



DataONE Search and Metadata User-Experience (UX) Study Results

August 2018



Prepared for:
DataONE Leadership Team

Prepared by:
Rachel Volentine
User-eXperience Lab Manager
Usability and Assessment Working Group

User-Experience Laboratory
Center for Information and Communication Studies
University of Tennessee
uxl@utk.edu
865.974.9865

TABLE OF CONTENTS

Overview & Top Findings	3
Project Methodology	4
User Profile	5
Usability Findings & Recommendations	6
Issue Severity Key	6
Task and Questionnaire Results	7
UXL Manager Critique of Search	17
UX Issues: Search Interface	22
UX Issues: Metadata Display	27
What's Next?	30
Appendix A: Usability Test Scenario	31

OVERVIEW & TOP FINDINGS

The goal of this research effort was to assess the usability of the DataONE Search and its metadata information. The DataONE Usability and Assessment Working Group conducted a remote usability test that aligned with the 2018 DataONE Users Group meeting in Tucson, Arizona. This study is intended for product development.

TOP FINDINGS

Some of the search facets do not function correctly.

Some search facets, for example location, do not function as the user expects. Locations, such as Yellowstone National Park and Puget Sound, are not location facets for some data sets and do not pull up all the relevant data sets. When the filters do not pull up all the data sets that clearly meet the criteria it creates confusion and mistrust in the accuracy and relevancy of the search results.

The auto suggest with drop down list does not always provide useful and accurate information.

While users like auto suggestions to help improve the accuracy of their search, the DataONE Search suggestions are not always accurate or useful. During the UX study users received suggestions that did not provide any results, and also received suggestions that did not seem related to their search term. Additionally the number next to the auto suggestion did not relate to the number of search terms the user received.

Most users can create a multiple facet query to find relevant data.

The UX test showed that most users interact with multiple facets of the search to find data.

Users do not scan the entire metadata record.

Only a few users scanned through the entire metadata record to complete a task. For many of the tasks, participants did not go past the files in the dataset. Additionally many participants said they need to download the metadata record to find metadata information. This implies that users may not be aware of all the information contained on the page.

Users confuse the metadata and data downloads.

One of the tasks asked participants to download the data for a record, and only two users downloaded the data file. Four participants downloaded the metadata record thinking it was the data. Providing more distinction between the metadata and data downloads will improve the usability of the metadata page.

PROJECT METHODOLOGY

Ten users completed the remote UX test. Recordings were collected July 16-August 5 2018. During the test users were asked to perform four tasks in the DataONE Search. Each task focused on a different Member Node that represents different metadata displays in the Search. The UX test scenario can be found in Appendix A.

Users' errors, comments, time on task, and screen recordings were recorded through Loop11, a remote usability software.

TESTING ENVIRONMENT

Users completed the UX test on their own computer and in the browser of their choice. Their preferences are listed in the User Profile section of this report.

USER PROFILE

Users were recruited by e-mail. Eight users attended the 2018 DataONE Users Group Meeting, and two users were contacted by a Brazilian associate of DataONE. Participant 4 did not get screen recorded due to poor internet connection. Participant 9 used a dual-screen monitor and selected the wrong screen to record, so the DataONE search does not appear in the recording. Participant 4 and participant 9 data are excluded from task analysis but their questionnaire responses are recorded.

USER #	OS	BROWSER	DISCIPLINE	HOW OFTEN SEARCH FOR DATA
1	Mac OSX 10	Chrome	Information Science	Occasionally
2	Mac OSX 10	Chrome	Computer Science	Never
3	Mac OSX 10	Chrome	Electrical Engineering and Computer Science	Occasionally
4	Mac OSX 10	Chrome	Library and Information Science	Never
5	Windows 10	Chrome	Ecological Modeling	Weekly
6	Linux	Firefox	Economics	Occasionally
7	Windows 10	Chrome	Life Sciences Librarian	Occasionally
8	Mac OSX 10	Chrome	Information Science	Occasionally
9	Mac OSX 10	Chrome	Data Management	Daily
10	Linux	Chrome	Calculus, Astronomy	Daily

USABILITY FINDINGS & RECOMMENDATIONS

ISSUE SERVERITY RATING KEY

The symbols below are used to rate each usability issue or finding in terms of its importance or impact to the usability of the Search. Ratings are based on the number of users who had the problem and also how seriously the problem affected the user or kept the user from completing the task. Usability issues are also determined by how they measure to the usability guidelines recommended by usability.gov¹ and Nielsen Norman Group's 10 usability heuristics for user interface design.²

Each UX issue documented in this section was assigned a priority level based on the following criteria:

Low: User may experience insignificant time delays or mild frustration, but will be able to complete the task.

Medium: User may experience noticeable delay or frustration, but will be able to complete the task with added effort.

High: User will experience noticeable delay or frustration, may not be able to complete the task.

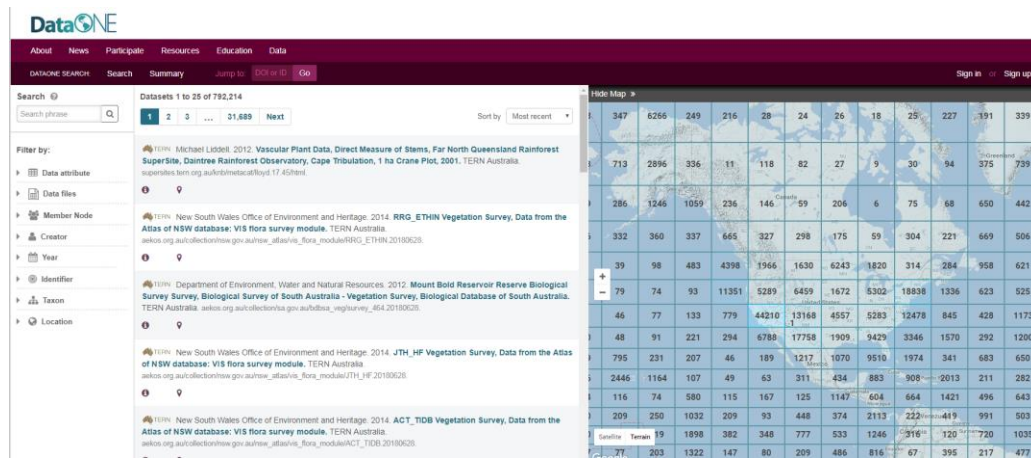
Success: Users were able to easily perform their desired task

Data: Describes summarized data of interest

¹ <https://guidelines.usability.gov/>

² <https://www.nngroup.com/articles/ten-usability-heuristics/>

TASK AND QUESTIONNAIRE RESULTS



1. Task One: Choose your own adventure

Data

The first task asks, “Choose your own adventure. Think about a topic you are interested in and find a quality data set that is relevant to your topic. Once you have found a quality data set you may end the task.” Five of the seven participants successfully found a data set that was relevant to their search query.

The purpose of the task was to familiarize the users with the Search, and allow them to browse data that is familiar. During previous UX studies, users would sometimes express concern or confusion because they are searching for data outside their area. This task allows users to evaluate the Search from an area they are familiar.

Participants chose a variety of topics including:

Participant	Search Topic	Did user find data set?
1	Flora in New Mexico	Yes
2	California and CO2	No
3	Fungus and Taiwan	Yes
4	<i>Video Error³</i>	
5	Sao Paulo	Yes
6	populations and economy	Yes
7	<i>Poor video quality</i>	
8	water depth and LTER	Yes
9	<i>Video Error⁴</i>	
10	Chromosphere	No

³ Participant 4 did not get screen recorded due to poor internet connection. His tasks were not collected, but his questionnaire responses are recorded.

⁴ Participant 9 recorded the wrong computer screen so the DataONE Search is not shown in the video.

Only two of the seven participants did not complete the task. Participant 2 did not complete the task because he had difficulty adding filters. He limited his search by LOCATION: CALIFORNIA and then tried several times to search for carbon dioxide or fossil fuels, but did not like the results because he kept trying new filters and phrases. He ended the task before pulling up any metadata information. Participant 10 searched for chromosphere and received zero results. The participant did not try additional search terms.

