**INITIAL COMMENTS**

**H\*1**

So who are the women in the official records?

**H2**

**Comments about the first visualisation:**

Yeah, I think this is really fascinating and I think this is the sort of thing that people will really love to be able to explore further because the story is always that the each of the universities, you know, Edinburgh, Glasgow, Aberdeen and St Andrews had hinterland from which they took students. And so the idea is that you would go to your local university and this is really amazing because this really shows just how Angus Forfeit. Perthshire is really and Fife are really the catchment areas of St Andrews. You know, OK, we've got people from that. So other places as well. But this is real mega attachment and then a little bit

of London, I guess again that's medical students. I love that one. That's fascinating.

**Comments about the transformation visuallization:** So that's these dotted lines here that represents Liz's work. And that's why it stops there. Because it's not OK. It doesn't

go up to the full 1897. Yeah, it's all seems very logical.

**Comments on the Transcription of LLA’s:** OK. Yeah, good. So the Lorraine helped you to fill in. Excellent. So this is Loraine’s spreadsheet then. That's amazing. Wow. Is this the LLA ID number? This is a transcription of the register. Can we have that back, please?

That would be really brilliant. That would be fantastic to put alongside the registers, cause

it would be it. It's in excel. Is it? So it will be easily searchable rather than having to play

through all the registers. That would be amazing. That would be really good.

**Comments on adding LLA’s:**

Who look. So this is adding in the LLA numbers in those periods. Gosh, what a difference those women make. Now so can I just be clear between 1892 and 1897, it includes the women who were registered on the ordinary degree programme, doesn't it? So

that presumably is these higher numbers here. They're still here in this bottom section where you've got the higher point coming in. So what you've got are these are extra. These are LLA ladies who are the extras. Wow, that's amazing.

**Comments on LLA’s in the content vis:** How much more varied their places of origin are? That's right down in the South of Ireland. And there's the sisters. That's interesting. We do

tend to have, you know, the association of being the family University.

**Comments on where to find information about where to find additional information about the LLA’s:** Did you use? Did you use the registers of general counsel at all for graduates after? OK, so when you graduate as a student, you become a member of the registered General Council and there is an annual register kept. And where it gives a current address and it's a new register for every year they write it out by hand every single year. And and I've used it recently to try to track where a particular woman went. She was an early graduate and she appears it gives her address. So she lives in Saint Andrews at a particular address. And then she moves to a different address in Saint Andrews, but identified that this particular lady stayed in the town all her life and she was down as being a teacher for the first few years of her life. But then she got married. And after she was married, there was no career allocated to her. Which is interesting, but that that would be the only the only place that I would find

that I think you'd be able to find consistent information after graduation, because those

are the typical records which the university keeps. To track the paths of the students, post graduation.

**Comments on the links to other Universities:**  Stockholm. Yeah, very interesting to see.

How much links there is between other universities by the relevant size of the dot,

presumably? Is that right? So most people go to Edinburgh then Aberdeen, Oxford, Cambridge and London.

**H3**

**Comment on going through the screenshot like a book:**

Question. So I've got a picture here, but what I can't do is just go through it like a book.

I just have the selected page that you've chosen? It's what I would like to do as a historian, I would. I would very much like to be able to do that. I think that would get me into Ooh, that's interesting.

**Comments on Smart’s family:** I didn't know that Robert Smart’s family got roped into the whole thing.

**Comments on spelling and database:**

How are you dealing in the database with? So if you look at George Galloway and he's

categorized as George Galloway, but actually he sees that the way he's spelt in the original

was rather different, how is that?

**Comments on Matriculation Roll:** Matriculation gives role gives you a nice sort of thing to hang it off because Issues that I've encountered with trying to deal with data on 16th century people. I've done some stuff to do with property records and the names are massively

inconsistent, massively and there's quite a lot of issues around, definitely establishing

whether you're dealing with the same person and things, and also trying to record the many

different forms of spelling and things like that. And it starts to get into a bit of a

mess, but at least this so you're like, OK, look here. This person matriculated at this year

and that gives us sort of clarity to the data that isn't there when you're dealing with,

like, multiple property documents. You thought you find, like, certain if you're looking at the

borough, Saint Andrews, you get more than one person with the same name cropping up.

If you look at the university now I'm sure there's more than one person with the same name.

**Comment on going through the screenshot like a book:**

I have to say, it’s tantalizing. You have these screenshots and I would love to go

explore the data more. Sorry, that's not helpful.

**Comments on LLA’s:**

I thought what's really interesting is that I'd have to confess probably with prejudices and

vibed from the university. I knew of the LLA scheme as very, very important for women's

education. But I had thought of it as something peripheral to the university and it's really

fascinating to see far more people doing LLA scheme than those who are actually studying at the university conventionally. And that's something I was oblivious to. So thank you!

