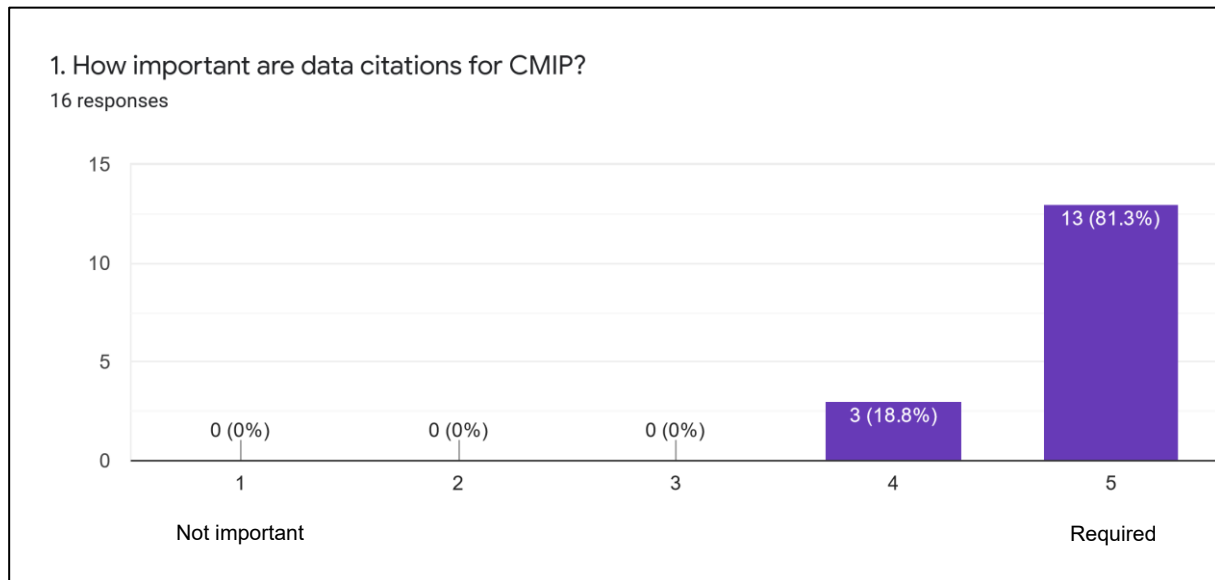
The CMIP6 Citation service (<http://cmip6cite.wdc-climate.de>) created a survey to collect feedback on the service, its tools and possible future directions. The CMIP6 citation managers were invited to participate between 25 May and 15 July 2021. 16 out of 78 citation managers responded. Taking into account that several invited persons are either replacements or represent modeling centers without CMIP6 contribution, the percentage of responses is in the usual range. As 16 is a small number, care has to be taken in the interpretation of the results and setting priorities for future developments.

Contents

CMIP6 Citation Survey Results	1
1. General Questions on importance and granularity of data citations	2
2. Feedback for the Citation Service and its tools	3
2.1 GUI	4
2.2 API	5
2.3 Citation Search	7
2.4 Documentation	7
2.5 User support	8
2.6 Overview question on level of satisfaction	9
3. Integration of data citations and data usage	10
4. Feedback on alternative options	12
5. General feedback on Citation Service	14
Appendix A – Survey questions	15
Appendix B – Individual answers	17

1. General Questions on importance and granularity of data citations

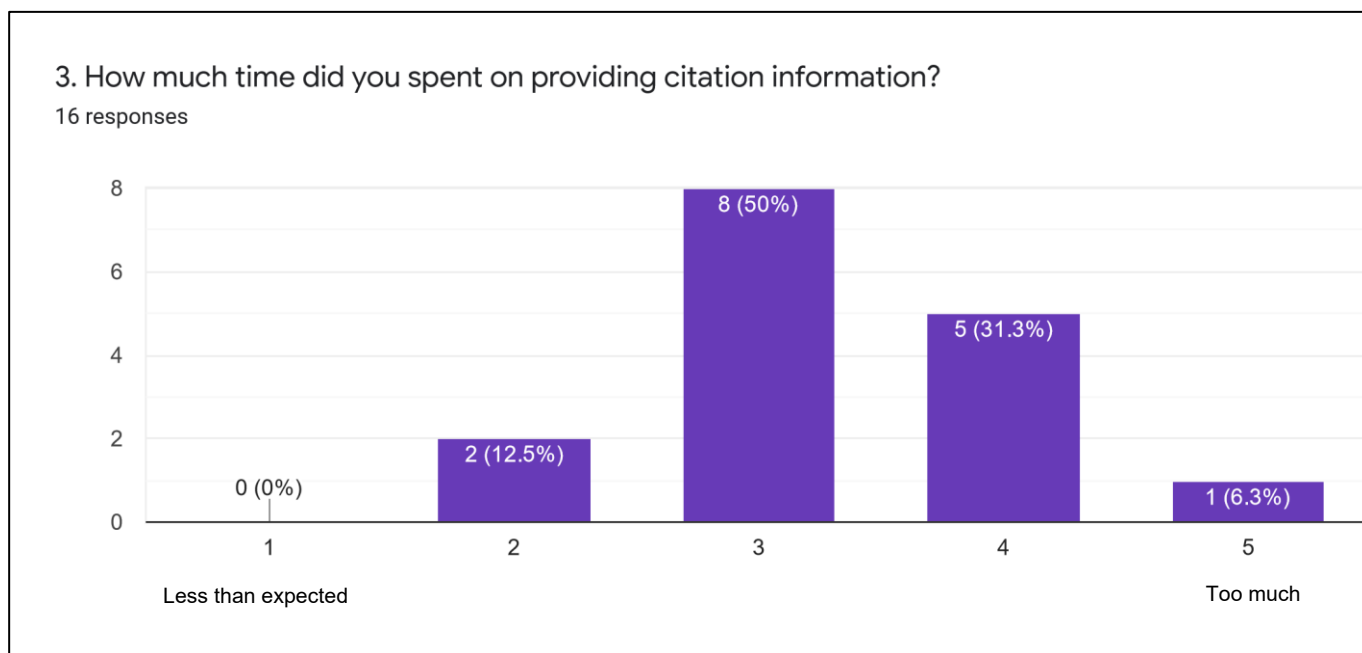
Data citations are regarded as important by all citation managers, for the majority (80 %) it is required for CMIP (question 1). 2/3 wish to have data citations on model contributions to a MIP and 50 % for each experiment (question 2). As these are the granularities of the current CMIP6 data citations, participants are mostly content with the provided data citation granularities. Four (25 %) want to have an opportunity to cite individual datasets and one to cite data of ensemble members.



2. Feedback for the Citation Service and its tools

The second set of questions (3-9) investigates the level of satisfaction with the Citation Service tools and applications as well as its documentation and support.

The time spent on providing data citations was about as expected for half of the participants (question 3). For the other half with a mismatch between expectation and actual time spent, a tendency to spending more time than expected is observed. Most time was spent on getting familiar with the system (question 3a in Appendix B – Individual answers). One participant regarded the GUI as easier to use and another participant judged the GUI as nearly unusable, explaining to be a Linux person.



The Citation Service provided two applications for citation managers to maintain their data citation information, a Graphical User Interface (GUI) and an Application Programming Interface (API) for script applications. The GUI was required to enter basic data such as author details (person), affiliations (institutions) and referenced article details, before the API could be used. The API was required for those participants, who wanted to control citation information on the experiment granularity. Lately, the GUI was changed by enabling the citation manager to control, which entries should be displayed in the GUI incl. experiment entries.

