

Transcript

ERIC Semi-Annual Update
June 2016

Slide 1

So now to begin with the annual update. As I mentioned earlier, I am Erin Pollard, the Project Officer for ERIC in the U.S. Department of Education. I'd like to welcome you to our semi-annual ERIC update webinar. This is going to be a longer version of the presentation that I gave at the American Library Association and Special Libraries Association conferences earlier this month.

Slide 2

Today I am going to begin with an overview of who ERIC serves, the size of our collection, and give some general facts and figures. Then I am going to talk about the accomplishments that we have made in the past year and some planned enhancements for the future. At the end I will take questions. If you have a question, as I mentioned before, please use the chat box on the side. I will either work the answer into the presentation or address it at the end.

Slide 3

ERIC is a collection of over 1.6 million education resources that has been building since 1966. We have journal articles, research reports, fact sheets, conference papers, and books in our collection.

Slide 4

In the past year, we have had over 12 million people from 242 countries use ERIC. Although we are a U.S. based resource, over half of our users are international. We have five main audiences.

Academics

The first are academics – these are academic librarians, professors, and students. I am going to be overgeneralizing for all these groups, but academics tend to have access to subscription databases and are looking for peer-reviewed journal articles. Full text isn't as important because generally these users can get the full text through their library subscriptions.

Researchers

The next group are researchers. They are looking for the latest information, preferably peer reviewed, but it is generally less important if the document is a journal article, conference paper, or report. Like academics, these users typically have access to subscription databases.

Educators

When we think of educators, we think of K-12 teachers, principals, and curriculum specialists. They often have little to no access to full text, so they are relying on ERIC to provide it. They are also less interested in journal articles with technical language and long econometric equations, and more interested in user-friendly, high-quality materials.

Policymakers

Policymakers, such as those in state or federal departments of education or legislatures, are also looking for high-quality, user-friendly materials. They need to make decisions quickly and rely on ERIC for current information from grantees and researchers across the country.

General Public

Lastly we have the general public. These are generally parents looking for information. They want full text and generally aren't looking for complicated journal articles.

So when we work on any new project or initiative for ERIC, we keep these audiences in mind to try to balance the needs of everyone.

Slide 5

ERIC gets its content from over 1,800 different publishers. Our providers range from big journal publishers, like Taylor & Francis and Wiley, to small publishers that may produce a single journal article as a part-time function. We also have a ton of non-journal publishers such as education associations, universities, research centers, and federal and state government agencies.

We added 87 new sources in 2015. These sources include 59 new journals and 28 new publishers of reports, conference papers, and other documents. In addition, last year IES grantees contributed 80 peer-reviewed full-text reports as a result of the IES Public Access Policy. These are what you would commonly call pre-pub, or pre-published versions, like users would submit in their institutional repository.

Slide 6

Now I'll describe ERIC's accomplishments over the past year.

Slide 7

So far this year, we have processed 23,324 new records. Eighty-three percent of these records are for peer-reviewed content and about 32 percent have full text available for download. So of these, 26 percent are peer-reviewed full-text documents. We've been working to increase the blue section of the pie – the work that is peer reviewed and has full text.

Slide 8

The way that we authorize this new content is through our selection policy. In January we released a new version of the selection policy that had one major change.

Slide 9

When ERIC first introduced a peer-review flag, the only documents that were eligible for inclusion were journal articles and work published by the Institute of Education Sciences, the parent organization of ERIC. However, this was causing confusion because some documents that were clearly peer reviewed did not have the peer-review flag. What we did was completely rewrite this section of the policy to clarify how we defined peer review and extend the option for all work to get this designation, as long as it was peer reviewed.

Slide 10

So now, when you select the “peer reviewed only” check box, both journal articles and non-journal entries will pop up. This leads into a frequently asked question – “Why don’t we have a checkbox for just peer-reviewed journal articles?” Right now users can find just peer-reviewed journals through two clicks – the checkmark at the top and then filtering by document type. We feel like this best meets the needs of multiple user groups. Students can filter their results if they have a specific requirements for their assignments, but researchers and policymakers can easily see the results from all of our peer-reviewed literature.