**H4**

**Comments about Elizabeth Garrett:** Elizabeth Garrett, wasn’t she a doctor?

**Comments on Smart’s transformation:** So it’s alphabetized and also it gives you all the information of what they studied and the years, and what they did afterwards. Wow. What a legacy.

**Comments on XML:** So is this sort of like a general kind of widely usable computer language or is it a specific one for this?

**LLA graph:** Wow, so there was a lot of LLA’s

**Searching for Maria Imandt:** There she is. Maria Isabella Imandt.  Her father was Peter and he was a friend of Karl Marx and because of the revolution and everything and if he was caught you

would be executed. So he flared, over to Scotland and then he ended up in Dundee for

Some reason and then he taught at the high school. And then, Mary Ament she did the LLA really early on. She went around the world for DC Thompson as a journalist. The first woman journalist to do so. It was her and Bessie Maxwell. And so she went around the world

in 1892 and 1890 maybe and she sent back articles and things. She had a camera with her and it was to send back information on how women around the world fared. You know, so it was way ahead of the of the time 1892. It was really, really early on and then she came back and she

Lectured about their adventures. So that’s not here [in the records] with men, they're all sort of information about what they did and why and nothing on here and she's so interesting. And she's only one woman, I bet all the other ones have got such a great story. That’s why they were all doing that because they thought it was unfair. She writes about the suffragettes and she

goes to the court cases where they were, they were arguing for equal rights and all this. And she writes about it. She's a very, very clever woman. So how did the distant learning work?

**H\*5**

**Latin names and abbreviations**:  It was more that you have got William in short form here, but then you've got I think that is. The Latin version of William. But then you also have like Andrew. Which is, you know, absolutely not a Latin version of anything. So there is an interesting like variation in some people go full Latin, Jacob for example. I'm assuming no one referred to him as Jacobus. But then you also have or or Charles.

**Wanting to see the original version of the 1413 matriculation:** I kind of wish you also had this is the medievalist who goes. I'd like to see the original. Having the incredibly horrific, you know, 16th century script wouldn't make a lot of sense of like ohh yeah, that's why it took time.

**Length of time for 1413 transcription**: That took longer, but that makes sense because they are much harder to read.

**1579:** I'm assuming that's also not very fun, period to try transcribe, given the four of them, we're doing it for a very long time.

**1747:** I can't. I cannot believe the amount of work he put down. Smart I mean.

**Comparison**: You have these massive information gathering periods and then actually getting it into databases is very small.It shows how much work just gathering information is.

**People icons:** We got a fourth person.

**Family members working:** I have to admit, I really love the fact that you've got these, you know, enthusiast archivist, academics, and then their families just get roped into it.

Like if you want to see Dad, this is how we're gonna do it. That would be great now, yeah. Transcribe this page first.

**LLA not included:** Of course not, because they weren't doing actual degrees according to academics.

Of course not, because they weren't actually gonna do anything with their degrees. They were just gonna get married.

**Missing lines:** Well, why do they not have a line out of curiosity?

**Questions about remote studies:** How does that work when communication takes months?

**LLA’s birth places:  I**'m also entertained by the fact that the birth places for the LLA's is

much more diverse than for the men. Now of course, some of that is the fact that the men actually traveled to Saint Andrews and therefore, you know the ones coming from Cork are less likely to go to San Andrews. But it's still an interesting Difference between you know,

Everywhere to more like Scotland

**D-tour:** Diderik Cappelen

**CS2**

So I'm amused that you've gone from names to just we.

And all the previous ones. I sort of like librarian Alice Crawford or Robert Plant. And then with point where it gets to you, doing stuff, you just go to entirely neutral.

It's probably not relevant to your study but you should take more credit for your work.

**Comment about Smart:**  It's amazing that they did this and one guy did this for 52 years.

**Highlighting people who worked on it:** Ohh, you highlighted him down here, but it's really cute.

**Comment about LLA’s:** Was it really that big a jump? Oh, my goodness. And you didn't know about this till you started doing the research?

**Career LLA:** That might want to be like noted somewhere, because at the moment it's sort of implied that all the people that did the LLA like if you weren't here to explain, I'd be trying to. I'd be drawing all sorts of interesting mental conclusions about why the LLA resourced.

**Birth LLa:** Is there a way to like switch off the lines?

**Clustering of people who actually went to St Andrews:** It's interesting how much. How kind of more clustered the male students, the people who actually went around, St Andrews? I mean, I guess they had to actually get there, which was nontrivial.

**CS3**

**Comment on layering:**

I like the layering of the history of the time. And then the way that it's kind of combined with this storytelling to explain like, you're not just presented with all of the information at once,

it's really cool how it it builds up in such a sense that you start to see firstly, it tells a really interesting picture on the levels of details of the various pieces of history. But also it shows just how much work and how many processes have gone into creating this data so that it's

usable in the digital format now. And that's super cool. I don't think you would get the

same appreciation of the like the work and the gravity without this build up without the story

so yeah, that's good use of storytelling for sure.