Questions 4-5 gathered feedback on GUI and API. The Citation Search (question 6) is a service more targeting at the data users than at the data providers.

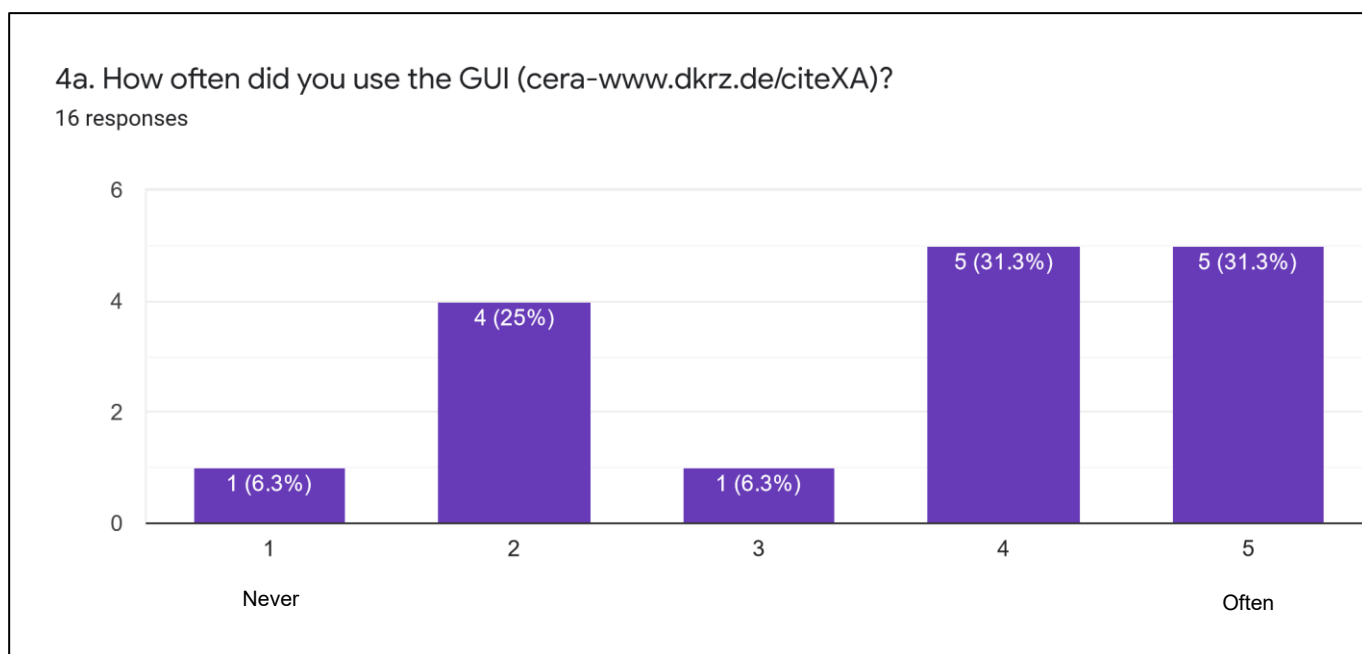
2.1 GUI

Except for one, all participants used the GUI (question 4a). As the use of the GUI is required to provide basic data, the one answer might be a citation manager who took over from a colleague who had entered all basic data or this person functions as replacement for the main citation manager. About 2/3 used the GUI fairly often or often.

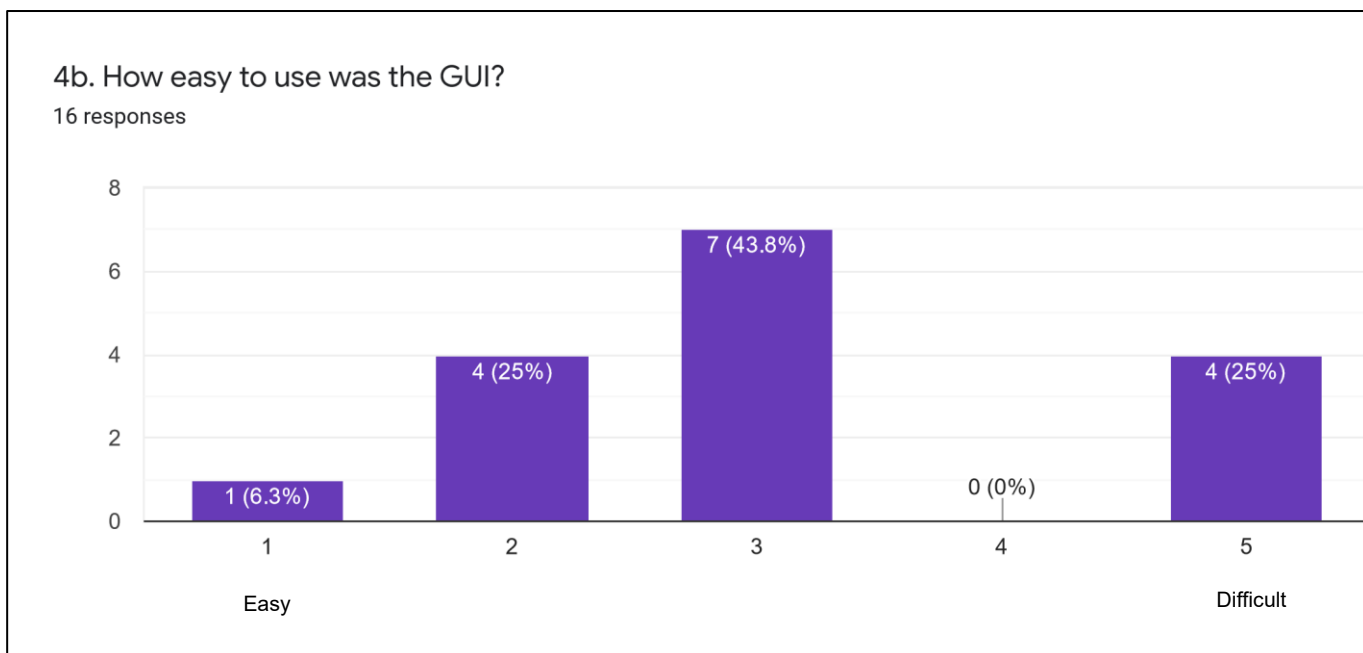
The question about the GUI's easy-to-use was answered by 3/4 with easy or neither easy nor difficult (question 4b). A quarter regarded the GUI as difficult to use, one of these four commented that he/she encountered no problems (questions 4c and 4d in Appendix B – Individual answers). Comments indicate difficulties with updating author lists, which was not allowed; instead, a new list based on the existing one had to be created. A second problem specified were missing citation entries due to an incomplete model registration. A third expressed that the linking was too complicated and that he/she missed edit/remove options without specifying that further.

Suggestions for improvements are:

- Allowing the update of author lists
- Using the GUI for all citation information¹ (including experiments)
- Giving-up the GUI in favor of the API and just provide a web-based control functionality
- Adding a filter option for the selection of persons of a specific institution



¹ The possibility to maintain all citation information in the GUI was added during the survey time, when the citation service was re-opened. An additional functionality allows the citation manager to control the visibility of individual citation entries.