Further, there are complications with this based off of the IES Public Access Policy. Many of our documents are final peer-reviewed manuscripts, so they are not journal articles yet, they are the pre-published versions. We note that there is debate over whether this counts as journals or does not count as journals, and this is up to individual professors and their assignments. So because this is murky, we feel that two clicks is the best option, and I’d be happy to talk about that more in the Q&A.

Slide 11

In addition to the major selection policy update, we also made a major Thesaurus update. The Thesaurus is the backbone of ERIC. It is our controlled vocabulary and how we index our content. We updated it this past October to capture changes in the terminology, outdated words, and new concepts.

Slide 12

When we update the Thesaurus, we work from a pool of candidate terms that have been suggested by ERIC users and indexers. As we review these terms, we ask:

- Is there literary warrant? Literary warrant means that the term appears in ERIC content. Terms with a high level of literary warrant – 1000 hits or more – are considered to be priorities.
- Is there a need? If we’ve received multiple requests for the same addition or change, we consider those requests to be a priority.

We also take into account if the change promotes indexing consistency, reflects the language of our users, and brings our Thesaurus into compliance with NISO standards.

Slide 13

This process took almost a year and in the end we made a handful of updates. We added 19 new descriptors, like Common Core State Standards and Low Income Students. We also added 28 new synonyms, and made 17 changes to existing terms, including the descriptor change from “mental retardation” to “intellectual disability.” These changes are done to help users find what they are looking for by matching the language that our users are currently using.

Slide 14

Another project that we are working on involves improvements to the ERIC identifiers field. The identifiers field is a historic field that contains information about a wide variety of topics. Currently, all identifier information is in one field that we made searchable in 2015. The problem with the field is that it is a hodgepodge of information that makes it very challenging for users to find what they are looking for.

Our solution is to break this one field into separate fields. This will allow more precise searching and make it possible to filter using these fields instead of just search. It will also add consistency to the identifiers by making them a controlled vocabulary. We think these searchable fields will be a huge benefit for researchers.

Slide 15

So the first change we are making is to geographic location. This field will contain the country name for non-U.S. countries, cities for the top 100 most populous U.S. cities, and state for smaller cities.

Slide 16

For example, we previously had “Alabama (Southwest)” as an identifier. This will be mapped to Alabama and show up with all other Alabama fields. The problem with Alabama (Southwest) is that this level of specification was not consistently applied across all records, and southwest Alabama is fairly subjective. Plus, we don’t think researchers would think of using this term when they’re searching our database.

Slide 17

Similarly, we had “Virginia (Fairfax County).” Now this will be mapped to Virginia because we did not consistently fill in the county for all of our records. This way if you are looking for work done in Fairfax County, you can use the Virginia location code.

Slide 18

Next we created the tests and measures field. In the legacy field there was a character limit that caused some creative truncations. This was problematic because it is very hard to search for

something when it has been truncated in several different ways. We also removed the author's name from each test because this wasn't done consistently. We wanted to make the information in each field as useful as possible to our end users.

Slide 19

So historically we had abbreviated the National Longitudinal Transition Study of Special Education Students using the truncation highlighted. This is a truncation I had never seen before and I don't think researchers would think to use. By spelling out the whole name, it will be much more useful to users.

Slide 20

We also removed parenthetical author names to increase accuracy and consistency of our records.

Slide 21

Lastly, we consolidated test names. This is just one example of a single test with 15 different variations. Users would enter one variation of the name and find results, so they wouldn't think that they would need to search the other 14 different ways. Now we will have one single term.

Slide 22

Next is laws and legislation. Like for the other fields, we updated truncations and standardized names.

