Script: ERIC Update: *Increasing Access to Full Text, Peer Reviewed Education Research*

<<slide 1>>

I am Erin Pollard. I am the project officer for ERIC in the Institute of Education Sciences in the US Department of Education. Today we are going to give you a preview of the presentation given at the ALA annual conference in Las Vegas on June 29, 2014.

<<slide 2>>

As a high level overview, today we are going to talk about what changes have been made in ERIC in the past year, what changes are coming, and how public access will increase the amount of peer reviewed full text that is in ERIC.

<<slide 3>>

The biggest change that you may have noticed was that we launched a new website on August 2nd! This is our day one website, which was fairly bare boned, but live and functional! We have added several enhancements over the past 10 months, but this was our first version.

The reason why we redesigned the website was to transition ERIC to cloud servers. This will save taxpayers millions of dollars a year and we can use the funds we saved to better enhance ERIC and have IES produce more peer reviewed, full text research that will appear in ERIC.

Moving to the cloud also allowed us to hire a small business to operate ERIC, something that is a key priority of the administration.

In the transition to the cloud, we needed to make sure that the database and the website was compatible with the new servers and new government policies, so we did a complete overhaul of everything– the website, the database, and the indexing software. This took time and resources, so our day one product was far from complete. But it was functional.

<<slide 4>>

One of the biggest benefits of the cloud transition is that even if the government is shut down due to a lack of appropriations, there is no technical reason why ERIC would need to be shut down. In about 3 weeks the transition will be complete and this will allow ERIC to continue operating, even if IES staff and contractors cannot. Unfortunately, we were not fully migrated in October and we had to shut down the site for 17 days. I realize how inconvenient it was for the entire ERIC community.

<<slide 5>>

So now that we will always be able to have the site up and running, let's talk about the new features. This website was designed to reflect current ERIC users and changes that we have seen in our usership over time. First, we have noticed that over half of our users come in from Google

and less than 10% of our users come in directly through ERIC.ed.gov. Therefore, we needed to have a website that was easy to use even if you knew nothing about ERIC. We wanted the website to be intuitive.

This website has a couple key changes that aren't apparent to the naked eye. First, everything is a permanent link, which means that there is no more permalink button. This means that users can copy and paste the URL of the article they are reading and that URL will work when they come back to the link later. This saves lots of frustration for our users who would copy and paste an ERIC URL and then would find that the link wouldn't work when they tried to access it later.

The second thing was that we redesigned our database to run searches much more quickly and using intuitive logic. You no longer need to learn complex searching strategies to produce the results that you want in ERIC. These strategies often still work, but as I will show you later, are not needed.

Finally, users can easily limit searches to peer reviewed and articles with full text available from their initial search. This is a common request from users.

<<slide 6>>

Today we have made lots of enhancements that I will talk through more in detail in the next few slides. It is important to note that the website isn't done– it will be a continuous project over time– but the bulk of the major developments are now live or very close to being live.

<<slide 7>>

We made a number of changes to the ERIC search.

First, we made the URLs stable. That means that you can bookmark your favorite search and then come back to it and it will automatically run again. This will save our novice users a lot of frustration– it was one thing we received a lot of complaints about.

When you search, the search term will automatically appear in the box. This way you will know what you searched for and can modify your search if you aren't getting exactly what you want.

Next, there are peer reviewed and full text only buttons. Users can limit their searches to only have results that show those options. These are our most frequent requests, so we wanted to make limiting your search for these fields as easy as possible for users.

On the right you will see indicators for peer reviewed, download full text, and direct link. This enables users to click directly to the PDF without going to the results page.

On the left you will see limiters. By clicking the limiters users will be able to narrow your search results to those specific criteria. Users can limit results to only those published in the last year or last 10 years they are trying to get recent research.

Finally, in the actual search results you will see the title, authors, source, year, part of the abstract, and the descriptor. It is not all of the metadata that ERIC has, but it is enough for users to learn if this publication might be helpful. If you click on the title, the full metadata appears.

<<slide 8>>

Each page has a publication with the full metadata on it, as well as the ability to download the full text if we have permission from the publisher

<<slide 9>>

In developing the simplified search box, we no longer have the advanced search feature. I know it was really popular, especially with this audience, but we have found that many users would end up automatically searching the advanced search box when our new general search would ultimately give them better results. Still, we wanted to provide information to some of our more expert users on how to do advanced search and created some search tips here.