2.2 API

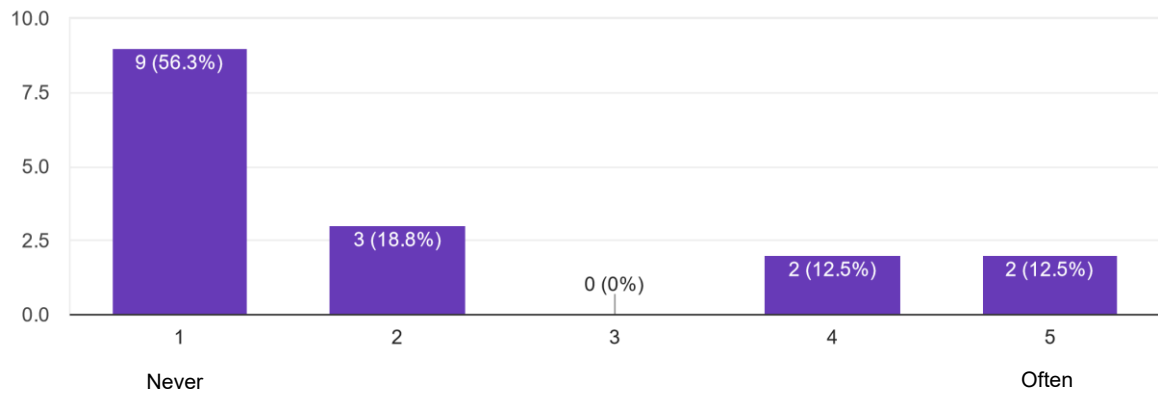
The majority of participants never used the GUI (56 %), another three used it rarely (question 5a). Only a quarter responded to have used the API fairly-often or often. Among the nine responses on the easy-to-use question, two never used the API (question 5b). Good half of the responses indicate that the API was neither easy nor difficult to use, for the other nearly half the API was fairly difficult or difficult to use. Most individual responses complained about server-side problems (questions 4c and 4d in Appendix B – Individual answers). Further raised problems were difficulties to understand the error messages and to figure out, which parts of the citation information can be altered by the API and which not.

Suggestions for improvements are:

- Enhance the functionality of the API to edit more or even all citation information
- Replace the JSON input file by command line options
- Explain the relationship between API and GUI better

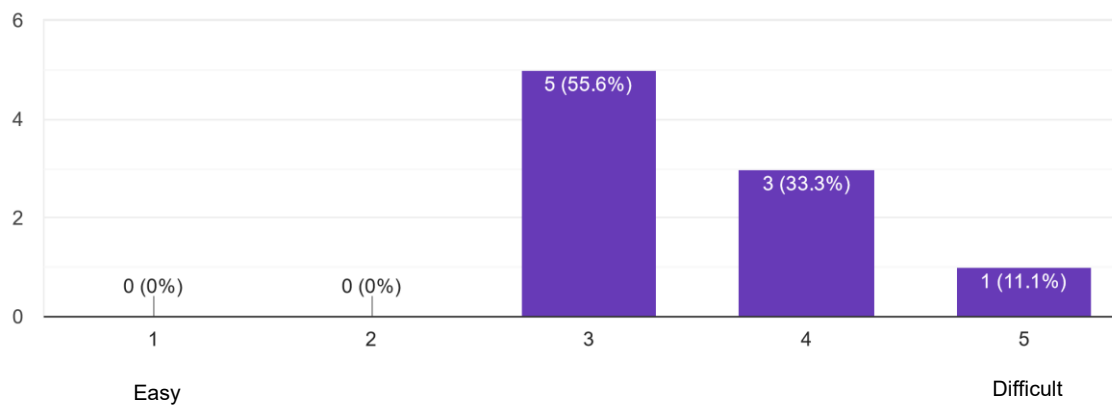
5a. How often did you use the API?

16 responses



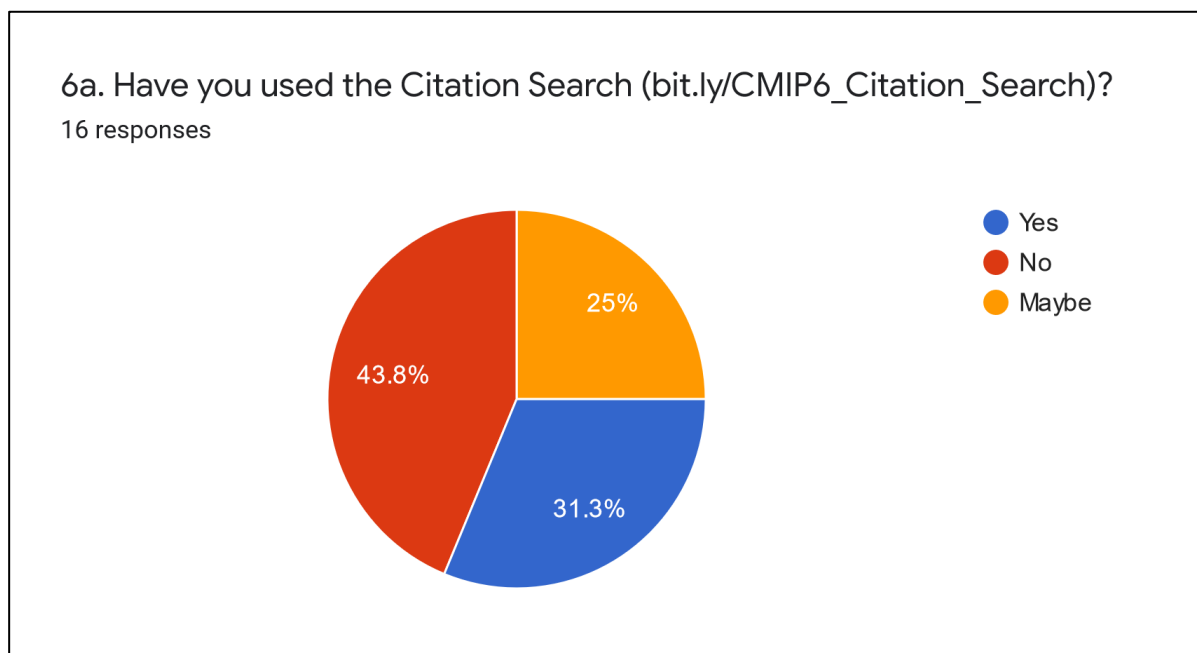
5b. How easy to use was the API?