**Comment on the path:**

So what, uh, what differentiates whether someone gets a path or not in this visualization?

**CS\*4**

**Smart’s transformation division: “**What do the rectangles mean?”

**Inquiries about the LLA exam results**: “What do the subject numbers mean?”

**LLA comment**: “Some of them did a lot of subjects. “

**CS5**

**Comments on transcription:**

And the time it took for the transcription is that including the entire process of creating the printed book that you could see on the right?

So the transcription processes, is that Typing up all the signatures. Or was it also like handwriting the signatures again in a different organization or something before

they would be typed up and printed?

**Comments on Smart:**

So in this part of the timeline, what did the boxes represent?

**Comments on Digitization:**

And then these two visual representations are they just indicating that at different

data formats? Or is there a significance to like the number of horizontal lines for

everything?

**Comments on the people:**

I like the people at the bottom too indicating like who was involved.

**Comments on LLA:**

Interesting this remote learning that long ago.

**Comments on Study:**

And so with this would this be like a research visit during their time at Saint Andrews? Or

is this like after their time at Saint Andrews they did another degree at another university?

**INTERVIEW QUESTIONS**

**What for you is the main ‘story’/point of the series of visualizations?**

**H\*6**

Well, there's the I suppose it's the background of the project that you're working on, so

you've got your narrative of the progress. And then the potential of how this data can be represented in different ways.

**H4**

You're getting so much more information out there, aren't you? Yeah. And you're

getting a you're you're getting a wider story about the university. And also, well, how badly women were treated in the university as well.

**H2**

And are you talking about the whole both, both these pages and the the circular thing?

OK, let me go back to the circuit to the spectrum thing. And I thought this was really

impactful, actually seen the impact of the arrival of women. And I like the idea of

seeing the continuity of the record creation visualized in as a circle. And that's really

helpful because I think that gives a sense of the fact that this is a an ongoing task which

covers a whole set of years of the university's history, but also is tackled in various times by different people, and I think this this gives a sense of continuity and effort and endeavor. So I really like seeing that portrayed in a circular fashion and already and being able to visualize which aspects of the work on the collections was done at at what particular time. And I think

it's really helpful to draw all that together. And for you to be able to show that. So if you've got the whole circle, I really like that. I think you need. It needs explanation. And the way that you work through that gradually and then steps to show it building up was really helpful.

I think it might be quite difficult to interpret all those different things. Coming at that full diagram called. I mean, even if you had labels on it saying what they all are, I think if you were to just screenshot that and present that as a thing, that would be quite complicated. So I think the way you built it incrementally was really helpful. And another thing I really liked was the linking in of the original source data that you've used to build the schema. I think the

idea of having examples of the record types and of the original sources is really beneficial because that helps you to understand a bit more about what kind of information would

be looked at? And what is then put into your background databases. Perhaps less for me, but more for other people. I I don't need to study those pictures. So for example, where you where you had your well, any of the books Not those, because there's a printed, but the handwritten books. So if you had the the register of the LLA’s or the handwritten this stuff for

example I, you know somebody else might be fascinated and might get hung up on looking at this and thinking, wow, that's really interesting. What is that telling me? Who are these

people? Why are they signing? Where are they coming from? And you kind of get sucked into the record. I'm happy to leave that because I'm familiar with those sources. So that's not so much of an issue really. But I love I love the way you've kind of built it in a circular

fashion. I think the the, the maps and the, the geographical coding of everything in relation

to where people go, where they came from. I think that's a really powerful tool, so I'm just going back to the end so that I can get back to those. I think that's very powerful. I think that actually is a way of getting a sense of the information in a new and different way. I like. I like the idea of being able to interrogate the data myself in this fashion. Whilst at the same time knowing that I've got the confidence of being able to go back to the alphabetical version if I need to. Umm so I I would. I would say that while this is really great, really interesting and enables you to engage in a kind of drop in and analyze way. For some purposes, you still do

need to be able to go back. To the original record, so this couldn't be used easily I guess, to look for unnamed individual. So this is allowing you to to draw the bigger picture and to

enable people to get bigger, higher level patterns and trends and places of origin for

all these different people. In a way that has never been possible before, because we've

only ever had individual itemized data at the local individual level, which is very

valuable and very significant for one person research. So the benefit of visualization is that

it allows you to take that base level data and tell new stories about the data and show things

in a new and different way, which really makes it much more impactful for people who don't need to know about the specifics of the records. They need to know that information that's in the system is validated by the fact that you've gone back to the originals, which you've clearly demonstrated. So this seems authoritative. You haven't made it up.