Five of the seven participants started the task by entering text into the search bar. The other two participants started by limiting the search by geographic region using the map-based query and then used the search box or year filter to refine their search. Only one participant used one facet (search box) to complete the task. Most participants engaged with multiple facets to search.

Three of the participants downloaded the metadata during the task. Two of the participants struggled to find a program to open the data.

Participants engaged with the autofill suggestions as they typed into the search box which improved their results.

None of the participants scrolled passed the first page of results.

2. Which metadata fields are important

Data

After completing the first task, we asked which metadata fields are most important to help determine the quality of data. Participants had a variety of responses (see below).

<i>Participant</i>	<i>Response: Important Metadata Fields</i>
1	abstract, geographic region, years
2	No response
3	Organization, creators, abstract
4	I don't know
5	I'm not sure if I found the data I was looking for. I found a metadata schema, probably an XML file, but no data. I am going to try to download another file that was available.
6	OBSERVATION, OBSERVATION_EST N_MISSING to determine what the population in each area is and how many observations have missing data
7	I opened the Excel file, but I don't understand what I'm looking at. The field codes don't seem to be defined. I'm not very familiar with looking at raw data, so this may be an issue of my own making.
8	Data Files, Location, Year, Member Node

- 9 | The dataset I selected looked relevant and credible from the discovery-level metadata; however, I was disappointed to not find any details about the data structure/attributes, or methodology and processing details, even in the attached metadata document. I would not use these data unless I had those details and found them satisfactory. Should these not be required of datasets contributed to DataONE? Otherwise, how can potential users gauge fitness for use?
- 10 | *no response*

3. Downloading Pangaea data

Data

Barrett, Pamela M, Neumann, Rebecca B, and Gawel, James E. 2018. Aqueous arsenic, oxygen, temperature, and plankton arsenic content in urban lakes in the Puget Sound lowland from September 2015 to August 2016. PANGAEA Data Publisher for Earth and Environmental Science. doi:10.1594/PANGAEA.884327.



[Copy Citation](#)

Files in this dataset				
Name	File type	Size	Downloads	
Metadata: Aqueous arsenic, oxygen, temperature, and plankton arsenic content in urban lakes in the Puget Sound lowland from September 2015 to August 2016	http://www.isotc211.org/2005/gmd-pangaea	17 KB	9 views	Download

Distribution	
URL	Format
https://doi.pangaea.de/10.1594/PANGAEA.884327?format=zip	application/zip

The second task asks, “Search and select a data set published by the Pangaea Member Node about arsenic content in the Puget Sound. View its metadata to determine if you can download the data.” Six of the seven participants found a data set that matched the task criteria, but only two people were able to download the data.

The purpose of this task was to assess if users had difficulty understanding the different terminology for data files in the Pangaea metadata. Pangaea uses “distribution” instead of “Files in this dataset”.

Five of the seven participants started the search by filtering my member node.⁵ The other two participants searched “Pangaea” in the search box.

After limiting the search to Pangaea, five of the seven participants went on to limit their search by “arsenic” and “Puget”.

One participant used the location filter to limit by Puget Sound. While Puget Sound is a location filter for other data sets, it is not a location filter for Pangaea data sets. That participant removed the location filter and used the search box for ‘puget’.

⁵ Participant 2 skipped this task. Participant 4 and participant 9 had video errors. They are not included in the analysis. Their open-ended responses are included.

Regardless of whether they limited through the filters or search box, they received three results that matched the task criteria. Six of the seven participants found a dataset that matched the task criteria. Participant 10 selected a dataset that was not located in the Puget Sound.

Participants confused the metadata and data downloads. Only two participants tried to download the distribution URL (participant 1 and 3), while four participants downloaded the metadata record thinking it was the data (participant 5, 6, 7, and 10), and one participant did not download the metadata or data (participant 8).

None of the participants were able to find an application that could read the data file. For example, participant 3 had to google .tab files to learn about what application can open them.

After the users completed the task we asked if they were able to download the data. Eight of the participants thought they downloaded the data, when the video playback showed only two participants selected the right file to download. Three participants mentioned they were confused by what program to use to view the data. They replied:

<i>Participant</i>	<i>Response: Able to download data?</i>
1	I think the distribution would have worked. I was confused by the file type and couldn't open it on my computer.
2	No response
3	Yes
4	I was able to download the data, but the interface wasn't really clear about what was what. The title of the metadata contained the word 'metadata', but it's in XML. I searched within the file and found the word download, but I wouldn't interpret what I found as answering the question "can you download the data file?" I also couldn't tell that the data file was a data file until I downloaded it and opened it in a text editor. The interface label is "Distribution" - what does that mean in this context? The table cell label is "URL" - I can see it's a URL, thanks very much. The right-hand cell is labeled "Format " and says the MIME type is "text/tab separated values" which is a good clue, but very jargony none the less.
5	I think so... just do not know what to do with a file that has .tab as extension
6	Yes I was able to download the data
7	I was able to download the data, but I wasn't sure what program to select to view it.
8	No
9	I was initially unable to find any relevant datasets by selecting Member Node=Pangaea and using the search box to find arsenic content Puget Sound. I then deleted the text box search, and kept Node=Pangaea and Geographic=Puget Sound. Still 0 results. I then

- tried leaving Member Node=Pangaea and did a text box search for arsenic, and finally got 3 results, all which were relevant to arsenic content in Puget Sound (why these didn't come up in my text box search of arsenic content Puget sound is a mystery). I chose "Water temperature, oxygen and aqueous arsenic content in urban lakes in the Puget Sound lowland from September 2015 to August 2016" and was able to open the metadata, find the DOI to the dataset landing page from Pangaea, and successfully download the CSV of the data
- 10 Yes I've got the metadata of the first time of my search.

We also asked, "Are there any terms you did not understand" or "any other comments?" Two participants mentioned they were confused by the metadata and data downloads. Their complete responses are below.

Participant	Response: Terms I did not understand
1	Distribution was unclear.
2	No response
3	Very Smooth!
4	"Distribution" and "URL" didn't say much. Make it clearer where the metadata file is. Make it possible to view the metadata in-line in the browser, esp when it's XML. Action labels: try "Download metadata" and "Download data" or "Download dataset(s)"
5	I think it took a while to start the download and so i was not really sure if i interacted properly. But after the download was shown as completed, i was OK
6	There were multiple matches and I wasn't sure which one to choose
7	Unable to open raw data. This would take me a little more time to sort out why and what program to use.
8	No but the download button next to the metadata made it look like you could download a data set when you could only download metadata
9	The Geographic search on Puget Sound produced zero results. Something is not working correctly.
10	No

4. Images of glaciers from Arctic Data Center

Data

The third task asks, "Search and select a data set from the Arctic Data Center that contains photographs of glaciers. Find what glaciers are included in the photographs." Five of six participants completed the task.⁶

⁶ Participant 2 and 9 did not complete the task. Participant 4 and 6 had issues with their video recordings. Their data is excluded from analysis, but their questionnaire responses are recorded below.

The purpose of the task was to assess how users interacted with a long metadata record.

All five participants who completed the task started by filtering by the Arctic Data Center Member Node and limiting by glaciers and photos or photography. One participant tried to use the data attribute filter to limit the search by photography, and even though the filter auto-populated with photography there were no results (see UX issue #15).

Participant 10 did not complete the task because he filtered by creator, “arctic data center.” While there are 71 results none of them include glacier data, and the participant did not realize that creator is the wrong filter.

The datasets selected by the participants all had over 100 images, and the metadata records were over 100 printed pages. The participants all expanded the data files table to view all the files in the dataset, but only participant 1 and participant 8 scrolled past the table of files. However neither participant scrolled to the end of the record. Besides looking at the table of data files, they spent time scanning the abstract.

After completing the task we asked them to name at least one of the glaciers shown in the photographs, and for overall thoughts about this metadata display. They had a variety of responses, including “It would be nice to collapse the many “other entity” fields into a list of entity names, and have the boxed areas expand only as they are selected,” and many comments that they were not sure of the glacier location or the subject of the images. Their full responses are below.

