

Integrations with ArchivesSpace - ArchivesSpace and ARKs

Wed, 4/28 1:02PM 59:51

SUMMARY KEYWORDS

arks, ark, url, record, create, identifier, objects, seth, aid, finding, archives, question, california digital library, update, naan, archival, links, webinar, urls, persistent

SPEAKERS

John Kunze, Seth Shaw, Christine Di Bella, Jessica Crouch



Jessica Crouch 00:07

Welcome everyone to the sixth webinar in our webinar series on integrating with ArchivesSpace. Each webinar in this series will highlight an integration with another application used in archives that ArchivesSpace members worked on or requested. Our sixth webinar in this series we'll explore how to integrate ARKs, Archival Resource Keys, the persistent identifiers designed by the California Digital Library to support long term access information objects into your ArchivesSpace implementation. In this webinar, John Kunze will describe the ARK design goals and current landscape. John created the ARK identifier scheme at the California Digital Library and leaves the ARKs in open initiative. With a background in computer science and math. He wrote Unix software that comes pre-installed with Mac and Linux systems and contributed heavily to internet standards for identifiers like URIs, archiving, including BagIt and metadata including Dublin Core. After Seth Shaw will describe the University of Nevada Las Vegas libraries adoption of ARKs for finding a persistent identifiers in 2017, and the workflow from them. Seth is an application developer for the UNLV libraries dedicated to Special Collections and Archives. Seth was formerly the electronic records archivist for Duke University and an assistant professor of archival studies at Clayton State University.



Jessica Crouch 01:19

Finally, Christine Di Bella, ArchivesSpace. Program Manager, will conclude the webinar by demonstrating the new functionality related to implementing ARKs in ArchivesSpace that became available with the with ArchivesSpace version 2.7. This functionality includes the option to use internal or external ARKs in exports or OAI-pmh harvests for resources and archival objects. We will hold all questions until the q&a session at the end. If you type a question into the chat, it will be read by me and answered during the q&a. You will notice your microphone and camera were muted when you entered, we ask that you keep your microphone muted until the q&a session. With that, I'll welcome John, Seth and Christine, and John, if you'd like to start us off.



John Kunze 02:02

Thank you, Jessica. I'm just still finding my way a little bit with the presentation mode and Google Slides. Thank you for your patience. So getting started the the I wanted to mention that the California Digital Library where I work is housed at the University of California in the office of the president. We take care of and represent the 10 campuses in terms of their digital library needs and that represents hundreds of museums, archives, galleries, etc. Five medical centers, five law schools three national labs, and they're responsible for 10% of the US research output. Part of this our job is to be extremely domain agnostic. So our solutions tend to be fairly general. So why do we care about ARK identifiers, while the average lifetime of a URL was once said to be 44 days. At the end of its life, a URL link breaks meaning it gives you the dreaded 404 not found error that most of us have seen. Irritating as that may be it's extremely awkward when trying to renew a grant. And it's a minor disaster for memory organizations like archives, libraries and museums.



John Kunze 03:27

Among the many kinds of web links, a persistent identifier is a link that in principle keeps working far into the future. Services that do discovery and interlinking between artifacts and research prefer persistent identifiers because of their stability. So ARKs are meant to be persistent web links for any kind of thing, digital, physical or abstract.



John Kunze 03:53

So anyway, surprisingly, few vendors support persistent links to content. Especially when you change vendors and all of your all your old content links are broken. ARKs are meant to be persistent, reliable links for any kind of thing that 3.2 billion ARK identifiers have been assigned by over 550 institutions located across the world.



John Kunze 04:29

Who is registered to create ARKs? Well, a wide variety of memory organizations from archives to libraries to publishers, including nonprofits, for profits, government agencies, etc. A few of them are listed here on this slide.



John Kunze 04:48

What does an ARK look like? At first glance, it's a particular kind of URL bearing an internal label. How can you tell if you have an ARK? Well, here's the internal label. Then there's the assigning authority number. This identifies the organization that created or assigned to the ARK. The thing that the ARK is assigned to appears to the left of that appears to the right of that, sorry, the hostname over on the left makes it actionable, which means it's something you can click on or in order to get to the thing. It's also known as the resolver. ARKs are unusual among persistent identifier solutions in allowing organizations to run their own resolvers. The part after the resolver is the core globally unique identity, which doesn't depend on the host being available or even on the future of the worldwide web. So that core unique identity could go into the future and be stuck after some future protocol independent of HTTP.



John Kunze 06:03

What are ARKs used for? Well, here's a list in descending order according to numbers of ARKs assigned. The heaviest user of ARKs is the largest genealogical research organization called FamilySearch International. Next is the archive of all the source files for mainstream published content. Technically, they have over 40 billion ARKs, but apparently there are only 100 million of them that we're meant to be interested in. The Internet Archive has been assigning ARKs to scan books and texts since 2008. The Smithsonian is rapidly expanding their use of ARKs from 15 million I understand with a target of 100-150 million. The French National Library was an early adopter, and here's a wonderful image depicting the biblical ark and of course it has an ARK.