9 responses



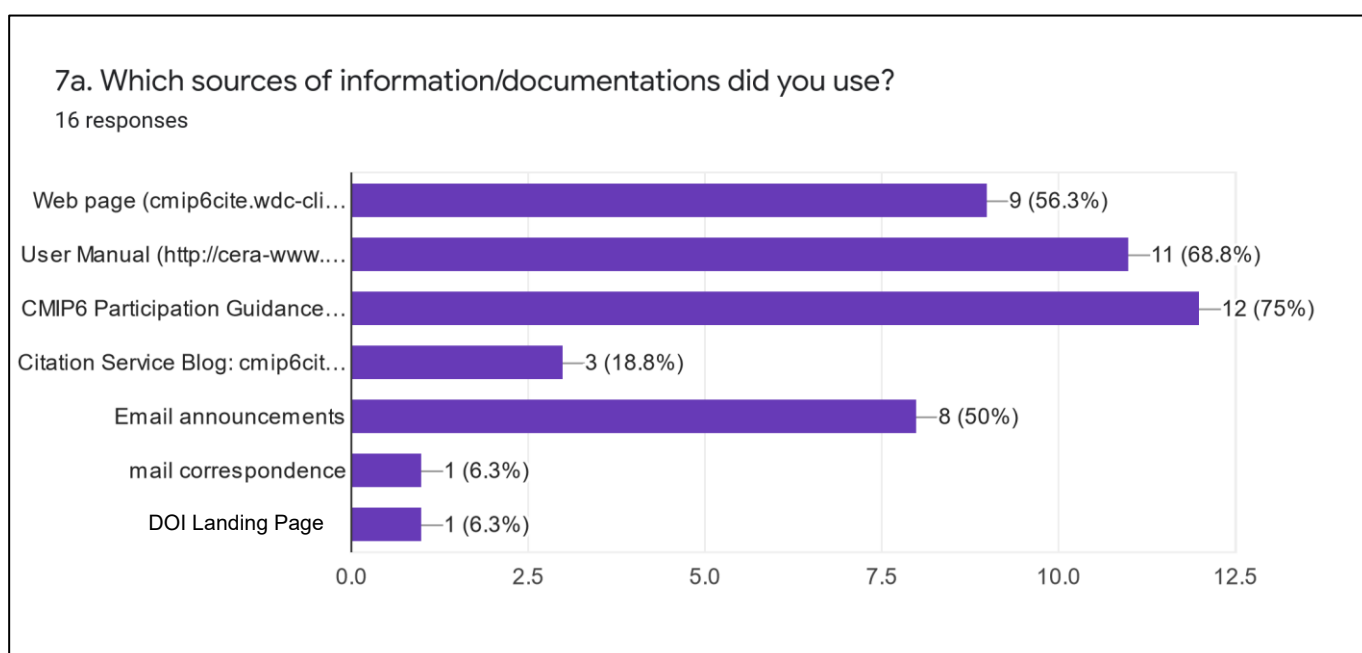
2.3 Citation Search

The Citation Search (question 6a) is a discovery service for data users. The citation managers as data providers do not need to use it or even know about it. Nevertheless, a quarter of the participants is familiar the Citation Search, 44 % do not know it and 31 % are unsure.



2.4 Documentation

The Citation Service and the CMIP6 project use a wide range of different platforms to disseminate information (question 7a). The major sources of information on the Citation Service were the general CMIP6 Participation Guidance on the PCMDI webpages and the dedicated user manual of the Citation Service used by 75 % and 69 % of the participants, respectively. More than half of the participants regarded the web pages of the Citation Service and its email announcements useful.

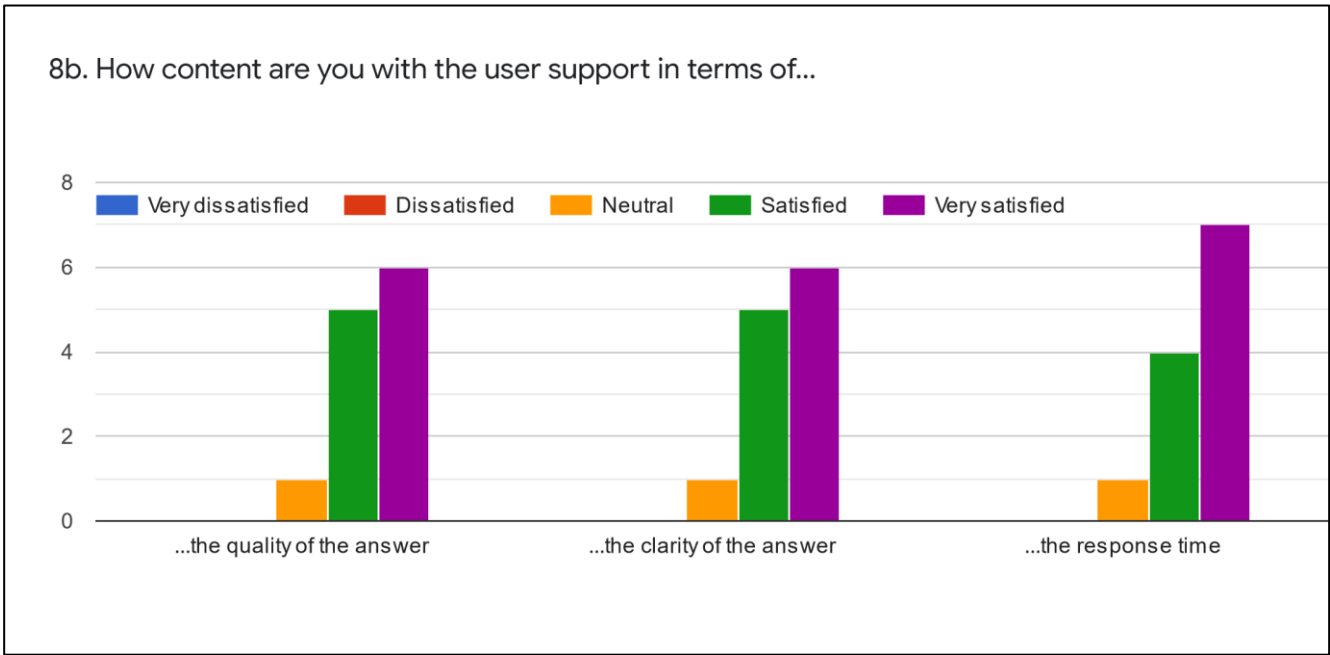
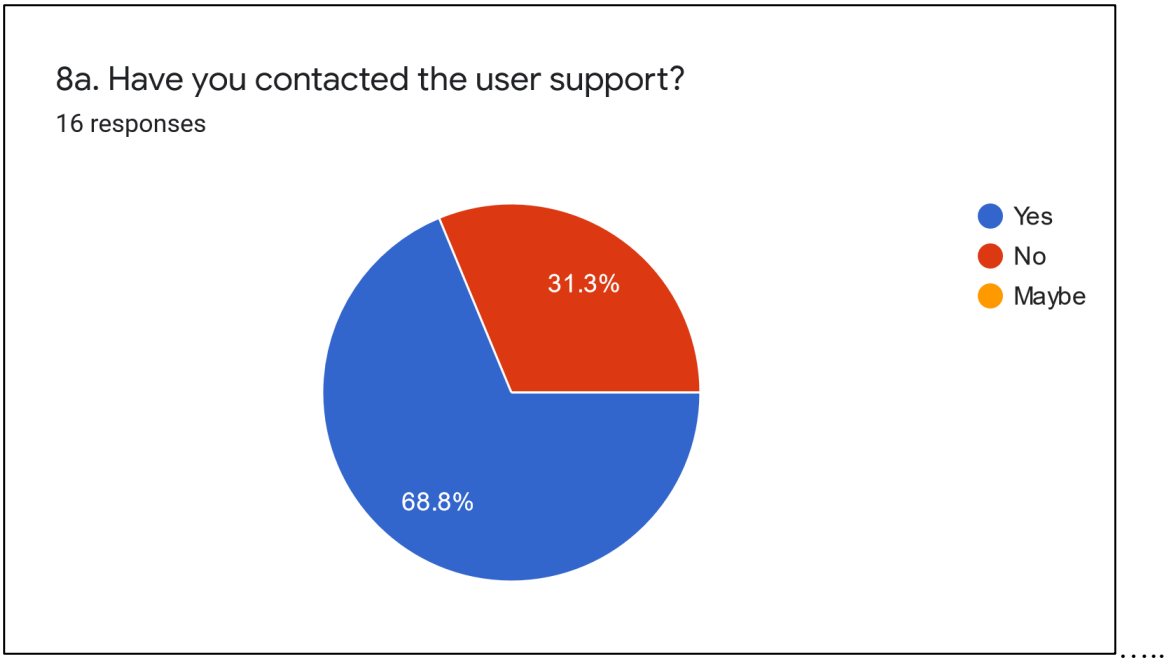


Individual comments on the documentation (questions 7b in Appendix B – Individual answers) pointed out

- that the different documentations are inconsistent and partly outdated and
- that the length of the how-to guide illustrates that the design is not straightforward.

