**Media Cloud Heuristic Analysis**

*A high-level analysis of Media Cloud’s learnability and usability from a novice user and general UX perspective*

* Submitted by Susan Stuart, Feb. 12, 2018

The purpose of this analysis was to document my impressions of Media Cloud and its user guides from a new novice user perspective, as well as a general user experience practitioner perspective, with an eye toward exploring and validating some of my hypotheses here via our novice user research studies. The analysis also provided a chance for me to simply become further acquainted with the capabilities of the tool, to capture early UI and documentation ideas for later consideration (and possible exploration), and to try to separate known bugs from other issues with the team.

User experience analysis is both an art and science; subjectivity will always be a component, and compromises abound in design strategies. But we can look to usage patterns and perceptions of similar applications, best practice UX heuristics, and leverage general cognitive research on how users and readers process complex information for the team to consider in improving usability and engagement.

1. **Methodology**

I reviewed the tools and documentation as a first-time user would encounter them in most cases, with a casual look at the site and UI, and then engaging with the available learning support—the current user guides. I also watched a demo by Natalie and tried my hand at a research challenge she provided. I found that my retention of the demo was low, and as the demo isn’t material users would encounter currently on the site, it won’t be a focus here. However, I think we can leverage this early analysis to develop new learning content structures in general, and aim test some of these in the user studies.

*Note: I used Dashboard for some of the tasks here when Explorer was unavailable for technical reasons; however, most of the conclusions focus on the Explorer UI.*

1. **A note about data analysis for novice users**There’s currently a lot of interest in making data analysis tools more accessible to more people, including novice users. As I immersed myself in this project, I began encountering [more commentaries on the challenges that this goal poses](https://www.usermuse.com/blog/can-data-discovery-tools-make-analysis-easy/?ref=quuu&utm_content=buffer3009a&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer), which confirmed my own suspicions about potential limitations (actual or perceived) for novices. Further, this analysis was conducted with the knowledge that Media Cloud’s primary strategic goals still focus mostly on the experience of expert research analyst users, as well as the consulting service Media Cloud provides for a fee.

Still, as I understand it, the option to learn, and, at least, to assume a basic level of data analysis confidence with Media Cloud should be available to everyone, and I agree that’s possible, but requires the exploration of simplification strategies. Novices are much tougher users to engage in general, so considerable thought will need to be given as to how much to gear the overall experience and instruction toward them.

This analysis largely focuses primarily on instructional support and secondarily on UI, but also touches on usefulness of current or potential product features.
2. **Initial impressions**

The design of the non-application pages of the Media Cloud site is visually appealing and my impression was that it introduced the product well, but we may want to look for opportunities to even better capture and summarize the features and benefits, as discussed below. In any case, this analysis focuses not on the overall site but on the Tools themselves.

1. **User engagement and expectations overall**

Users tend to impatient in understanding the benefits of applications, and are reluctant to put effort into learning them until they are fully convinced the benefits will be worth it. When being introduced to new tools, they quickly expect to:

•  understand the basic concept functionality (especially features and benefits)

• feel intrigued enough to learn more or try the tools out

• feel confident that they are the right users for the tools; they can use and benefit from them with a minimal amount of effort

I believe that while Media Cloud provides a good level of intrigue and a fair summary of features and benefits, confidence probably falls quickly for most novice users, as it did for me, when trying the tools out. Some of this was due to usability issues, but much of it was because I didn’t know what to look for, and **unlike many applications,** **data analysis tools often do not provide the kinds of outputs that can be designed to “jump out at you,” providing a clear message with only one interpretation**.

So as a user, I experienced a constant hope/ disappointment cycle, waiting for that big “payoff” on the outputs screen that I get with most software applications, but never feeling sure I was getting any notable result with Media Cloud. Also, while the input/ output sequence in the interface is fairly simple, the actual tasks involved as a whole in using Media Cloud may involve a substantial offline or off-site effort, including formulating research questions, identifying key events, and so on. The Media Cloud documentation, whether appearing in the navitorial prompts, online help, or in user guides, could benefit from integrating more instruction and guidance of the overall offline/ online task flow in one place, rather than in the somewhat scattered and incomplete documentation that exists now.

I agree with the article mentioned in Section II above, that instruction should focus on three things: how to formulate the right research questions, how to understand the underlying data (what’s included, what’s *not* included, and implications of both), and how to master a basic level (at least) the techniques for recognizing what’s significant and otherwise analyzing data.

1. **Could Media Cloud require the novice user to have to “dig” less?**

An expectation I had when first trying Media Cloud, probably from becoming accustomed to applications serving up data they want me to care about, was that the tools would surface data points that stand out in some way, rather than just providing functionality for me to manually “dig” for results. While this crosses over into the feature suggestion category of the analysis, I feel that some simple functionality could be built or expanded upon to make a greater impact upon landing and during the first-time experience for the user in an evaluative mode.

Media Cloud could, upon landing, provide saved “moments” of data (borrowing the Twitter analogy only; not the label). While to do this programmatically might require AI functionality (?) for some features, curation of data “moments” by MIT/ Harvard experts (and perhaps, associated or “guest curators”) to get users started, I believe, could help considerably with engagement—and perhaps even be the only function many novices need.

These data “moments” (similar to the ideas of saved “topics” in Topic Mapper or “sample searches” in Explore) could perhaps have an intriguing summary, more engaging visuals (such as media images, not just dataviz), callouts and integrated external information such as event timelines to help illustrate why the saved data points to a unique conclusion people might care about. And since Media Cloud is really about *news media* (a point that could probably benefit from more reiteration at the highest levels of the site), the experience might be enhanced if the saved topics/ curated data were be organized under a familiar news taxonomy (politics, business, technology, science, etc), to reflect a familiar mental model, encouraging the option to browse, rather than just “dig.” Perhaps the user could then subscribe to “alerts” on these categories (or subcategories), as many novice users might not have a reason to frequently visit Media Cloud.