John Kunze 06:59

ARKs are important for people for identifying people, especially historical figures. For example, here is an there's an ARK for Ada Lovelace. ARKs are also important for abstractions used in the Semantic Web, things like diseases, vocabulary terms and time periods, they get ARKs. So you may have heard of other persistent identifier solutions, that's the digital object identifier DOIs or handles, maybe PURLs and or URLs. By ARKs. If you wanted a very terse and somewhat snarky summary, you can Google for 10 persistent

myths about persistent identifiers. That sort of, it's a complex abstract subject. And so they're here. So here's a sort of a peek into that. There's also more information at the FAQ on the ARKs in the open website, which I'll point you later on.

J

John Kunze 08:04

So, one of the things that distinguishes ARKs, as I mentioned, is a flexible resolution. So this is one of the few that's not a silo. You're able to, you're able to resolve or serve your ARKs, either centrally, and that's through this thing called n2t dotnet. Or you can do it through your own web server. Part of this is an outgrowth of this sort of commitment to openness that you'll find at the University of California and I've worked there a long time, I've been influenced by this history of openness. Having worked on Berkeley Unix, which was open source before, you know, 10-12 years before the word open source existed is open access that we're doing in the CDL. There's, there's a fairly well publicized termination of journal subscriptions with Elsevier and the whole push towards open access to all of the publicly funded research. And since 2001, we've been doing these non paywalled and decentralized persistent identifiers called ARKs.

J

John Kunze 09:23

Now there, there's still probably some more comparisons one can do between ARKs and DOIs. DOIs are well known digital object identifier, and it's the publishing industry solution. Unfortunately, they require a membership. They have per identifier fees, and they have quite rigid metadata requirements. You can assign DOIs to people, but you have to find out how to specify who the publisher is of that person. Um, ARKs are a cultural heritage solution. They have no fees or membership required. They have highly flexible creation and metadata policies. So you can assign ARKs to things that don't yet exist, which is a very important thing to do when you're preparing documents to be released all at once. You can tell there's all sorts of examples given in the FAQ.

J

John Kunze 10:21

DOIs have a nice advantage in the way that they're interconnected to the through the publishing ecosystem, and ARKs are taking take advantage of many of those, those things. ARKs are also stable. They're linked to metadata. And they're found in various strategic places like the data Citation Index, they're in Wikipedia articles, wiki data records, Internet Archive, and are also found in ORCID researcher profiles. ORCIDs are identifiers for researchers. Also people but work it only does living people, not historical things. Moving along, so the the commitment that ARKs have made is that there's basically the ARKs scheme will not charge fees to create or use ARKs. That doesn't mean

to say that this is free, a free activity. There are vendors who will support things for fees and many of you are acquainted with vendors. ARKs, the ARKs scheme will not limit the number of ARKs you assign not to limit the kind of content that you identify or not, and will not require metadata. And hardly, it may sound odd that they don't even require persistence. In other words, it doesn't require all sorts of sleights of hand and bringing, you know, it doesn't require a sort of hiding reality for you to delete ARKs that are just were just mistakes.



John Kunze 11:53

Importantly, the ARK identifier scheme will not require the use of any particular resolvers. So nobody gets to control access to your server if you don't want that.



John Kunze 12:07

So I'd like to, this is the concluding slide here. For me, my part, if you'd like to get involved in ARKs, it's really very much of a community affair. We if you want to get started learning more, I just suggest you Google ARK identifiers FAQ, that's easier than remembering the URL. You want to start assigning you have to you have to obtain that one of those five digit numbers to identify your organization. And that's that that URL there and n2t dotnet. Since late 2018, we have been involved with LYRASIS, and DuraSpace before that, to to support open infrastructure for ARKs, there's a very small amount of infrastructure. But it still needs to be completely community owned at the moment. It's sort of partly owned by California Digital Library. And it shouldn't depend only on that. So we're trying to have it mature into this completely community owned situation. And so that's at ARKs in the open.org, which I invite you to visit and to support us and get involved in one of our working groups. So with that, I would like to turn it over to Seth and I will stop sharing.