The advanced search tips is not a replacement for the advanced search and we aren't opposed to bringing it back if there is a pressing need, but we want to make sure it is designed in a way that will be useful and won't undermine the new search technology the site is based on. If there are specific searches that you cannot replicate with this logic, but could before, please let us know. We will work on a way to make it possible.

As I mentioned earlier, this system was designed to search smartly, so many of these searches may not produce better results than searching more simply.

<<slide 10>>

For example, if you want to search for charter schools and Race to the Top, you may be inclined to search "charter schools" AND "race to the top". However, if you just search those terms without the quotes and Boolean searching you get very similar results. The website automatically pulls in results that has the terms that you search for close together and in the order that you typed them. Those are the results which will automatically appear on the front page of the search results.

<<slide 11>>

Similarly, if you just put in quotes around phrases, you might not get as good of results as if you just straight search. You will notice that the number of results that each search produces varies widely, but the most relevant results appear on the first pages of all searches.

<<slide 12>>

Many users feel that they should limit by field name whenever possible and using the field limiters is the best way to do this. So if they are searching for Ruth Curran Neild's work on dropouts, they may want to search: author:"Ruth Neild" descriptor:"dropout". That search would produce these four results.

<<slide 13>>

However, searching "Ruth Neild dropout" would produce the same four articles, as well as three articles that are heavily based on Ruth's work. This may be a better way of searching for most purposes. It isn't that the field limiters don't work, they do work very well. They just aren't needed in most cases. The search is smart enough to find what you are looking for.

The one time we have found the field limiters to be particularly helpful is if an author has a last name that is very common in abstracts, such as Young or Reading. For that case, limiting "young" to the authors name is really helpful in producing better results. Without using the field limiter most of your search results will be about young students and young teachers.

<<slide 14>>

In redesigning the site, we also developed a new thesaurus tool that we hope the users find really useful. By clicking on the thesaurus tab on the ERIC search box, you can search the thesaurus. You will then have all matching ERIC terms appear below. If you click on one…

<<slide 15>>

…you get a whole slew of information. First, you will get the general category that it is in. If you click that….

<<slide 16>>

You will get all of the terms in that category, including the dead terms (they are crossed out and in italics)

<<slide 17>>

You will also get broader terms and then narrower terms.

We also have the "use this term instead of" list, which shows synonyms for the term.

We also have related terms and you can search the collection using this descriptor, which will show all of the reports tagged with that descriptor….

<<slide 18>>

… Here is what the search result would end up looking like.

<<slide 19>>

Elsewhere in ERIC, in the notes page, you will find our most recent news, as well as our new topical newsletters. I'll talk more about this in a little bit.

<<slide 20>>

We have a list of frequently asked questions– in most cases, if a user has a question, they can find the answer here. We update this on a regular basis, so if there is something we are missing, please let us know.

<<slide 21>>

One of the new features of on the site is the downloads page. Prior to this website redesign, we would allow users and vendors to apply for a license to use ERIC's data for their own purposes. This was an expensive and time consuming process and we thought that it was unnecessary

given the administration's efforts in open data, or to make government data freely available to the people. It is the tax payer's data and should be made available without a license.

We ended all of our licenses and put ERIC's metadata on our website for users to download, free of cost and without restriction. We update the files monthly– normally around the 15th of every month and anyone who wants to use our data can download it here. So while our website is updated weekly, the files that vendors use will be roughly a month old.

One advantage of this feature is that it allows all sorts of users to take ERIC's information and create cool things with it. If you know of an interesting way that people are using our data, please let us know. We would love to highlight their project.

<<slide 22>>

Another major task that we tackled was rewriting the selection policy. On January 2nd we adopted a new selection policy that governs what is in our collection.

<<slide 23>>

It is important to note that we didn't do this alone. The new selection policy was written in conjunction with staff who had been working on ERIC for years, as well as new faces to the ERIC team. We also worked with a Collections Advisory Group comprised of nine experts in library science from around the country. They worked through the various issues surrounding the Selection Policy and then provided recommendations on what direction ERIC should go in.

<<slide 24>>

These are the members of the collections advisory group. As you can see, they are well respected individuals from across the country and the discipline. They worked with the ERIC team to develop a new policy.

We decided to draft a new selection policy to be more transparent. Much of what is included in the new policy is no different than how ERIC was operating in the past. However, details and information is now publicly available for ERIC users to know what types of materials are included, and not included in ERIC.