Again, it seems that Topic Mapper and the sample searches are an attempt to address the need to see more intriguing results immediately, but I recommend exploring this idea further. The sample search screenshots on the landing page of Explorer feel large and distracting to me without saying much, and the browsing of saved topics is a start, but lack organization, and users might be more intrigued by elaboration/ commentary by topic authors. Both topics and sample searches also drop me into the complex output UI, so we might consider an interstitial explanatory page for these saved instances.

1. **The distinction and flow between Explorer, Topic Mapper and Source Manager doesn’t feel worked out.**

As a new user, I thought that I would want to try and use both Explorer (or Dashboard) and Topic Mapper because their features would be different, but, in fact, there seems to be considerable overlap in visible outputs, with the main distinction being in the data itself (RSS vs. spidering)—but even that isn’t communicated upfront. The taglines/ descriptors for the two tools were vague and confusing to me, both at first and even when I actually understood their functionality.

* Why do the two exist separately? One assumption was that Explorer is for novices, Topic Mapper for experts. I wondered if Topic Mapper had everything that Explorer does and more… is there ever a reason to use Explorer if you’re not a beginner?
* Does Explorer/ Dashboard really yield a “quick overview” (as it says in the tagline/ descriptor)?
* Topic Mapper was also called “an extension of Dashboard” in the guides. It felt to me like that was the intent, and therefore why separate these products at all? Why not just give users the option to keep deeper with optional functionality in a single flow?
* The Explorer descriptor/ tagline refers to “topic,” which adds some confusion.
* The Topic Mapper descriptor/ tagline references “how language is used” as being a distinctive feature, but Explorer actually has “language” in the subnav. Doesn’t it also show how language is used?
* The user must immediately decide whether to create a topic or a query, so they must know that a topic includes a query, but these two appear at first glance to be either/ or choices.
* Also, Rahul mentioned in an e-mail to a listserv that one of Topic Mapper’s key features was creating subtopics, so you could avoid the Boolean keyword searches. Where is this functionality? I only see query functionality at the highest level, at least. It would be interesting to know what the Boolean search equivalent of a subtopic is—just an “AND” between two terms…? Given that subtopics are an easier concept to understand, the presence of this functionality again begs the question of why it’s being saved for the more
in-depth, presumably “advanced” tool.
* What is the significance of the data being based on RSS only vs. spidering? Just more data? is there any risk that you’re “over-reaching” in Topic Mapper in terms of a relevant search? I recommend discussing the reach of these two data gathering methods—pros and cons. Then provide a specific (as possible) recommendation on where to begin.
* One of the guides states about Topic Mapper’s spidering: *“since hyperlinks deteriorate over time we recommend you don’t go back in time more than one year.”* Learning this limitation adding further confusion when I considered the positioning of Topic Mapper as the more robust tool.

Source Manager is intuitive in terms of what it is and why I should care about it, but although it exists mostly as a reference/ database area, I felt I could benefit by a deeper explanation of sources—what is and *isn’t* included (TV transcripts, podcasts? more specificity would be great), as well as some coaching around how I can evaluate when to include which sources, and being direct about the limitations of Media Cloud queries because of any source limitations. One user interview validated my own instinct that sources would be one of users’ very first considerations (he immediately asked how far back the data goes), yet Source Manager appears last visually in the order of the tool components. While it wouldn’t make sense for Source Manager to be a primary focus, I recommend we explore ways to better address users’ natural questions about sources upfront. I also think users would benefit greatly from a taxonomy to make the sources easier to browse and find, both from the Source Manager landing page, and within the task flow of setting up a query or topic. Finally, it would be helpful to know how common suggesting a new source is part of the process of using the tools. How do you know when a new source needs to be added to adequately address your research needs? Is this primarily a feature for advanced users?

1. **Visual design, general UI, and performance issues**

Given that the Media Cloud team is currently at work on a backlog of bug fixes and performance issues, this analysis will not spend a great deal of time in this area (except to indicate when I’ve encountered what I think are bugs in the task flows in section V and Appendix A, for clarification’s sake). But task-blocking issues I experienced included:

• Long load times—sometimes the long wait alone was enough to make me give up on a task, but initially I thought these indicated bugs. Loading indicators could benefit from some messaging around what’s going on (e.g., “this will take some time,” “crawling 50,000 sources,” progress bar and estimate if possible, and/ or a link to click to be e-mailed when the results are ready, especially if the team is interested in increasing accessibility for low bandwidth users).

• The pulse graph appearing squished unless you size the browser to full screen
width.

• Not being able to easily browse and select sources in the flow of a query; sometimes this appeared to be due to bugs, but it was also just difficult to know which terms to search in the source finder dialog when trying to get a comprehensive and targeted list. I had a nagging worry that I was missing key sources I should be including.

* Getting logged in and out randomly and at awkward points; it seems to me that Media Cloud should almost never log me out, as privacy doesn’t seem like a huge concern with this application.

Something that wasn’t a task blocker, but worth mentioning as a stumbling point: many times I didn’t know output/ results were appearing below the browser fold (as on the Explorer initial query search landing).

Regarding visual design in general, refinement of text sizing, contrast, and general appearance for explanatory text might be valuable and will be kept in mind as improvements to navitorial and help text are suggested. For designs that haven’t been implemented, it’s suggested that the team be cautious around the use of red and green to delineate between distinct states and functions, as 8% of the population is color-blind and these colors are the most problematic for them.

1. **Task and learning explorations**

Appendix A contains an informal “diary-like” record of my journey of trying a research challenge with the tools. A few failure points were:

• Not being able to find some features, such as word trees.

