Minerva26

From “We Can’t” to “Here’s How” — A Practical Discussion on Hyperlinked Files in Discovery

Kelly Twigger: 0:11

Welcome to the Meet and Confer Podcast. I'm your host, Kelly Twigger. I'm a practicing attorney and the CEO and founder of Minerva 26, a discovery strategy platform that helps litigators create and validate discovery strategy by connecting rulings, rules, and workflows so they can make faster, defensible decisions that protect credibility and control costs. If you're afraid of what to do with ESI, we're here to help.

Kelly Twigger: 0:39

Today I'm joined by Arman Gungor, who also has two roles in the e-discovery space. Armand serves as the vice president of forensics at Meridian Discovery and is also the founder of Metaspike and the creator of Forensic Email Collector and Forensic Email Intelligence. Arman has been right in the middle of the hyperlinked files and modern attachments battles since 2021 when Judge Parker's decision in Nichols v. Noom first pushed his work with Forensic Email Collector into the spotlight.

Kelly Twigger: 1:12

We're going to unpack what hyperlink files really are, what Google and Microsoft actually give you today, how tools like MetaSpike's change what's reasonable in discovery and what all of this means for your obligations and your strategy. If you've ever looked at an email with a link instead of a PDF and thought, what exactly am I supposed to do with this? This episode is for you. Let's dive in. Arman, thanks so much for joining me today. I really appreciate you being here. Thank you.

Arman Gungor: 1:42

Likewise. Thanks for inviting me.

Kelly Twigger: 1:44

We have a an important topic to talk about today that I know is something you became quite famous for in 2021 when Judge Parker wrote the Nichols v. Noom decision, and the plaintiffs argued that Noom should use your forensic email collector tool to collect hyperlinked files in that particular case. And the court actually declined to order them to do so because they'd already done a production. But

I think it was the first time for me that your tool and you became in the spotlight of this hyperlinked files issue. So I really appreciate you being here today.

Kelly Twigger: 2:24

And I wanted to talk to you about this because I think one of the challenges that we have in Discovery now, because it's mostly all ESI, is this notion that technology keeps moving forward. And we as lawyers, our obligations keep changing. We keep needing to understand more about the technology and the technical aspects in order to make uh crucial strategic decisions for our clients in terms of discovery. And hyperlinked files is one of them. So I will I'm gonna turn it over to you to let you describe the problem and I'll jump in.

Arman Gungor: 3:09

Yeah, thank you. Sure. So what we mean by hyperlinked files is when there's an email, we are used to having the conventional attachments of emails, which will be actual files. But what happens when somebody instead of sharing an actual attachment, they share a hyperlinked file in the body of an email. And that could be a deliberate choice, as in you know, somebody just clicks the share button and they share a file with someone, or it could be something that's enforced by the provider. For example, Google does this for, I believe, attachments over 25 megabytes. And if you try to attach something large, it'll become a Google Drive file, and then a hyperlink will be embedded in the email. So what happens if somebody shares a file? And it doesn't have to be a file, it will also be a folder. So if somebody shares a file or a folder, what do you do then? Does it become part of the email? Is it within the scope of discovery? What happens if you don't include that at the at the time that you perform the forensic preservation? So that's the challenge. I think the first part of the question, as in, you know, is it part of the email's context, is pretty straightforward. I think it is definitely part of the email's context. And if you look at those emails and somebody's talking about the attachment, and if it's not there, then that's that's trouble.

Kelly Twigger: 4:20

So that part is You know, that interesting point that you raise because Judge Parker, in that one decision I mentioned to you, said that it was not an attachment, but the courts since then have pretty much conceded that hyperlinked files that are referenced in a communication are attachments. Can I just clarify two points for you? One is we can have hyperlinked files anywhere, right? It's not just limited to email. They can be in text messages, they can be in WhatsApp messages, any kind of communication where you're sending text and you can include a link. A hyperlink file can be in any of those. Okay. All right. And I had a second point which I've completely forgotten, so I'll come back.

Arman Gungor: 4:59

Well, that's that's all right. But you know, when I was talking about that, I kind of danced around that attachment phrase because yeah, maybe technically they may not be called attachments, but what I said is are they in the context of the email? And I think, you know, the reasoning for how the production is defined and how we always talk about families and parent-child relationships, the fact that if it's an attachment or something else is not really that relevant. I think what's important is is it in the context of the email? As in when you review the email, do you really need to refer to that file to get the full context of what's going on there? And I think the answer to that is yes, you have to have that context to decide what's happening. And I've experienced this myself personally. I've had some cases like that where even not in a corporate scenario, individuals using regular free Gmail they would exchange some files as Google Drive hyperlinks, and when you come to examine those emails, or even in the discovery context, when you need to review them and make decisions, if those files are not there, you just let lost the whole context, as if somebody just dropped the conventional attachments of an email. So contextually, I think it's fairly straightforward. But as far as the production goes, should it be your standard operating procedure? I think that very much depends on scope and again context and proportionality. Because in a case where there are a few mailboxes and it's not overly burdensome, I think it's an easy decision. When it scales up, there are definitely some challenges, and they depend on the provider as well. Providers have different takes on this, they have different APIs and they have different capabilities, and there are varying levels of difficulty in getting those Google Drive or Microsoft OneDrive or SharePoint files out of the provider and linking them with the emails. So that part is a little bit open to discussion, and maybe it's like on a case-by-case basis. Parties need to figure out if it should be considered in context and part of the production. But that is really the challenge that we're trying to solve. And yeah, you mentioned that 2021 case and the landscape has changed a little bit since then. And I'm happy to see that the providers have kind of tried to adapt. And I saw that Google actually updated Walt to include that hyperlinked attachment support last year. So now, even if you don't use a specialized software, but if you're doing a regular Google Walt export, you could get hyperlink Google Drive attachments out of Walt.

Kelly Twigger: 7:29

So I'm going to stop you just to make sure everybody's on the same page because a lot of folks use Microsoft, right? So you either use Microsoft or use Google. Some people will use Google for their free email, so they might use it for their personal mail, but exporting it for purposes of discovery never kind of comes into play. So for folks who use Google Apps, of which Google Mail is one, there's an archive solution that can be turned on with Google Mail called Google Vault. And what you're saying is that last year, so in 2024, Google Vault added functionality to it that allows you to collect hyperlinked files directly from Google Vault. Is that right?

