**Carcanet Oral History Interview [03 May 2019]**

**Respondent: Fran Baker (FB)**

**Interviewer: Victoria Stobo (VS)**

FB: My name’s Fran Baker, I work as the Archivist and Librarian at Chatsworth now, but I was at the John Rylands Library for about twenty-two years.

VS: Could you tell me about your work at the Rylands? How you started? How you then became involved in the Carcanet Press Archive?

FB: Yes, so I started work in the late nineties, I suppose, on a temporary cataloguing, funded project, by HEFCE [Higher Education Funding Council for England] as it was then. I’m not sure what the name of that is now. Literally as a general cataloguer, and when the end of that contract came up, [redacted] I managed to get the role of Assistant Modern Literary Archivist] in about 2000, I think. That was when I really started working with the Modern Literary Archives, and a large part of my job was still cataloguing them. I did an awful lot of cataloguing in those first five or six years. Not just working on the Carcanet Press Archive, but on all the related archives around it, like *Critical Quarterly* and Grevel Lindop and various poets who published with Carcanet or had connections with them, in some way. [redacted]With Carcanet, we’ve only got really good, detailed, item-level catalogues for the first two accessions, which go up to something like 1980. The subsequent – I don’t know what number accession we’re on now, eighteen or nineteen or something?

VS: I think the accessions have gone up to… I think they’re in the thirties now?

FB: No, they can’t be! They can’t have taken in –

VS: I think they’ve… I think there’s one really big one –

FB: Accession three, that’s the monstrous one.

VS: Yes, and that’s what I’m cataloguing at the moment, and I’m not sure… there does seem to be a range of accession numbers and I don’t know whether they’re breaking them into smaller and smaller chunks? I’ll need to have a look at that document again. But the box list is really, really long.

FB: I know, I did most of it! Well no, Karen our collections assistant, did a huge chunk of that as well. Just getting that documented was a massive step forward. I think the situation with that was that when Stella arrived, I’m not sure why the first two accession were kind of in a relatively sensible state, and accession three wasn’t, but [redacted] all the original order is kind of fragmented. So, you’ve got sequences that make sense in some places, but they’re all broken up and you might find something related to the same book or the same correspondents in a completely different box. And actually getting your head around that accession to do a proper archival arrangement of it would be a complete nightmare.

VS: Yes! It’ll be interesting to see – we’ve ended up targeting the cataloguing based on specific writers, and that can kind of fit with the existing arrangement for Accession one and Accession two, so I think we’ll probably maintain that for this project, make the EAD [Encoded Archival Description] available and then, I guess, as time goes on and the rest of Accession three is catalogued, hopefully that will still be a relevant structure.

FB: Yes, I mean it’s the sort of thing that can only ever be done in small chunks. Which is kind of the way that I started approaching the *Guardian* archive, because it was so huge. We managed to get different pots of funding to catalogue different little bits of it. Cataloguing the whole Accession three would take years and years and years.

VS: Absolutely, yes. Can you talk about the custodial arrangement while you were at JRL, so how it was selected and appraised, and how much involvement the staff at Carcanet had in that process as well?

FB: Well it was interesting because a decision, as far as I understand it, had been made right back in the 1970s when we first started taking it, that we would only take – I say we, I’m not here anymore – we would only take the author correspondence and the book production files because they were considered to be the most significant things. [redacted] And so we’re sort of lacking all those kind of administrative files with the minutes of meetings, and the financial stuff. [redacted]So, that pattern continued when I was the Assistant Modern Literary Archivist, and I continued that when I took the archive on. So, it was just taking annual or bi-annual accessions in from the office, and I know Stella was involved in setting up that process, so she obviously spoke to the staff there. So they knew what to do with their files, so we would go to box things up in the office and they’d already have it quite well-organised, like the author correspondence, the secretary would sort out and it would all be alphabetical, which did reflect the way she kept it. And all the book files, all the manuscripts and proofs, and whatever, all relating to the same book would be kept together, and we just listed those alphabetically.[redacted] So, in terms of appraisal, and selection, yes, I followed that set pattern for selection and didn’t really do any appraisal, because we kind of knew what we were taking, and that we wanted to keep it all. So, I suppose, yes, the staff were involved in that sense in that they prepared the stuff for us, so it was all working to a very set pattern. For digital, that was altogether different because it was totally new when we started taking the emails, so we submitted a survey to the – we decided, initially, to target Michael [Schmidt] and the editor at the time, who was Judith [Willson], simply because Judith produces the book files, and also has really important author correspondence. And Michael, obviously, is the lynchpin of the whole operation. So, we targeted them because we felt it would be a continuation of the same pattern, if you know what I mean. They were continuing to do the book files in a very hard-copy way, still. Although you’d get manuscripts and proofs attached to emails, all these significant ones that had annotations or anything would still be printed off and put in a box. It was really just the correspondence that we were lacking. It very much started tailing off when we’d take those annual accessions, there’d either be no correspondence files at all, or very few. It was obvious that we were just not taking in a huge chunk of their day-to-day correspondence with writers, which is what prompted the digital project. Shall I go into that now?