Slide 23

For example, there is some unique truncating going on for the Perkins Act. We have standardized this name to the full official name of the Act. That way search terms like "Applied Technology" and "Vocational" will return results.

Slide 24

The next big thing we have worked on is to fully integrate ERIC with a newly redesigned IES website by adding hyperlinks to useful resources.

Slide 25

First, we are adding a link to the IES website for all IES products. For example, this is a guide from the What Works Clearinghouse. And if you click the link to the guide, you will find an e-reader version, a study summary, and videos from the authors. Many IES products have companion pieces, so adding these links will help connect users to them.

Slide 26

Second, we are going to add links to funding information for work funded by IES grants. This will enable users to see what their tax payer dollars are being spent to fund and how a particular article feeds into a larger body of work.

Slide 27

Third, for studies that meet the What Works Clearinghouse standards, we will be linking to that study's page on the new What Works Clearinghouse website that will be launched this fall. The Clearinghouse reviews studies that examine causal impacts to determine if the methods that they used are rigorous enough to support their claim. So if a study said that a certain curriculum causes students to learn 10% more than what they were doing before, the Clearinghouse would look at the methods behind that study to see if they were sound enough to support that claim. The problem is, a lot of these studies have pages of economic equations, literally Greek, that is not user friendly to users who have not taken graduate-level statistics classes. The Clearinghouse breaks down all of the information in the study, such as the setting, who was studied, how the study was conducted, and the findings, and puts it into a nice, user-friendly website. This will be tremendously helpful to students and I cannot wait until it is live. We think that this will be incredibly, incredibly useful.

Slide 28

The last link we are adding is related to author identification information, specifically ORCID, SciENcv, NIH eRA Commons, and NSF FastLane. These Author IDs take users to third-party websites to find additional information about the author, such as background and other publications they've written.

This solves what I call the "John Young" problem. There are thousands of John Youngs, so if you are looking for work by a particular John Young, having a specific author code really helps. And it really helps you figure out what other pieces the author has written if you're trying to figure out their body of work.

Slide 29

In addition to the new hyperlinks, we are also continuing our work to restore the PDFs that were taken offline because they are not readable. These are the documents marked as "PDF Pending Restoration" in the collection.

Background

We are converting these documents to make them searchable PDFs. However, this is not a quick nor straightforward activity given the quality of some of the documents we are trying to restore. To understand some of the challenges it is helpful to remember that when ERIC was created 50 years ago, it was primarily a microfiche-based system. When ERIC went online, the microfiche was converted to PDFs to make it easy for users to access documents. However, the technology for PDFs 10+ years ago was different than it is today and the PDFs were made as images. This meant that they were not machine readable, which enables searching and is necessary for individuals with disabilities.

Optical Character Recognition (OCR)

To make the documents searchable, ERIC used Optical Character Recognition (called OCR) software to scan the PDFs. This process takes the existing image and then assigns letters, numbers, or punctuation to create a readable text layer that a user can search, copy, and paste. The text created by the OCR process is known as alternate text.

A large number of ERIC's historic documents were typed on a typewriter, converted to microfiche, and then converted to PDF. As long-time ERIC users may have seen, these documents were pretty hard to read with a human eye. Machines (via the OCR process) could not do much better.

Candidates for Restoration

Some of our documents are so faint that they are not readable to the human eye nor via the OCR process, and therefore are not a focus of the current effort to restore unreadable PDFs. Other documents that are in better shape still may have words misspelled, truncated, or missing letters. They require manual editing to correct the text and punctuation to match the image. Making these manual edits can be challenging. We don't want to infer what the author is saying or inadvertently change the alternate text associated with the document.

Long Timeline

Given the challenges that exist in these records, restoring the PDFs to be readable is a time-consuming process. We are combining the use of automated tools and manual edits to make restoration of these previously unreadable documents as efficient and timely as possible. These efforts are part of an effort to be performed over a number of years to restore as many of these unreadable documents as possible.