2.5 User support

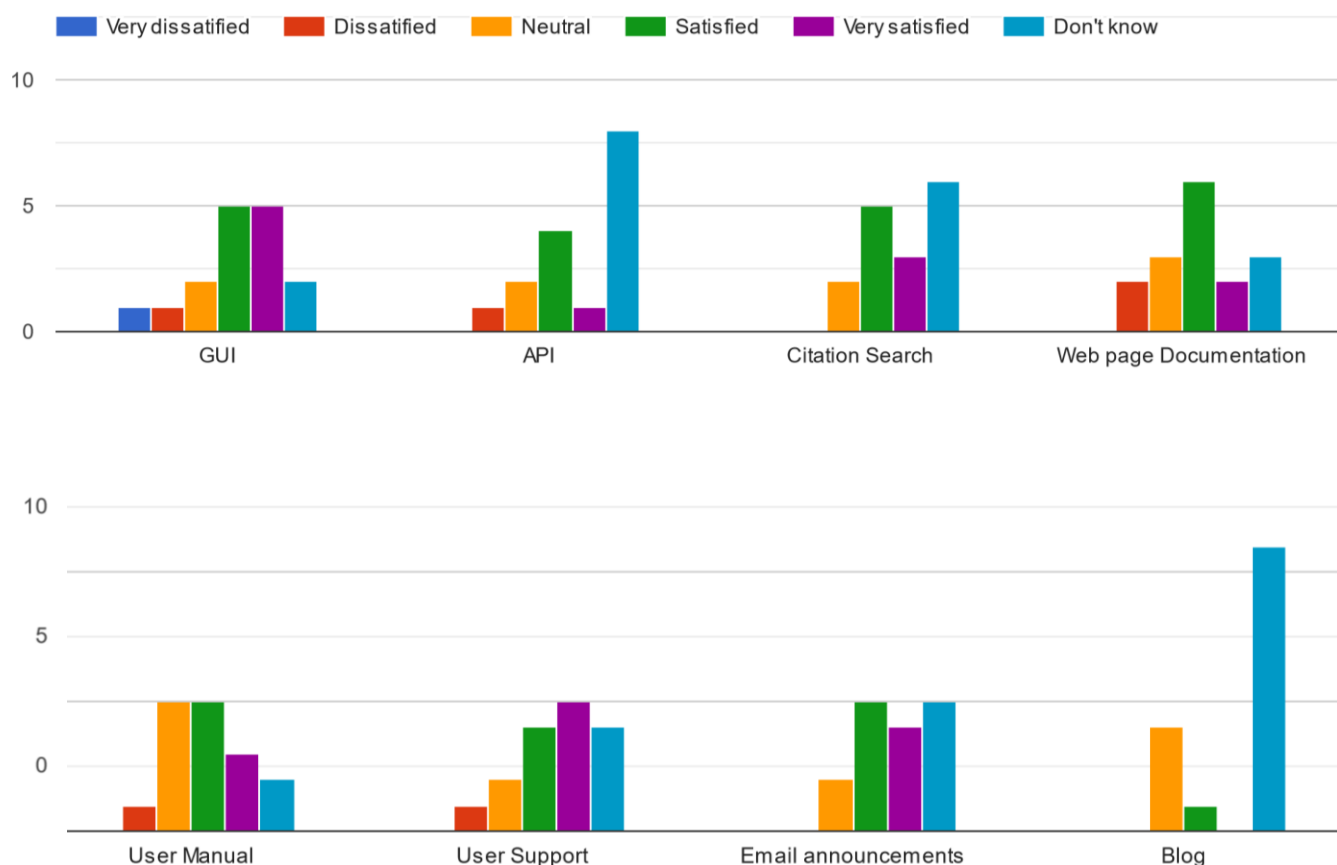
2/3 of the participants contacted the user support (question 8a). The majority was either satisfied or very satisfied with the support in terms of response time and the quality of the support (question 8b). One person was indefinite about the user support.



2.6 Overall rating for the different components/aspects of the Citation Service

The ratings for different components/aspects of the Citation Service show a diverse picture (question 9), which summarizes the responses to the previous questions. From the applications for the Citation Managers, the GUI is rated positive by nearly 2/3 of the participants with two negative votes (one dissatisfied and one very dissatisfied). Few users only used the API, therefore only a good half of the responses rated the service. Among these nine votes, five were positive, two neutral and one negative. The Citation Search and the Blog were directed to data users rather than data providers. Thus, they are not well known by the participants. The eight votes for the Citation Search were neutral (2) to positive (6). The documentation of the web page has a wide spread of rates with six satisfied responses, two dissatisfied and two very satisfied votes. The rating for the user manual is neutral to satisfied/very satisfied, with one dissatisfied vote. Nine participants are satisfied or very satisfied with the user support and the email announcements. One person is dissatisfied with the user support, which is not reflected in the answers for the previous question 8b.

9. How overall satisfied are you with the following core tools or services of the Citation Service?

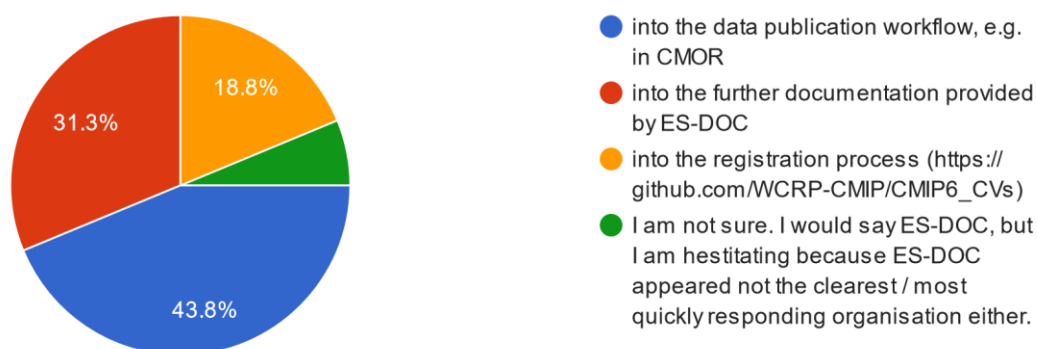


3. Integration of data citations and data usage

The next set of questions 10-12 gathered feedback for possible future directions and priorities. Among the given options for a better integration of the Citation Service into the CMIP6 data infrastructure (question 10), slightly more participants are in favor of an integration into metadata services ES-DOC or the CMIP registration (about 56 %) than of an integration into the data workflow (44 %), e.g. CMOR. This illustrates the dual function of data citations, identifying data collections and documenting data. From the nine votes in favor of a metadata integration 2/3 prefer ES-DOC and 1/3 the CMIP6 registration.

10. We investigate options to better integrate the citation service into the CMIP infrastructure. Where should the provision of citation information be integrated?