<i>Participant</i>	<i>Response: glacier's name</i>	<i>Overall thoughts</i>
1	Oblique glacier? Not as obvious but I scanned quickly.	It seems really detailed, which I really like. But its a lot of scrolling and you don't know what you are going to find--so is it worth scrolling? I would like an overview of the parts of the metadata or tabs I could open and close to view information.
2	No response	
3	AP81K1	In this situation it could be nice to collapse the many "Other Entity" fields into a list of Entity Names, and have the boxed areas expand only as they are selected. It is a bit cumbersome to scroll through the metadata when there are so many pictures and each one takes up most of the screen.
4	Linnébreen (Linné Glacier)	It seems fine. I wonder how a user could determine the correct sequence of photographs or easily stitch them together in an animation. I didn't examine the EML file - the answer might be in there. I had no idea where this glacier was. I had to scroll all the way to the bottom to find it was in Svalbard. Some clue nearer the top about location might be good.
5	I just saw 2 pictures... am not sure if they were the same or different ones. I can not even name the files i downloaded	Hard to say if the glaciers photos are from the same or different ones. Do not have information about the glacier

	(photos) as i just picked alleatory two of them	
6	Unknown. Could not find location names with a quick search the metadata. Would have to download the dataset and inspect.	Hard to find any of the location name values. Would be good to have summaries of possible categorical variable values (such as the glacier names)
7	I didn't see any glaciers in the photos. I saw clear cutting and Mt. Rainier in the distance.	I would like a description attached to the tiff files in some way
8	unknown just says Oblique Glacier Photo	display ok but getting to Photographs by selecting data attribute did not work. Just had to scan results
9	no response	
10	I could not find one.	Hard to pick only pictures in this database.

5. Dates data was collected

Data

The final task asks users to, “Search and select a dataset published by USGS about gas and heat emission in Yellowstone National Park. View its metadata to determine the dates the data was collected.” Five of seven participants completed the task.⁷

The purpose of this task was to assess how easily users could find specific information (e.g., date) in the metadata record.

Participant 10 selected the creator filter, and could not find relevant results. He did not complete the task. The other participants started by filtering by DATA SOURCE: USGS SCIENCE DATA CATALOG.

Two participants (participant 1 and 9) used the location filter to limit by Yellowstone. While there are 30 results for “data source=USGS and Location=Yellowstone”, none of those results include heat and gas emissions data. Participant 9 ended the task due to errors with the location filter.

Participant 1 also had difficulties with searching for gas and heat emission. The participant kept typing *heat* and gas emission which returns zero results, but gas and *heat* emission returns 10 results.

Participant 1 along with the other four participants completed the task by searching by member node and then keyword search for Yellowstone and gas and heat emission.

Participants easily found the temporal coverage information in the metadata record. Participants either scrolled to the temporal coverage metadata, searched by CTRL+F, or found in the title.

⁷ Participant 2 and 7 did not complete task. A video error occurred for Participant 4. Those participants are excluded from analysis. Participant 4 questionnaire responses are included.

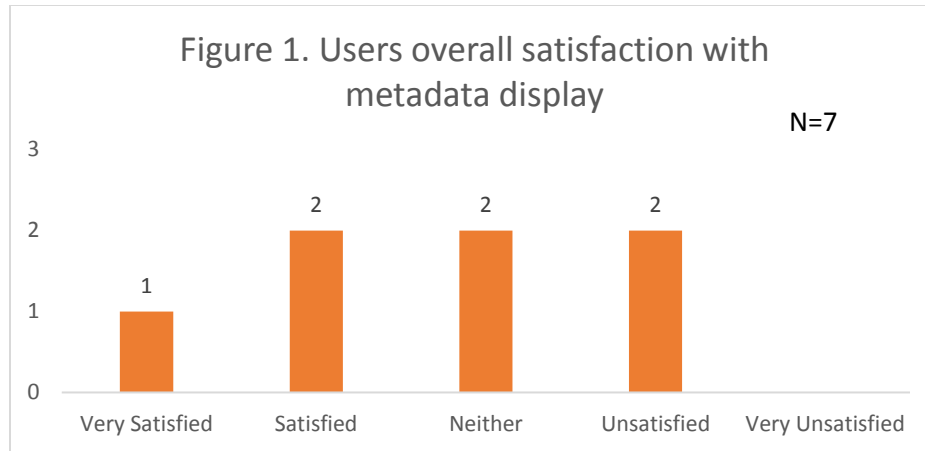
After completing the task we asked for the year(s) the data was collected and for any additional comments. Overall participants were satisfied with how the temporal information is presented, but two participants commented on the issues they had with the location filter. They replied:

<i>Participant</i>	<i>Response: Year(s)</i>	<i>Comments</i>
1	2016	the search was tricky. The location filter did not seem to ork. I got yellowstone results but none of them included heat and gas emission. When I removed the location fitler I was able to get heat and gas emission data from yellowstone. I was also confsued by the suggestion pop up as I was typing in the location filter.
2	<i>no response</i>	
3	2016	Very easy to find
4	2016	It[’s great the title and abstract contained the collection dates
5	2016	I could visualize this information in the website and not in the metadata file I wish to inform that i downloaded the metadata file but had not any software installed in my computer to open it. It was presented to me an error message from Windows... And this error message was in portuguese (that is the language i installed the windows...)
6	2016	
7	<i>no response</i>	<i>no response</i>
8	May-October 2016	Location is a confusing field. I had to put Yellowstone in the normal search box which did provide suggested text of "Yellowstone National Park." Location facet may not work.
9	<i>No response</i>	
10	I could not find results.	I think these tasks require some previous knowledge about them. It is easier when you know something about it and one can replace some topics by related topics.

6. Overall satisfaction with metadata display

data

After completing all four tasks the participants were asked to rate their satisfaction with the metadata displays from very satisfied (1) to very unsatisfied (5) (See image 1). Seven participants answered the question, and their responses ranged from very satisfied to unsatisfied (see Figure 1). Their responses implies that there is room to improve the user satisfaction of the Search.



7. What would the user change?

Data

The last question of the UX test was, “If you could change one thing about the metadata displays what would you change?” Their responses are below. While they had a variety of suggestions, they all centered on making it easier to find the relevant information from moving information to a more prominent position or collapsing sections so it is easier to scroll.

Participant	Response: What I would change
1	Nothing. Maybe the geographic region should be higher.
2	No response
3	Collapse metadata sections when there are many (>5-10) of the same item (e.g. "Other Entity") in a row, so the user can easily scroll through the items by name and expand them when needed, and also to allow the user to view the entire contents of the metadata without having to scroll so far.
4	Better labels for the Pangaea data. They were so confusing and jargony.
5	make them more readable and easy to access
6	Location information (as a map) should be closer to the top
7	No response
8	Facets. If something turns out not to be available there should not be a number next to a subject facet of results when terms are filled out, yet do not show and results when submitted.
9	The metadata themselves often do not contain sufficient information to support re-use. Users should see data structure/features and lineage/methodology/data quality as displayed attributes, which, if absent in the metadata, should be

displayed as empty in the metadata display. This will save users time to know that not enough metadata has been supplied to support responsible reuse of those data.

10 I am not familiar with the topics searched.

A few of the participants also had some final comments before ending the test.

<i>Participant</i>	<i>Response: Final comments</i>
1	There's a lot of content. That's good thing. I'd like a way to see everything on page page (not below fold) then view more to see details that are significant to me.
3	Looks great!
4	It would be helpful to provide answers to everything in your questions in the display so that a user doesn't have to download a file - esp. xml - to try to determine if it's relevant.
5	As informed before i did not had a software to visualize the metadata. I opened some files using Internet Explorer but this was not the main idea, i guess...
9	Search is very deficient operationally. Faceted search produces no matches even though they clearly exist in the metadata for those facets. The full text search is highly suspect; searches fail for a combination of keywords that actually exist in the titles - probably the highest weighted fields in the metadata - which is a huge failure.
10	As I said before it would be easier to search for familiar topics. This would allow me to use other keywords in the search.

UXL MANAGER CRITIQUE OF SEARCH

The User-eXperience Lab researcher identified a few issues while creating the UX test and during pilot testing. The issues influenced how the UXL researcher developed some of the tasks. The issues are described in this section.