And and and you have cleaned the data. I mean yes there are flaws, but basically you've

cleaned the data and you've made it as robust as you can with whatever caveats you want to put in place. So what's exciting about this for me is that it enables new stories from old data.

**H\*1**

Well, I guess for me it's the story of the manual Labor and how long it takes to produce

these outputs. The number of different people involved. Working over successive generations, often bringing in family members and mixture of students and archivists.

The the job is really a sort of team effort over a century or more which is something that I don't think is often apparent to the end user. And I think if anything, we often quite it's

almost seems embarrassing to say that you spend a lot of time and money repeatedly on

the same set of records because it makes it sound like you didn't get it right the first

time, or that you're just retreading the same ground. Whereas I think this does a good

job of explaining the sort of layering. Hopefully with the intentions that you're making it better and more usable each time, right, and adapting your technology and that kind of thing.

So I think it does a good a good job of telling that story that we don't really tell well

Otherwise.

**H3**

I would say that there are surely two points. One point is the story of how the

data made a journey from these manuscript records through various forms of publication and

digitization, and I think it does quite a good job of reminding people of the labor involved in such a project. So I see that's one story. The other story I'd see relates to rises and falls in

student numbers and also the relationship of official enrolled students in the university and the LLA’s. So that's what I’d probably see as the two main stories, I mean, I think from my perspective, I would have been very, very interested to see the data visualized for the

earlier periods as well. I do find that really interesting. That probably relates to the

periods I am myself interested in. My area of research is mostly 16th century.

I'm curious, does this spike does that relate to when they have the sort of early 19th century

reforms of the university?  Yeah, for me, the two stories are the labor involved in each stage of sort of editing or transcription or uploading to a new form of record keeping. And also that story about the relationship between the LLA and the traditional students.

**H\*5**

I think for me it's it's there's two stories here. There's the historical story, the story of the students, be they, you know, the male students who went to Saint Andrews, be it the female students who, never actually went but still studied officially. Be it the register the the handwritten original record. So that's the one part is the actual historical data. You can do analysis on. You've shown the possibilities of this to allow us to see the history and like

the bigger context rather than individual students and at individual times the ability to

do bigger statistical analysis. But also, what do we even call it? I wanna say like metadata. The history of the history of the register. Umm, the history of the history of studying the students of Saint Andrews. Here the the fact that this job of just transcribing handwritten stuff and then gathering information further information about these people takes so much

more time and so many people compared relatively to the modern effort, which admittedly, probably has more people attached to it, but for shorter periods of time, like I was only there for a year at most. And none of the work that I did that you've done would have been

possible without those. What is it, 60 years of preparatory work? Like without, without

smart and without the Andersons. Like none of this would have been possible, and so that

combination of the history of the history, the historiography, I guess. And the actual history is a very interesting and valuable tool. And I think with with this kind of particularly with this kind of data analysis where you've taken, you're so far removed technically from the original

source because I mean, the original sources for this are what smart, collected and and

the original university registers. You're so far removed from those sources that

actually having this tracking of what happened here is just as important as the actual data.

Uh, now, now smart did a very good job. He referenced all of his stuff like, you know, in

that sense, it's not really a problem, but I at least I found when I was working with medieval history that you couldn't just know the source. You also had to know how the source got to you and this is knowing how the source, in this case a massive data set, came into existence. And I and I think that's that's a very important thing to to keep track of and to show and also you know for funding reasons be like look this took 60 years to

gather. We're gonna need more money. Are you looking to continue? Yeah, I mean, obviously, you'd probably want to scale it differently from smart, too again, spent what, 30 years working on it? But but I think if you you could probably use this to kind of

justify that this is gonna take time and it's gonna take effort and we're gonna need, you know, a significant amount of money.

**CS2**

I think about the kind of records transcription with a nice little flourish of look at all these forgotten women at the end. Yeah, it was nice seeing all the people that worked on this project for such a long period of time.

**CS\*1**

Well, it's interesting because it shows the history of the University of the students of the university and what we know, and it gives a good idea of how records and what was collected over the different time periods actually changed. And what we were able to

gather from it because of that. Because of the way it was recorded. It emphasizes very

much like the LLA records, for instance, emphasize how having an additional source of information can just really mean that for certain range of time you have a lot more data

and a lot more detailed data to an extent. Because you've got not just the fact that the LLA’s weren't that detailed and in themselves, but they added a lot more

information in the facts that you knew that they were LLA students. So that gave us a

whole new element to add to all the visualizations so the point of it was just

showing how much the amount of information you could get out of data and how it changed over time and how many different sources of different types of data there are and how

they could all combine together to make a more complete set. And it's just it's just interesting

to explore the history of the student population of the university.