16 responses

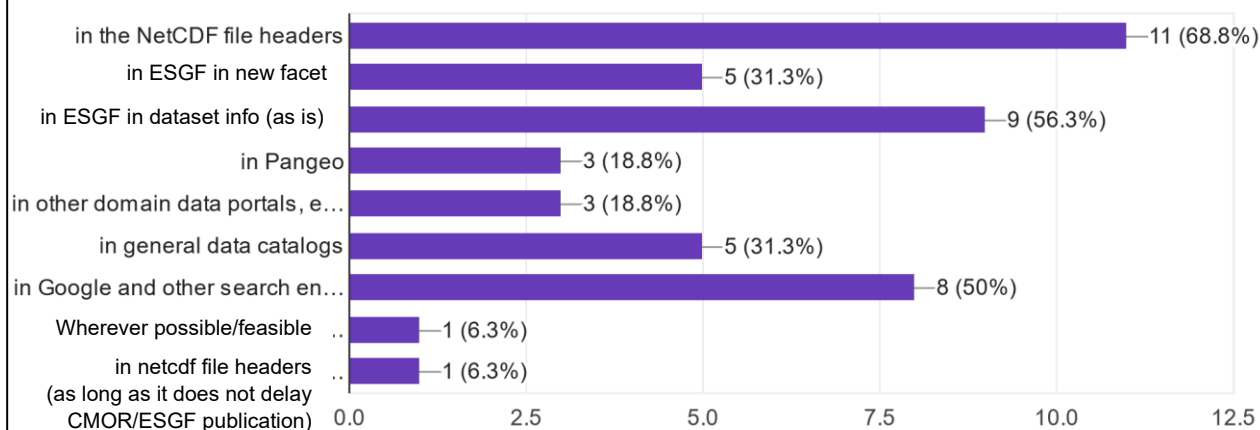


More than half of the participants would like to see data citations in the NetCDF file headers² (69 % - 75 %), as part of the dataset information in the ESGF portals (as is; 56 %) and discoverable in Google and other search engines (50 %; question 11). This illustrates that data citations should be available in both, project-specific and general services or as one participant puts it “wherever possible/feasible...” 5 participants each are in favor of a new data citation search facet in the ESGF and the integration in general data catalogs. Less important to the participants is the availability of data citations in Pangeo and further domain data catalogs (3 votes each).

² Currently, the furtherInfoUrl is part of the global attributes in each NetCDF file header. This points to a web page hosted by ES-DOC, which includes data citation information.

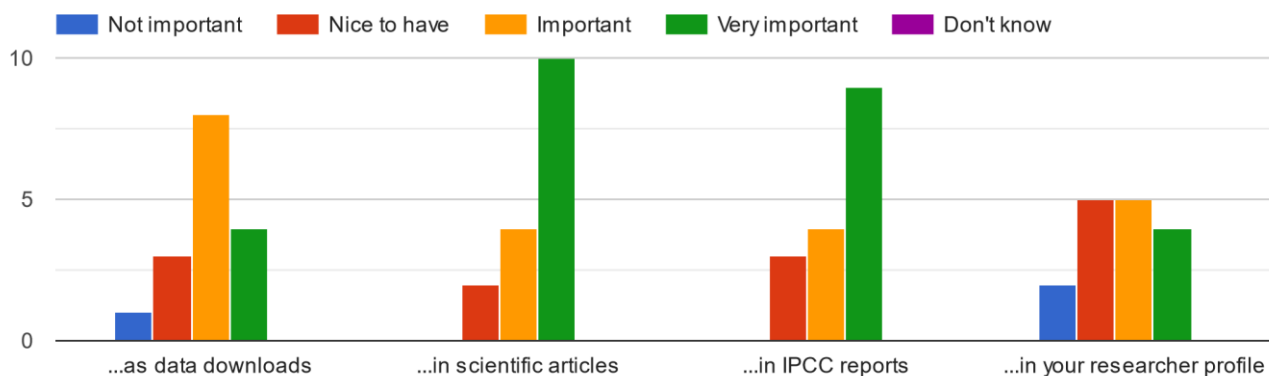
11. Where need data users to find data citations?

16 responses



Question 12 on the importance of data usage information shows that data references in scientific articles and in the IPCC report are very important to the majority of participants (10 and 9). Data download information and inclusion in the researcher profile is less important for the participants or even not important.

12. How important is data usage information...

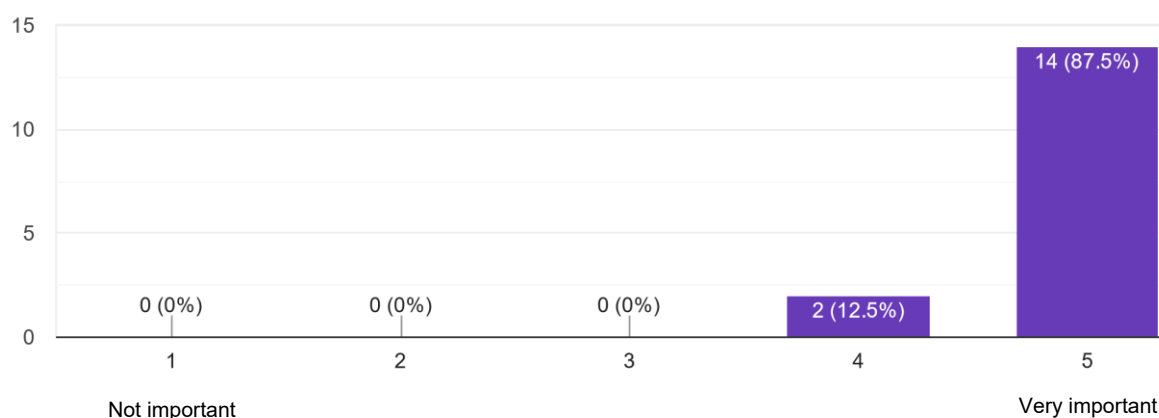


4. Feedback on alternative options

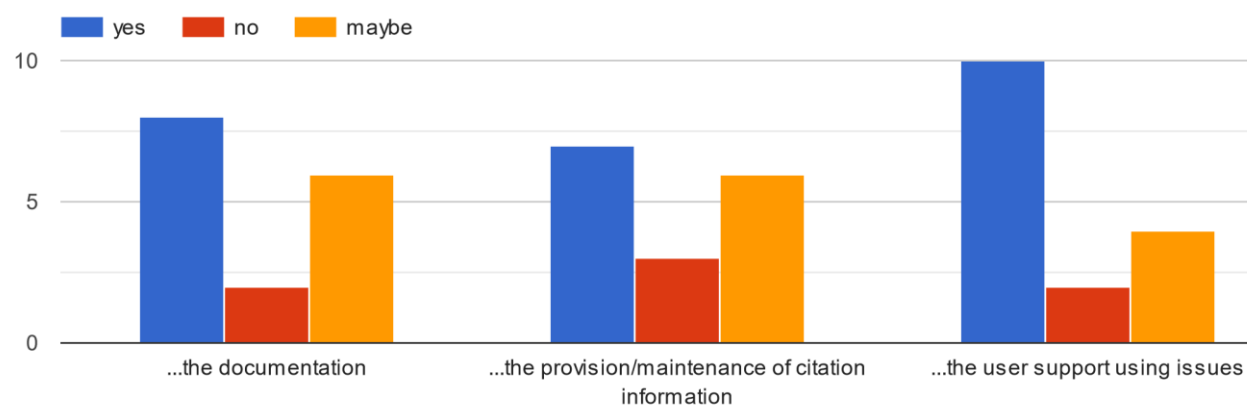
The last set of questions (13-15) gathers feedback on alternative options for data citations. Responses to question 13 clearly show that the participants wish to have DOIs as persistent identifiers. There are no indifferent or opposing votes. The majority agrees that GitHub is an option (question 14), esp. for issue tracking (10). The responses for the use of GitHub for maintaining the citation information are less clear with seven yes, six maybe and three no votes. Many participants seem to be indifferent about the platform for maintaining citation information but to be open for a change.