• Not being able to figure out how to compare data with the type of media source as the highest delineator; for example left-wing vs. right-wing narratives, or U.S. sources vs. international sources, which, to me, would seem to be one of the most important strategies for research. I expected the tool to let me focus less on the individual search terms and more easily filter/ compare data by source collections. But I suspect the solution to this issue may lie in doing something not immediately intuitive and requiring a little extra work—such as setting up several queries with the same search terms but different source collections.

* Receiving an error on the number of seed stories required to create a topic, but no explanation on how to select the minimum 500 seed stories (adding more collections didn’t seem to help).

Below are key points on each tool.

1. **Explorer**
* The single search query box is well-placed and looks easy; it’s less confusing than Dashboard’s input UI. But to me, the expectation is that the search will be quick and will yield easy results, and now we’ve lost the implication that comparative searches are (often? usually?) what’s most useful. Also, the functionality to add a second query term doesn’t appear until you’ve already run a search on the first term, which is limiting. I recommend exploring other options to strike a balance between simplicity in getting started with search and providing more robust query functionality upfront.
* “Attention” as a category was not impressive or easy to understand at a glance as “people/ places” and “names/ organizations.” Sample stories were difficult to scan, and sometimes individual stories seemed irrelevant; I wasn’t sure what to do about this.
* “Attention Over Time”—do you only care about this if there is a sudden peak? Or can you tell me something about how much media care about this issue comparing total number of sentences to other topics overall?
* I didn’t understand Total Attention. I usually got a solid circle. When would this not be the case?
* The word clouds might benefit from some visual enhancement; regardless, it would make more sense to me to first define what they are (e.g., "other words used commonly in stories with ‘climate change’.”) Then the user would click to explore that data in a different way. Such as, show me only the stories that include "students”; is this the same as clicking the word and having it added to the query string with the Boolean “AND”? It took me a few times to even notice the query being modified in the UI this way, and to understand the subsequent word cloud.
* Could (and should) it be said that clicking word clouds can actually be used to build beginner queries?
* It occurred to me that I would like to compare a word cloud for each query (for comparative queries) side by side, and have the system highlight anything interesting.
* I wasn’t sure what “compare top words” the gray word cloud at the bottom, was.
1. **Topic Mapper**
* I wasn’t sure where to start on Topic Mapper page because there are various things that call to my attention—create topic, search by topic, view these topics that I’m not sure who saved and aren’t really categorized.
* A key initial question was what the difference between searching for a topic and creating a topic? Am I searching saved topics? If so, move search closer to topic list and say so. Otherwise, this makes me think I’m doing the same thing as Explorer, and that adds more confusion as to why Explorer/ Topic Mapper are different.
* Nothing came up when landing on Topic Mapper when I wasn’t logged in. (bug?)
* I recommend reworking the filtering (personal topics, saved topics, etc)—what do these really mean, which should be the default, how to navigate, etc.
* The system said I didn’t have permissions to view Starred Topics list (even though I was able to star). (bug?)
* On closer inspection I saw names of the people who have created topics, but I wasn’t sure why they’ve shared them with me. It would be great to see a summary of findings, at least for that point in time—why these topics were significant to save and share. Also, I’m not sure whether to start with someone else’s topic and modify it (will I overwrite?) or should I start fresh with my own?
* It was difficult to know how to get back to Topic Mapper landing page. UI says “go to Topic homepage” and I thought that’s what it meant but it took me somewhere else. I clicked the tab again, even though this was not intuitive, to arrive at Topic Mapper home, then saw starred topics (instead of personal topics) for some reason.
* At one point I wondered if the date choices were contextual to the source selected.
* In the results, I wasn’t sure at first why I was seeing the timeline; somehow it dominates the page and isn’t immediately intuitive—is it clickable? Is the assumption I’ll want to play with the dates this way to explore results? Am I looking at results or a slider? Then I saw that I needed to choose monthly or weekly first to make the timeline clickable. Showing and hiding timespans needs careful handling, should the user forget or not see what’s shown or hidden.
* I think this page needs better headers/ chunking, and an implied order logic for sections. This page also looks like it’s trying to have a responsive design, when I was told it’s not responsive. The visual “blocks” make some things larger/ more emphasized than should be, given their relative importance.
* The saved topic “2018 Italian Election” is categorized as personal. Why is it personal? Did the author save it for himself? I would assume that it would be private then. What is a “public topic” then vs. “personal”? I had a lot of confusion and trepidation around the privacy and sharing of topics. I wanted to know my topic was private by default unless I chose to share it.
* In some results, there was quite a bit of data not available–Top Themes, Top People, Top Organizations, Geographic Attention.
* Not sure about the Download Links and Word Maps—should they appear in separate window? I got a blank screen when clicking on the .gefx map link and then couldn’t get back to the topic. (bug?)
1. **Source Manager**

The main issues of providing a better summary of what’s included, what isn’t, and guidance on selection has already been covered. Additionally:

* I’d like to understand the rationale for categorization better, to see if we can improve it for the novice user without breaking anything for the expert. I’m not sure how helpful “featured collections” are. “Country collections” makes the most sense, but other categories might be helpful as well. “Custom collections” is vague, especially for the primary way of organizing sources.
* “ABYZ” and other collections jargon needs to be defined for novices.
* The search, auto-suggest, filtering and selecting of sources in the context of a query was problematic. It would be great to have a better way to browse all sources by region or other category, in-context help for choosing sources, and better auto-complete for finding sources by keyword (although this function may be altogether unimportant for novices). I was left with a nagging worry that I didn’t select all the sources I should have because I wasn’t given easy access to browsing the whole database easily, and I wasn’t given coaching or explanations on sources.
1. **High-level documentation issues**