8. Author search

Medium

The screenshot displays the DataONE Search interface. The top navigation bar includes links for About, News, Participate, Resources, Education, and Data. Below this, a search bar is present with a 'Search' button and a 'Summary' button. A 'Jump to: DOI or ID' link is also visible. The main search results area shows 'Datasets 1 to 15 of 15'. The search filters are applied to 'young, adam' and 'young'. The 'My Search' section is highlighted with a red circle. The search results are sorted by 'Most recent'. The results list includes datasets from Dryad and Arctic Data Center, such as 'Young, Adam M., Higuera, Philip E., Duffy, Paul A., and Hu, Feng Sheng. 2016. Figures. Dryad Digital Repository' and 'Adam M. Young. 2015. Observed, modeled, and projected fire, climate, and vegetation data from Alaskan tundra and boreal forest ecosystems, 1950-2100, Arctic Data Center'.

The creator, Adam M. Young, has authored datasets in Dryad and Arctic Data Center. If the user searches for him through the creator filter (YOUNG, ADAM), the user only gets results from Dryad, and ADAM YOUNG returns zero results. If the user searches for “young” through the keyword filter the user can receive results from Arctic Data Center and Dryad. There seems to be Member Node differences between what information is available for certain Search filters. Filters need to function consistently for datasets from each Member Node for users to

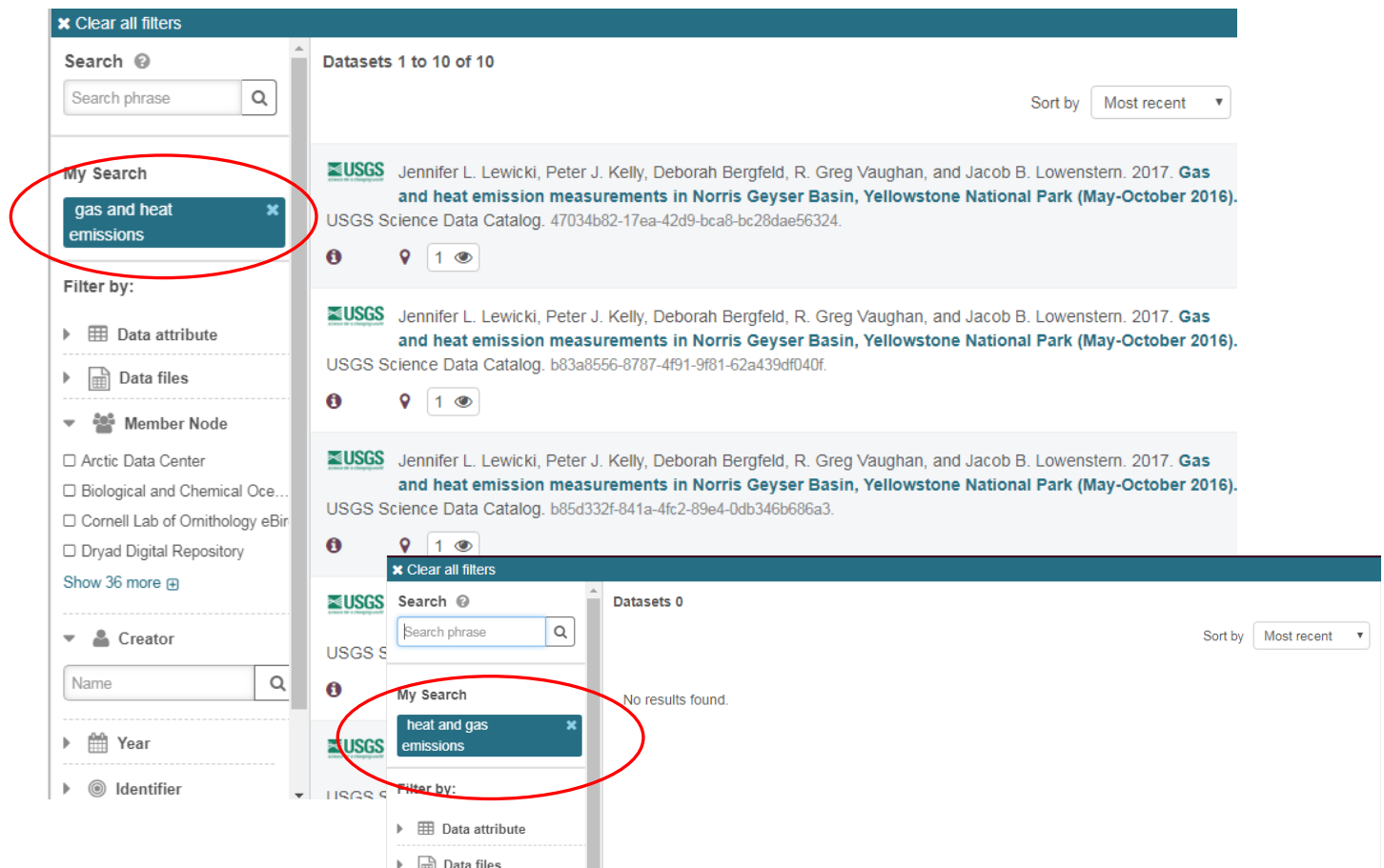
find all the relevant data. This task ended up being removed from the usability test because users struggled with the task during pilot testing.

Recommendation:

- The creator filter needs to function consistently for data from each member node.
- Evaluate which metadata fields are being used in each filter and compare between member nodes.
- Creator filter should provide the same results regardless of the order of the author's name.
 - If the order matters then that needs to be expressed in the filter.

9. Word order changes results

High



How the user orders a search term changes the results. For example, if a user needed data on heat and gas emission from Yellowstone geysers they must enter "gas and *heat* emission". If they search for "*heat* and gas emission" they receive zero results. While this created issues during pilot testing, we kept the task in the UX test. Since the task scenario had the correct

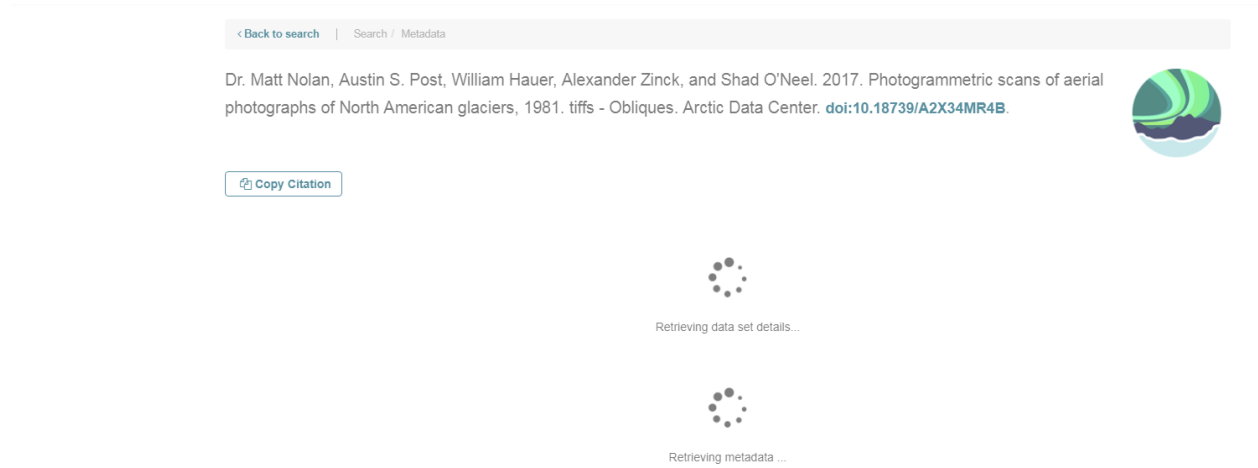
order of words, it is not surprising that only one participant during the study tried to search for the incorrect order.

Recommendations:

- Allow the search to search for phrases regardless of the word order.

10. Delay in loading metadata

High



There is a significant lag between when the user clicks on a data set and when the details load. It seems to range from 5 seconds to 30 seconds. Those times all occurred on the same university network. During the usability test, participants also had significant delays from a few seconds to over 20 seconds.