**CS3**

Yeah, good question. I think there are, you know, going back to the point that I made earlier

about the work and the effort involved in the transcription processes, I think the

storytelling aspect of that really drives how much work goes into these different processes

of digitization and you know, as a as someone that works with data in data visualization I sometimes kind of take it for granted that data is readily available, and yet there are clearly topics like this where a huge amount of effort needs to go in to make the data available for creation of, you know, even a simple visualization would take months of human work to create the transcription required. So that's really impactful for me. I think the wider story as well on just how small a piece of history women in higher education takes up and also like the very fact that you know, I discovered there is a different book for all of their records and they weren't first allowed to study there. Like there's a message about women in academia there, which I think is also really interesting to see visualized here. And that's

compounded with the efforts in digitization and like the accuracy of the records and stuff, right, so like not only did they have less opportunity to study there but also there’s simply

less record of women's involvement in academia, and I think that's, you know. It's it's a really quite powerful message actually just that, that suppression of their presence in academia.

So that those two things. Effort involved and the role of women, right, in academia and in history.

**CS\*4**

It is a story of the data rather than the story of the university and it’s interesting to see how people have dedicated years and years getting the data into useful formats. Do you know what purpose or uses the early attempts transcribing typing up names and dates? Did they have an initial purpose or they just thought, oh this will be useful one day.

*They turned handwritten stuff, which is difficult to read into print so that everyone*

*can read Everyone can read the print, so it opens up the space for the*

*Information to more  people.*

But you don’t know if there’s any early uses of the data.

*It was record keeping. To keep track of information about students.*

And LLA staff is very interesting because there's so many people there who were discriminated against. The scale of it. And people [LLA] making the best of it by

getting some degree rather than not getting it a degree at all. That must be a surprise to a lot

of people.

**CS5**

I think there's two parts I guess. So I think initially had started out as just this process of working with older records and  transforming them into different formats so that further research on them can be done more easily with the tools available to us with digital technology these days and there's also this sort of build up of different people involved

in creating these records that you kind of see going on as well. And then I think towards

the end I think to me, I mean this was I am a woman. I'm also doing work on gender bias, so I'm quite intrigued by the kind of existence of this whole other record. Or set of records that that was found during your work and the story that kind of comes out of that of all these other women. Like this spike is just really quite impressive to me. There has been a lot of people that were very excited, a lot of women that were very excited to be able to enroll in a

university, even if they couldn't go in person. And so that to me just opens up a lot of

future research directions and says a lot about what isn't being included and the

kind of information that we just don't have that limits our analysis and limits the ability to be

certain about the kind of analysis we can do on on historical records.

**CS\*6**

That's an interesting question. Because the sort of like my personal preferences were oh, look at the cool women stuff towards the end, but I guess also probably I would read it more

As the difficulty of actually getting information out of records, right, not knowing what

isn't there. So this discovery [LLA] is  really cool and Is missing for quite a long point of this overarching story because old records are complicated and not very accessible. Does that

make sense?

**How do the transformations differ? What do they have in common?**

**H\*6**

Would you describe this [content vis] as a transformation? I find this [content vis] easier to understand. This [provenance] is not so intuitive for me, but that's maybe just me.

Yeah, maybe I'm too simple because this is the I like this one [original circle].

They were all graduation and matriculations were they? The data is becoming more complicated.

**H4**

From written sort of published very simple published in pages to an awful lot more inclusive and much more wider information. What they have in common is that it’s the same information.

**H3**

I think it's interesting how much more quickly you've been able to do the recent digitizing.

I think honestly, coming at it from a historian's eye view, I think that indicates on the one hand how computers are quite quick at processing data. Am I correct in thinking that you were able to automate some of it? It wasn't all keyed in manually. Did you say? Yeah. And I think that's really interesting. What you can do. I do think it's slightly deceptive because I think I am not at all surprised that it took Robert Smart, the amount of time it did to do that work. Alphabetizing and looking at the rest of their experiences and things, and I think for me as a historian, I'm very conscious that most historical records are not in the form that computers want the data in. And that actually, most historical records are in a forms that the

data is quite tricky to use generally and that there is a huge amount of human labor involved in that. And that's just incredibly time consuming. And when you're trying to do what Smart is doing, not simply taking one source and transcribing that and editing it, but trying to link it up to other sources. The moment you start to do that, it becomes a huge task in terms of research and I think it's massively to his credit what he did. But I think there's potentially from

historians perspective, a really, really deceptive story coming through in the way the

data is portrayed here, which is that like ohh, as soon as you go to digitization it all comes

simple and it's not. The only reason why that digitization was possible was because of the

decades of work that Smart had done if that makes sense. That would be a take away that I

would have even though I think what you've done is a really, really interesting visualization

Of how long it actually took for each project, and I think that's a really interesting story in

Itself. And I think the point you're making that there's all this layers of work to get to the

ultimate digital resource that you've got, I think he's a really interesting point and I don't think we spend enough time either as the public or as historians talking about the layers of work that we are reliant on. But I do also think that there is potentially something slightly misleading.