Arman Gungor: 8:10

That's right. There is definitely some nuances there. It doesn't get you 100% of the way there because of, I would say, two main things. One of them is the way Google Vault is doing it now is a very efficient way, and you can export a number of custodians all in one go, kind of similar to what you do with Purview. And whatever Google Drive files are associated with those mailboxes, the related Google Drive files will be exported all together as a single instance. So if that Google Drive file has been referenced 75 times here and there, you're gonna get one instance of that. And there's a reference file that allows you to kind of match it up with where it should go.

Kelly Twigger: 8:50

Now this so the reference file allows you to match up the communication. So let's say in this case, the communication is an email to the actual file. And if the file, if that if there are 75 times in the collection, there are 75 different emails in which that file is referenced, it will only be collected one time, but there is a ref, what did you call it, a reference ID that you can match with the communications to the file so that all 75 communications would have that reference ID to that one file.

Arman Gungor: 9:26

That's right. So which makes sense from an efficiency standpoint. There is no point in recollecting the same Google Drive file 75s, which we did with attachments over and over and over again.

Kelly Twigger: 9:37

So this is actually more efficient.

Arman Gungor: 9:39

Yes, efficiency-wise, that's perfect. One part they're missing is that they don't offer a way to reconnect those things for you. So that's left to the user, and not every user is able to do that. So when you want it to be ready for e-discovery ingestion, you have this set of emails and you have this set of unique instances of Google Drive files, but they're not really in any way connected directly. So you have a couple of options. One of the options is you would have to write your own process or scripts or whatever and reorganize them in a way that has some type of connection that is compatible with your e-discovery solution. When we saw this, we jumped on it as well and we built that into forensic email collector. So we wanted to take that as a starting point, as if you're collecting from Google's APIs directly, but start with a Google Vault export and reorganize it in a way that kind of mimics a forensic email collection. So everything is packaged. There are Google Drive attachment packages where there's the parent email and the conventional attachments and the Google Drive attachments in one place for that parent-child context. But that was one of the things that they're missing. So that support is there. It's not a hundred, it doesn't get you 100% there if you don't have the practices and what you've got to have, you can't do it yourself, right?

Kelly Twigger: 10:55

I can't export from Vault. And I'm fairly savvy when it comes to those things, more so than most litigators. But you can't I can't do it myself unless I've written some sort of script that allows me to take that ID and link it back to the email communications. Yeah, and that's one of the things that forensic email collector, your tool, does.

Arman Gungor: 11:13

Yes, if you have if you had just two, three emails, then you could probably link it yourself manually. But if you have millions of emails, then you definitely have some type of automated process to do it. And it's not an insurmountable problem, really. It's solvable, but it does have some challenges because part of the problem also is that not problem, but part of the challenge is that the emails come in containers, so you don't get individual emails, but you get either PST files or inbox files, which are essentially a bunch of text-based emails combined together, like concatenated. You have a question there.

Kelly Twigger: 11:47

Yeah, I do because I want to make sure. So inbox, the inbox container is what comes from Google, just so listeners understand the difference. So the inbox container is what comes from Google, and the PST is a Microsoft container creation, right?

Arman Gungor: 12:01

Yes, but Google also exports in PST format to to accommodate users who want that, I think. So you have both options when you do, right?

Kelly Twigger: 12:09

It had always been inbox, was Google's standard functionality, but they've since added PSTs because Microsoft users wanted it.

Arman Gungor: 12:16

I think it's been a few years that they've had that functionality, but yeah, initially it was Mbox only. And the reason for that is clear on Google's back end, these emails are MIME emails, so they're text-based emails. So in when you stay in that realm, then the container format for that is Mbox, which is essentially a bunch of MIME emails just concatenated together with a little bit of formatting sprinkled in there. So for Google to go from that to PST is kind of artificial in a way, because there is no reason for Google to have PST output.

Kelly Twigger: 12:46

What's the difference between an Mbox and a PST?

Arman Gungor: 12:48

PSD is Microsoft's format, and they have a whole different take on this. They have essentially individual mappy properties, is what they call them. So it's like a container individual what properties? Mapy. So Microsoft's specification and architecture for email, essentially. And in a MAPI container, we have a bunch of different properties, and they are stored in some type of fashion that's not really defined in the specification. So you could have it be anything you want, but Microsoft's take on is is to use the compound file binary format, which is essentially what used to be the Microsoft Word files and stuff like that back in the day. So that there's essential that container, and inside that there's a bunch of streams, and each stream contains a mapping property. But the gist of it is that there's a bunch of different compartments in an email, and each compartment, each property stores a different aspect of the email. So one of those properties could be, for example, the headers of the email, one of those properties could be the text-based body of the email, one of them could be the HTML-based body of the email, and a ton of others. And presumably the reason for this is that Microsoft's architecture is a lot more complex than just email, right? So you have Outlook. In Outlook, you have emails, you have calendar items, you have contacts, you have the possibility to add warning buttons into email. So there's a lot of rich stuff there that doesn't really fit with the traditional MIME standard. So you can't take all that rich functionality and turn it into a MIME email without any extra bells and whistles. So to be able to store and house all this, they have to have a database type system, and that is essential their map containers. And when they export, they want to preserve all that stuff that they have on the server. So they put it into a uh PST mapping container that houses a bunch of mapping properties and individual messages in that in essence. So for Microsoft, this makes sense because they have that kind of architecture on the back end and they kind of reflect that when they export it out. For Google, when you export individual email, you see that they have them as individual MIME messages. So for them, it makes sense to export Mbox files because they're just concatenating a bunch of them and getting them out to you.

Kelly Twigger: 15:01

Make it really, I'm gonna break it down. Google, simple. Microsoft, complicated.

Arman Gungor: 15:06

Well, you could maybe you could boil it down to that. I mean, Google has a lot of complexity too, but in different ways. But from a compliance standpoint, I think that that's a fair assessment. Google Vault and Google's take on email seems to be a little bit simpler and easier to grasp than Microsoft because of all the stuff that they've interconnected. There's exchange, there's teams, there's there's a lot of stuff going on there.