Appendix B contains an informal “diary-like” record of my journey of learning to use Media Cloud through the available documentation; this included the Introduction to Media Cloud, Advocates Guide, Dashboard Guide, Topic Mapper Guide, and Query Building Guide. A few summary points of recommendations for improvement:

* As much as possible, I recommend we explore ways to provide more
in-context instruction (perhaps a demo mode or other types of online help). This should help give novice users the support they need to get started without searching for separate documentation.
* Ideally, the navitorial copy on each page should be reviewed and edited in tandem with any new instructional materials to ensure it makes sense in the context of new support, and is as clear and concise as possible.
* As previously mentioned, novice users could benefit from a bigger picture introduction to the tools, including an explanation of offline tasks (such as how to formulate a good research question) that helps educate them on appropriate uses of Media Cloud in news media research, as well as more guidance in interpreting results. The Advocates Guide is a good start but didn’t go far enough, in my opinion, on some strategic tasks and data analytics assessment techniques.
* Vagueness was a problem occasionally in the documentation language, much more so than industry jargon, although even words like “outlets” should be defined. (Example of vagueness: *“how issues are being talked about in the larger media landscape”—*are people actually talking? What is the “larger media landscape” and what does it not include?)
* Occasionally, verbs needed to be more active and specific (example: does a user really want to just “browse” sources?).
* More consistency is needed between the procedural steps, explanatory text and visuals. It’s much easier to stick to one example visually when explaining material, and the documentation sometimes switched use case visuals in the middle of an explanation.
1. **Further ideation**

While beyond the scope of this project, I wanted to mention some general thoughts on user experience and product strategy novice users of Media Cloud. The civic moment we’re in, where news media is more top of mind for more people than ever, seems to represent an opportunity to increase the awareness of Media Cloud to a much larger community, who will most certainly be primarily novices of media research. Could Media Cloud serve ordinary citizens who don’t have specific in-depth research needs, but do have concerns about media influence?

Media Cloud’s word clouds and other types of visualizations, perhaps visually enhanced and with other salient information integrated (like events), might provide a way to potentially engage all internet users on their own media browsing history, or on sudden peaks or anomalies detected in media topics. This functionality could be provided in the form of a browser plug-in (or app), perhaps, but the goal is quite similar to that of the data “moments” I suggested featured on landing—how can we both serve and get people more engaged with critical media issues, possibly including their own media reading habits? Getting a wider citizenry more deeply engaged with media influence would seem to have potential for a significant civic impact—but careful consideration and exploration of engaging new UX concepts would be essential.

**Appendix A – “diary” of working through research challenge**

**NATALIE'S CHALLENGE (in red):**

* Try to look at media coverage about climate change before and after Trump's comments [(some initial context from a preliminary google search)](http://www.independent.co.uk/news/world/americas/donald-trump-misunderstands-basic-facts-climate-change-piers-morgan-interview-a8181381.html)
* Previous findings have showed that right-wing media never used the term 'climate change', and rather focused on the environment - has this changed?
* See if there are any differences in language used before and after, then see if there are any stark differences in the right-leaning media sources before and after as compared with the left

Natalie helped formulate this research question for me based on my interests in environment and climate change. I actually think it's fairly challenging to formulate a meaningful question - and this particular one might be even more specific than I would try to formulate off the top of my head. That may or may not have been for a reason, but guidance on formulating a good question would be the first thing I want, once I decide I'm happy with the depth/ breadth the sources and decide to try the tool. Or—I would like to know what to expect if I’m just exploring a topic area, what to look for that might be different there. This second use case seems to not as easily lead to salient results, so I think recommendations should be clearly discussed as to whether or not a casually browsing of a topic is a good use of the tool.

What I did:

1 - searched outside of Media Cloud for the date of the event (Trump's comments in a particular interview)

2 - Found the date and checked the query date boxes in Dashboard (Explorer was down). (If we integrate coaching/ in-context help, we can coach to start here if you're researching impact of an event).

* **Previous findings have showed that right-wing media never used the term 'climate change', and rather focused on the environment - has this changed?**

3 - So then I add a second query box - for a comparative study - "climate change" (query 1) and environment (query 2). There was no in-context help to remind me what to do with two-word phrases, but I think I remember that they should be in quotes, though I'm not sure (if two words without quotes is equivalent to adding a boolean 'AND').

I struggled with narrowing media collections by partisanship. The "+" button function wasn't intuitive (I thought it would be an expand function). I immediately just tried the search, which didn't work (no results - bug?). I thought then I should click a blue link to add "liberal partisan sources" but doing that takes me to Source Manager where (after being forced to log in), I can see the collection but the "search on Dashboard" button does not add it to my current query, but takes me to a new query page. I recommend revisiting this entire flow and UI. There seems to be much available in Source Manager that needs to be available contextually to a query though more robust search, filter, and add functionality.



Finally I figured out that clicking the "+" button added all the collections to the query, and I can click the "X" circles to eliminate the ones I don't want. So I added "US Media" but removed "[U.S. Partisan Sources 2012 - Libertarian](https://sources.mediacloud.org/#collections/8878294)" from "climate change" and environment queries. I'm not sure if I should have removed the other ones as well... and why they aren't already part of US Top Online News.

 **• See if there are any differences in language used before and after, then see if there are any stark differences in the right-leaning media sources before and after as compared with the left**

The before and after did show a small peak in Pulse. Comparing the right-learning media sources with left was harder. Am I supposed to investigate this in sample stories themselves? Doesn't seem worth my time (especially as stories are not marked as to which collection they belong). I expected the results to show me findings within each collection, which would help me to make this comparison. Do I need to download the CSVs to do this? I try, but I get an error ({"error":"error 10 from memcached\_set\_multi: SERVER ERROR"}).