**H\*1**

What they have in common so. I mean, they're all actually fairly manual processes. They

all pretty much involve a person or a team of people working in detail over time on

something. The intention I think is to make the data more usable, more accessible and to allow different kinds of interrogation of the data by different audiences. Making use of the forms of communication of the day, whether that's a printed catalogue or a database or a an

image on the screen, or a visualization. What's different about them? I mean, aside from the fact that they literally different forms of data, they look different and they take different forms and they cover different periods. I would say there are fundamentally similar.

**H\*5**

I'm a big fan of this one. This one's really important [LLA graph]. But also I wasn't really aware of the Andersons. That was unfamiliar to me. New information.

Uh, interesting information. But it's one of those things with, like, the Andersons and

with smart is, you know, like logically you've been told 30 years, right. They worked on

this for 20 years or 30 years or whatever but seeing that represented as time.

Makes it a lot clearer. Just how much time we're talking about, and of course for me,

knowing that I'm like a tiny line in here somewhere. It it definitely makes that point.

I yeah. So I think for me that's one of the surprises and it's not a surprise technically. Like logically, I knew the information, but it's a surprise in seeing it like that makes it

very clear just how immense the job actually was. And I guess also in that sense

the the change from the the first source that you had pictured which was the

handwritten roll up to this coded would have made no sense to the poor student

from 1867. And honestly, probably makes no sense to half of the people who look at this.

Is also very interesting. And also the way that you have the different ways of database we

got that database, you got this one you've got the different kind of visualizations of the same technically the same information. Bit different ways of dealing with it. And then if I can go back to this one [content] I also found this one. The distinct difference between

the people who went to university to get a career I mean they all got careers, obviously.

I mean, these guys also got careers, but the amount of people who did the career

versus the amount of people who actually went to do other universities. We're connected with other universities and it is a very interesting difference. And obviously you know there's a lot of stuff you can do. This is just a start, like you're just scratching the surface. There's

so much you could do with this. Just seeing the place names cause I've looked at

the place names a million times You know I've read through all of these records and.

Having it visualized as actual places, making it clear that you know obviously for the men you know most of them are from here. It also makes the outliers, the poor sods that came in from Shetland or you know, the one guy. Sorry, two guys from Germany. It makes them

much more clear how, you know, I came in just under an international student and obviously, you know, there were other Norwegians, I came to Saint Andrews with people I went to high school with. So, you know, not very similar but coming to Saint Andrews as an international student, as possibly the only student from I mean, this guy was Norwegian. These two are the names suggest they might have been English. Just born in Germany. But you know, coming in as the only person from Aberystwyth, you know that maybe the only Welsh person in your class. Is gonna be a very different experience when you're that far away from home, when it isn't just a phone call or a plane or whatever. It's actually a bit of work. Let's say he died in 39. So, like, he's there for the trains, right? But in the earlier period, there weren’t even trains, how did you get all the way up there? By boat? It makes it very visual how far some of these people traveled from what they grew up with. And I think in if you had the whole map from what I remember like the the amount of people born in the colonies in one shape or form of course is is massive as well. Now obviously for them it's a different life story because they probably would have ended up in England anyway because, British family, but still having been born somewhere and moved so far, and sure, most of them probably like this guy went back home, right? He was away for a couple years and he came home. But it kind of illustrates the mobility that we have a tendency to forget existed in the pre digital world. In  in a similar way too, I went to a museum in Oslo recently because they have an exhibit on animals in Viking and Iron Age chart. And multiple of those art pieces are either actually Frankish or inspired by Frankish belt buckle things. Which suggests that you know either someone's down here. Or someone or things from down here are making their way all the way up here. And that kind of connection is easy, easy to forget when we think. Ohh, you know, they didn't have airplanes and everything took so long. Why would they? But actually, there is a huge amount of mobility. At any period. And I appreciate the visualization of that and also with these women and as you say, there are a lot of internationals of like the the sheer bureaucracy needed to get these women through this education is very fascinating, even, of course, with this particular map, it's largely British. There's only a couple. I with with you know, the full colonies it it's, it's just an interesting.