VS: I think we should probably go into it just now.

FB: So, I sent out this initial questionnaire, which was kind of about their record-keeping habits as regards email, which was very interesting. It became clear they did things like auto-archiving, but they didn’t really know what that meant, or where their archived PST files would be sitting on their computer. And we did discover that [redacted]someone had accidentally deleted two years’ worth of Michael’s email. And they’d upgraded computers and everything since then, so that was gone. And the correspondence started to get smaller, right back to the – there’s a period in the late nineties when the correspondence was starting to tail off a bit. But when we went in to actually take the first tranche of emails, the earliest ones we could find were 2001. So, we were obviously missing a chunk from the late nineties that has just gone. [redacted] So we did the questionnaire about how they kept their own stuff, which we felt a little more informed after getting that back, but really it wasn’t until we got in and sat down with them and went through it with them that you get a real handle on what they’re doing. And in terms of selection and appraisal, for that, I’d had this notion that we could go in, sit and talk to them, and identify folders in their email accounts that we would not take. But in the end, that proved to be so time-consuming and complicated that we just agreed with them that we’d take everything. And then we would appraise out the stuff that wasn’t of long-term research value, or that was personal after transfer. So, I think it’s a sign of how much they trusted us really, to let us do that. So, we just took, for the first accession, we went and did a search for any existing PST files that they’d archived, that were either on their network, or their desktops, so there were five of those. And then we did a harvest of their live email. So that was a huge accession, and it literally was taking everything and appraisal – how did I do it the first-time round? Actually, the editor, Judith, was unbelievably organised, her organisational talents, because I’ve never seen an email account quite like it before, everything – her inbox was tiny. Everything was filed away into author and book files, and she would even file her sent items. So, there wasn’t much in that folder either, because she’d filed away the important -

VS: I’ve never met anyone that files their sent items!

FB: It was astonishing! So really, appraisal for her was virtually zero. And she didn’t use it for personal stuff at all, so it was kind of a recordkeeping dream.

VS: She’s the model we should ask all depositors to manage their email.

FB: Michael’s, by contrast was just an absolute nightmare, because it included everything. [redacted] Things like Amazon receipts from buying books online – it was just everything. And because his filing structures were so huge as well, I literally had to – we hadn’t got the forensic email examiner software at that point, so I was just looking at it in this dummy Outlook account that we’d got set up, that wasn’t attached to a particular person, and I literally took screenshots. It involved about 60 screenshots of his filing structure, printed them off, and marked up for him to approve, the folders that I was thinking of deleting.

VS: That’s how big the filing structure was?!

FB: There were multiple PST files that all had the same structure, so some of them might have had a lot fewer emails in, but they still had the same – and obviously it changed over time as he added more files, but I felt I should do it for all of them. So, I sent that back with highlighted, saying ‘We’re thinking of getting rid of this one,’ it was all a bit manual. Which he signed off and agreed. I still kept - you can only do it at folder level, you can’t possibly appraise any lower than that. [redacted]

VS: So, the sorts of things that he signed off on, would that be purchase receipts and things like that?