13. The CMIP6 Citation Service uses DOIs as persistent identifiers. How important is the DOI registration?

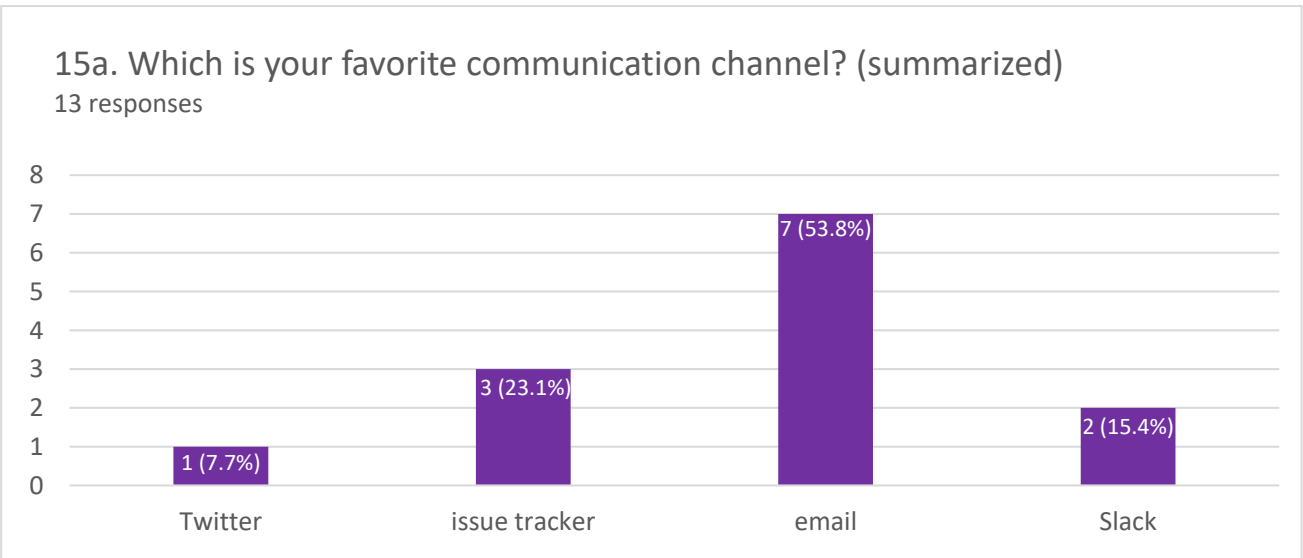
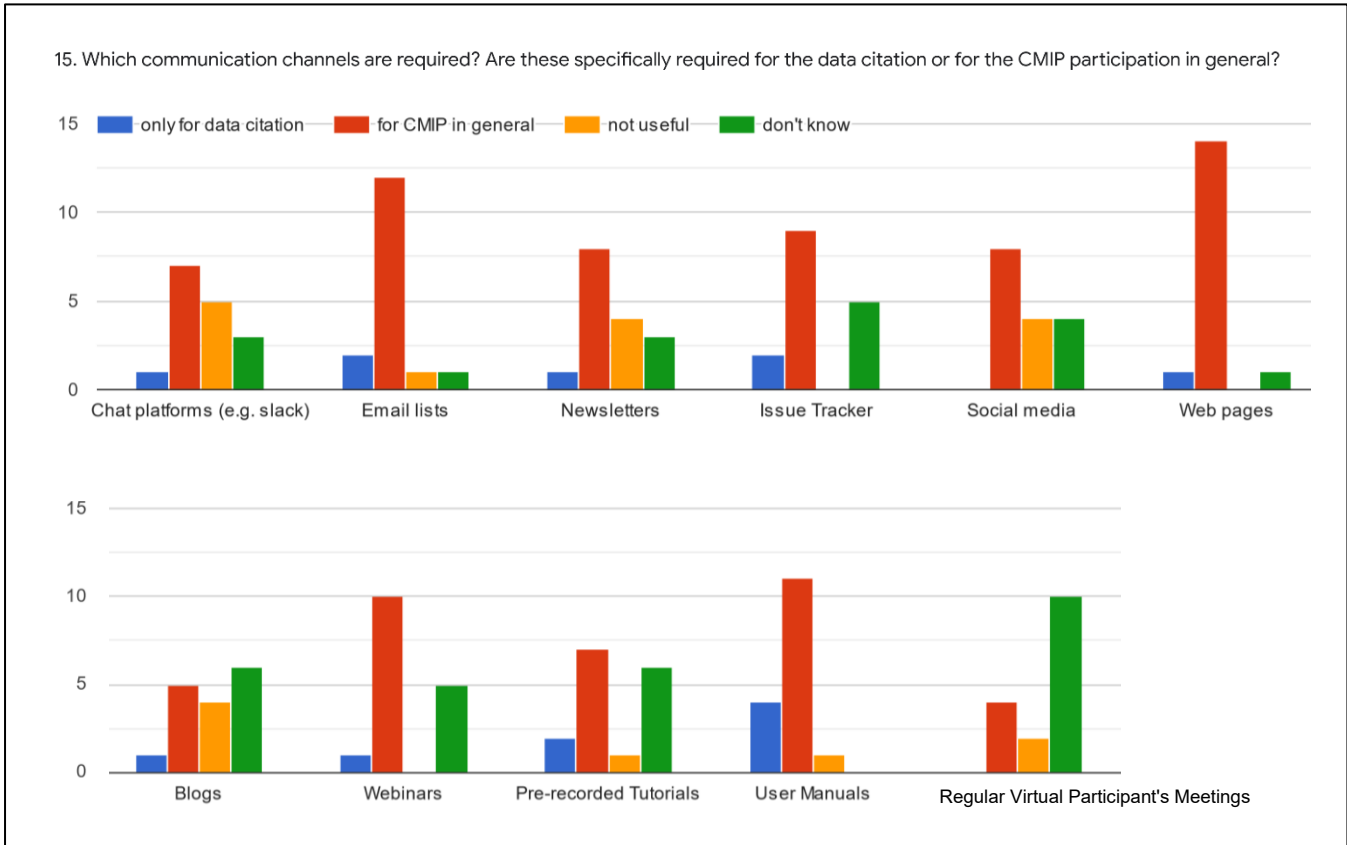
16 responses



14. Could github be an alternative for...



The feedback to preferred communication channels for the Citation Service and CMIP in general (question 15 and 15a) shows that a majority favor traditional communication channels for CMIP like email lists (12 of 16), web pages (14), and user manuals (11), which the project already provides. The open responses about the favorite communication channel stress the importance of email communication. As further communication channels, more than half of the participants would like to have newsletters (8), an issue tracker (9), social media (8), and webinars (10). Opinions are split (7 pro, 5 contra) the usefulness of chat platforms like slack.



5. General feedback on Citation Service

Finally, general free text likes and dislikes are gathered (see responses to questions 16 and 17 in Appendix B – Individual answers). Among the likes, the value of having data citations was expressed, which enables to give credit to data contributors. One response emphasized that the change of author lists within a limited timeframe was important.

Among the dislikes, some recommendations are provided:

- Provide shared accounts to enable model runners to enter their information themselves;
- Extend the service to other projects hosted by ESGF; and
- Improve workflows.

The need to decrease the required time to get familiar with the system was expressed by several participants. One participant expressed his/her frustration with the GUI again.

6. Summary

The survey shows the importance of data citations for CMIP data using DOIs and a general overall satisfaction with the provided Citation Service for CMIP6. Data citations should be disseminated as widely as possible. There is a wish for adding citation information in the NetCDF file headers.

For the maintenance of the citation content depending on the local workflow both applications are required, GUI and API. Improvements of the documentation in quality and harmonization and the wish for a less complicated / easier-to-use application is expressed.

The opinions on future directions are diverse: About half of the participants favor a better integration of the citation service into other documenting services ES-DOC and the participation registration (GitHub) and half a better integration into the data preparation workflow. Webinars and transparent issue tracking are regarded as useful additions to the important traditional communication channels email, user guides and webpage documentations.