**H2**

I would say that the base level was simply a transcription of the list that was there in hard

copy, so it was just a list of names and what it was doing was printing a list that was held in

manuscript to save people having to go to the original manuscript registers to enable people in those days to get a printed book with a list in it, which is indexed, which enables you to

find the person you're interested in. Gradually, as the as time has gone on, if you look at Bob

Smart as the opposite extreme to that. He is researching into each of those named individuals and what he has done is he's taken as his starting point, then list of the names from the register but then he's taken those as the people that he has to research and he has then gone on and he's done different aspects of research. He has researched within the university's own record and brought together things that are relevant from other places

alongside the core list of names within the universities own archives. But he is then also

gone to extra places as appear in his bibliography, so things like for a medical student, he's

gone to the medical register and he's found out that subsequent career from the medical

registry, he's used obituaries, he's used newspapers, he's used the alumnus Chronicle. So he's he's brought a lot of extra resources to bear which has enabled him to build a bit of

the life story of the individual as well as just the bare bones of when they were in the university. So the  Bobs biographical register on which you base this is a much fuller record, but you haven't taken all the information out of Bob's register and visualized it here. You've restricted yourself to certain themes. So if you applied this methodology to the extra data that's there, you could tell other stories. But my question would then be how consistent was Bob in the amount of research he did for each individual and was it possible to always find out information about a medical student? And of course, it probably wasn't. And was it always possible to find out where ministers went? That's probably easier, because

if they were in the Church of Scotland, the fasty of the Church of Scotland, the list of

of ministers does give you a really good way in. And that's a  reliable resource.

But, sometimes I guess for Bob it would be spot check. It would be, you know, if it happened to be somebody came across who'd gone on to be a student somewhere else and there was

information about them at Cambridge or whatever he could incorporate that into his register.

But I don't think he would necessarily systematically have been able to find the same level

of detail about everybody in particular categories of student. Umm, the next progression into

the database format is apart from the LLA work which you've done where you have

actually created new data because you have transcribed original registers and you

yourself through Lorraine's work and Liz's work, you've amplified the base data to make it consistent to include the LLA’s, that's new in your particular study. And but other than that, you're basing your visualizations on data which was created by other people, and which you have taken as accurate and as the base for your work. So what you haven't done, and it's was never within your remit, is to check any of that information and make sure it's right. You've taken that as right because we know that Bob was extremely diligent and careful and reliable and it was 40 years of his work. So he, as much as possible made it right. And so I think I think what I'm saying is there is a, there's a a sort of a curve on the amount

of information about the individual that's produced. But that it's not necessarily the same

curve for every student and every type of student, because sometimes the records don't

include that information and they'll be stars about whom Bob could have written a whole book with the amount of information he had on them. And there be other people about

whom he knows very, very little. And I think there's always the problem that the underlying

sources don't necessarily always tell you everything you want to know.

**CS3**

I like the both the visual representation and the images which are presenting seem very consistent. I really like that there is in my mind a kind of three stages to the information. One of them is just the hand records which were probably practical at the time, but which have no

functionality for searching or organizing like it's just as the records came in, then the next

stage that transformation into a more organized kind of historical archive I guess, which is where you've got the either the printed records or they're starting to be categorized, or

they're put in some kind of format for for reflection and understanding. And then the

final transformation, which I think is it takes various forms here like the XML documents, the

databases, blah blah blah. But the idea there is that the digitization enables new insight

to be generated above and beyond the original organization and so this is where you can build these visualizations from the from that enhancement simply by adding structure.

**CS\*6**

I guess they differ in a whole bunch of different ways, right? Ohh, that's a cool one. Hey, you can see this person was involved in both of them. One of the things that I was gonna say was the number of people involved differs. Some of them are very small efforts of like, one person being interested in things. Some of them are lots of people being interested in things. There's a bunch of transformations between different formats, both in terms of digital formats like moving to XML json, sort of different. Different markups and structured data. And some of the transformations are more about structure and different ways. So if you go back to the earlier ones, some of them  Just layed the data more nicely, some of them have more information in them than others. And in common, They are all a step forward in each of the transformations the data is slightly more accessible than it was in the one before. Some of them loads more accessible and some of them are a bit more accessible, but it's always it's never a step back.

**CS5**

I think. I suppose they differ in terms of the time that they take so especially from the period where things were being transcribed by hand and then printed versus being able

to have information digitally and going from CSV or Excel file to a XML format or XML to database format. That kind of thing is much faster. And then I guess what would be

similar is that it seems like all these transformations are about improving access to information, making it easier for people to discover the records and do further research on them.

**CS2**

You mean sort of the manual, then the digital, then the database? I mean, they kind of progressed to the most recent they kept progressing to the best data keeping that was available at the time, you know, handwriting meant to sort of alphabetized and then digitizing them and then putting them into structured data and then to databases.

**CS\*4**

Looking at the big circular thing sort of at this stage you're using different dimensions and things and for lots of different things. University in the center and then the number of students and lines coming out of the center and then you've got further out, you've got the dates and the history of the dataset. The work was done on it. You've got the names and years and subjects [original records] and that's pretty much all you’ve got there then. And other people later on consulted a lot more records to find out what they did later and where they went and lots of cross referencing of other sources.