Seth Shaw 13:41

Thanks, John. Alright, here we go. So I'm gonna be talking relatively briefly about our own adoption of finding aid ARKs. I will mention as we get started that we decided made the decision to adopt ARKs before I was even hired here. So this is one of the first things that we worked on after I arrived to start creating ARKs for our finding aids in 2017. We'll do the background. When you put a finding aid online, the URLs could go all sorts of places. If you have a social media presence, you might be posting a blog post saying, hey, check out this new collection, this finding aid that we created for this cool collection that we've received. It could go on your Facebook, your Twitter, your, whatever social media, you could be creating collection guide listing all the various finding aids that apply to a particular subject. Your OAI-pmh harvests may pull in records of your finding aids, it's to

help people you know find your collections amidst others. You might also be exporting MARC records to put into your library catalog to find these. And you know, hopefully, maybe some researcher actually cites your finding aid for some reason. Giving them a link to put in their, their citation would help bring other people to finding aid. And of course, you have the ever present browser bookmark, or researcher says, Hey, I'm going to be working on this collection. Let me bookmark this, to see to come back to it occasionally to see as I'm figuring out what I want to get access to.

S

Seth Shaw 15:32

So these URLs go all over the place. But as we just mentioned, URLs are subject to link rot, the 44 days number that John just mentioned, what happens is, at least for finding aid systems, you might start with a local database of finding aids, or maybe some local EADs. But then you move systems, right, you, you change systems, they get better the infrastructure adapts, and you might have a URL for your local database. And then you move to a PDF, which has a different URL than the original finding aid. And then you move to an EAD, which has yet another URL. And then maybe you use the ArchivesSpace user interface or some other interface for providing online access to your finding aids, well each time you change systems, more often than not, that creates a new URL. And a new URL means that the old URL is either broken or you have to maintain this redirect from the old URLs to the new URLs, which can you know, after a few iterations, gets kind of frustrating and difficult to manage and annoy. And that's even just for the URLs that you create. What about updating all those URLs that you put on your social media that people put in citations that aggregators got. I really don't want to, and in many cases, I can't go update old URLs. They're just out there. So an ARK allows us to say, Hey, I have a persistent identifier, that if you visit this URL, if you use it in your social media, then I will always be able to get you to the right finding aid. So we're not gonna have to update all of our blog entries, we're not gonna have to update now go back and edit, can't even do the editing old posts and some social media forms. So yes, we're gonna start with our persistent identifiers, and lead with those whenever we create URLs to make sure that people always get to the right place. For example, when we started this ARK adventure, we were using ArchivesSpace to create our descriptions, but we were exporting PDFs to put on our website. And you can see the URL bar, you know, we had a finding aid folder with the PDF URL. But in our finding aid, we always have the statement saying, here's the persistent URL to our finding aid. And here's our ARK. So you will always be able to find us finding our description of this collection, if you find this ARK and follow it. Now, you might be asking, so it's great. How did you get this in here into your finding aid, or into your EAD? So we use that did not advance like I wanted it to just like, where do we put it in ArchivesSpace.

S

Seth Shaw 18:40

We use the finding aid data section and use the EAD location field. Admittedly, this is not semantically correct, because this ARK does not take you to the EAD, is just taking it to wherever our web accessible description is. But we weren't generating EAD for distribution. And this was the most convenient spot to put it in ArchivesSpace. So we said you know it's not 100% correct. But you know, it's going to work for us well enough. It also means that we don't have to worry about configuring a user defined field and updating the interface and doing all that background to try to fix that. We'll just stick it right here. And we will move on with our lives for the time being.

S

Seth Shaw 19:31

So of course we have the EZID service I didn't should have said of course, we do use the EZID service, which is really handy. They have a nice web interface for updating the metadata about a particular object that has an ARK. And also if you hit Edit, it will allow you to change the URL that you're finding it is that should we ever decide to choose change systems so we use the EZID service. And we have this web interface to do this in. But even better. So actually, before that, so what's the workflow for this? Well, we could complete the finding aid, use the web interface to build the ARK, and then copy that ARK into ArchivesSpace and then generate a PDF, and then copy the finding into the web server and then update the corresponding links, online database, and our library catalog. We can do it. It's manual, if we have to kind of annoying, but we'd like to automate things, right? So what we did is we created a semi automated workflow, this is what we actually do. Our Technical Services folks, they create a list of completed finding aids. In part, this is for telling me that they need to be published to the website. In part, it's also to tell the catalogers for our library catalog that they can harvest MARC records to put into WorldCat, and then pull into the library catalog. So we use this for workflow and periodically they say, Hey, we have a batch of finding aids that we'd like to put online now. I say, Okay, let me run my script. So I just get a list of the resource identifiers from this spreadsheet that we have. And the script for each of them will mint an ARK via EZ ID, if we haven't already put one in ArchivesSpace, for one reason or another. It then also updates ArchivesSpace resource with that ARK. Again, if we had to mint one, then it'll generate the PDF via the ArchivesSpace API, and then copy that finding aid to our web server. And then I have a supplemental script for updating this older online search database that we have to allow us to search these descriptions. And I've linked that script down below if you're interested in what I did. If you do look at it be kind this was written before ArchivesSnake existed. So I didn't have a handy library to do it. So if I was going to do this, again, I would certainly have used the ArchivesSnake, but it wasn't there. So we didn't.