Slide 30

We have made a huge push to increase our communications activities to a wide variety of users over the past year. You can find all of this work in the multimedia link in the ERIC website. So this is changing gears a little bit, talking about our new outreach activities.

Slide 31

ERIC just released three new videos this spring, including one that describes how ERIC selects new sources for the collection. We've received lots of questions about how this process is done. While all of this information is in our selection policy, this is a quick 3 minute summary of our process. We use videos like these because we find that many of our publishers and our users are not reading the selection policy, so to be transparent, we put out here in a video format.

Slide 32

The second video is one I am incredibly excited about. It is a step-by-step guide on how to use ERIC to write a research paper. This is a video that you absolutely want to watch, put in your

libguides, and share with your students. In fact, several professors who we showed this video to have put it in their syllabi for freshmen students.

We developed this video with the help of professors who regularly teach undergraduates and see where they struggle in writing papers. I highly recommend you checking this video out and telling your students to watch it.

Slide 33

The last video we released is called “About ERIC.” It is essentially a beginner’s guide to ERIC. It provides information on ERIC’s mission and sponsorship, what’s in the collection, who uses ERIC, and provides links to other resources for more information.

Slide34

We will be continuing to build our video library by adding a number of new videos throughout the coming year. These videos are intended to address the varying needs of our audiences, but especially the needs of ERIC’s novice users. Several of the videos will explain how to perform a specific task. We are also considering three other short FAQ videos to respond to frequently asked questions to the ERIC Help Desk. If you have ideas for videos, please either put it in the comments box of the chat or send it to EricRequests@ed.gov. We would love to know what you think would be helpful.

Slide 35

We will also be creating more infographics that explain concepts or convey metrics using creative designs. ERIC has released two infographics so far: our year-in-review for 2015 which depicted the size of the ERIC user base, content statistics, and new products that were developed during the year; and “How a Record is Added to ERIC” which shows the flow of new content into the database. You can see both in the FAQ area of the ERIC website. We plan to provide many more of these infographics throughout the year. And again, if you have suggestions or ideas, please send us a note.

Slide 36

We also plan to host more webinars like these to keep you up-to-date with ERIC. At least one of these webinars will describe and clarify a key area of confusion for ERIC novice users, but we are still working on topics. If you have suggestions of ways that you think we could provide a helpful webinar, please let us know. And we plan to host these in the fall, winter, and next spring. Additionally, if you think this is useful, we will do ERIC semi-annual updates in the January timeframe and June timeframe.

Slide 37

Now we want to open up to any questions that you have on ERIC. Please use the chat box to ask All Panelists any questions. While you are waiting for us to answer, please follow us on Twitter and like us on Facebook. Additionally, if you think of a question after this presentation,

or think your question would best be answered offline, please email us at ERICRequests@ed.gov. We can answer your questions that way as well.

Questions and Answers

Question 1:

Thanks for your video! I note that they are over 40 minutes long, and I think our students would appreciate being able to access the specific information needed. Would it be possible for you to release timestamps for the different sections of each video?

Answer 1:

I was hoping you did not see that. What was happening was we made these slides before the video was live. These videos are not 40 minutes long – no one would watch them. About ERIC is, I think, 3 minutes and 33 seconds, the How to Write a Research Paper video is 8 minutes and 37 seconds or 40 seconds, which is long, and longer than most students watch, but it was as short as we could get with all the steps necessary. And we thought of ways to break it up shorter, but we could not find a good way to do it. But if you have suggestions, we would be happy to provide timestamps. And the video How ERIC Selects Resources is also in the 3 to 4 minute range. We do have longer videos such as the recordings of our webinars, and what we did for the Thesaurus webinar, was to break it into 10 minute segments. They weren't 10 minutes on the dot, it was based off of sections. If you have ideas of videos you think are too long, please let us know. We'd be happy to shorten them.