*Trying the challenge in Topic Mapper*

I'm not sure how to start here - create a new topic, search topics or filter the topics first to browse? So I search and there are a couple of climate change ones, but not as specific as I want for my question. So I create a new topic.

But I'm surprised that I'm creating into a public database (it won't let me create a topic with the name "climate change" as it's already taken). So I'm already a bit put off, because I don't really want to share what I'm doing with anyone. Why isn't there a private option?

"Climate change" OR environment I decide my query will be, as there are not two separate query boxes to do a comparison - I'm unsure if this is right. I'm also not able to select any sources (bug?)



Typing in “us” or “uni” yielded nothing. Finally typing in "united states" yields a long list - but is this really everything for the US, and how do I choose?



I add "Geographic Collections: United States - National" - So I search "partisan" and was able to get to the liberal and conservative collections - so I add those too (bug: the dropdown selection window narrowed for some reasons to half the width above so I couldn't read the collection titles well).

So I think I have the equivalent query now in Topic Mapper (and it occurs to me that mentioning you can do equivalent queries in both would be helpful).

The page shows me a preview and prompts me to make sure my topic looks right... I'm not quite sure how to do that. Again, the solid pie chart under seed stories seems strange, and the narrow Pulse graph just looks a little odd visually, although I can confirm a peak there on January 29 that is helpful.

Confirm page seemed unnecessary, and then I got an error.

But good news - after I did the above, Explorer was back up! So I start at the top left search box, which feels convenient and easy, and hope that " "climate change" OR environment " is the right search string. It takes me to a page where I could add another query, and I wonder why that was buried down a level. I agree with letting that be an option rather than a default (as in Dashboard), but I'd like to have the full query functionality up front. The fact that it's asking me to choose a color connotes that comparison querying (contrasting colors) still should be the default.

Adding collections still has the same number of problems for me. I'm not sure the difference between collections and sources, and expect to get a list I can browse when I click "All" vs "Geographic." There are load time issues, but the main point is I don't know where to start. I think using this window to browse and filter might be a better default, particularly if there is a logical taxonomy and some in-context help here for finding the right collections/ sources.



-----

"Partisan" yields a list that looks good— I'd like an "add all" button.

Something that I notice after I add my media and before I do the search is it's already seemed to have surfaced some results for me below, which I didn't really notice— was I supposed to use these to evaluate and refine? And again, the same questions there.

After the Search:

Internal Error (red bar) at top of page, but I still get results.

The Attention Over Time graph appeared really squished, and Cindy helped me get it to spread out by adjusting my browser window to full width of the screen.

The results themselves are similar to Dashboard, so the same strengths and weaknesses (questions), but People and Places seems like a category that's more interesting and helpful than Sample Stories—mostly because it is easier to read and answers the "who and where" questions in a simple manner.

Word Space takes some staring... interesting things in there although the contrast issues I think make the cloud overall pretty difficult to parse.

I like the other word cloud layout better, it's just more visually appealing... I would like to compare my word cloud queries side by side— and have the system point to anything interesting...??

But mostly, I still wonder why I can't compare by partisan collection categories— how does anyone figure out how the left vs. right talked about climate change after Trump's comments, except by going through the stories CSV?

**Appendix B – “diary” of working through user documentation**

Advocates Guide

I like that the Advocates Guide is short.

Media or news media? What jumps out at me in the first summary: “across the globe,” “50,000 resources,” “big data.”

*“Tracking media coverage is simple -- understanding the underlying narrative is far more complex. Simple alerts monitor how often an article is shared, liked, or retweeted; whether it’s trending on Twitter or has its own hashtag. These metrics can help confirm how widely a message is being disseminated.”*

Are you saying Media Cloud provides these alerts for the user to set? Where?

*“How does my messaging influence today’s zeitgeist? How does my message resonate within*

*the newscycle? Can I frame my argument in a way that is more likely to resonate and influence people?”*

All sound useful.

But the following sections seem repetitive and vague as to exactly what Media Cloud does.

*“Dashboard section of the website allows users to quickly and easily identify both the news sources and specific stories that are primary influencers within the media and among readers for a partiuclar topic.”*

That sounds pretty general, at least at first glance.

*“Keyword search yield interesting findings in frequency, prevalence, and context”*

Are these key concepts for identifying salience in the outputs?

Pulse—number of sentences, peaks and valleys—says it’s not nuanced enough to draw conclusions. Should the inference be that I should always use this in conjunction with other parts of the tool?

*“Word clouds depict terms most frequently used”*

So is the inference that the combination of sentences *with* terms yields a special insight? More specifics on what to look for here?

*“These queries can also be specified for certain outlets (or types of outlets so that you can understand the narrative being used by particular sources).”*

What are “outlets”?

*“For example, as seen in Figure 2 below, Media Cloud can create word clouds that compare word uses from both conservative and liberal media outlets for the user’s chosen keyword.”*

These media self-describe as conservative or liberal?

Word trees—order of words, common words preceding and following—more information as to how frequent terms are being used (context).

*“they also provide insight into how specific media outlets cover stories differently.”*

How? Do I see a word tree per outlet?

*“This is particularly helpful for understanding how readers of different media sources are consuming information on the same topic and why different audiences might have fundamentally different interpretations of the same story or data.”*

I’m not understanding how one word tree would show that. Need more details.

*“Source Networks*

*Graphics such as those shown above put into perspective how stories are shaped by*

*reporters and the specific language they use.”*

I understand this point, but not why this section is labeled source networks. Also, I’m thinking… is this just written word (no TV transcripting?)

*“Media Cloud’s suite of tools also allows the user to learn more about how these issues are being talked about in the larger media landscape or the framework for their coverage.”*

Another vague statement.