We will be revising the current policy every 2 years. It will not be a complete re-write like this version (that will happen every 5 years), but instead a tweaking on areas that we can improve on. All of these changes will be in consultation with the Collections Advisory Group.

<<slide 25>>

It is important to note that this is the policy from January 2014 onward. We are not applying the policy retroactively to anything that already existed in ERIC. Nothing will be removed from the collection.

In drafting this policy we wanted to keep ERIC's pace of indexing to be about the same as it was in the past– 4000 records a month--but to shift the focus slightly. We are now focusing on getting material into the collection quickly, so there is not an unnecessary delay between when an article is released and when it is available to you.

We are also making a shift in quality, we want to make sure what we are indexing is the highest quality educational research and records are a prudent use of taxpayer dollars.

<<slide 26>>

As we set out on this process, we at IES had four overarching goals. The first was to make sure that the articles that we included in ERIC were education research and were on one of the topics were are required to cover by law. The legislative topic areas include the 16 topics covered by the former ERIC Clearinghouses as well as closing the achievement gap, and education practices that improve academic achievement and promote learning.

Below is our definition of research and this does mark an important shift in the ERIC collection. We are no longer indexing just articles on education, but the articles have to be research. This meant we lost some great sources that were education related, but not research (like *Education Week* and *The Chronicle of Higher Education*). However, we gained some truly fantastic research in return, much of it with full text access.

<<slide 27>>

Our second goal was to increase the amount of full text documents that are not peer reviewed. We want to ensure that this material is rigorous and relevant education research. For the most part, these sources would have gone through a considerable review process, but it might not be the kind that meets the threshold of peer review.

Some examples of these documents would be think tank pieces, white papers, reports from school district research offices, and work published by federal technical assistance providers. We are currently indexing some of those, but we would like to increase the amount of material like this.

One interesting observation is that a lot of these documents are not marked as "peer reviewed" because we do not assign peer review status to non-IES, non-journal sources. (IES publications are legislatively required to be peer-reviewed, which is why we can designate them as such). However, in our review process, it was clear that the quality of the non-journal sources was often just as good, if not better, than the peer reviewed journals. Our non-journal sources are rigorous and relevant research, just like our journal sources.

<<slide 28>>

Third, we would like to limit the citation only indexing, or the material without full text, to only those resources that are peer reviewed or are of substantial rigor and relevance. If we cannot provide full text access, we want to ensure that the citation alone will be useful to users conducting research.

Finally, and in some ways our most exciting goal, is to increase the number of peer-reviewed, full text documents. We plan to do this through two ways. First, there is an increase in open access journals that we will try to index. Second, there are a new series of federal requirements that require all federal grantees that are funded by research dollars to publish their final, peer-reviewed manuscripts in repositories like ERIC. I will be talking about that in a lot greater detail later.

<<slide 29>>

When I mentioned our goals, I talked about shifts in the collection. One of the most important things to note is the scale of the change we are hoping for. Above are our goals– this is what we aim to do, but it isn't a promise or a guarantee. It is simply what we are working to and monitoring our progress against.

About 14% of the articles published in 2011 in ERIC have free full text. Of those, half were peer reviewed and half were not peer reviewed. By 2015, we would like to increase the proportion of full text document to be about a quarter of new ERIC entries. We would like to keep about half of the articles being peer reviewed and half not being peer reviewed.

As for the citation only articles, we would like to reduce the non-peer reviewed, citation only articles from approximately 9% to 2%. We would also like to reduce the amount of peer-reviewed, citation only articles from 77% to 73%.

So what does this mean in practical terms? We want to increase the number of full text documents from about 560 a month to 1000 a month, or approximately 5,280 more full text documents a year than what was in ERIC before. It is a big shift, but it is only a shift. It is not a total overhaul.

<<slide 30>>

Once the selection policy was complete, we went through every journal and non-journal source individually (and there are a lot of them!) to see if it fit the new selection policy. Many did, but some did not. We then contacted each publisher to let them know about the decision and posted the list of approved sources on the website.

We will revise this list every 6 months and do a complete re-analysis every 5 years. At the 6 month mark we will review new sources for consideration, so if there is a source you think we should add, please email it to us at ERICrequests@ed.gov

We will also remove any sources that are no longer publishing or sending us their material during the 6 month review. At the 5 year mark we will go through every source again and compare it to the selection policy.

<<slide 31>>

We also have launched a brand new style of newsletter, in the form of the IES newsflash. If you are not receiving them, you can sign up at ies.ed.gov/newsflash. We are under NCEE. Every 6 weeks or so we send a topical newsflash that highlights some key resources in ERIC on a popular issue, as well as provides information to users on how to search the ERIC collection for more information.