Kelly Twigger: 15:28

Yeah, just listening to that last two minutes that you gave me, which for me being highly technical is interesting to me. I'm guessing everyone else is just a lump over their heads. Because it just that's what makes this, that's what makes eDiscovery so hard for lawyers, is there are so many responsibilities that litigators already have, and being able to understand this technical information is difficult. And it's just another job. And so folks like you who provide forensic services in litigation support who are solving these problems are so critical. But I think just to take it back for a minute to the high level for the lawyers, because when we used to collect email with physical attachments, which very rarely exist now, we had an email, then we had the actual attachment. And when they showed up in our review platform, we would have the email actually there, and then the very next document would be the attachment, and we'd have metadata that showed the parent-child relationship, and we'd have all the metadata for the attachment as well. Now that we're collecting documents at hyperlink files, how are we able to deal with the parent-child relationship? You talked a little bit about that ID structure from Google, and I want to talk about it from Microsoft, but the version is another issue that we come into with that relationship. And that is if I sent an email to you today and I send a link, let's say I send a link to the notes for us to prepare for this podcast, and you go into those notes and you make a change to that, um, to those notes, and so there are additional edits, and you send me an email back and say, Yeah, I've done some updates. Let me know what you think. If we go and collect those emails, when we collect both of them, let's say they're a Google will give us an ID that will allow us to connect that same document to both emails, but won't the version be the version that exists at the time of the collection? It's not gonna allow me to collect the version that existed when I sent it to you and then the version that existed when you sent it to me. It's only gonna collect the most recent or what we call the contemporaneous version.

Arman Gungor: 17:51

So before we get into that, let me backtrack and finish that last thought we had because the reason we got into Mbox versus PSD was one of the challenges in the Vault export, which I was saying that they come in a container. So the challenge that's associated with that container is that you don't have individual messages to go and tie to the individual items in that reference file. You have a container full of emails, so you have to also figure out how within that container the individual messages tie with those messages.

Kelly Twigger: 18:18

So it's another level of complication with the PS.

Arman Gungor: 18:21

There are some breadcrumbs there that we use, and there's a good 100% accurate way to match them up, but again, another challenge to solve basically. The second challenge was what you're just getting into now, which is the versions problem. And Google Vault currently does not do this. So if you get the hyperlink files from a Google Vault export, it only gives you what we call Google Drive attachments, and which would be the current version of those Google Drive files. But it doesn't give you any older versions that fit the email. But what you mentioned is true. So if you have a situation where there's an email and there's a hyperlink file, and the email is let's say sent a few years ago, and the legal team is interested in seeing what that file looked like back at that time as the email was being sent, then you get into revisions in the Google world and versions in the Microsoft world, similar concepts. So the good news is the providers track those versions of the emails. So when you talk to, let's say, Google Drive API, you're able to say, here's my file ID, and I want to see what revisions of this file you have, and then you get a list back, and then you can see what the dates of those revisions are, and you can figure out okay, this revision probably goes with this email based on the timing. So you can pick the appropriate revision, or you can get all of them if you feel like seeing exactly how that file changed.

Kelly Twigger: 19:40

So if I'm a if I'm a receiving party, I've received a production from you, and in that production, and I'm gonna assume for purposes of this that you and I have agreed as counsel that you'll provide content to me, you'll you'll provide metadata that tells me if there's a hyperlink in a message, in a communication. I can then filter in my review platform once I've loaded your production by the documents that have hyperlinks in them. I can then decide, reviewing those documents, which hyperlink documents I want. And in addition to telling you which hyperlink documents I want, I should tell you that I want which version I want.

Arman Gungor: 20:29

That's true. Usually, in my experience, this mostly happens at the time of collection. So as opposed to a two-step process, the parties decide, okay, we want the Google Drive attachments to be in scope. And then when collecting them, we want to either collect the current version for all of them, or we want to choose the specific version that goes with each email, which we call the last revision before the sent date of the parent email, and you automatically get that version and you have that in your database. But in some cases, internally, it might make sense before production to do like a two-phase process where you don't actually get into all revisions and all Google Drive files and perhaps make an initial collection and turn it over to your internal legal team for preliminary review, and then they identify hey, okay, these are the emails where we are interested in the modern attachments, and we want to see this revision based on date, for example. Then they can get back to the to their technical team with a list like that and say, hey, we want you to go back now and pull these specific revisions for us for these emails, and then you can load that up.

Kelly Twigger: 21:35

Okay, so two options then. One the parties agree at the outset either to receive contemporaneous versions or the version that existed at the time the email was sent. If that's the option the parties take, is there a manual process? And I'm thinking about proportionality considerations. Is there a manual process for you as the provider to provide though to give me the versions that were created at the time the email was sent?

Arman Gungor: 22:09

It's automatic. So forensic email collector does that assessment for you. It'll take a look at the date of the email, it'll get a list of old revisions, evaluate their dates, and pick the right revision based on the timing correlation there. If you wanted to do it manually, then yeah, there would have to be some type of process around it to kind of automate that.

Kelly Twigger: 22:30

So forensic email collector will do that for you. It will give you the version that was sent at the time the email was sent. But is there built-in capability within Google or Microsoft to do that currently? Or do you have to have an outside tool, like for instance?

Arman Gungor: 22:44

Yeah, I don't think they do that. Google Vault just exports today's copy of the to the modern attachments, and I believe the same for Microsoft right now. So there is no you know, specific revision for an email at the moment.

Kelly Twigger: 22:57

Okay. Just I'm just gonna clarify one more time, just because I know everyone listening, this is gonna be an important point for them. And that is that if you want to be able to collect the version that existed at the time a communication was sent, you're likely going to need to use a third-party tool like a forensic email collector to be able to do that, because that functionality is not native to Microsoft, whether that's purview or to Google.

Arman Gungor: 23:26

It's not built into their ready-made discovery tools. It is native to them in the sense that it's part of their API. So if you know how to talk programmatically to their systems, you could do it yourself too. But it it requires you to develop some software basically. So if you're looking at off-the-shelf products, then you're right. They're built-in e discovery tools, don't do that currently.

Kelly Twigger: 23:45

Okay, great. Thank you for that clarification. I'm gonna ask you one more thing, and that is because a big part of this podcast is trying to help lawyers learn the technology and the terminology so they can speak it and feel more comfortable with it. And we've used the term API a couple of times. Can you explain what an API is?