*“One of Media Cloud’s more informative and innovative tools allows users to identify a network of sources. This tool can be found in Topic Mapper.”*

“Identify a network of sources” sound a bit vague and unexciting, rather than innovative.

*“Media Cloud uses a source networking tool to identify and depict media ecosystems for a keyword based on two criteria: 1) influence and 2) language use.”*

“Source networking tool” “identify and depict” and “media ecosystems” all vague.

*“Influence is determined by factors including how many articles in the database link to the story*

*and how frequently it is shared through social media (such as through Facebook), and tracked through other various digital analytics and social networking tools.”*

Would be nice to see this in bullet points. “determined by” vs. “tracked through” – difference?

*“Users can use the Media Cloud website to depict this information in chart form (see Figure 4). Because of the complexity of this search tool, it might take a day or two to receive the results.”*

OK, now I realize that this image corresponds with “Source Networks” as a section—might be helpful to show it first or closer to the title. Does it really take a day or two to receive results sometimes? What might cause this to be so?

*“Users can also collaborate with Media Cloud to produce certain types of information in a graphical form. For example, Figure 5 below depicts the news sources that are most influential in terms of weblinks at a particular point in time.”*

Does “collaborate with Media Cloud” mean the team is going to consult with me? Is there another way to get the Gephi visualizations so I don’t have to hire you? Also “in terms of weblinks at a particular point in time” needs more explanation too. As I read on, I see this is a snapshot of a moment in time, and may be compared with various other points in time (week, month, year in), but would be good to know why that would be useful, when that methodology should be employed (to answer which kinds of questions). Wouldn’t these vary considerably based on any particular day? Do they sort of aggregate weblinks for a period up to the point the query is run, or do you define the period of time as you run the query?

*“As Figure 5 also illustrates, the map’s use of colors can help highlight “subcommunities” of content producers covering a particular topic. In this graph, the color green illustrates research and evidentiary-focused sources, purple covers political and mainstream news outlets, and orange is policy-focused sources.”*

Bullets would help; callouts in the screenshot of the graph too.

*“Advocates can use graphs like these to maximize opportunities to influence the narrative by identifying the media sources that are most commonly linked by other sources and therefore central to the development of a particular narrative.”*

Seems repetitive?

*“These maps are most helpful for following stories that are unfolding online where information is shared by linking to other, key stories.”*

I don’t understand how that’s different from any other output of Media Cloud.

*“Media Cloud can also work with a user to prepare a similar graph that depicts how prominent sources communicate a specific narrative by clustering sources that share similar language together. The use of similar word choices indicates these media sources are using a shared narrative and framework for discussing a specific issue. The more words two sources have in common, the more tightly they will*

*cluster together, pushing away sources that use different language.”*

OK, feels like a 3D map – axis of sources and words, etc… would like to see an example of this as well, and common conclusions drawn (e.g. “Left-wing media had more of a common narrative with evidentiary-focused resources when discussing climate change, than did right-wing media”)

Dashboard guide

The query building coaching... are you considering this an advanced task? Need to see examples to reiterate principles, and then have a practice/ assessment.

Source selection didn’t work for me—I think this is an area that needs more focus, as there is some complexity in selecting sources—how do I know if I’m selecting the right ones, the full list of what’s available, why doesn’t it suggest if I just want “Europe” etc, organize by region, how else? But overall, it’s just buggy right now, or, at least, I couldn’t figure out the flow of getting “US top media” out and other stuff in.

*“Clicking on a point in the graph will generate a word cloud with the terms used in the sentences matching your query on that particular period.*”

Period? Point in a graph is a single day, so what is the period?

*“By clicking on any of the words in the cloud you can generate a new search which includes the original search plus the word you clicked on.*

With “AND”?

Would be good to know what downloaded file data is good for—even if only advanced tasks, would just be good to have an idea.

Sentences or stories? I think stories but it would be interesting to learn a bit more about why the change was made, how it changes the interpretation of results…. Entire story allows for a wider net, any buts…??

*“For searches of our US Mainstream Media and Europe Media Monitor collections dated in 2015 or later, Dashboard will also provide an attention map with information on the countries more frequently mentioned in your query results--based on a sample of sentences. The more intense the color, the more frequently that country is mentioned. By clicking on a country you can narrow your search to articles that mention places in that country. You can also download the underlying data in CSV format by clicking on the menu in the upper right corner of the map.”*

Confirm this is old info… sounds like stories about/ CLIFF-CLAVIN engine works with all sources and really is about the countries rather than just countries mentioned…

*“Highlight an area by clicking and dragging on the chart to re-run your search, limited to that span of time, as done in the following example looking at how terror\* and refugee\* are spoken about together:”*

Does this still work? Also, in examples… best to stick with one example until you’ve indicated a change has taken place through a procedure you’re demonstrating.

*“Click and point on the graph to see a word cloud for that query during that day, week, or month:”*

Again, how do I know I’ll get day, week or month?

*“Compare topics*

*The graph allows you to create multiple queries and easily observe the points in time where keyword mentions peaked and declined. By clicking on any of the points on the graph, it also lets you visualize the comparison between keywords in two or more media collections, or comparisons between two or more keywords in the same media collection.* ***Looking at this before, during, and after the peaks is especially helpful.”***

…“because a peak indicates a surge in your topic, and surges are often accompanied by changes in the narrative afterwards. So – look to see if there is a change after the peak. Here’s more about how to do that…”

*“Compare topics across countries or different media ecosystems*

*It is possible to take a quick glance at how a certain topic is discussed between media collections e.g. how migrants are featured in the mass media in Mexico* (trabajo / work, frontera / border, derechos / rights, indocumentados / undocumented, refugiados / refugees) compared to the United States (illegal, undocumented, refugees, muslim, ban):”

So am I to assume you use different language terms when comparing across countries?
I wouldn’t have thought of that... oh no, the term “migrant” only was used—was it translated?