Appendix A – Survey questions

1. How important are data citations for CMIP?
2. On which data collection granularity are data citations required?
3. How much time did you spent on providing citation information?
 - 3a. Can you explain your answer of question 3.?
- 4a. How often did you use the GUI (cera-www.dkrz.de/citeXA)?
- 4b. How easy to use was the GUI?
- 4c. Did you encounter problems or errors using the GUI? Please describe.
- 4d. Do you have recommendations for improvements of the GUI?
- 5a. How often did you use the API?
- 5b. How easy to use was the API?
- 5c. Did you encounter problems or errors using the API? Please describe.
- 5d. Do you have recommendations for improvements of the API?
- 6a. Have you used the Citation Search (bit.ly/CMIP6_Citation_Search)?
- 6b. If yes, please share your user experience with the Citation Search?
- 7a. Which sources of information/documentations did you use?
- 7b. How can we improve the documentation? What is currently missing?
- 8a. Have you contacted the user support?
- 8b. How content are you with the user support in terms of...
 - ...the quality of the answer
 - ...the clarity of the answer
 - ...the response time
9. How overall satisfied are you with the following core tools or services of the Citation Service?
 - GUI
 - API
 - Citation Search
 - Web page Documentation
 - User Manual
 - User Support
 - Email announcements
 - Blog

10. We investigate options to better integrate the citation service into the CMIP infrastructure. Where should the provision of citation information be integrated?

11. Where need data users to find data citations?

12. How important is data usage information...

...as data downloads

...in scientific articles

...in IPCC reports

...in your researcher profile

13. The CMIP6 Citation Service uses DOIs as persistent identifiers. How important is the DOI registration?

14. Could github be an alternative for...

...the documentation

...the provision/maintenance of citation information

...the user support using issues

15. Which communication channels are required? Are these specifically required for the data citation or for the CMIP participation in general?

Chat platforms (e.g. slack)

Email lists

Newsletters

Issue Tracker

Social media

Web pages

Blogs

Webinars

Pre-recorded Tutorials

User Manuals

Regular Virtual Participant's Meetings

15a. Which is your favorite communication channel?

16a. What did you like about the Citation Service?

16b. What did you dislike about the Citation Service?

17. Do you want to share any additional feedback?

Appendix B – Individual answers

3a. Can you explain your answer of question 3. (time spent on citations)?

1. We ran experiments on behalf of another institute's submission so it was fairly easy because they had to do the bulk of the work.
2. API takes me slightly longer time because I was not familiar with it. The GUI was easier for me.
3. I spent more time to create members in every project.
4. It took time to get familiar with the system
5. Sorry to say, but the GUI has been the worst/unhandiest GUI I ever have worked with (I am a linux based worker). The API itself was acceptable (far from perfect) but often the server was down and thus the API could not be used. Which makes the installation hard, because one does not know what's the matter.
6. interface was fiddly, and not intuitive (to us) but once we got the hang of it, it wasn't too hard to do the minimum
7. Felt like service was fairly smooth, took a little time (especially at start) but on balance was happy to put that time in to be able to provide citation information for our data providers.

4c/4d. on GUI problems and recommendations for improvements

4c. Did you encounter problems or errors using the GUI? Please describe.

4d. Do you have recommendations for improvements of the GUI?

1		Author lists could not be updated, only new ones created; an update option would be nice; the possibility to add/edit all citation information with the GUI would be useful
2	No	
3	I just wanted to fix author and/or contributor list, I had to create new one instead.	
4	Nope.	
5	No, It is good.	Maybe UI will improvement.
6		Experiment granularity rather than only MIP granularity
7	Yes. Too many. One very very annoying thing is that one cannot remove/edit wrong items. The linking is far too complicated for what it is. I would recommend a total revision of the structure and ease it from the perspective of the user, for who this is a infrequent additional task. The user should be able to do this task in limited time, especially the reading and understanding of what has to be done must be short. I recommend a linux command line based API (beside a GUI if you like) which can do all without the GUI. In my case I must be able to script the entire stuff	We'll replace it by the API and only have a GUI or web space in which the final result can be watched.

	because of the many different MIP experiments.	
8	Most common problem I encountered was usually when model/exp had not yet been entered by modelling group (not a fault of the service itself).	Minor but would be helpful to be able to filter author lists by organisation before adding to a list you create. Found searching for a long list of authors in the build step a bit tedious.

5c/5d. on API problems and recommendations for improvements

<i>5c. Did you encounter problems or errors using the API? Please describe.</i>	<i>5d. Do you have recommendations for improvements of the API?</i>
yes, there was sometimes a server-side problem and no upload of citation information was possible, but this was fixed quickly intermittent connection issues	
At first, some parts cannot be overwritten by API. The problem was that I didn't know which part can be overwritten and which part cannot.	Should better API can set as much part as possible, including author/contributor list from scratch (is it already possible?)
No	A more detailed introduction to explain the relationship between API and GUI? Maybe Just for me :)
No	No
The server was mostly not working, almost every time I had to contact you to inform there were problems.	Extend it and make all tasks possible from the API and enable that all these tasks can be managed from the command line, such that all can be scripted.
N/A	N/A
Had a few errors I couldn't understand but I think this could have been related to service going down in January.	For any new versions, could be helpful to build from command line instead of having to modify json text. I found that a bit tedious, would have been easier and less prone to errors to do this more systematically perhaps. Especially when doing many together, which was the case in our earlier publishing days.

6b. If yes, please share your user experience with the Citation Search?

1. I found it quite straightforward to use.
2. Nice to have. Not sure many know about it though.
3. It is easy to use.
4. I believe I only used this when I was unsure if we had previously added citation information.

7b. How can we improve the documentation? What is currently missing?

1. Please carefully read and try to follow your instructions, there are quite some clear errors in it. A more centralised point, it was in the beginning unclear what was outdated what was the official starting point / the maintained updated docu
2. too much emphasis on beautiful data models and not enough on UX. How-to guides are essential, but if a 'how-to' explanation of something apparently straightforward has 5 steps, it's not well designed. For the most common use cases, the how-to should be super-straightforward.

16a/16b. on general positive/negative feedback

16a. What did you like about the Citation Service?	16b. What did you dislike about the Citation Service?
ability to support large data volumes and the ability to change author list until a particular timeframe	not sure. may be the accountability aspect wherein only the citation officer is entering all the info and not able to request the model runners to do it themselves. Support for any project hosted in ESGF and a process to figure out funding based on project would have been great.
1) provide a new opportunity to quantify modelers' contribution except for the scientific publication 2) easy use	nope.
Good leadership and follow up from Martina	
The help offered to get the citation right	Time needed to get familiar with the system
	The GUI was a horror to work with. Even when one in the end perfectly understand how it worked it was a horror to work with it. I am again sorry to say it, but honestly it has been the most frustrating job I ever have done in my professional work.
that it exists	how hard it is to find just what you need and building the database
Database approach for data entry; was useful to avoid duplication and variations of the same entries; also having data citation available as a service	Very minor - the workflows being a bit clunky and/or tedious but was not a deal breaker

17. Do you want to share any additional feedback?

1. Furthering use of data citations is very important for recognition of people working on the models, who might publish fewer papers than those working on analysis of the resulting data. It is a very valuable service. Thank you for your effort so far, and look forward to a new and improve version.
2. Really appreciated having this available. Great service and also hope it's recognised as something important to support going forward. Thanks to all behind the scenes supporting and providing this service. Much appreciated!