

**From:** Bowman, Randal  
**To:** [Shulman, Stu](#)  
**Cc:** [Marcia Cash](#)  
**Subject:** Re: This round of comments sorts much like last one  
**Date:** Thursday, June 15, 2017 4:59:44 PM

---

Checking in the morning is fine. I searched for terms in quotation marks to get the exact term, and the words used are the title of the bucket, and, without the quotation marks, of the datasets. There were a couple of cases where the quotation marks didn't transfer as the title of the bucket, but I used them in all cases. The "n" and numbers reflect near-duplicate items combined, and you know about the duplicates.

With respect to Jennifer's issues, I'm not sure I understand them, but it seems that at a minimum, if Adobe is not built into DiscoverText, coders would need instructions on how to switch back and forth between it and their Adobe software so as to not have major delays.

If the issue is something we are doing in uploading written comments received here, we can possibly change how we do that. For example, I've told them to stop uploading post cards and to instead put them into "for" and "against" boxes, and then measure the cards when the comment period closes. They have already determined how many post cards in an inch, or something similar, so that is actually faster than coding them.

On Thu, Jun 15, 2017 at 5:46 PM, Shulman, Stu <[stu@texifter.com](mailto:stu@texifter.com)> wrote:

They should be included, but I would like to look at where things stand in the morning and review your handy work.

I have also been talking with Jennifer today. I think she was hoping the OCR would handle hand-writing; it will not. I reiterated my view that PDFs should be read using Adobe software.

Stu

On Thu, Jun 15, 2017 at 4:54 PM, Bowman, Randal <[randal\\_bowman@ios.doi.gov](mailto:randal_bowman@ios.doi.gov)> wrote:

Between combining duplicate and near-duplicate files, and using some of my search terms from last time, the list of uncoded items, using the "items uncoded" filter, is less than 8,000.

Search and Browse Project De-Duplicated  
**Through June 13** ([Project De-Duplicated](#))

Showing 1 to 1,000 of 7,922 total **Filters applied**

There do not seem to be any particular patterns to the remaining comments. I believe the coding team should be able to handle these in a week.

Question: are the comments in the smaller duplicate and near-duplicate groups included in the 7,922 uncoded items, or are they segregated and in need of being coded by group or combination of groups? If so, I can do that in the morning while the coders start work on the single comments.

--

Dr. Stuart W. Shulman  
Founder and CEO, Texifter  
Cell: 413-992-8513

LinkedIn: <http://www.linkedin.com/in/stuartwshulman>