S

Seth Shaw 22:18

But this is great. This means that the technical services stuff, all they have to do is say, hey, it's ready to go. I can do a whole bunch of them all at once with very little workflow. And now we have a persistent identifier for all of our finding aids that's in all of our PDFs. And that we can put in our blog posts, social media, etc, etc. Now, what we would really like to do is just hit the publish all and let the system take over the rest of it. Which we could do through some, you know, doing some Ruby work in the background to create a plugin to do this. But for what we are doing, it was easy enough and quick enough to do the Python script. That's just what we did. But also, we are really looking forward to seeing what the ArchivesSpace community has done for integrating ARKs internally. So we are actually excited for what's coming forward so that we can adjust our workflows to fit with the system a bit better, and to make this even easier for our staff.

S

Seth Shaw 23:27

But yet, there's more because we are using this ARKs system, should we ever have to move systems again, like we're currently publishing PDFs. Come this summer, we plan to move away from those PDFs into a new online finding aid system, again, based off of ArchivesSpace. So how are we gonna update all that? Well, we can batch update, ARK URL pairs and all the metadata via the EZ ID API. I even already have a script for this because we use it for our digital objects already for minting and updating ARKs for digital objects. So I can take a CSV file with the ARK, the various metadata fields, and the URL that I want it to point to. And I can batch update all my fields. So when we migrate our finding is from PDFs to this other system. All I have to do is have a list of ARKs and what their new URL is, and hit the API and it'll be done. So really easy. Thanks, EZID Thanks, ARKs. And now I'm going to turn it over to Christine so that we can hear about how ArchivesSpace is integrating this more directly rather than our back of the office setup.

C

Christine Di Bella 24:53

Thanks, Seth. So I'm going to share my screen and show you what we've got. So I'm really just going to, so John gave you great information on what ARKs are, Seth gave you great information on why they use ARKs and what they're doing currently. I'm going to focus very much on ARKs in ArchivesSpace itself. And unfortunately, I can't really get I can't really get into like all the technical mechanisms of how it works on the back end. There, we could probably pull in someone from the technical staff who could do that, if that is needed later on. So if you have questions of that nature, I'm happy to log them and pass them along, but unfortunately, I'm probably not going to be able to answer them myself.

C

Christine Di Bella 25:38

So ARKs in ArchivesSpace. ARKs, as Jessica mentioned, at the top of the program, these this was an optional feature added for version 2.7.0. We with as we do with a lot of things that come out brand new in ArchivesSpace. We start small with it and gauge what the interest is. And then we're very interested in hearing if or how people would like this feature to grow. So right now it can be used with, ARKs can be created for resources or archival objects. It is completely optional. You don't have to use the ARKs feature in ArchivesSpace. If it's not something that's going to be helpful to you. There is more information about using ARKs in ArchivesSpace, available in the ArchivesSpace user manual. And if you're a member of ArchivesSpace, if you're a member of ArchivesSpace, feel free to get in touch with the program team about accessing the user manual, if you haven't already.

C

Christine Di Bella 26:41

ARKs in ArchivesSpace, we have two ways that you can work with them. One is that you can use ArchivesSpace to generate the ARKs. So essentially like the kinds of things that Seth and John were talking about for creating those special URLs, those special identifiers. ArchivesSpace can do that, based on certain configuration settings that you've identified. Those were what were terming internal ARKs. And then ArchivesSpace can also be used to record ARKs that are generated by another system. So the kinds of things that Seth was showing, now we have a slightly better field that you could put those in if you wanted to, Seth, those we're terming external ARKs. In the system, external ARKs, take precedence over internal ARKs. So if you have both an internal ARK for record, and you have an external ARK for record, ArchivesSpace assumes that the external ARK is what you want to use. And that's what's going to appear in some of the outputs of ArchivesSpace internal ARKs, resolve to record in the public user interface. So the expectation is if you're not using if you're using the ARKs feature, and you're not using the public user interface, that the external ARKs field is going to be more useful to you. When enabled ARKs, appear in some places in the ArchivesSpace interface, and they're also included in exports and OAI-pmh harvests. And actually the requests for ARKs functionality really came out of the OAI-pmh work that happened a few years ago. That was part of the original specification, we were able to return to it a little bit later on.

C

Christine Di Bella 28:33

I mentioned that there are some configuration options for ARKs in ArchivesSpace. So these are things that are set in the configuration file on your ArchivesSpace server. Usually, if that's not you, who maintains that server, that's going to be a system administrator or somebody else in your IT area. So if you want to use ARKs, and you're not

the person who maintains the server, these are some configuration settings you're going to want to alert that person to.