FB: Yes, and it was at folder level. So, I was literally just giving him a list of folders and saying, ‘This one? This one? This one?’ He did at least file a lot of Amazon receipts into a folder. I don’t know why he kept them! [redacted]

VS: [redacted]

FB: But partly also because we’ve got his personal archive, or we had some of it and we’ve now got the rest of it, so I was trying to think beyond just his work for Carcanet, and if we’ve got his hardcopy personal correspondence we should be taking his personal correspondence in Carcanet. So, we did keep some. [redacted] And you know, we did make sure it was going - they signed a thing saying, it’s going to be completely embargoed. [redacted] And it did make me realise in terms of selection that we were actually capturing a lot more in the email of the business side of things than we had ever been capturing in the – because we’d made this agreement with them not to take business files. It struck me that now, through taking the email, we’re taking a lot of those business-related files, as attachments. Minutes of the Board meetings, financial spreadsheets, that kind of thing. Which did make me think, we really need to start thinking more about this stuff and taking the hard copy stuff that’s associated with that as well. But then, it just came to an end, and I left, so I didn’t get a chance to pursue it. I know there was a ton of financial stuff that I went to see just before I left, and was thinking, we really ought to be taking some of this, but I left it for Jess. So, yes I suppose that’s selection and appraisal, really.

VS: From reading the project reports, I know that you identified the email archive as the JRL’s first big born-digital archive.

FB: I’ve just realised I forgot to say something about selection. What we didn’t do, and it was another area I picked up on as a gap, because we focused on Michael and the editor, we weren’t taking the marketing officer and the publicity person, and also the head of finance, so it made me think about that, and we really ought to be taking the marketing stuff, because that’s around how they present their writers to the world. And often how the writer wants to be presented to the world. So that’s something else we didn’t select at that time but should.

VS: When the project completed and you left the JRL, how much time was there in terms of developing processes and policies for collecting born-digital material at the Rylands? The question makes the link between what you learned from the email archive being the first big born-digital collection, and then how that’s gone on to influence other processes at the JRL?

FB: I suppose the email was – obviously, we’d been taking in digital stuff for decades, because we’d got floppy disks, Amstrad disks, you know, all that kind of stuff. But it was really the Carcanet that kicked everything off in earnest. And that was only because this tiny pot of funding came up from JISC that was for case studies in digital preservation that was supposed to kick start things. It funded seven weeks’ work. [redacted] So, we then carried on for about eighteen months, internally funded, just doing it for the equivalent of a day a week of my time and the software developer’s time. For me, that was really, really productive, and it was really great working with a qualified software developer, and we worked at addressing these problems and learning to talk each other’s language. After that, things… well, I suppose, linked to that, and it was pre-Carcanet, but we had set up, as a result – because I was involved in two joint projects with the Bodleian, one of them was called the Paradigm project, that was digital personal papers, and one was called CAIRO which isn’t so well known, but it was about developing an ingest tool for digital archives, it didn’t really – we did tons of work, but the development side didn’t really… So, as a result of that, I’d done quite a lot of trying to make the case for having digital preservation represented in our work somehow. I took a paper to the leadership team and we set up a kind of digital preservation steering group – a strategy group, initially. So we developed a strategy and then it turned into a Steering Group, one strand of which was continuing with the Carcanet, there were lots of other strands around legacy material, dealing with stuff that wasn’t on media but was being saved in folders and all that kind of thing. [redacted]We did acquire Preservica, which was great.[redacted], and I thought Preservica was the best option from the ones that we corresponded with, they came and did presentations and quotes. And I did think they were the best option. But I understand, well, I think the email is now, at the point when I was leaving, we were still struggling with…. The email was really well preserved, we’d worked out all these systems for breaking it down to email level, created all this metadata like PREMIS and descriptive metadata in EAD, as a result of automating, the metadata extraction tools we were using would squirt information into these EAD templates, and it was stored in this beautiful, shiny repository called FEDORA, and then they decided to not pursue FEDORA, I think mainly because it requires a lot of software development expertise and resource. So, I can understand why they wanted to buy a commercial product, but then there was the whole issue of how do we get this stuff from there into there, and it was in a very specific structure, so…

VS: I did wonder, because when I read the reports, oh there is EAD for the email collection, but I’ve never seen it.

FB: Yes, because it’s sitting in FEDORA. We never got it so it presented prettily. So it’s literally in raw EAD, so I did narrative – we treated it, because we’d taken in Carcanet as accessions, and we kind of arranged it like that, because it was easiest, so I treated each accession of email in the same way. I allocated an archival reference code to each PST file, did a narrative overview in the scope content, I called it sequence level, top level describing what was in that particular sequence of email and then the developer worked out a way of extracting date information, quantity information, number of folders, number of emails, digital size, so that’s all sitting as far as I know, still in FEDORA.

VS: That is amazing! That level of metadata extraction into EAD is great. Was it able to pull things like the subject line?