The time delay is very frustrating when looking at multiple data sets. Usability guidelines recommend keeping the response time under 10 seconds. Ideally the response time is 0.1 seconds, but a 1 second delay still allows a user to feel in control even though they can sense a delay. After 10 seconds the user's attention wanders, and they are likely to leave the site.⁸ This is a major usability issue that creates an unpleasant user experience and could result in the loss of users.

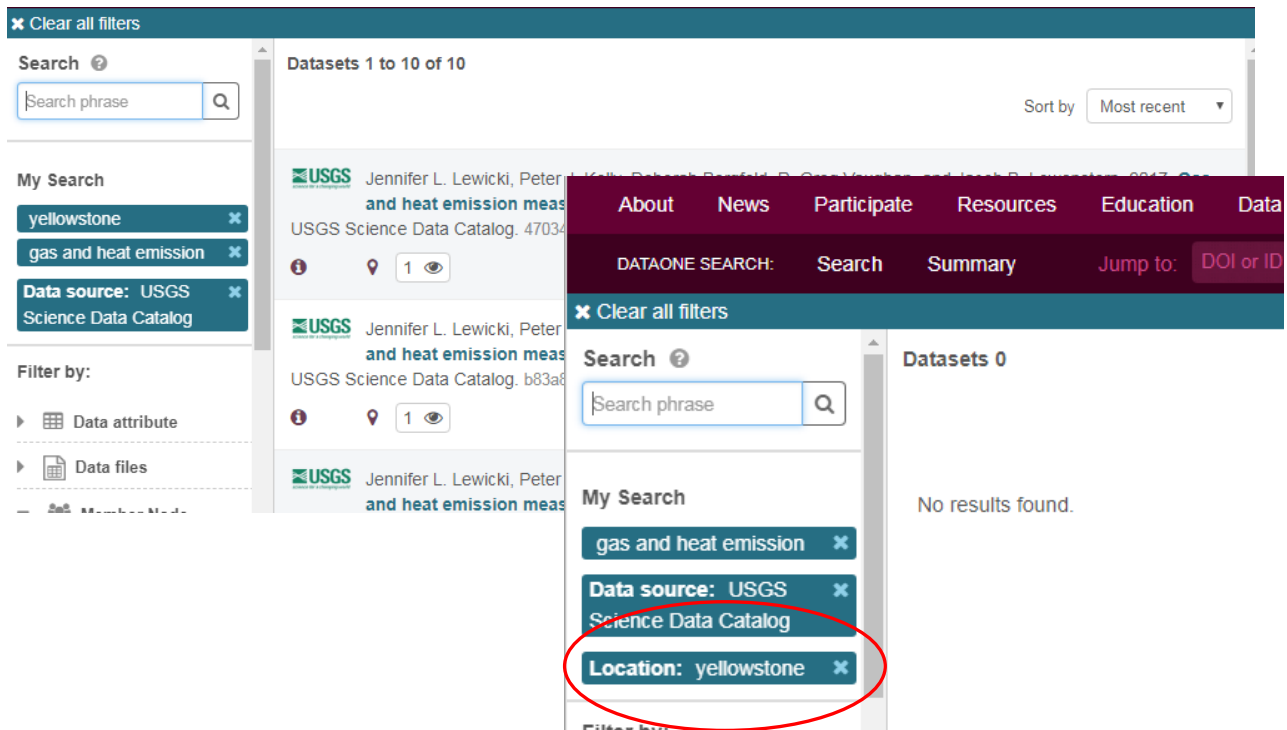
Recommendation:

- The load time for the metadata display needs to feel "instantaneous"

⁸ <https://www.nngroup.com/articles/website-response-times/>

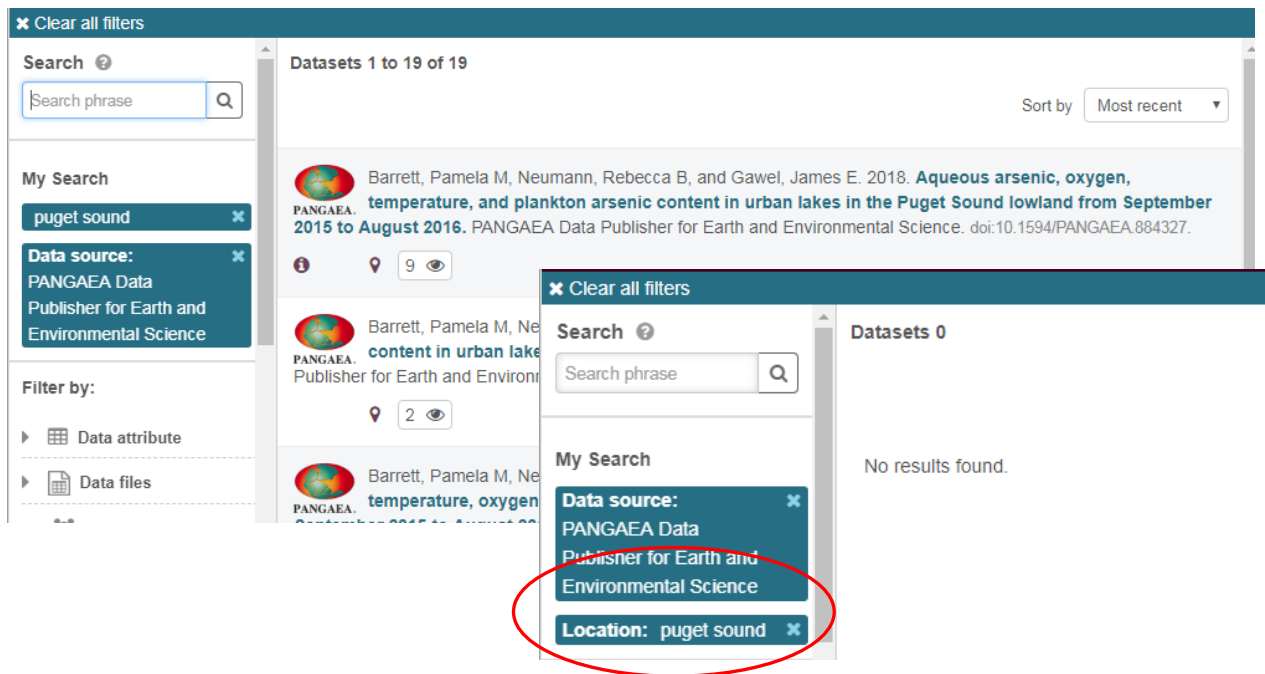
11. Location filter

Medium



Similar to issue #8 with the creator filter, the location filter does not function the same way for each member node and for data sets within member nodes. The irregularity of how the location filter functions causes usability issues. We encountered two issues with the location filter:

1. We asked users to find a data set from USGS about heat and gas emission from Yellowstone National Park (see the above image). There are USGS data sets with the Yellowstone location filter, but they do not include heat and gas emissions data. The user must remove the Yellowstone location filter and search for Yellowstone and 'heat and gas emissions' through the search box.
2. We asked users to find a data set from Pangaea Member Node about arsenic content in Puget Sound, but when a user selects the MEMBER NODE filter as PANGAEA and then uses the location filter to limit by "Puget Sound" they receive zero results (see image below). Puget Sound is not a location filter for Pangaea. If the user uses the location filter to filter by Puget Sound without specifying the member node, they receive 84 results from Dryad.