C

Christine Di Bella 29:03

To use the feature ARKs enabled is kind of the key. So you need to set ARKs enabled to true. And then that's going to show up, make some things appear in the staff interface. And then the second thing that you're going to get going to want to do if you want to create true ARKs, true valid ARKs, is you're going to want to get a NAAN from the California Digital Library from their service. And then you're going to want to indicate what that NAAN is in the configuration file. And then the last thing that you can specify in the configuration file is the URL prefix for internal ARK URLs that defaults in the configuration file to the public user interface URL, since in most cases that's what it's going to be. Once you set ARKs enabled to true, then a few things are going to happen. ARKs will be automatically generated when new resources and archival objects are created. There will be a way for you to input external ARKs when you're in edit mode for resources and archival object records there will be a field there for external ARKs. And you can also generate ARKs in bulk for all resource and archival object records in all of the repositories in your ArchivesSpace by running a background job called generate ARKs in bulk. I wanted to point out that when you use the ARKs feature in ArchivesSpace, it doesn't do anything to your URIs itself. So those those things that look like repository slash two resources slash one that identify the different records in your ArchivesSpace. Even if you're using the ARKs feature, those will still work as they did before. You can just choose though, to use the ARKs in different kinds of outputs, and get the benefits that John mentioned for using ARKs.

C

Christine Di Bella 31:14

Alright, I'm going to show you a little demonstration of how this works in ArchivesSpace. So and this is going to be a very little demonstration. But it'll give you an idea of how simple it is to use ARKs in ArchivesSpace if you're interested. So the first thing I'm going to show you is what happens when you create a brand new record in ArchivesSpace. Now, all of this is what happens once you've set ARKs enabled to true, if you still have ARKs enable to false, these things that I'm showing you won't be there. But I'll create a brand new record, a brand a new resource in this case, fill out the minimal amount of information that we need to make a valid record in ArchivesSpace.

C

Christine Di Bella 32:06

As you see down here, there's a field that you if you don't have ARKs enabled, that doesn't

show up in your staff interface record for a resource right now, external Ark URL. So that's where you could put an ARK that comes from another another way that you generated an ARK, not something that has been generated by ArchivesSpace, you can use that there to specify that. I'll finish up making this record, you'll see that there isn't a place in the edit view of a record to put in an internal ARK. And that's because an internal ARK is generated by the system, you can't actually like input an internal ARK. So that's why it doesn't appear as a field that you can change in the edit mode of record.

C

Christine Di Bella 33:12

Okay, there's a couple more fields that are required. But I'll give a plug here for using the default values feature of ArchivesSpace. And I didn't make you sit through me creating those. Okay, so we're still in edit mode of ArchivesSpace, still see external ARK URL is blank, you still see that there's nowhere on this view where you can see the internal ARK for the record. But if I go over to view mode for this record, you will see that there's something different there. Okay, this is view mode, read only mode for this record. So at the sort of the administrative information for this record, the bottom of the basic information section, you see that in addition to the URI, there's also an area for the ARK. So meaning that there was an ARK created when I saved that record. If you take this ARK, and you put it over into the browser address bar, you'll see that this resolves to that record. And you will see though, that in the address bar, it it's the regular old URI for this record, rather than the ARK.

C

Christine Di Bella 34:42

So that's what happens with the brand new record. I'll show you what the a record that's already in here is and I did not plan this at all. But this is actually the same record that Seth showed because UNLV had given us some test data at one point to use in different kinds of demonstrations. So I mentioned that when you enable ARKs, records that are already in your system will not get ARKs until you run that background job. So I just wanted to make sure that you saw this in action. So even though we have ARKs enabled, this record predates the enabling of ARKs, so there's no ARK over here. If I want to have an ARK on this record, I'm going to have to go into the background jobs and run run a background job over the whole thing. So and this is part of this is so that you have some flexibility in determining sort of, if this is using ARKs or something you want to do from a particular point forward or if it's or if you're really using external ARKs, and you don't need sort of the system load of having ARKs created for everything. There's different reasons why you you might not want to have at the time that you upgrade to 2.7 or the time that you enable ARKs in the configuration file, why you might want to wait to do that. So we've handled this instead by giving you a background job that you can run at the time that

you're ready to have ARKs for everything in the system. So this background job is very basic. And all it does is run across all the records in all the repositories of your system, and add ARKs for resources and archival objects that don't already have an ARK. If they already have an ARK, it will skip those. And also it will create them for all records that are in there, but it won't touch the external ARKs field at all. So if you have external ARKs in there, it will leave that alone and external ARKs still take precedence. So even with an external ARK and an internal ARK in there, it will, your external ARK will still take precedence, regardless of whether you've run this background job or not. That is, that is it for the demo. So I'm going to stop sharing now. For that, I'm going to put up one slide for Jessica and Jessica will handle the q&a and welcome any questions that you have.



Jessica Crouch 37:25

Thanks. Thank you, everyone. Yeah, we'll go ahead and open the floor to questions and comments. If you'd like to ask a question, you can feel free to unmute yourself and just ask it. But I see you already have a question in the chat. Feel free to just put your comments in the chat. And I will read any questions that come through there. So Harvard University Archives asks, what happens if you merge or delete a resource or archival object?