FB: I’m trying to think. Well, the subject line, the EAD metadata was different from all the other types of metadata, so we had a whole, every PST file, and in fact, every individual email had different data streams in FEDORA. So, you’d have a PREMIS data stream, a FITS, this tool called FITS that pulls out technical metadata and things, so we had that saved. We had the EAD, we had some Dublin Core, and then we migrated each individual email to different formats, so we kept it in MSG, then we had an EML, an XML version of each individual email, and we had another developer come in, because they put in some resource for this, and he built (and I don’t know if anyone still knows how it works) but he built me a kind of curatorial interface to sit on FEDORA, which looked like an email inbox, that allowed me to search on everything like – because the tool we were using, Aid4Mail, to break everything down, extracted all that metadata, so yes, I could search on things like subject line, sender, recipient, and sort everything.

VS: That’s amazing!

FB: I know! And then we got Preservica and abandoned it.

VS: I’m using Preservica at the moment to try and make a hundred-email selection and I’m really struggling because the search functionality isn’t amazing –

FB: You could try doing it in Paraben’s Email Examiner. [redacted] Actually, when we were talking about appraisal, I forgot about that, so the first time I was doing those zillion screenshots, the second accession we took in, we’d got Email Examiner and I did write some instructions, but I was using that to, because it can create reports, which was really handy, so I was using it to flag folders we didn’t want to keep, and then we could run off a report of those folders and give it to Michael, which was a lot easier. But that seemed to be quite – you could have a go, I’m sure it was quite good at searching, rendering and everything, it was brilliant. It would render everything perfectly, you could view all the attachments, and because it’s forensic it’s not having an impact on the stuff.

VS: That’s the thing, when you see things in the viewer that they provide in Preservica, it looks like an email, and when you open the messages in Outlook, that’s fine, but sometimes, I don’t know what it is about the interface in Preservica but I just find it’s not as intuitive as I was expecting it to be. And also there doesn’t appear to be a way of – when you get search results back, they’re not itemised, you just have to go through all of them, which doesn’t really help much for finding things. But then that’s not really what Preservica is for?

FB: It should allow you to search for things, really.

VS: I think it’s just, either that or I need to spend more time getting to know Preservica and develop a better search strategy, but it’s interesting that the Paraben tool is there.

FB: That’s if the license has been renewed.

VS: We’ve been looking at ePADD.

FB: Yes, that was something I hadn’t got to at all, and that sounds really good.

VS: The experiments we’ve run on it so far, it looks fantastic, it’s whether we need a PC or laptop with a huge amount of RAM to run, for the size of the archive.

FB: Do you have to convert everything to MBOX?

VS: Yes.

FB: That was something we abandoned at an early stage, because when we had the first developer in [redacted], we were using different tools and I’d read that MBOX was supposed to be a preferred format, and we just kept hitting hitches with, there seemed to be different versions, different types of the MBOX specification, and none of them seemed quite right. We ended up abandoning that. Just because we had so little time, anyway.

VS: It’s definitely a format that we would migrate to purely for ePADD, and that would just be for curating and managing access to a certain extent. I’m not sure whether, with MBOX, then you migrate to that, what metadata you’re losing in that process.

FB: Well, you’d worry about things changing. That was an eternal problem that we seemed to have was doing checksums at every stage to see if things had changed, and just – I can’t remember, I think Outlook we’d stopped by making it completely standalone. Did it still…? I suppose, because we were using Outlook for appraisal and compacting the files anyway, so it was going to change inevitably because we were deleting folders at that point. Some of the tools we used, we found that simply attaching the tool to the PST, would change its checksum readings. You have to be really careful about that, because you just don’t know what’s being changed as a result of what you’re doing. Even when you’re only trying to look at it, not actually change it. And yes, with the MBOX we were using a tool called Message Save and we didn’t get on very well with that at all. So we went for Aid4Mail, but I suspect that part of it was we hit a problem and I think that was when we were using Message Save, it kept, things just weren’t running properly or processing, we realised ultimately it was to do with file title – you know, file path lengths. Because we had the email stored in a folder, something like Master, Archive, Do Not Delete On Pain of Death, and the tool was failing because our file path was too long, it was really silly.

VS: It’s funny when a tiny little detail like that will throw it off entirely.