Arman Gungor: 24:03

Yes, it stands for application programming interface. So it's a special interface that a provider offers for usual software developers to be able to talk to their systems in an automated way and to do things there. That's what we use. For example, when we talk to Google, we use the API to make specific requests, such as, hey, give me this file or give me this email. What's the data on this email? Kind of automated communications with that service.

Kelly Twigger: 24:29

Great. Thank you very much for that. I'm gonna switch a little bit because part of what prompted our discussion was the discussion that you had with Rocky Messing and Tom O'Connor, I think it was almost a year ago now, about this very topic and about some of the capabilities. And when you and I were talking, prepping for today's discussion, you said there's been a lot of changes since that call that we need to talk about in terms of the capabilities of the tools. What are those updates in terms of what's changed in the last year? And what do those continuing changes mean for us as litigators and what we need to be thinking about in the hyperlink files area?

Arman Gungor: 25:13

It is constantly changing, it's an evolving area, as you might expect. So on the Google side, like we mentioned, there's support now natively from Google Vault to export hyperlink files. So that was kind of a big change, and we kind of matched their speed there and we built FEC around it to support the same thing. On our side, there have been some additions in the model attachment area in terms of Microsoft. We added first support for acquiring OneDrive files and SharePoint sites entirely without related to emails, so essentially just collecting OneDrive as a whole or SharePoint as a whole. And very recently, we also added support for OneDrive attachments of emails or SharePoint attachments of emails. So kind of the same path that we took with Google, replaying that same playbook on the Microsoft side and starting with the OneDrive attachments, and then we will continue to add the advanced features to bring it on par with what we do for Google. And a couple of things we did recently is also around that a mirroring of cloud attachments with their emails. What we initially did was what we call drive attachment packages, which was essentially a bundle of parent email conventional attachments and drive attachments and revisions all in one package. And now we took that a step further recently, and we do something called staging now, which is essentially an extra preparation step that takes kind of cherry picks either the original email or the modern attachment bundle, depending on whether or not that email had modern attachments. So let's say that you have five emails, only one of them had a modern attachment, and the other ones did not. So in the output, you would have four emails as just emails, and then one of them would come with the modern attachments, essentially making it ready to ingest right away. Before they were in two separate places, and it took a little bit extra preparation step for people to reorganize them and get them ready to throw into some review or rediscovery tool. So we made that a little bit easier, mostly for small to medium-sized firms. I would say our large service providers have internal processes that already do this automatically, so I don't think they needed that. But smaller service providers or law firms or small companies that do this kind of stuff internally, they don't have the resources to build a process around it, usual. So we're trying to help them get there pretty much in a plug-and-play fashion. So those are the main differences. I would say the most remarkable thing is probably the OneDrive attachments, which is something that a lot of people have been creating for. So that's new, and we'll see where we'll take it from there.

Kelly Twigger: 27:53

And those are all that's all functionality that's in the forensic email collector tool, right? You've been continuing to develop that as technology has changed with Google and Microsoft.

Arman Gungor: 28:02

That's right. The first part that I mentioned, Google Walt has support, that's definitely their own thing. But outside of that, yeah, all the staging packaging and OneDrive attachments are all in forensic email collector.

Kelly Twigger: 28:12

And essentially what you just described in terms of the stag, the putting everything together in a package that I can then load into a review platform, will that show up the same way that physical attachments used to in a parent-child relationship with metadata that supports it?

Arman Gungor: 28:28

That depends on how the review tool handles it, because it's not exactly an email anymore. That's probably something that's worth clarifying too, because a lot of people have this question. And the question when you collect modern attachments, how do you output that? Do you uh insert them into the original email or do you keep them outside? And I think that's an important consideration. And the answer to that is we do not put them in the original email. And we have a few reasons for that. One of them is that once you open up that original email and insert modern attachments into it, then you make very substantive changes to the email. All the metadata changes, my boundaries change, the digital signatures don't verify anymore. So you just create a whole new email, is which is not desirable. The second thing is when you consider the context of information exchange and productions, if this is not clearly communicated, it could be a misunderstanding. Because if I received an email like that with a bunch of modern attachments, I might be misled into thinking that that's how that email originated, but it's not really case, it's a constructed email that's pieced together from different sources. So that's another potential source of confusion. But probably the most compelling reason is that cloud attachments can be very numerous and very large. Let's say that an individual shared not a Google Drive file, but a Google Drive folder in email, and the folder contained 10,000 items totaling two terabytes in size. If I took that 10,000 items and two terabytes and shoehorned it into that one email, they're gonna all have all sorts of processing problems on the As you throw that into your review tool or do anything with it, really. So it's really almost unfeasible to do that. So we chose a different path. So instead of modifying the email and putting them in there, we create essentially zip packages where in the zip there's the original email. Like I said before, there's the conventional attachments that are already in that email. And then the cloud attachments and revisions are also in the same zip container. So if you have a review tool that automatically treats zip containers as a family, then that would be pretty much a no-brainer. You'll be good to go. If you have any discovery tool that treats zip containers as folders and doesn't really care about the parent-child relationship in there, then you might have to do some extra work there. So that depends entirely on the solution. I know that a lot of our users work with the mainstream e-discovery hosting providers, the big names, and they all have found ways to do that in an automated way to link that to have that parent-child relationship. So I think for the most part, it's an unissue now. Maybe in the initially it might have been a bit of a challenge because it was new. But I think probably hosting providers have already encountered this enough that now they know what to do with it.

Kelly Twigger: 31:13

So yes, if you have your own litigation support team, and yes, if they've worked with these issues and come up with solutions for them. But let's say that represents about 5% of the population of litigators, right? That the other 95% are going, okay, I don't know what the questions are that I need to ask here. And so let's clarify from a just a high level as a litigator. If I'm talking to you as my head of litigation support or my person who's responsible for going and doing the collections, what are the questions or what are the things that I need to tell you so that you do the collection in a way I want it done? And the flip of that is what do I need to negotiate with the other side to be able to get to what I call hyperlinked files and you're calling modern attachments?

Arman Gungor: 32:05

Yeah, I think like the general description of you know what you need on the output side is important. That's gonna come down to the rest of your processing stack for the most part. I think it's gonna come down to you know what type of structures your YouTube recognizes as an attachment family.