Christine Di Bella 37:50

So with as long as the original record stays there, as long as they're sort of the original thing that you were merging into, that wouldn't be a problem. But if you're completely round tripping records, and you've deleted the original record, then that ARK will no longer apply to the to the archival object, you'll get a new ARK for the new archival object. ArchivesSpace treats it treats a delete as a delete. So there's still an ARK, there's still a table that has the former ARK. But there isn't right now, the ARK, it still doesn't stay the same. If you're creating essentially a brand new archival object out of two, it has come up with the resource, so I answered that in terms of archival object rather than a resource. But with a resource, you definitely have the option of sort of keeping the if you're just looking at ARKs, at the top level at the resource level, we would recommend that rather than like, deleting completely the resource and starting from scratch, we would recommend like keeping that high level top level of the resource and only inputting things at the lower levels. And then your ARK is gonna stay the same at the top no matter what. And that would that would work better for that kind of workflow.



Jessica Crouch 39:28

Or if you really needed to delete that resource record, you could just use an external ARK, right?



Christine Di Bella 39:33

Yes, oh definitely, yeah, absolutely. So an external ARK would be a better choice in that kind of situation.



Jessica Crouch 39:41

Question for John. If you have an item, say a medieval manuscript for which you have a catalog record, and that you have digitized and put online would you recommend having one ARK for all of those things actionable via different services or three ARKs for each? The latter seems like a requirement for expressing relationships and linked data.



John Kunze 40:04

Well, you have choices that you can you can actually do, you can have one ARK that leads to that is gets you to a sort of a landing page experience from which the other ARKs are linked. And that's often done where you have an ARK to an overall object and you have some sort of sub objects, maybe variants of it, or parts, components of it. And that is a common, common practice. So for the linked data applications, you can use the sort of extended, the sub ARKs, and for the sort of landing page experience, you can publish the overall ARK.



Jessica Crouch 40:52

Thanks.



Jessica Crouch 40:55

Next question. We have a lot of question in the chat. Can you batch import ARKs in ArchivesSpace? Christine?



Christine Di Bella 41:02

So if you have batch import, external ARKs, ARKs that you already have? Or basically, my answer is probably going to be the same either way. So I think the API is probably the best way to do that. If you if you have a lot of existing records with either existing ARKs, or at some point, you've gone through, and you've created ARKs in ArchivesSpace. But you've made some modifications or something like that, generally, the API is going to be best for that.



Jessica Crouch 41:38

Yeah, thanks. And if you want to follow up on that question, feel free. Thank you. Yeah. Next, if one were to move to a different system, how difficult is it to move internal ARKs to an external resolver? So that they would exist independently of ArchivesSpace?



Christine Di Bella 41:58

I would think, I mean, if would think it wouldn't be that difficult in the sense that the ARK that's being created, it has those characteristics that John mentioned, you know that it has a certain identifier, it has a certain format. I don't know, John, if you want to explain, like, is there any way that it would need to be registered outside of ArchivesSpace? Because that ArchivesSpace connection no longer exists? It would definitely need to resolve to something else if you weren't using ArchivesSpace anymore. But the core of the identifier, what makes it unique to that bit of description, that record of description, probably could stay the same, even at a different system.



John Kunze 42:47

Right? I'm not sure I know enough, right now. I don't know enough about the ArchivesSpace implementation, the ARKs that we were seeing in the demo looked, you know, localhost. And they were sort of short, like the 9999 slash two. That's an ARK that suggests it's not a sort of a it suggests with as five nines, that is not a real ARK. that's a that's a NAAN that's been reserved for that. So to promote that to, you know, I probably have to know a little bit more about the situation. But generally speaking, an internal ARK that is sort of well formed, there's no reason why that that name can't move around with it. Wherever, wherever the object moves. The big challenge with any persistent identifier scheme, not just ARKs is when objects move, you successfully find a new home for them. It wants that critical requirement has been met. Then, given a well formed, persistent identifier, it's just a question of updating some table somewhere. Well, at the wherever the resolver is. So I think I'd have to actually talk to Christine a little bit more to understand that. So I'm sorry, I can't make it. It's only half an answer. Really.



Christine Di Bella 44:28

Yeah. And I will say in terms of the 9999. I got lazy in the demo. And that configuration setting that you saw that is specifically for putting in a NAAN in that wasn't it, that's the designated not a real NAAN. So if it had a real NAAN, yes, that it should look like anytime that you're using. If you're expecting to use ARKs and make a valid ARK, you would need to put in that configuration file in that setting. your actual NAAN, 999 won't cut it.

J John Kunze 45:03
that's helpful. Yeah.

J John Kunze 45:09
Okay, thanks.