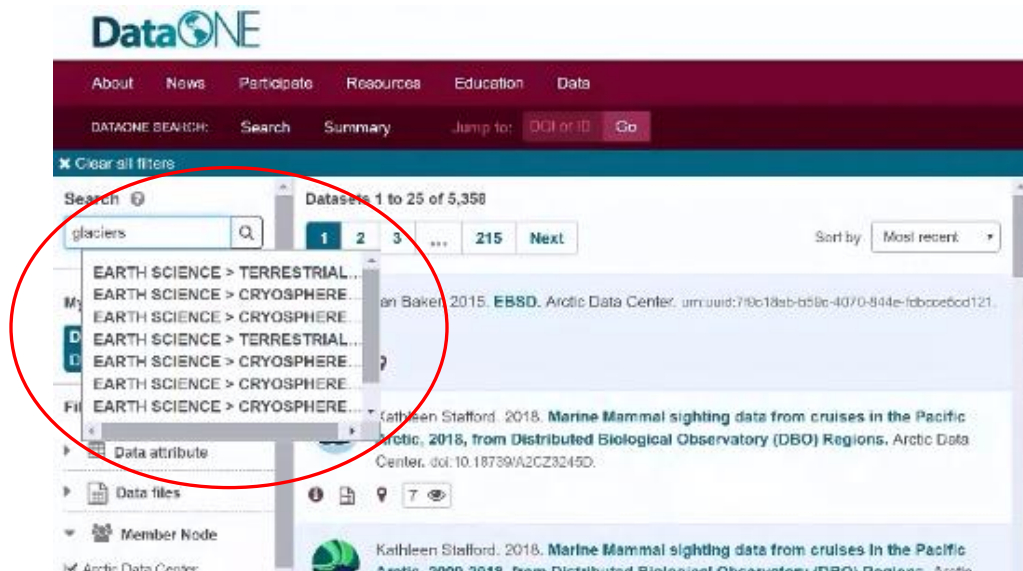
Recommendation:

- The filters need to function consistently for data from each member node. If a filter only pulls data from certain member nodes it provides poor results and leads to usability issues.

UX ISSUES: SEARCH INTERFACE

12. Auto-fill suggestions window

Low



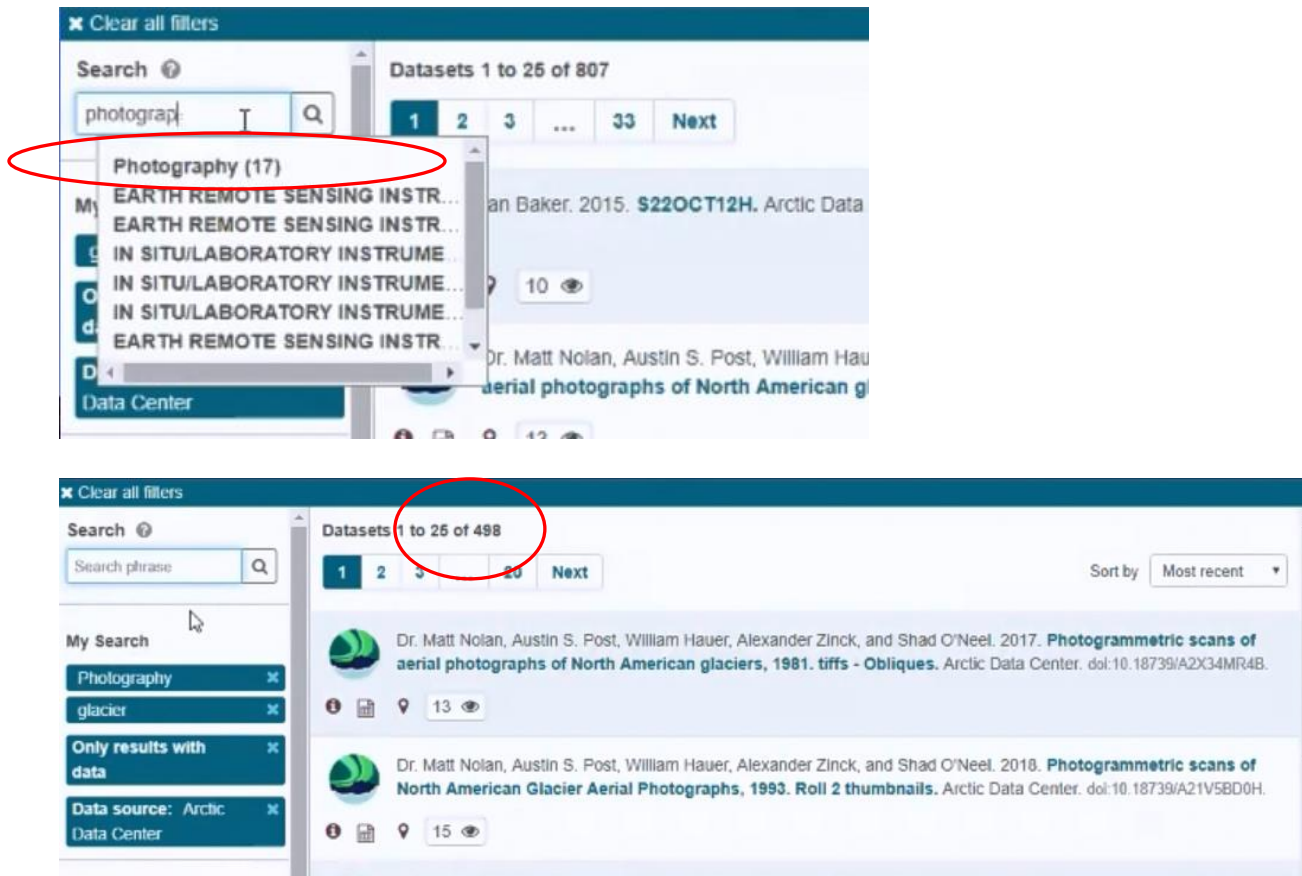
Some of the auto-fill suggestions do not fit in the window. While there appears to be a page scroll, the scroll does not allow the user to view the entire suggestion. This makes it impossible for the user to know which search term they are applying.

Recommendations:

- Increase size of the box to adjust to fit the entire suggestion.
- Or improve the functionality of the bottom page scroll to allow the user to move the page left and right.

13. Number of search results

Medium



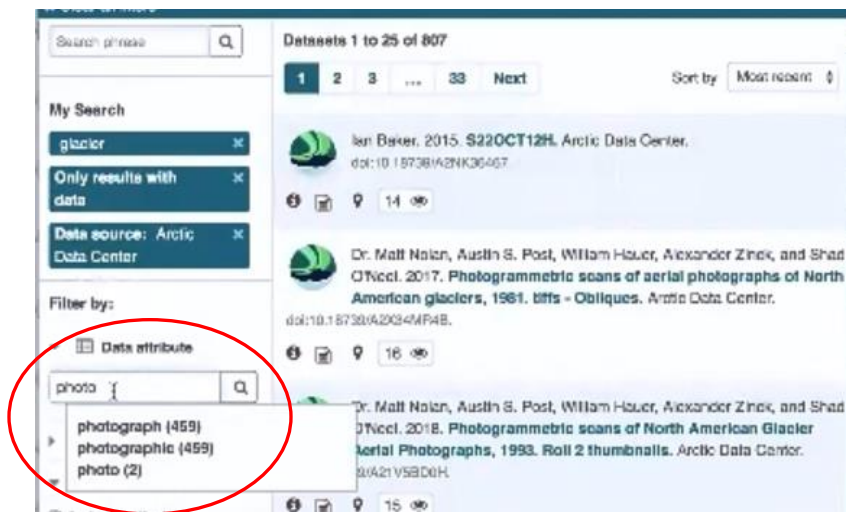
When a user selects a search phrase with an associated number of results (e.g., photography (17)) they expect to receive only 17 results. The DataONE Search, however, returns 498 results. This error occurs throughout the search, and the numbers by the filters do not represent anything in the current interface. One participant expressed his frustration, “If something turns out not to be available there should not be a number next to a subject facet of results when terms are filled out, yet do not show any results when submitted.”