We also send out periodic "blasts" through this service when we have new news– such as when the PDF restoration is complete or online submission is now available. More on those in a minute.

<<slide 32>>

If you can't get your question answered through the FAQs, you are always welcome to contact our help desk via phone or email. The contact information is available on this page.

In addition to questions, please feel free to send us your feedback!

<<slide 33>>

We are now going to switch gears and talk about two brand new developments in ERIC. The first is the PDF restoration process. As background, on August 2nd, 2012 we found that there was highly sensitive personally identifiable information in a few of our documents.

To give you an example of the concern, think of a graduate thesis from the 1980s. The author's social security number used as a student ID and were both put on the cover page of the paper. So if you Googled the author, their social security number would pop up as the first result in Google. Because of the security concern, we pulled all of the PDFs offline and had to scan each one for personally identifiable information in order to be able to release it back online.

Where we left last year was that we had released a portion of the documents, but it would take about 5 more years to complete and we would release a large batch of documents in the beginning of October, then approximately 4,000 a month going forward. That ended up not being entirely accurate. We decided that our plan wasn't feasible– it was really expensive and took too long, among other things.

We ended up finding a better (and way less expensive) solution and I am happy to report that this project, or at least the first phase of it is done. We have released all documents that are machine readable and do not have personally identifiable information. They are on the ERIC website right now.

<<slide 34>>

You will still notice that not every document is released– there are two kinds that we aren't releasing yet. The first are the ones with social security numbers. There are 650 of them, or about 0.2% of the collection. We have to figure out how to deal with them in a way that does not involve risk.

To try to answer a question that I know will be asked, because the microfiche is still out there we cannot simply black line the documents. If a user found a document that is black lined, like the one above, they could go to the microfiche, to that page, and get that document. They could then use that information for identify fraud. Therefore, this isn't the best solution.

We are still exploring other options, but for now the PDFs are not available online and they probably won't be online anytime soon. We will be removing the "PDF pending" identifier for them, but their citations will remain in the collection.

<<slide 35>>

We also have some PDFs (approximately 23,000 documents or 8% of the collection) that are just plain not readable. This is actually one of our most readable from a recent batch. It is a horrible copy from microfiche and we fairly positive we can get a better copy.

What we are doing with these is not releasing them because they are not useful or readable. There's no social security numbers or other PII, but you can't read them. We will be marking them as "pending restoration" and then will be going through them one by one to re-scan them and make them into fully searchable PDFs. This will be a long and expensive process, but it will ultimately be better for our users.

<<slide 36>>

Any exciting new feature that we are working on is the online submission feature. This should be live in the next few weeks. When we make it live, we will accept all kinds of non-journal materials that meet the selection policy, as well as journal articles that are the result of federal funding or if it is submitted by the author.

We will give preference to work funded by the Department or other federal agencies. This is because we are required to index this material by law and by budget authority. So as you hear me talk about federal grantees as a special class, that is a change due to the open access requirements that we are currently implementing. This will be a shift in ERIC, but an exciting one that will increase the amount of peer reviewed full text.

We also give preference to professional conference papers, theses, dissertations, and practicum papers.

<<slide 37>>

The types of materials we are not going to index through the online submission system are full or partial journal issues. What has happened in the past is that some journals will submit an entire issue of a journal through online submission. In those cases, the journal really should go through the review process that Fern described earlier. What can happen is if individual authors submit their articles through the online submission system.

We also are not indexing stand alone lesson plans– those are valuable, but not education research. We will index evidence based teaching practices and a document analyzing lesson plans.

Similarly, subject matter content is not considered education research. An article on the Pythagorean theorem and how it is derived is not eligible, but an article on evidenced based practices to teach geometry would be eligible and a great addition.

Finally, web pages, websites, blogs, and the like aren't eligible.

<<slide 38>>

The online submission system should be live shortly. I am going to go through a brief demo of how the site works. When you log on you will see the requirements for submission.

<<slide 39>>

You will then have to certify that your document meets the requirements to be indexed. We put in these questions to help users make sure that their document is eligible before spending the time needed to submit.

<<slide 40>>

You then are asked to fill out the title, authors, abstract, and the type of document you are submitting. We ask about program funding as part of public access requirements, which I will talk more about in a little bit.

<<slide 41>>

We then ask for publication details, a PDF of the publication, and contact information.