Question 2:

Can you demonstrate an ERIC search with the new descriptors? Such as tests, scales or geography?

Answer 2:

No I can't, because the new descriptors are only available on the development server, and I'm not logged in right now. And we only have 90 test records populated. What will happen is the new identifiers field will be filtered on the side, just like it currently is on the ERIC search engine. On the left-hand column in ERIC, you will be able to filter using these search terms. You will also be able to do a search. We got the development prototype of this yesterday so we're still refining how it will look, but we will do either a video or some kind of outreach to make sure that you can understand how these are used.

I believe I have a screenshot here. So, on this page [ref, slide 15] the identifier will be Location:Australia. That will be the search term that we will use, and Australia will appear in the location description along with any other state that is relevant.

Question 3:

The next question is about APIs, whether or not an API is available.

Answer 3:

The answer is no, not yet, but it's something that we are really working to have. We want to do this in the most useful way possible. Our Plan A is to make our version of the API the same as other products in the Institution of Education Sciences, such as The National Center for Education Statistics product -- in the same format. It's taking some time to coordinate, but we are working on it and it's something that we know we need and we want, but it's taking some time.

Question 4:

When will the new updates go live?

Answer 4:

So this an important question that I realize I never stated in the presentation. We have 1.6 million records in ERIC, so it's a huge database. We expect to finalize the field structure in the next week or two and start sending over batches of these new records to build up. We're hoping that by sometime in the next month or two, we will be able to launch this live on ERIC. This is on the eric.ed.gov site. Then, if you use ERIC through a vendor such as EBSCO or ProQuest, we will put a new download file up.

Those individuals will have to download the new file structure and build a new system, so there's going to be some delay. You will see it live on ERIC first, then we'll work with partners that we know use ERIC, help them get all these updates ready in their system. I'm hoping that this will be live by September, hopefully sooner. Given the size and scale of our collection, we want this up as soon as possible, but we want to make sure it's done right, so we're trying to balance both of these.

Question 5:

I was asked if I had any estimate on the API application.

Answer 5:

There are just too many balls in the air for me to give you that. I don't know. I would say probably at least minimum a year.

Question 6:

Are there any plans that enable the export feature into Excel files in addition to the current version that we have which is what PubMed uses?

Answer 6:

That is the first time I am seeing this request. The answer is no, we don't have any plans now. But I don't see any reason why we couldn't do this and why we probably couldn't do this very easily. We will put that on our consideration list and if you have more details or more thoughts about this, please send us an e-mail at ERICRequests@ed.gov and we will put that on our wish list of something for the developers to think about how to do. Something else that is on our

wish list that we are working on, we aren't going to promise that we're going to have, is trying to find a way to do links to holdings, seeing as that's a common request from users. They want to know if their library has access to a PDF if ERIC can't get it. We're working on that and we're just trying to find a way to do it.

Question 7:

So I see a question here asking me about the difference between EDs and EJs and just asking for some more clarification about why we can't have a peer-reviewed journals only button.

Answer 7:

The reason for this is that the IES Public Access Policy, which is going to be expanded both ED-wide and government-wide, requires our grantees to submit their final peer-reviewed manuscripts to IES through ERIC and to make those available to the public. When we make them public, we make them public as ED documents. So these are not journal articles yet, but often the exact same text. And what we have found is that sometimes some people consider these to be journals, some people consider these not to be journals. To create a button to say this is a journal or this is not a journal, you need a clear answer – there is a lot of gray muck there. You also run into the same issue with IES-funded reports, where they're not journals, they're peer-reviewed documents, and are something that could be okay and so you don't want to filter them out. So because there is so much area of interpretation that we think it's better for users to use a two-filter approach until the field comes to consensus over what is and is not a journal article.

If questions come up, please feel free to e-mail us at ERICRequests@ed.gov and, in the meantime, please go to eric.ed.gov/?multimedia and check out our videos. We really think that they'll be helpful.

Alright. Thank you everybody.