Kelly Twigger: 32:21

So the That's tech that's complicated, that's litigation support, people hosting data, strict litigators. What are the things I need to say? So it seems to me I want the email, I want to know which version I want, whether I want the contemporaneous version or whether I'm willing to take the most recent version or the version that exists at the time of collection. What are the other things that I want besides version? Do I want metadata?

Arman Gungor: 32:50

Yeah, some sort of thing. Another important thing is you want to define or decide on how you handle folders, because there's a distinction between an individual file modern attachment and email, and then there's a different scenario when there's a folder. What do you do then? Do you ignore that folder, or do you want your tool or process to actually dig into that folder and extract everything that's in there? Do you want that to do it recursively if there are any subfolders that are in there? So that's an important distinction because, like I said, like somebody could share the root of their shared Google Drive for the whole team, and that could be really large.

Kelly Twigger: 33:25

So that's another that then so what I'm trying to articulate for listeners is do I need my team to look at the data first and understand what the issues are that I need to provide guidance on? And if this is the other side, how am I gonna ask them? Am I gonna say, hey, look, guys, I want you to go in here. I want to know, here's what I want to do on versions. If there are folders at issue, I want you to identify those, and then I want us to meet and confer about those specific issues. Is that a good approach?

Arman Gungor: 34:03

Yeah, I think in the initial meet and confer, the two things that I would try to decide on is are modern attachments part of this? And if so, are we interested in the current versions of these modern attachments or the contemporaneous versions like at the time that the email was sent? So that's the initial decision to make. Once we have that, internally you might have to do a multi-step process where you do like an initial assessment to see maybe the volume of cloud attachments you have. Are there any folders there that have been shared and how much stuff there is potentially in those folders? And are revisions-wise, you know, are there any specific revisions that are really relevant that should be produced? So there internally, you might have to do like multiple takes, and then there might be another meeting confirmed to confer to reiterate some of those points or discuss further. But initially, definitely you want to discuss modern attachments and the version issue.

Kelly Twigger: 34:56

If I'm working with a corporate client and I have access to their Microsoft email, how can I go into the email in appropriate custodians' mailboxes? And I'm limiting that knowing full well that we're moving away from custodial discovery. But let's say I'm going into specific custodians' mailboxes, how can I determine which emails or how many emails have an attachment, have a hyperlink to file associated with them?

Arman Gungor: 35:24

Yeah, that's that's a good question. So there's a way to access your custodian of choice or all of them using Microsoft's authentication mechanism. So authentication-wise, that's easy. But in terms of metadata and how to quickly determine which ones have cloud attachments, that's something that I do not recall seeing as a search term in Microsoft's tooling. Like uh Google has this, for example, where you could run a search and identify which emails have cloud attachments. I don't recall off the cuff if Microsoft had that, I believe not. If they don't, then you might have to do some type of initial metadata only acquisition to essentially get some information about those emails and then go from there.

Kelly Twigger: 36:07

But yeah, I don't think what metadata fields would tell you that there's a hyperlinked file included in the text.

Arman Gungor: 36:15

Yeah, I don't actually in in addition to metadata, I think to be on the safe side, you might even have to get the body of the message because there's gonna be the hyperlinks that are in the body.

Kelly Twigger: 36:24

So very so very difficult, if not impossible, then to determine how many messages you might have in an email collection that have hyperlink files in Microsoft's email.

Arman Gungor: 36:34

Off the top of my head, I didn't have a good solution for that, but I I will do some research into that to see if Microsoft has a mechanism to search for that.

Kelly Twigger: 36:42

I think what we're trying to do here is identify what the issues are, right? And that's what comes to what you and I discussed previously, which is the proportionality considerations, right? What are a party's obligations with regard to hyperlinked files when the email service that they use doesn't allow you to filter by which documents contain hyperlinked files, so you even have a sense as to how many are there. And then to be able to negotiate from there is very difficult. So it does feel to me, based on the technology as it exists right now, that in dealing with Microsoft email, that a number of the types of processes that parties have undertaken in case law is to do a sampling approach, right? I will provide you the email, you identify of the documents containing hyperlinks, which files you want, you come back to me. At that point, you might even be looking at which version you want of those files. And then those files are provided. And many parties, there are many examples in the case law of parties doing that sampling approach, right? We'll come to you with X number of requests a week. You get two weeks to respond to that, you give us those files, then we can give you X number of requests. Usually it's 20 to 50 every two weeks or so. I think the most recent example of that is the NRA Uber case. So do you know, Arman, if Microsoft or Google are going to continue to make additional updates to their technology to help resolve this problem from a discovery perspective?

unknown: 38:17

I don't know.

Arman Gungor: 38:18

First hand knowledge, but I can only imagine that they will, because this is an emerging issue. And from what they've done in the recent past, I think it's safe to say that they're expected to keep moving in the same direction. But yeah, I mean what you're saying is true, and there are a lot of challenges. There are more challenges than we covered. For example, in terms of identifying which items have cloud attachments, Google has some mechanisms to run some searches to identify them. But when you actually validate those searches, what the searches return is not always exactly true. When we do our own assessment of which items have drive attachments, we find more than what Google returns. So it's you can't really blindly rely on that. But it could be a good starting point to like get the ball rolling and have a rough idea of what volume to expect. The other trouble is this: let's say that you have a bunch of emails and they have cloud attachments, but there's no guarantee that you're getting those cloud attachments because when you request them, you might run into a few issues. One of them could be that that attachment is no longer there. So what if it was inserted that would then subsequently deleted? There's still the hyperlink, but the file is gone. The second thing is you may not have the permission to access the file or download the file any longer. Perhaps you had it and you don't you lost it later, or perhaps you never had it and somebody sent you an attachment that you didn't have access permissions to download. So there's definitely quite a few moving parts and challenges. But that's not to say that you know, well, this should not be part of e-discovery. I think it's just good for people to be roughly aware of you know what type of challenges they might face when they go down this path and have some potential solutions and workarounds for them ready, or at least know that you know these will need some discussion with opposing.