FB: Collecting – I suppose we wrote born-digital into our collecting policy, but we didn’t want to make it a separate thing, so we said, we will collect in these subject areas and we will collect in any format, whether it’s in a paper or a digital format in those particular areas, so we didn’t have a separate collecting policy for digital, because that wouldn’t really make sense.

VS: Yes. Are there any other significant born-digital archives that have been accessioned since you took custody of the Carcanet Archive?

FB: Yes, so Anvil Press, the poetry publisher, we took in an entire Mac and a PC and separately I took in their website archive, and I don’t know what that looks like, he transferred it to me, just in a file, I think, and I just put it on the network, it’s in one of the digital preservation folders, because I realised when I was doing the box list for Anvil that their website had gone. I was initially using it because it was really useful for looking up publication dates and titles, and suddenly it wasn’t there anymore, I thought we really should have taken that shouldn’t we? The Mac and the PC, I don’t know how far they’ve got with that. They were going to be transferred up to the Main Library and imaged. Because they were using a really odd email program that I’d never heard of before that was really obscure, and I can’t remember the name of it now. I’d never heard of it, and I spoke to people who knew about email and they’d never heard of it, so that’s going to be an interesting challenge. That’s the other biggest one we’ve taken. Things have probably arrived since I’ve left.

VS: I don’t know whether you were involved in the deposit negotiations with Carcanet, around the digital archive, and whether there was a significant difference between that and what had been agreed for the paper archives?

FB: I was, and I can’t remember whether we wrote up a new – I think I did, write a new deposit agreement. [redacted]

VS: [redacted]

FB: [redacted]

VS: We’ve covered a bit about the cataloguing process already, so we’ve touched on cataloguing the paper archives and then we did touch on cataloguing the digital collection a little bit, so talking about what’s available in Fedora, but then how I guess there wasn’t ongoing resource or support for Fedora –

FB: No, and it was a bit – so we were also using, because we had to for technical reasons, the EAD schema rather than the DTD, which is what we use for normal cataloguing, which meant that we didn’t have the nice stylesheet, to attach to it to make it look pretty. But my idea was that that could go onto Elgar or whatever as series level descriptions of the email, at some point. But it would need to be, perhaps, converted to the DTD, and I don’t know how you do that. You need to find it first. I hope you can. It was based on trawling through each PST file to get an overview of kind of what was being talked about, and we had those useful automatically extracted information for dates, and also file extensions. So, we put in - I can’t remember which field of EAD – but we sort of extracted what the – I mean obviously you can’t say it’s valid instances of all those formats, but we had a big list of, ‘These are the file formats represented as attachments’, dot doc, dot whatever. So there is some useful stuff in there.

VS: Presumably Fedora has been maintained, just not supported?

FB: I think they’re still using it for certain things, but you’ll need to talk to Ian Gifford’s team about that.[redacted] And it was all a bit clunky and we were doing absolutely everything that you can do for digital that we could do, anyway, but getting all that technical metadata and preservation metadata, we just didn’t find any ways, so we got a PREMIS document but we had to customise it for our own local, certain local requirements and technical stuff, we had to create a kind of, locally validating version of PREMIS, so it probably wouldn’t validate properly against the normal PREMIS schema. And the developer also did drop down menus and things for certain elements to save time, because the PREMIS side of things was very manual. Which it ideally shouldn’t be, but it was copying and pasting checksums in every time you did a checksum or did a virus check, copying and pasting the information. But they’re all there. But the problem is, a lot of that technical stuff will, in a way, be irrelevant if now, you know, the masters are being taken in by Preservica and dealt with completely differently. That’s it really, for cataloguing the digital element, following that treating each harvest as an accession and each PST as a series, which reflected the way the paper side of things – well, the first two accession that have been catalogued properly, are arranged. It’s just box lists, but the box lists are generally really good, and Karen...

VS: They’re very detailed, they’re excellent.

FB: It’s just getting it into EAD that’s the massive undertaking.

VS: How was the collection used while you were at JRL? That covers both paper and digital, although I know you weren’t able to make much use of the digital -

FB: No, I suppose the digital wasn’t really used by anyone other than us. So, there’s nice visualisations using the metadata from it. That was using Gephi and I don’t know how complicated that is, but Oscar, one of the research institute people, was saying he found it really straightforward, based on – admittedly his body of data is miniscule, but he just had a spreadsheet and said you can run it quite straightforwardly if you’ve got it in structured form. And that could be fun.