*“Compare topics across types of media sources*

*You can also compare across media types. The use of different words to refer to the same topic can expose the tendencies or biases of each media discourse regarding actors or processes e.g. comparing results of the search term “USAID” in the separate ‘conservative’ and ‘liberal’ collections in US coverage might give you the following high level insights (however, note that collections are built with specific analysts and problems in mind - “conservative” may not mean the same thing to all users, and media outlets coverage may change over time).”*

OK, but these “types” are actually just different sources. So maybe filtering sources by “type” or some more specific categories would be helpful.

*“Click a query label at the bottom of the chart in Pulse to hide it. That can sometimes make it easier to select the exact query and word cloud you wish to explore while comparing different ones. For example, "Shuar US" results have been hidden here:”*

Sounds like this is just a visual cleanup function.

*“Look at topics in various countries and languages”*

I thought we already did this?

*“Find news published in different regions of the world. Editors and authors can take advantage of the Media Cloud to identify relevant sources for their reporting e.g. in a news organization if an editor wants to research and write a story about the parties that could form a governing coalition in the Netherlands, her or she may find that the GreenLeft party (GroenLinks) is a key part of coverage. The writer can then conduct a Dashboard search to explore what the party's positions have been in the past:”*

Hmm… identify relevant sources for reporting new stories?

Parties that could benefit from a a governing coalition

GroenLinks – is this a key party? (What do you mean by “key part of that message”?)

“Search what party’s positions have been in the past”—this seems like a big task in terms of figuring out key words to search.

*“Find breaking stories or stories from which controversies began*

*Dashboard can help you to identify the first article that talked about a topic within a period of time, in this case for the ‘Project for Civil Society Macedonia’ it was in "Veterans Today" in April 2016: “*

OK, that’s easy enough—maybe just add a callout to show that here’s the first story on this query search... as long as you’re confident enough that this is the only search phrase you need to use.

*“Track how a topic evolved in the media*

*By looking at the various stories across time and their concurrent word clouds, one can track coverage of an issue like the Dakota Access Pipeline or “Standing Rock” in North Dakota (image includes annotations):”*



OK – I think these callouts represent a user who did research on key events that correspond to peaks? **It seems to me this is powerful and should be explained.** Because otherwise you think you need to do something in Media Cloud to give you these callouts/ events.

*“Observe the international coverage of a source of the geographic focus of a specific topic*

*You can track the geographic coverage of a particular news source or collection e.g. is US or UK news talking more about the Syrian crisis? Does the LATimes focus on covering Russia as much as the New York Times?”*

Hmm… but over time or at any one particular time? They probably peak and valley separately, maybe it’s useful to see how they differ and are similar with events as a backdrop.

Topic Mapper Guide

This was a bit buried, but helpful:

*The main difference between a query in Dashboard and a topic in Topic Mapper is that Dashboard will only look for stories in our database and will produce a smaller number of outputs for analysis (see our Dashboard guide). Topic Mapper will launch a crawler to follow the links in the stories it finds in our database (that’s why it takes some time to generate a topic) and will thus:*

OK, this is the best explanation I’ve seen yet. So Explorer searches stories from the collections, but Topic Mapper searches those plus crawls links in those stories out to those stories… but is there any risk that you’re “over-reaching” then in terms of a relevant search?

Intro to Media Cloud

*“Media Cloud is a web-based, open source platform that tracks media conversations across the globe.*

*The platform aggregates news content from over 25,000 digital sources from over 200 countries in multiple languages, providing analysis on media attention and influence.”*

Media attention vs. influence—influence isn’t something talked about a lot yet, but that would also be an interesting category to break down in the instruction.

*“Use Dashboard to see immediate results on how a specific topic is spoken about in digital media*

*Investigate a specific media topic in-depth, seeing the most influential sources, stories, and how language is used.”*

“how a topic is spoken about” is vague and a casual expression… it has no specific meaning, and doesn’t help me to distinguish from “how language is used” in Topic Mapper.

*“Browse the media sources and collections in our database, and suggest more to add.”*

I want to do more than browse, especially at first— I’d like a summary of what’s included and what isn’t. Then I’d like help in selecting the right sources for my research question.

*“Topic Mapper is an extension of Dashboard”*

Hmm….

*“TYPES OF ANALYSIS*

*News cycle attention Influential stories*

*Key events*

*Influential sources*

*Frequently used words*

*Network analysis of inlinks/*

*citations*

*Frequently used phrases*

*Language framing*

*Social media data*

*Topic classification”*

It’s confusing that you can visually match some of these to the interface but not others.

*“The system currently contains digital sources from over 200 countries around the world, and collects content via RSS feeds from these on a daily basis.”*

OK, so RSS is what how sources are monitored, spidering happens in Topic Mapper only, expanding on the RSS data.

*“The Media Cloud definition of a “news” source is wide, including both traditional newspapers and innovative formats such as comedy. The system contains only textcontent, including TV and radio news feeds via transcripts.”*

OK, I was wondering about the transcripts— I think this is a big point and should be made more prominent (also knowing the percentage of tv/ radio/ podcast (?) transcripts)

*Key takeaways:*

*The right wing media ecosystem altered the broader media agenda:*

*- Immigration is the key topic around which Trump and Breitbart found common cause; just as Trump made this a focal point for his campaign, Breitbart devoted disproportionate attention to the topic and*

*was widely shared on social media*

*- Mainstream media, though critical, focused on immigration as well*

*- Though framing of immigration differed, media organizations and activists can use this information to critically reflect on how agendas were set, and develop strategic plans for covering future elections*

Consistency - These conclusions don’t match the graph and source cloud… no mention of immigration.