Recommendation:

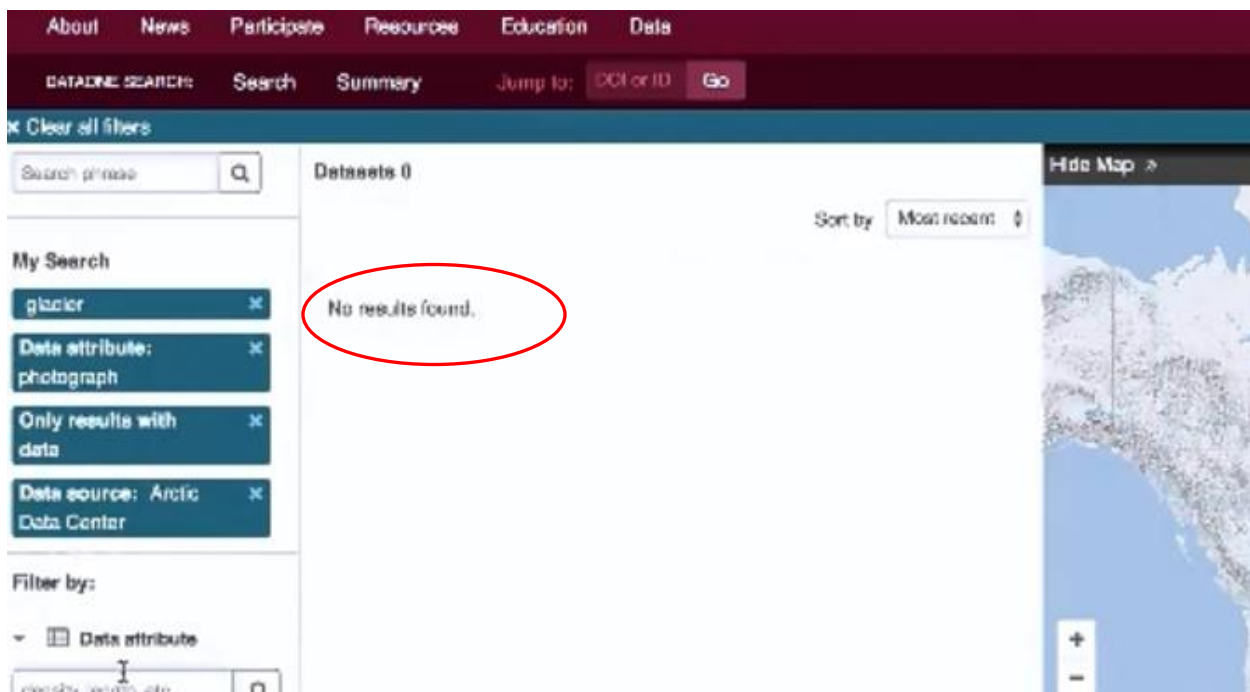
- Improve how the autocomplete dropdowns are being generated.
- If cannot be fixed then the numbers need to be removed.

14. Search provides auto-fill suggestions without results

High



When a user starts to type photograph in the data attribute filter, they receive autosuggestions for “photograph (459)”. This implies there are 459 results with a photograph data attribute, but when the user selects the autocomplete suggestion they receive zero results.

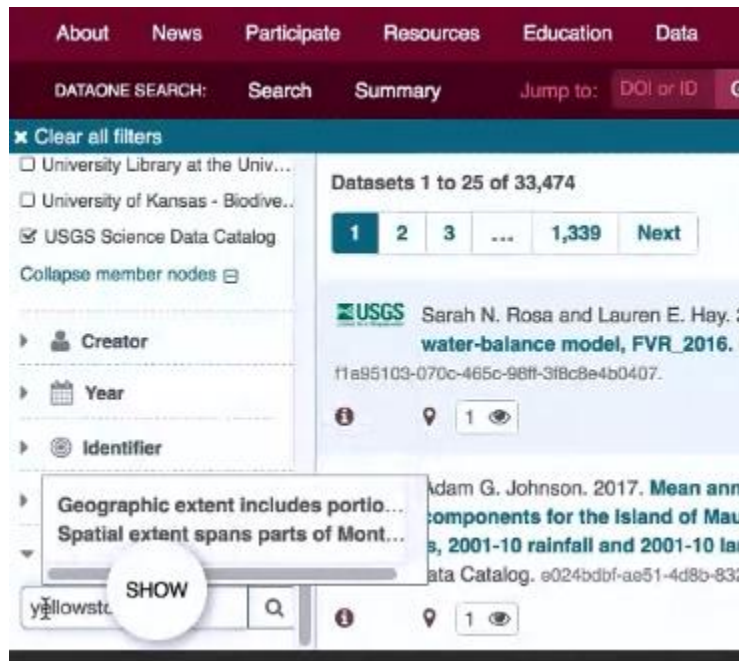


Recommendation:

- This is tied to the previous UX issue. The autocompletes need to be corrected to represent accurate data.
- Autocomplete suggestions should only suggest topics with results.

15. Search provides auto-fill suggestions that don't seem related

Low



Note: The "Show" bubble is part of the Loop11 Interface

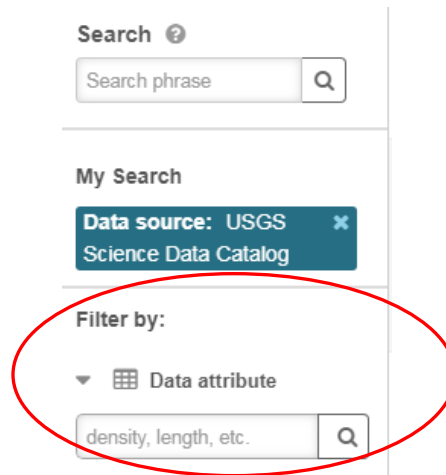
Participant 1 typed Yellowstone into the Location filter, and the auto-suggestions do not seem relevant, and because of UX Issue #12 the user cannot view the entire suggestion to determine its relevancy. After the task participant 1 said, "I was also confused by the suggestion pop up as I was typing in the location filter."

Recommendation:

- This issue is related to the previous two issues. The auto-suggestion drop-down needs to be corrected so that only accurate and related suggestions appear.

16. Purpose of the Attribute Filter

Low



The screenshot shows a search interface with the following elements:

- Search** section with a search bar containing "Search phrase" and a magnifying glass icon.
- My Search** section with a blue button labeled "Data source: USGS Science Data Catalog" and a close icon (X).
- Filter by:** section, which is circled in red. It contains a dropdown menu with a grid icon and the text "Data attribute". Below the dropdown is a search bar containing "density, length, etc." and a magnifying glass icon.

Users could not use the data attribute filter to successfully filter the results. Users did not seem to have a grasp on what is considered a data attribute. All their attempts to use the attribute filter were unsuccessful. It is not clear if the attribute filter is functioning correctly or if the filter needs more specification to help users successfully use it.

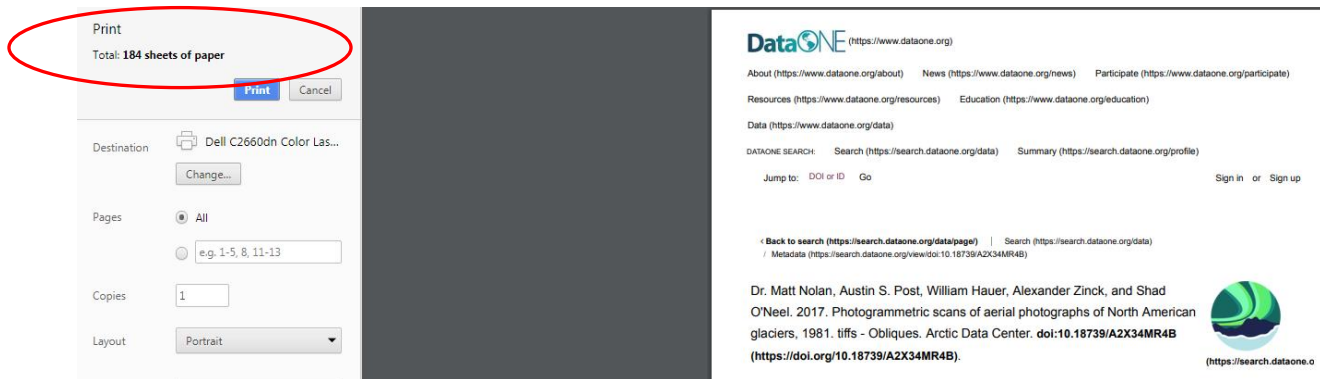
Recommendation:

- Users need more clarification on how to use the data attribute filter.
- Consider switching the placement of the data attribute and location filters. Users used the location filter more than they did they data attribute filter.

UX ISSUES: METADATA DISPLAY

17. Decrease length of metadata displays

Medium



Some of the metadata records are over 100 printed pages. In some cases, the user has to scroll over nearly 90 pages of “other entity” information before they find metadata information, and as they are scrolling they do not even know if there is additional information below the entities. In the study, most participants did not scroll passed the data table. In task 3 where they looked at glacier photographs, none of the participants scrolled past the “data table, image, and other detail” information, even though some of them scrolled for a long time.