VS: We probably will look at doing more visualisations with the samples that we decide to use.

FB: And they’d be smaller as well.

VS: They’d be smaller, but I wouldn’t mind playing around with the metadata a bit more to do more visualisations. It’s just deciding on what it is we actually want to visualise.

FB: And why. It’s fun doing, but what purpose does that actually serve? [Laughter]

VS: We have to work out the questions we want to ask of the data first, and I think those’ll gradually – there are three broad themes that we’re pursuing at the moment [outlines themes].

FB: [redacted]

VS: [redacted]

FB: [redacted] The future, yes. So, in terms of use, the paper tended largely to be people with an interest in a specific writer. That was the majority of the kind of enquiries we got, which is why that was something I tried to focus on for the digital side of things. Because most people are interested in searching on a writer’s name and pulling up stuff to do with the writer. [redacted] The difficulty with having it, with use, the first two accessions, that was fine, but the rest – it hasn’t been vetted for GDPR, so it was that process of having to read through absolutely everything before it could be issued. And make decisions about closures, and that was just a nightmare. [redacted] So yes, it probably would get a lot more use if it had been fully catalogued and they were online, and people could come in and there wasn’t an issue with having to read everything. And I suppose… it was being used more in seminars for students and that kind of thing at the point when I was finishing off here. We had a few seminars focused on either Carcanet writers or Carcanet as a Press. So that was good. It’s certainly one of your most important collections.

VS: Other collections deposited at the Rylands because this is where the Carcanet Press Archive is held – is that a big draw?

FB: Yes, because we’ve got papers of lots of different writers who’ve published with them, and I think Michael’s a really good ambassador for the Rylands because he tells people, ‘Our archive’s here, and they’re wonderful,’ and you know sometimes writers come and they say, ‘Michael has suggested I approach you about this,’ so absolutely, and I think Anvil was really entirely to do with Michael, because Anvil was winding up and they were taking over the list, and it was Michael’s suggestion that their archive ought to come here, which makes sense given that Carcanet are now taking on all their authors. So yes, lots is the answer to that.

VS: How many of those were deposited while you were at JRL? Were they already in the archive?

FB: Some were already here, I’m trying to think what – there was the *Critical Quarterly* archive, Brian Cox, who published with Carcanet and worked with Michael, so that was part of the initial set of Modern Literary Archives… certainly, over the years, I’ve been involved in them, we’ve taken in Grevel [Lindop]’s papers, Elaine Feinstein, that’s a fantastic archive, and loads of little accessions: someone would say, ‘Oh, I’ve got these letters that were sent to me by so-and-so published by Carcanet, do you want them?’ So yes, lots and lots of smaller collections, as well as the biggies.

VS: [General topics] I was interested in what you said in the reports coming out of the project, about the switch from written correspondence to email correspondence, and what is included in email that would have been left out of written correspondence, but also what’s been left out of email that probably would have been included in written correspondence, so for example, email replacing phone, but I guess the classic letter-writing style in email –

FB: I suppose you do get lots of emails saying, ‘See you at…’ or ‘yes’ or something, but I suppose if that’s part of a thread, preserving it makes sense. I suppose one thing that I noticed with the Carcanet, talking about replacing the phone, but also replacing just walking a few yards into another office and talking to someone face-to-face. We were capturing loads more stuff sent between the staff at Carcanet, often talking about their writers, that we wouldn’t have got before, because it would have just been said face-to-face or there might have been memos that just weren’t retained, so I actually think, with the email in relation to Carcanet anyway, we were getting a much more complete picture of the Press, apart from the sections I’ve mentioned we weren’t taking email from. A much more complete archive then we were just with those two series of records that we were taking.

VS: I find that really interesting, because I feel like when we talk about email correspondence there’s almost like an assumption that, oh, there isn’t really much of value in that?

FB: Oh god, that’s not true at all. And it’s interesting, I did some interviews with people about email archiving, and one of them was, someone who had worked at Carcanet, and I was saying, ‘What are your views on the quality of email correspondence as a research resource,’ and she was quite, I can’t remember which writer was it, she was saying, some writers, she felt, having their email was a much fuller picture of that person, but in some cases, it just wasn’t at all. Some writers who would go into great detail and be eloquent in their letters weren’t so much in their archives because it was seen as a – one thing she made the point of, and I thought that was interesting, was that when email first came in, it was seen as this very quick and easy… it wasn’t a replacement for correspondence, it was just a quick way of talking to someone, you’d write quick notes. You would still write a letter for more meaningful correspondence. But that slowly changed over the years, now that email is the formal form of corresponding, you know. Informal is text, or social media, or whatever. So there’s a sort of change over time with the quality, so now it is a higher quality research resource than perhaps it was in the late 1990s, when people weren’t really using it that much.