J Jessica Crouch 45:11
Um, how do you track deprecated ARKs?

J John Kunze 45:19
I can take a quick stab at that. I mean, it depends on the system. Of course, I've been involved with the EZID system that surface mentioned. For a while and, you know, deprecated stuff happens. So, to deal with it in EZID, there's a flag is set. So, as Christine mentioned, the important thing is to keep a record of what the ark was once bound to. So you don't forget that and, but you can set a flag that says, basically, this is a tombstone, when someone tries to click on that ARK they should get that's not there. But here's, here's what it was bound to. So that's kind of the advice that most people tend to follow in spirit about what's different implementations that I'd say. Okay, thanks.

J Jessica Crouch 46:17
Is it possible to configure ARKs to include the use of data set ID in addition to the NAAN? In a large organization, there can be multiple ArchivesSpace implementations, and we'd like to be able to configure each to use different data set IDs.

C Christine Di Bella 46:34
So right now in ArchivesSpace, no, that is not possible. This first level of implementation is definitely very simple. In terms of the complexity in implementing it that way, we did look, I will say we did look at some options for making this repository specific say, because within a within a single ArchivesSpace, there may be different repositories that are actually different physical entities that have gotten different NAANs. And so we could definitely see that use case. But it added substantial complexity to this from a technical standpoint. So we decided to post you know, we decided to see what sort of the level of interest was in this feature and how people were using it before sort of jumping into that. Data set might be another thing, where, right now we couldn't say what the level of

complexity would be for implementing it, it would definitely be more complex than what's there now. It's not supported by what's there now. But you could certainly make a feature request for that.

J Jessica Crouch 47:54
Thank you.

J Jessica Crouch 47:58
We've got several more questions. But while we're chatting, Seth, can you post the GitHub links to your batch updating scripts in the chat. And of course, all of the slides from today will be included with the recording of this webinar. And I know Seth did include those in his slides. Another question for John, should the ARK be displayed, and or the URL of the record page instead of redirecting to another non persistent URL?

J John Kunze 48:29
Yeah, that's a very good practice. That is, the burden is on the sort of the maintainer of the objects, the provider of the objects, it's really easy for the persistent identifier to get sort of separated or lost from the original object. The browser makers have this habit, when they redirect to a URL that may be very persistent, they'll still always display the least persistent URL. So you get to the thing you want and your first instinct is to bookmark that page. And of course, you're saving exactly the wrong URL. That's why best practice for archiving institutions is to, if you can prominently display the firm lane for the persistent URL on that page, the way Seth showed us with, you know, here's how to link to this page, because it will frequently be wrong in the address bar. There's probably more to say on that subject. There's a kind of a tension there with the browser makers who don't and also some security experts. So there's a there's an argument that has to take place to see if, you know the browser makers can be persuaded to somehow help us

C Christine Di Bella 49:51
on that, John, right there that is very helpful feedback for us, because as I showed in the demo like it, it is Different thing that shows up in the browser, the ARK doesn't show up in the browser. But currently, our template for the public user interface doesn't show the ARK. So that that alone, like that sounds like an improvement that, from a best practice standpoint is would be important for us make to the system.

S

Seth Shaw 50:25

We're also planning on in many cases, using our ARK as our local URL that we display, rendered, the resolver will be different. But as long as we can keep using that ARK as our URL, as we were systems, that bookmark will least be semi permanent, as long as we're still using the base URL for it. Even though we still publish using the the actual resolver, that things are listed as.

J

Jessica Crouch 51:03

Great, thanks. So this next question, of course, the three presenters are welcome to answer but if anyone in the conversation if this applies to you, please feel free to chime in. Do you know of an organization that's generating ARKs for several different systems, including ArchivesSpace, preferably one where there's a NAAN for the whole organization.

S

Seth Shaw 51:23

So I'll quickly answer where we are. So we had showed just in this particular webinar, how we're using it with ArchivesSpace, but we also create ARKs for all of our digital objects in contentdm. We also have some ARKs for some of our web archives in archive it. So we already do have ARKs and a few different places. And one of the nice things about all of our ARKs and contentdm is where the progress of migrating our digital asset management system content dm two and Islandora eight site and all of those ARKs, are coming with us. And so when we go live with the new site, I'll be able to use that batch update script to say, Here's all of the new URLs for all of our ARKs, batch update. All of the ARKs that we've already published for these digital objects will be automatically updated. We're actually getting a lot more use out of it, promulgating our digital objects in our fighting games, but within the context of this particular webinar. Yes, we are using it for finding aids as well.

J

John Kunze 52:36

I can take a quick stab, I'm not sure. It almost seems like there may be two questions in here. But if one question might be, if there's a there's a whole there's one NAAN for the whole organization is if there are like sub organizations. In this example of the French National Library, very large organization has NAAN, one name is running already number. And they have seven completely autonomous departments, some of which don't like even like each other. It's like they've extended their NAAN a little bit to create something that we've called shoulders, basically, it's a kind of like, an extended NAAN. And they just have the discipline to assign their ARKs separately within that namespace. I don't know if that gets to the heart of that question.