*“Framing differs depending on the source of information:*

*- Traditional newspapers focus on the conflict between religious authorities advocating traditional contraceptive methods such as abstinence, and public health approaches of using condoms or IUDs*

*- Online, youth-oriented sources focused on celebrities and pop culture supporting the use of modern contraceptives*

*- Health organizations and non-profits working in Nigeria can use these insights to find partners, design campaigns and build advocacy efforts”*

Consistency - The images here seem to indicate that articles were examined separately rather than relying on the visualizations. It would just be nice to know… is there a “traditional” vs. “youth-oriented” label, or did the user know in looking at the sources?

*“Research Question: Who were key information authorities for coverage of the Ebola epidemic, and did that impact the global health response?”*

Consistency — again, the conclusions here are not visually illustrated.

*“Re-Entry Post-Incarceration Documentary*

*Research Question: Has this documentary, and accompanying public outreach, changed coverage of re-entry after incarceration in digital media conversations?”*

Is developing new collections often part of the process? When do you know when this has to be done?

Media Cloud Task Analysis

1. **Formulate your research question(s)**
2. Identify topic
	* Brainstorm subtopics
3. Explore investigative possibilities:
	* Identify dates for key events linked to your topic
	* Identify key people or organizations linked to your topic
	* Identify key media sources (news and social) you might be interested in
	* Identify helpful comparisons
		+ By time
		+ By events
		+ By country
		+ By subtopics or themes
			- Theme types in Media Cloud:
		+ By media type
		+ By readers of specific online publications
		+ By readers of partisan publications
4. Think about attention, language, and influence
	* how *much* the media is discussing your topic
	* how they are discussing it
	* who is most influential in those conversations (both in terms of sources and subjects)
5. Optional: determine your hypothesis
6. Write your question(s)
	* Sample formats
7. **Learn about Sources**

1) Plan ahead by determining which sources are right for your question

a. Major collections

b. Regional collections

c. Languages

d. Partisanship

e. Special collections

f. Individual sources

1. **Explore and refine your initial results with Explorer**
2. Formulate your query or queries
	1. Enter boolean search
	2. Choose date range
	3. Choose sources
	4. Create optional comparative queries to answer comparative questions
3. Run queries
4. Analyze and refine the queries
	1. Attention (how much the media is discussing your topic)
		* 1. Attention Over Time

• Option: if you don’t get peaks, try refining your query by changing date ranges or sources

• Option: map key events against the peak graphs

• For non-event peaks, explore what else could be driving that
coverage—both the rise of the peak, and the fall-off. To answer this, browse stories to see if you can find clues in the ways the narrative was shaping (see step 4)), other events that weren’t an obvious focus for your topic.

• Option: map comparative peak graphs over one another

• Option: click the peaks to see sample stories for each (below the peak graph)

* + - 1. Total Attention: if you’ve run multiple queries to compare, you can compare Total Attention (total number of stories) for each query
			2. Themes: Look for unexpected themes indicated.
			3. Sample Stories: use these for optional step 4 below
	1. Language (how the media is discussing your topic)
		+ 1. Top Words: look for unexpected terms or terms that might reveal more clues to help answer your question in the word clouds
			2. Word Space: look for any distinct separations between subtopics
			3. Optional: click a word in the word cloud identified in #1 to rerun the query
			4. Optional: add a word from Word Space to your query and rerun
	2. People and Places (the people, organization and places being discussed the most in the stories – part of measuring influence)
		+ 1. Top People: browse the figures discussed the most
			2. Top Organizations: browse the organizations discussed the most
			3. Geographic coverage: see what regions of the world are getting attention
1. Optional: browse stories for additional terms to add to your query
	1. Find these via Sample Stories, in the Attention tab
		* 1. Download for more complete review, more sorting options
			2. Identify the top stories and look for patterns?? (Can you do this in Explorer?) Examples: regions, frames: contentious frames are often the media’s subtopics
			3. Identify most common data sources media points to

a. Optional: review related datasets from trusted sources

* + 1. Example: UN, CDC, ?? (list of common other sources to review)
	1. Add common terms discovered in stories to refine your queries
1. Optional: use a search engine (like Google) to search identical queries (with restricted date ranges) and compare for gaps or further indications of influence from search engine algorithm rankings
2. Write out the answer to your research question, and cite visualizations or media references to support it.

**D) Dig deeper to validate your answer using Topics**

- deeper assessment of influence of specific media outlets through rankings number of stories from specific sources, social media shares and inlinks

 - what are the big players saying? Stories, language?

 - compare media attention peak graph to overall peak graph – when did they start paying attention to issue?

- compare inlinks to shares (shares make things trend and drop off quickly – inlinks tend to remain influential longer)

- explore results of referenced non-news sources and digital-only sources

24:50

choosing terms: make sure you delineate from other events, topics, etc – get specific as you can with your query

CSV – had to label day one, day two, etc – created own visualization in Excel

Guide:

* + - * + Explorer only (overview or transition to Topic Mapper?) – they need to practice, and this is the best tool for it (unless practice with other people’s topics) (30 hours)
				+ 1 draft, 1 refinement
				+ graphical?
				+ case study examples
				+ can I get my hands on Word copies of other guides?
				+ Usability test protocol (4 hours)
				+ Participants recruitment, schedule + test (10 hours)
				+ Have them read aloud and react to intro material, then perform key tasks.
				+ Need to identify these
				+ Report – 3-5 pages – will include: (12 hours)
				+ Brief summary of protocol
				+ List of participants
				+ Goal of the study
				+ Tasks, with a “grade” and any associated recommendations for improvement in documentation and/ or UI (UI for later)
				+ Refined guide (15 hours)