VS: It would be interesting to see whether, depending on the correspondent, some email archives are more voluble, more meaningful, richer than others.

FB: I’d say, probably, a publisher’s archive, publishing writers, is almost inevitably going to be a high-quality email archive. We don’t really have other things to compare it to, not within the Rylands anyway. Similar archives to compare it to, until we manage to get our heads around Anvil… that’ll be interesting.

VS: [Working with depositors, Art360 App] Advice we might provide to depositors?

FB: I suppose there was an old-fashioned view that you shouldn’t try and intervene in the way people keep their records, but I think that’s kind of gone by-the-by with digital. I think you can, well, unless you’re an in-house archivist for a company or something, you can advise, but that’s all it’s going to be isn’t it, and most people are going to ignore you, I can’t help but think. Because all they’rei nterested in is their day-to-day business and life, they’re not really interested in how easy it’s going to be for an archivist to deal with it in the future.

VS: Is it easier just to have a giant bucket and rely on search or is some form of organisation or filing more beneficial in the long-term, and it’s whether users can see the value in organising it. I tend to be a lot more organised with my institutional accounts, so I do tend to file things in folders, although I don’t file my sent items, and I find that really interesting. Whereas with my own personal email account, it’s just a giant bucket that I search.

FB: [redacted] I started off doing what I do at work and setting up all these folders, and now I think I’ve got a thousand emails in my inbox, and I just get to the point where I can’t be bothered. For me, in terms of appraising those archives, it was really helpful that they’d used folders. Because even Michael, he had a billion folders, but he used folders, and particularly for the particular writers and publications, he had a folder for everyone. And that was just invaluable when you’re trying to manage something, and make it make sense to other people. So, I’d be in favour of folders, but whether you can persuade people to – or file their sent mail, no one’s ever going to do that!

VS: I think in a work situation, if you do have to find something reasonably quickly, I would still argue for filing, I guess.

FB: [redacted] I try and file things at work, although things come in so fast you just lose them. But I try. And that’s my favoured approach to managing it. And I suppose, also deleting stuff when it’s no longer relevant.

VS: Yes, I’m bad for never doing that.

FB: [redacted] Capacity is something that forces people to cull. But on the other hand, you wouldn’t really want to do that in a big way, because it means people are getting rid of the important stuff. One other thing would be advising people to create PST files and archive your stuff at intervals, but again, are they really going to do that? Some good practice on that would be useful because Carcanet obviously had been doing that, but didn’t really know what it meant, or where the file was sitting when they’d done it.

VS: The available forensic tools, obviously more tools will be available now than when the project initially started, but I was struck by what you said about how important having a software developer on the team was, how do we attract software developers to this kind of work and keep them – it’s always short-term contracts for short-term projects, so how can we make it attractive to developers?

FB: I know, it’s a tricky one, because they can probably earn much more money doing other things. I did find initially that we were talking a really different language – we didn’t understand each other’s worlds, but over time, that changed, and it became really productive. So, I think if you can develop a good relationship with a software developer it can be really, really positive. When so much is based on one developer, or one archivist, and then those people go – you try and leave behind as much documentation as you can. [redacted] I suppose attracting developers… it’s a problem, isn’t it? [redacted] We certainly found it hard to find a temporary developer at that very early stage of the funded project, I think we ended up getting two expressions of interest, interviewing two people [redacted]. And they were contract people, because they just worked on short-term contracts. Actually, it takes quite a long time to get your head round it and then actually make something happen.

VS: If you look at a project like ePADD, that’s six years at least that project’s been running for. To develop that single tool. The possibility of converting or tweaking existing legal forensic software, and whether -

FB: Yes, the paraben thing was designed for law enforcement, but I had come across, I think it was the British Library, we saw it in use. It was very much based on – everything is called a case; the terminology is very legal. But it did seem to – I’m sure it does a lot more than I actually ended up using it for, because I was just walking through the manual and thinking, ‘Well, how could this help us?’ I’d definitely have a go with that though. I don’t t know what other forensics tools are out there that aren’t archives- specific.