Kelly Twigger: 40:04

Yeah, you and you touched on a big important point, and that is essentially possession, custody, or control as it relates to the files that are at a hyperlink in a communication that is otherwise relevant. So if I have an email, like the email exchange we were talking about before, where I share those notes with you who go on your merry way, I remove your access to those notes. Someone collects your email and the link that exists is there, but you no longer have access to that file. And so the question becomes legally, are you obligated to provide it? And the answer is if you no longer have access to it, then it is not in your possession custody or control. So we've got this now overlaying argument or challenge on top of providing hyperlinked files. And I think so. There's the possession custody or control argument, which is an issue. And there's also another argument that Doug Austin raises regularly when we talk about hyperlinked files, and that is that preservation of hyperlinked files is very difficult because of the way that these systems are set up. And if you're put on legal hold, but you no longer have access to that document that I just mentioned, and I'm not on legal hold, then that document may never get collected or it may not get preserved. And say the document rolls off based on a retention policy that otherwise it should have been on legal hold. Is there a way that we're going to be able to can we deal with that right now? That preservation issue?

Arman Gungor: 41:45

Yeah, that's that's tough. And what you mentioned earlier is very true. And that has actually two components to the control and ownership and access issue. There's also the issue of let's say I have my mailbox and I have some Google Drive hyperlink files in there, and a number of them are mine from my Google Drive, but a number of them are from the outside, people that have sent me files. So they're not my Google Drive files, I just have access to them. So if I have access to them, we call these shared with me files. And both providers, both Microsoft and Google, have this concept. So are you obligated to also produce these shared with me files? If you if you don't have access to them, then you can say it's not feasible. But to the extent that you are able to access them, do you include them in your production? That's another challenge. And then there's that issue of legal holes, and that was a potential problem there as well, because Google Drive and the email data are two separate things. You might run into problems where maybe the email is held for retention, but the corresponding drive files are not. So you have a situation where, yeah, you have retained the emails, but the linked drive files are gone, and you can't really perform a full Google Drive hyperlinked export and production anymore. So that's another thing that needs to be set up.

Kelly Twigger: 42:57

Yeah, that's a huge problem because now for preservation, we can't just put a legal hold on email boxes the way that we used to. Now we have to put a hold on SharePoint sites, Google Drive, wherever the documents that will be linked to the communications might live.

Arman Gungor: 43:13

That's something that the providers will have to solve. So for Google, for example, when they allow you to prepare a hold on a mailbox, then there should be an option that says, do you also want to hold any hyperlink Google Drive files? And if so, they should be part of that hold automatically. Otherwise, it would be really difficult for the end users to handle that because the only thing you could do is hold the entire Google Drive. And that's even in some cases perhaps not feasible because of the fact that some of those might become from the outside. So you have access to them, but they're not yours to hold. So there's some retention challenges there as well.

Kelly Twigger: 43:49

But yeah, yeah, there are there are legal questions. There are significant legal questions. And I think we I just did a podcast last month with Judge Rodriguez from the district court in Texas, and we talked specifically about the possession custody or control issue and whether or not the the tests that we currently use for possession custody or control really hold up under the technological advancements that we've seen in the last decade or so because it's just the landscape has changed so substantially, and the use of mobile devices to communicate has just it's really muddied the waters of e-discovery. We got very comfortable dealing with email and attachments and how we've just blown it up and started all over. I there's a couple of things that else I wanted to ask you about. One is you you mentioned to me that there's a second tool that you've built at Metaspike called Forensic Email Intelligence or FEI. Can you tell us a little bit about what that tool is, what it does?

Arman Gungor: 44:44

Yeah, so it's an investigation tool that essentially kind of takes it from where forensic email collector leaves off and starts with already locally available emails. So not a collection tool, but working with email files, MSG files, PSTs, OSTs, that kind of existing email data. And then allows for email investigations. The main use case there is if you have a case and there's, for example, email fraud, somebody backdated an email, there's maybe business email compromise, a phishing email has been sent, a wire transfer has been made to a wrong party because of a man-in-the-middle attack or something like that. Which is quite common these days. So it is a tool that's designed to kind of deep dive into email and break it into its components and look at what we call tool marks. So you know, when an email client touches an email, it usually leave some marks in terms of hidden timestamps and different headers and things of that nature. So kind of interpret that data, correlate them, and then score the emails across the board so that you can load up, for example, millions of emails and then have the tool automatically score them. So in a large case, you have a starting point of which emails do I look at first if you know if I don't know where to start. That's a problem I personally had. I would be score them by what? Score them based on how egregious their problems are in terms of the internal inconsistencies in the email.

Kelly Twigger: 46:04

Oh, okay.

Arman Gungor: 46:05

So for example, the email has some digital signatures, but they don't verify, so that's a problem. Or the email has some attachments, but the date of the attachment is after the date of the email. That's another problem. Or the email has some hidden timestamps in its message identifiers and things of that nature, and they don't match up with the email's apparent date.

Kelly Twigger: 46:22

So all those different data points has a scoring system for each one of those problems, and the more problems you have, the higher your score is, and so those get elevated to the top to look at.

Arman Gungor: 46:32

Exactly. Yeah, because I would, you know, I've received some cases in the past where I would be asked to look at a specific single email, and when I have that, it's easy because I know what to do, and it'll take me maybe a few hours to get through all the details of the email, but I can solve that no issue. But when somebody gives me, okay, we've been receiving these email productions, and here's 200,000 emails, and we know this person is a fraudster, so there's a bunch of fraudulent emails here, but we don't know which ones, so figure this out. That becomes a whole different problem because then I don't have a starting point, and where do I even go? So this mainly solves that issue where you can throw them in and automatically score them.

Kelly Twigger: 47:08

So you just load all 200,000 emails into that tool, it analyzes them, creates that scoring system, and then ranks them for you to start looking at.

Arman Gungor: 47:16

Yeah, it also analyzes them contextually too, because there are also some problems in emails that only come to the surface when you look at the whole case. For example, let's say that uh, you know, one of the metadata fields in emails is called message ID. It's one of the traditional email headers that everybody is aware of. So it's supposed to be unique for every instance of every email. So if you have a specific email message and you made five copies of it, then yeah, they will all have the same message ID. That's not a problem. But if you have multiple emails with the same message ID but their contents are different, then that's kind of a problem. How is it that there are multiple emails that they say different things but they have the same message ID? And this is based on real cases I worked. We've saw some cases based on exactly this kind of problem where somebody would produce hundreds of fraudulent emails and they would make mistakes and they would reuse the same email template to produce multiple fraudulent emails, and then you would catch them by those similarities in their emails. So we also do that type of contextual analysis to see if there are any patterns in the whole project rather than just in the one email.