This should be live in the next few weeks and we will send out a newsflash when it is ready.

<<slide 42>>

Another thing which we are working on is the topic based section of the site. The ERIC topic pages will start being rolled out in the next year or so. They will contain a 2,000 word, Wikipedia like overview of the topics listed below. These are going to be peer reviewed and were written in conjunction with subject matter experts, but are designed for a general audience. These summaries will then be topic pages and subtopic pages that have links to related resources in ERIC to help users find more information on the topic.

The reason why we are having these pages is that we find a lot of our novice users come to our site with very little background on what they are searching for. Think of your average college freshman in Education 100 who is going to write their 5 page term paper on what the research says on improving education or a parent who knows a little bit about the topic, but needs a primer. These pages will help orient users to the topic so that they will have enough background to do a search.

As information professionals, you can appreciate the challenge of developing accurate, rigorous, representative, peer-reviewed summaries that will be brief and user friendly. It is a really interesting challenge and we are excited to see the results.

<<slide 43>>

Over the next year we will be having a couple of exciting changes. First, we will be reducing our call center hours to Monday through Friday, 9am to 7pm, excluding Federal holidays. We found that we were getting so few calls in the early mornings, late nights, and weekends that it really wasn't a prudent use of taxpayer money to keep the call center open. Instead, we will have a 12 hour email response window during off hours. With the cost savings, we are improving our user outreach.

We are able to have a booth at ALA and will be giving multiple presentations there. Second, we will be having two webinars a year. These will be announced in advance and then archived on our website, just like this presentation is. We will plan on having one on the online submission process and one on the PDF restoration process, but those are subject to change. Third, we are going to produce two videos a year. These will be basic and the topics we are considering are on how to do a basic search and what is ERIC.

The last development is the one that I think will be really exciting to our users. In the future, such as 2016 or 2017 we will be getting a API. As a result, the API can be used to power apps

and be combined with data from other sources. It can also be used to get far better metrics from those using ERIC data. Instead of downloading ERIC metadata from our website, they will pull it from our server. This means we will be able to tell how many people used a particular article, whether they searched ERIC through EBSCO or the ERIC website. And you can integrate ERIC into your other offerings.

<<slide 44>>

We are now going to switch gears and discuss public access and how that will increase the amount of full text materials in ERIC. Publishers permit indexing of their content in ERIC by agreement, but access to the full-text has always assumed that the users have a subscription. We talked earlier about the goal of increasing the percentage of full text articles in ERIC. One of the ways that ERIC will do that is though a change in the way that the government manages access to research that it has funded.

<<slide 45>>

The Federal government has decided that agencies need to be more systematic about collecting, preserving and making accessible the *publications* and *data* resulting from federally funded research.

A memo was issued by the Office of Science and Technology Policy a year ago outlining the principles and objectives of the new policy. Since then federal scientific agencies have been developing plans to meet the requirements.

They are grappling with the tension between the requirements for open access to federal research and the value to the scholarly community of publication in peer-reviewed journals.

<<slide 46>>

The policy principles describe why this new approach is important. Fundamentally, it is because scientific research is the foundation for innovation and economic progress.

<<slide 47>>

Federal agencies will need to advise their grantees and contractors how to make their published research available. The Department of ED has decided that ERIC will be the primary repository for access to research it has funded.

The National Rehabilitation Information Center (NARIC) is the library of the National Institute on Disability and Rehabilitation Research (known as NIDRR), a part of the Department of Education and it will be the repository for NIDRR research.

<<slide 48>>

The Department's Public Access Plan envisions that authors who have received funding from the Department of Education will deposit a final peer reviewed manuscript of their article in ERIC. There will be an administrative interval or embargo, probably for 12 months, during which ERIC users will go to the publishers' sites for the article and will still need a subscription for access.

For research resulting from grants made by the Institute of Education Sciences this has been happening since fiscal year FY2012, or October 2011.

<<slide 49>>

When the embargo period is over ERIC users will have access to the final peer reviewed manuscript in ERIC or to a publisher's version if the publisher has agreed to make it available.

This process will continue for IES sponsored research and gradually be implemented for research sponsored by other parts of the department. Different offices in the Deptartment of Education have different guiding legislation and therefore implementation may vary across programs.

<<slide 50>>

Thank you so much for listening today. If you have any questions or concerns, please feel free to contact the ERIC help desk at ERICRequest@ed.gov. Alternatively, you can contact me or Pamela directly. Our contact information is below.