VS: I guess it’s identifying them, playing around with them, seeing what works –

FB: It really helps to have a techie around when you’re doing that because [redacted] I’m one of these people who wants to read the manual before doing anything, and Yes, you probably need someone a bit different.

VS: [Some discussion of the MA in Archives and Records Management, Digital Records courses]. We get the students to download BitCurator, play with it, break it. We’re trying to get them to not give up despite things going wrong, find a different way –

FB: That was one of the things I really, really found with the project was that software developer and IT people see something going wrong as a challenge, or an opportunity whereas I’m just constantly defeated, ‘I can’t believe it’s gone wrong again!’ It’s a different mind-set, isn’t it?

VS: Yes, and it’s developing comfort with all of their, it’s that language barrier that you were talking about, developing a comfort with all of that terminology so that they can speak to their IT department about it, depending on whatever institution they end up in.

FB: What we found helpful was inviting people down here to look at hard copy archives, and to talk about how special they were, and how this is the equivalent of that, for today. Which I think was really helpful.

VS: And then how do we provide access?! Which is really my problem.

FB: Yes, so do you have a plan for how you will make them accessible?

VS: We’ll make a selection, a long list of 200, run it past Michael to get his consent, potentially run it past any third parties involved, and we’re still trying to decide how we actually make it available. There’s no way to make unredacted emails available on the frontend of ePADD, because they never considered when they developed it.

FB: Of course. For viewing you have to go to the archive to actually – I didn’t realise you couldn’t!

VS: No ridiculous institution is going to make their email available remotely! Actually, could you make it available remotely for us?

FB: Presumably Preservica has that front end?

VS: It does, what it’s going to look like to search… we’ve had a look at a couple of different institutions that use it as their front end for searching, and it’s little document icons – it’s trying to make email more visually appealing? So even if we end up making it available, the individual message files, or you know, a text searchable PDF of the entire lot, which depending on the researcher that’s accessing it, they might be quite happy with that.

FB: Yes, it seems to defeat the object, making it into a PDF.

VS: It does. But yes, at the moment I’m considering, if we make it available in as many different searchable formats as possible, then hopefully we should be meeting a range of researcher needs.

FB: It would be interesting to get feedback as well. When we asked people, we interviewed various people about how they’d envisage accessing email, and they didn’t – I think they struggled to imagine how it would work.

VS: Am I going to be in someone’s inbox?

FB: I suppose that makes most sense, really, is presenting it as an email inbox, because that’s how people experience email.

VS: Yes, like Elizabeth Gaskell’s inbox! If we could present it in an inbox, but then, if we’re giving them the individual message files anyway, they’ll be opening them in their own email client, so we don’t actually need to do that. But it’s whether a researcher accessing it will realise that that’s what’s going to open happen when they open these files.

FB: Does Preservica not let you select, display your email, the attachments?

VS: That’s the bit that I’m not familiar with yet, I need to have a play around with what the front end will actually look like, because the university doesn’t use the front end of Preservica.

FB: I think the only purpose we envisaged when we got Preservica for using their front end was the – well, it was actually any archives, because it’s all textual stuff, whereas most people are interested in images which is completely different. We did envisage using the Preservica front end, and that was one of the things that – I think – the Preservica webinar on email archiving makes it look fantastic! I mean I don’t know how far it reflects reality, but that was my idea, that we could do it all through Preservica’s frontend.

VS: Yes, and I don’t know whether it just needs – the other institution’s Preservica frontends that I’ve looked at are starting to look a little bit dated compared to other user interfaces – but Yes, that’s just something that we have to work with. And it may be the case that Preservica are about to update that imminently, so… it just needs a bit more exploration at this point.

FB: [redacted]

VS: [personal feelings about working with Carcanet over such a long period of time, with Michael]

FB: Entirely positive. He’s so involved with the archive, and engaged, and knows so many amazing people. I think we were really, really lucky to get that archive here, but it’s totally appropriate that it should be in Manchester, because that’s where he has made it such a big success. And the fact that we’ve acquired all these other writers’ archives as a result of getting that, I think it’s brilliant. It’s one of the most active, vibrant areas of collecting of the library, is modern literary now, and it’s largely due to Michael’s efforts.

VS: Thank you very much.

[ENDS]