Kelly Twigger: 48:21

I think that's fascinating, and the reason I think that's fascinating is because one of the things that I focus on is being able to leverage ESI to tell a story. And what you've just explained, using FEI with fraudulent emails is a beautiful way to be able to tell a story, right? With not a significant amount of cost from a legal perspective to be able to do that investigation and understand a story that otherwise with people and just looking at data with my naked eye that I wouldn't be able to tell. So if someone says to me, this email's fraudulent, I don't know where it came from, we can use your tool, someone can use your tool, you could use your tool to look at it and be able to tell us what is wrong with it, why we know that it's fraudulent.

Arman Gungor: 49:13

It's true. And you know, I would like your listeners to take one thing from this. You know, email evidence is obviously very prevalent everywhere, but unfortunately, because it's textual, it's also very easy to modify. So when you have key smoking and emails in a case, I would refrain from taking them with face value by default because a lot of times they are uh possibly modified. I've seen so many cases where there was email fraud and so blatant fraud that you wouldn't think somebody would actually do this and put it in front of a judge, but they do anyway. So it pays to at least quickly try to verify those emails' authenticity, and it's easy to do in most cases because they have digital signatures that are very easy to verify. If you have a receiver's copy of an email, essentially what happens is as the email travels from the sender to the recipient, the servers on the way along the way sign the message with their private keys. The technical details of it are not that important, but the end result is that at the end you have an immutable digital signature that tells you whether the email is changed or not since it's been signed. So it makes it really easy to verify okay, the email's message body and the attachments and a certain set of key header fields may have been changed or could not have been changed by anybody else after it's been sent and received. So that's a very quick check to do, and I think it's incredibly valuable, incredibly valuable to essentially have like as a first line of defense, you have a key email, have it checked or check it yourself, easy to do to see what that signature says. If it's bad, then that might warrant some further looking into to make sure you know you're not dealing with fake evidence.

Kelly Twigger: 50:54

One of the things that's that's fascinating, and I one of the things that comes up for us as litigators on a regular basis is timing, because it may be two days or a day or five hours before a deposition when we're sitting down with an email and we're looking at it going, this doesn't seem right to me. This just doesn't, there's something off here, and I don't I can't do what you can do with your tools or with your knowledge. But when I how long does it take if I called you and said, Arman, I need your help looking at this email? I'm gonna postpone this deposition for a week. Can you help me take a look at this? Tell me what you need. What is the timing associated with that? Like how quickly can you help someone?

Arman Gungor: 51:35

If you have an email, like an individual email you're interested, and if you have it in the right format, to do a preliminary check like this and verify digital signatures is press a button and you get the results. That's not really a full examination of you know every little bit in the email, but that gives you like a preliminary, very quick assessment. But the right format is important, and that is where we're going back to you know having the emails collected in the right format. And we had this discussion about, for example, Mbox versus PST earlier. It is important to stay in the right and original format of these emails because all those digital digital signatures are out the window once you convert the emails.

Kelly Twigger: 52:12

So if you have what is the right original format of an email?

Arman Gungor: 52:16

Typically, that's going to be a MIME message. So, similar to, for example, let's say you have a Gmail account, you go to gmail.com and you click on your email and you say download original, it gives you an email file. That's the original MIME email. Same thing with Yahoo! You say download raw email message, I think for Yahoo, and it gives you the same format. So that's that's the universal format called internet message format, and that's how the emails travel over the internet. So the digital signatures that are applied to the emails are essentially calculated based on that format. So once the email is converted to a different format, like going from email to PSD, for example, then the signature is totally invalid now, so you just lost that ability.

Kelly Twigger: 52:56

So in all this that because you said Microsoft doesn't store its emails as mimes.

Arman Gungor: 53:02

Yeah, exactly. So Microsoft breaks it apart when you export from Microsoft though. Initially they didn't have a way to do this, I think, but they have a way to store what they call the MIME skeleton of the email so that they're able to rehydrate and give you that MIME message when you make a request against their API. So Microsoft also does support MIME email collections, but there are some challenges there. One of the issues with Microsoft is that as part of that intake process, they modify the email contents. So when I receive an email as a Microsoft user, for example, when my received copy arrives in my mailbox, I find that the HTML body of the email is slightly different than what the original sender had composed because of some internal processes they have, and that interferes with the digital signatures. To a small extent, that is reversible because we know what that is. So if I have a relatively brief email with not too many bells and whistles, I can actually undo that and still verify the signature, which I've done in the past. But as the emails get more complex with a lot of attachments and lots of body and stuff, then that becomes very difficult. And I see that as a major problem and as a negative towards Microsoft's systems right now. I hope they they fix that issue so they can give you unadultered emails in the near future. But that's only Microsoft right now. I mean, pretty much all the other providers deal with MIME and they don't really change the contents of the MIME emails. So as a general practice, even if you are accustomed to using PSD format, which a lot of us are in the eDiscovery, it's the format to use for information exchange. If you deal in PST format, I would say also collect MIME copies of these emails as an insurance policy because you never know.

Kelly Twigger: 54:48

And you go back if you've preserved the emails and they're still existing on the Microsoft. I was gonna say server, but obviously it's not a local server anymore. If they're still existing in the Microsoft environment, you could always go back and recollect it in a MIME format, right? If necessary.

Arman Gungor: 55:03

Well, it might be too late though, then. But you know, that might be some situations where the email is no longer available.

Kelly Twigger: 55:09

I would say make it part of the email, preserved if it's on hold, and you could go back and recollect it in the MIME format if necessary. Yeah, because I don't think there's a way to do a second pass if you might start collecting something in MIME, as opposed to in their PSTs, the way that the because that's a standard export from Microsoft.