Providing a way for users to view all the metadata fields above the fold will help user’s better access the information.

Recommendation:

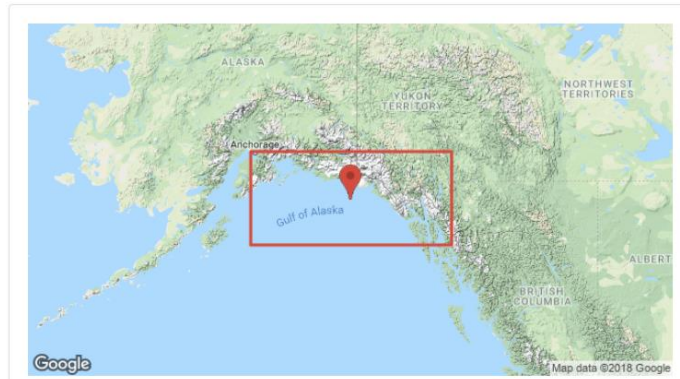
- Collapse the metadata fields, and users can open the fields to view more information if they need. This will be especially helpful for the large metadata records.
- Create a hyperlinked list of the metadata fields at the top of the record and allow users to jump to the necessary information.

18. Users would like Geographic Information to be higher

Low

Geographic Region

Geographic Description	North American glaciers		
Bounding Coordinates	North	61.282059	degrees
	South	57.111998	degrees
	East	-132.79654	degrees
	West	-150.325505	degrees



While users consider a lot of different metadata fields to be important in determining quality, there seems to be the most overlap with the geographic information. When asked what fields are the most important to help determine the quality of data, three participants mentioned geographic region or location. When asked if they could change one thing about the metadata display, two participants thought the location/map should be closer to the top.

Recommendation:

- Move geographic region information above the data table details.
- OR provide a hyperlinked list of the metadata fields above the page fold (see UX issue #17).

19. Users confused by 'download' metadata

Medium

Files in this dataset Package: resource_map_doi:10.1073/92X34846				
Name	File type	Size	Download All	
Metadata: science_metadata.xml	EML v2.1.1	280 KB	Download	
AP81K1_047.tif	image/tiff	185 MB	Download	
AP81K1_278.tif	image/tiff	135 MB	Download	
AP81K1_265.tif	image/tiff	135 MB	Download	

During the study participants often downloaded the metadata record thinking they were downloading the data. In task 2 they were asked to download the data for a Pangaea record. Only two users downloaded the data, while four users thought they downloaded the data but really downloaded the metadata. One participant said, "The download button next to the

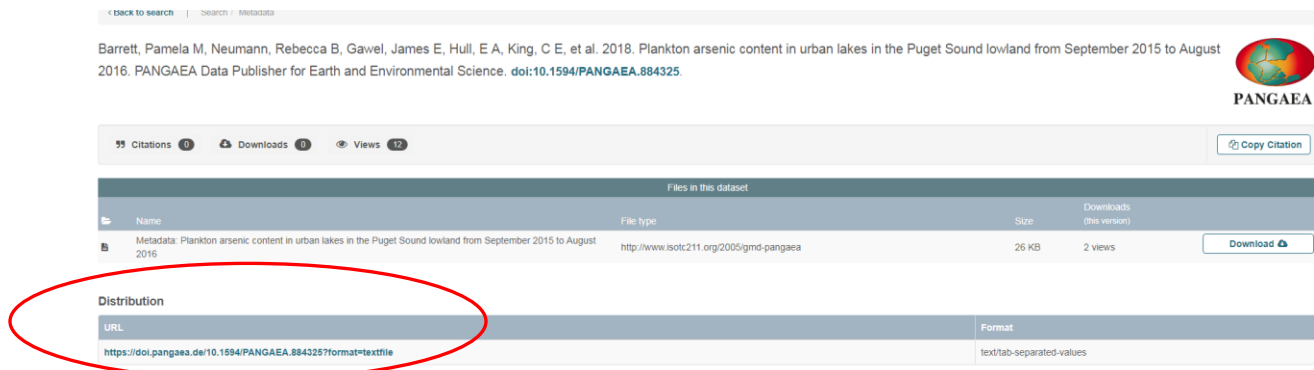
metadata made it look like you could download a data set when you could only download metadata.” Another participant commented, “The interface wasn’t really clear about what was what.”

Recommendation:

- Consider alternative ways to present the metadata record.
 - Replace “download” with “download metadata”
 - Move the metadata record to a separate table with a clear heading.

20. “Distribution” is jargon

Low



The screenshot shows a PANGAEA dataset page for 'Plankton arsenic content in urban lakes in the Puget Sound lowland from September 2015 to August 2016'. It includes a table of files in the dataset and a separate 'Distribution' table. The 'Distribution' table has two columns: 'URL' and 'Format'. The URL is 'https://doi.pangaea.de/10.1594/PANGAEA.884325?format=textfile' and the format is 'text/tab-separated-values'. The 'Distribution' table is circled in red.

Name	File type	Size	Downloads (this version)
Metadata: Plankton arsenic content in urban lakes in the Puget Sound lowland from September 2015 to August 2016	http://www.isotc211.org/2005/gmd-pangaea	26 KB	2 views

URL	Format
https://doi.pangaea.de/10.1594/PANGAEA.884325?format=textfile	text/tab-separated-values

Users were confused by the ‘distribution’ wording. In other metadata records the data files were included in the top table, but PANGAEA separates them in a new ‘distribution’ table. Users did not equate ‘distribution’ with downloading a data file. One user said the one thing he would change about the Search was “better labels for the PANGAEA data. They were confusing and jargony.”

Usability guidelines recommends using consistent labeling throughout a site.⁹ Users should not have to remember that sometimes they have to download data in the top table and sometimes they have to go to the distribution table. While PANGAEA may prefer the distribution labeling, the user considers it jargon and it does not improve the user’s experience.

Recommendation:

- Data files should be placed in the same place and labeled the same way in all metadata records.
- Remove the “distribution” table in PANGAEA records.

⁹ <https://www.nngroup.com/articles/do-interface-standards-stifle-design-creativity/>

WHAT'S NEXT?

NEXT STEP RECOMMENDATIONS

Review Search Engine Functionality

UX testing showed multiple issues with the facets, and there may be more issues that were not exposed during testing. We recommend reviewing the functionality of the different search facets from a technical perspective.

Continue Iterative Usability Testing

DataONE has conducted iterative testing of the DataONE Search since it was launched. The Usability & Assessment team recommends continuing UX testing on an annual basis and as new features are added to the search.

Conduct Eye Tracking Studies

Eye tracking studies can show where users look on the search and metadata displays. This data can add information to the current findings.

APPENDIX A: UX TEST SCENARIO

DataONE Users Group UX 2018

Demographics:

- 1) What is your discipline? _____
- 2) How often do you search for data? DAILY WEEKLY MONTHLY OCCASIONALLY NEVER

Tasks:

- 1) Choose your own adventure. Think about a topic you are interested in and find a quality data set that is relevant to your topic. Once you have found a quality data set you may end the task.
 - a. What fields are most important to you to help determine the quality of the data, and did you find them?
- 2) Search and select a data set published by the Pangaea Member Node about arsenic content in the Puget Sound. View its metadata to determine if you can download the data.
 - a. Were you able to download the data? If not, why?
 - b. Were there any terms you did not understand? Any other comments?
- 3) Search and select a data set from the Arctic Data Center that contains photographs of glaciers. Find what glaciers are included in the photographs.
 - a. Name at least one of the glaciers shown in the photographs? _____
 - b. Overall thoughts about this metadata display?
- 4) Search and select a dataset published by USGS about gas and heat emission in Yellowstone National Park. View its metadata to determine the dates the data was collected.
 - a. What year(s) was the data collected?
 - b. Any additional comments?

Post-Task questions:

Please answer the following questions based on your experience with all four tasks.

1. Overall, what was your satisfaction with the metadata displays? Scale of 1-5 (very satisfied-very unsatisfied)
2. Any final comments about the metadata displays you viewed during the test?

3. If you could change one thing about the metadata displays what would you change?