S Seth Shaw 53:31
I'll mentioned that this concept of shoulders, we also use that so all of our finding aids have an ARK that start with after the NAAN start with an F one, I think maybe it's an F zero. And all of our digital objects start with a D. Moving forward as well. So we also do that.

J Jessica Crouch 53:59
The Smithsonian is also using shoulders. They offered that in the chat.

J John Kunze 54:07
Yeah, Seth and Smithsonian are both using EZID so they kind of get we kind of forced the shelter concept. I think, Well, we've, we've exposed people to the shoulder concept and the there's like there are different but anyone could use them. In theory. They're just a concept that EZID has documented uses.

J Jessica Crouch 54:35
Question about ARKs in ArchivesSpace. Any plans for implementing nice opaque IDs in ARKs?

C Christine Di Bella 54:44
I would say that isn't currently on our radar. But it's another one where we'd be very interested in feature requests. We'd be interested in seeing a feature request and I see lower down there's a comment about Is there a working group currently being planned? There is not. But certainly, if there are a lot of community members interested in this, and it seems like there are, we'd be interested in, in hearing about what kind of effort people might want to put together either on their own or with some help from us. To talk about this more, this was definitely a very simple, not necessarily simple technologically to put together implementation, but definitely pitched at like, minimum viable product type thing for getting it into the application. There are many more ways to build on it. So if there are community members that are interested in doing that, I think that there's a lot of possibility for making that happen.

J John Kunze 55:47
And there's a strong possibility that there will be some money to fund a developer at CDL

to create a sort of a next generation noid. That's the software that generates this this form of the identifier with check digits and nice randomization that may come through. If it does, it'll be open source, probably in Python. Happy to talk to anyone who wants to help and Christine as well, if you could use that.

C Christine Di Bella 56:20
That'd be great.

J Jessica Crouch 56:26
Awesome.

J Jessica Crouch 56:27
I'm going back to the question of using shoulders for different departments or divisions, the Smithsonian would just like to point out that they like each other. You can like each other and still do that. Seth, I feel like you now have to say that you like each other.

C Christine Di Bella 56:44
It's a lot of pressure.

S Seth Shaw 56:45
Yep. No, we like each other but I manage all those systems that we use the shoulders for. So I hope I like all of the things I do.

J John Kunze 56:53
I should not have said I should not have said that. There are people who don't like each other. That was just that's wrong. That's incorrect.

C Christine Di Bella 57:02
Who knew this would be such a contentious gossipy webinar, it's very exciting.

J Jessica Crouch 57:08

I love it. I live for this. All right. There's a few more questions in the chat. We are at the top of the hour. So if anyone has to go I really appreciate you coming. But one question is the ARK stored in the external ARK field in a space exported in EAD? If so, what element?

C Christine Di Bella 57:27

I love this quiz. But it is the intention of both the internal and external ARKs is that they be included in in exports. And in OAI-pmh. As I mentioned, the original impetus for this was the OAI-pmh. functionality. So yes, it's there. And currently, at least in EAD2002. It's in the EADID URL. And I wasn't fast enough to get the EAD3 export out. But it's also in a field, tag in EAD3. And I can certainly look at that to see what it is. But Noah, I bet you can find that too.

J Jessica Crouch 58:14

Awesome. I think that is all the questions in the chat. Thank you everyone for participating.

J Jessica Crouch 58:22

Wow, we've got more people texting things in. FYI, there are already some issues raised on Atlassian for how and where ARKs are exported in EAD.

J Jessica Crouch 58:34

That's from Andrew.

J Jessica Crouch 58:38

But it sounds like there's a lot of interest about using ARKs and ArchivesSpace. And maybe there is there is a group there. One final question that I have for John is if I'm if I'm adopting or considering adopting ARKs, and I want to reach out to other users for insight or advice, you mentioned ARKs in the open is that where I should go to interact with other users.

J John Kunze 58:57

It's a place to start that'll it'll quickly point you to, for example, a mailing list the ARKs forum. That's the main place to for discussion. And there's a FAQ and places do ways to get involved and express your interest. get onto our working group.

-  Jessica Crouch 59:21
Awesome, wonderful. Thank you again to John, Seth and Christine for this webinar. And thank you everyone for attending and for staying a little late.
-  John Kunze 59:33
Pleasure.
-  Jessica Crouch 59:34
Thank you. Yeah, as a reminder, our next webinar will be April 8 on integrating with Alma, I hope to see you guys there. And have a great day.
-  John Kunze 59:45
You too. Bye.
-  Jessica Crouch 59:46
Thanks, John.