Arman Gungor: 55:28

The problem that I've seen is this usually case timelines are pretty long, as you know. So when the project starts, when there's a dispute, there's an initial preservation, and by the time the case is adjudicated and it comes in front of the judge, and there's maybe some concerns of authenticity and all that stuff, several years may have passed. So by then by the time you go back and revisit that, you may have lost context. And there's another issue when you preserve the emails, you we're not only interested in what's in the emails, but there's additional information about the emails, which we call server metadata, essentially contextually information about the emails that are not in the email itself. And to give you some simple examples, for example, identifiers of these emails are usually sequential. So you might have some sequence information from the email folder, right there, as in you know it's emails 5678, for example. But if you look at email number six, its date is completely out of sequence, so there's a bit of a discrepancy there in terms of the sequence of these emails. So there's a lot of clues there that lie in the server metadata. And by the time you come back and revisit that case in two years and try to dig into that, it might very well be too late. So all the preservation stuff, I would highly recommend to take care of in that initial phase of preservation and not leave it for like, okay, if there's a dispute, we'll deal with it later and in the future.

Kelly Twigger: 56:49

No, I completely hear you. And that's the second or third time that something that you said has prompted me back to another theme of mine, which is you need to be doing this very early. You need to be engaged with all of these data-related issues very early on in the case, that waiting, as we often do as litigators, until you get to trial to try and address some of these issues, you're going to end up in a world of hurt because the way electronic information can go away, regardless of your best intentions, to be able to keep it in place. And so that can be a tremendous problem. So I think that early planning, that early engaging with the information, the early focus on telling your story and what you need to be able to tell your story is so critical in terms of understanding but addressing all of these technological challenges that come with getting the information. And one of the things that I talk about a lot is that we really have two separate obligations in discovery. One is to find the information to allow us to tell our client's story. And that's perhaps your primary goal. But the second is to meet your obligations to provide information and discovery under the rules. And oftentimes we get too focused on meeting our obligations to provide discovery and not on telling the story. So I think what your this entire conversation has really emphasized for me is that you've got to focus at the outset about what is the evidence that's going to allow you to tell the story and how are you going to get that in front of a court. And you got to try to address those technological issues early. Otherwise, you could end up, like you said, with not being able to have either the form or the information at all to be able to recreate that story.

Arman Gungor: 58:43

It's true. And I think you can go even a little bit further back because if you are, for example, general counsel for a corporation, or you are essentially the corporation that's planning to have some type of litigation in the future, like possibilities. I think, like, you know, when you're setting up your IT systems, you're essentially about to settle into a workflow, and that might involve collaboration tools, it might involve cloud storage tools. I think before you really get started fully, it pays to kind of evaluate your stance in terms of e-discovery readiness. Because, you know, let's say you're going to use Provider X and they support modern attachments and they support a collaboration tool. It's really cool. But from day one, I would sit down and say, okay, what would happen if we got into litigation today and if we were asked to produce all this stuff? What are the steps that we need to take? So to do like a dry run, basically, and figure out exactly what would have to happen for you to be able to pull that off before you use something like confidential mode in Google or the end to end encryption in Google, for example, that's relatively new. Is it really feasible to have those features and at the same time be reasonably ready for litigation? If it's not, then you might have to restrategize before you really settle into that workflow. Because otherwise, by the time several years. Pass and you have accumulated terabytes of data and you know thousands of modern attachments and encrypted files, you might find yourself in a deadlock where you're ordered to do one thing, but the reality is you just can't do it.

Kelly Twigger: 1:00:14

Yeah, that's spot on. I think knowing what your capabilities are before you even get into litigation, but at the very least, identifying them as soon as your duty to preserve attaches or you reasonably anticipate litigation is so critical because otherwise we find ourselves agreeing to all kinds of things that we then can't meet. And we get in these huge long negotiations about what we're able to provide, whether it's hyperlink files or anything else. So this has been a fant a fascinating conversation, Armin. I'm so appreciative. Thanks so much for joining us on the Meet and Confer podcast. Can you, before we go, tell everyone whether if they don't have FEC or FEI and want to have access to those tools and training, how they can get access to those?

Arman Gungor: 1:01:00

Yeah, sure. Our website is Munaspike.com. Our software and pricing and everything is there. So we work with online orders, so you can go pick up a license there if you like. We usually have live virtual live training every six months or so. We have actually a forensic email collector training coming up in five days. So in those trainings, we essentially try to cover it soup to nuts from modern attachments to in-place searches, exports, all kinds of details. So it's helpful to kind of wrap your head around what the possibilities are and get involved in using the tool fully. We have another training that's on forensic email investigations that's actually a random neutral training. It's not based on our own tool, but it's just a general email investigation training program where we take you through my messages, kind of similar to the things we discussed. What are the email headers? What are digital signatures? What are some marks of fraud and manipulation of emails? What are the possibilities that providers offer for capturing server metadata, evaluating those practice labs and stuff like that? So all that isMetaspike.com. A lot of you might have me on their LinkedIn so you can reach out to me there if you like, and we'll go from there.

Kelly Twigger: 1:02:11

Fantastic. All right. Thanks so much, Arman. We'll have to follow up again as the technology continues to change and our capabilities continue to change, specifically as it relates to hyperlink files. But I'm so grateful for you being here today. Thanks so much for joining us and imparting all your wisdom.

Arman Gungor: 1:02:26

Thanks for having me. Thank you.

Kelly Twigger: 1:02:31

That's it for today's episode of the Meet and Confer podcast. Thanks for joining me, and a huge thank you to Arman Gungor for sharing his expertise on hyperlinked files, modern attachments, and the constantly moving target of cloud-based email and document workflows. It gets complex in the e-discovery workspace, and we're incredibly grateful to folks like Arman who are spending their time figuring out solutions for us to be able to find the evidence we need to tell a story with ESI. If this conversation has you rethinking how you're handling discovery or realizing how many assumptions you're making about hyperlinked files, versions, and access, that's exactly why I built Minerva 26. It's a discovery strategy platform that helps litigators create and validate their discovery strategy by connecting case law rules and real-world workflows, such as how to identify and handle hyperlinked files, so that you can make faster, more defensible decisions that shape your emotions, protect your credibility, control costs for your client, and let you get the information you need to tell your client's story and win. To see how Minerva26 can become the strategy hub for your discovery decisions, head to Minerva26.com and schedule a strategy session or request a demo. Until next time, remember Discovery isn't just a box to check, it is your litigation strategy. Leverage ESI to win.