I don't know if it's because of the number of data points that people had for each student. So here you've only got their name and date and what they studied, ‘dives’ or ‘pauper’ whether it had something to do with their funding. And the way the stuff is laid out it's sort of harder to be harder for a computer to read this and then later on it becomes more tabular and then.

When you see the XML or json is properly structured. Whether that has something to do with the history of data visualization. And the the closer we get to the present day, the more

information there is on people.

**CS\*1**

Well, some of them were just some transformations were just sort of very simple. Digitizing or not necessarily digitizing, but transcribing information as it is not organizing it in any way, just taking here's that data and just copying that down as is. No, no information learned from it. Just having it in a more readable form, so that's that's not really. It didn't make it easy to

analyze because it's just jumbled up the same way as it was written was originally when

it was written down. So some of the transformations were like that. They were actually not trying to get any additional data out.They were just trying to keep the copy of the data in a more readable format. That’s the original here [transcription] where it shows everyone that was mentioned before just in a more readable way. I don’t know if the original source was separated into these but to have them transcribed to this I’d imagine they probably did have them separated in this way. So then there were other types of transformations, such as when we go the alphabetization where it is actually getting some information out of the data, so instead of just just transcribing it, it's actually making it more parsable. So you could look for people you could look up a certain surname so it's sort of useful kind of data and transformation. Yeah, you could give a surname you could look up a relative, for instance.

And that's more useful resource to look up without having to rely on on dates. You know, somebody went to the University of a certain year. You could actually just look them up by name using this format so this makes it a lot more searchable. So that's the value of this

Transformation. The next one, yeah, obviously XML makes it possible to be parsed via computer, get out of it. It's not. It's not obviously as useful as database which came later, but it but it allowed it to be computer readable and can be understood by computers and because it can be understood by computers, you can get information out of that make databases out of it, which they did later on. And this means you can do anything that point that step, that transformation to the database that was the most valuable transformation because at that point you can just and you can analyze so much get so much unintentional data that wasn't. The data wasn't originally meant for. You can get so much of that information out of here now. So and so, so like percentage of students that were female versus male and things like that, which you couldn’t do with the original data or any of the previous versions very easily. With XML you could do a search for all instances of females, but it's not a simple process of just doing a database query. And then the visualization

that's also a very valuable transformation because it allows you to actually explore all the

the data in a visual way.  If you don't know exactly what you're looking for, and you just wanting to explore and see if you find anything interesting. I think the visualization transformation is very valuable for that. Especially the final one where it's obviously interactive and you can find out information by hovering above it but I assume there was some level of interactivity in the previous visualization one as well but but. But yeah, the interactive visualizations are very good for just exploring data.

And if you don't know what you're looking for.

**What gets added/removed from the records after each transformation?**

**H\*5**

Well. I think this is one of those cases where, as we've been doing today, you need the records as well, right? We've been looking people up because we go wait a minute. Why? Because the thing with this kind of data analysis, this kind of graphing and mapping and plotting is it gives you a very good overview. It allows you to see the big picture. It allows you to draw, you know the the big lines, the big trends, the you know, and you, you get a couple of the outliers where you go, huh? But for the most part, like London. Like they might as well not exist as individuals, because that's just a BLOB. And it's the same with. If I go

back a bit like these women, right?  I mean, obviously they're important as individuals, but in

this particular graphing, they are not individuals. That's not what they're there for.

Uh, they are there to show the overview and that's very important, because with history

Often pre digital tools this overview Isn't always there, and particularly from having worked

in medieval history where source material is, you know, sparse and spotted, and what you have is very little. A lot of the work you do is anecdotal. It is based on that one person or that

 one place or that one time. And so being able to use modern technology to do a digital

humanities work like this to get the overview is important. But it's also important to not lose

sight of the individual stories in it. Or the individual places, or the individual careers or the

individual like. If you focus too much on, for example, the maps, you might forget that.

Or you might not see that, for example, all the people in London are doctors. I'm not

saying they are, but you know, like there might be other patterns that you miss because

you've only mapped certain thing. And also. Like with this particular one, obviously this is, you know, one version and a small version and not, you know, a full version.

But here for example, I'd have to look up, look, look them up. Let's go find the one of the

men. So I'm actually doing the right one. Like, I'd have to look up John little to see where

he when he went to his dates, his career, his life, anything like that. I'm only getting kind

Of. Bare minimum facts whereas you know we probably know quite a bit more about John.

that isn't easily represented it it's not. I'm not saying you you should represent it. It's just it's the nature of this kind of large scale mapping means you leave detail, you lose the details and it's a discussion we've been having at work about AI. is the the importance of.

Accuracy and precision i.e. The details versus being able to get an overview that people can

actually manage. Because I've read all 14,000 of these, you know, however many

entries there were, and it's overwhelming. It's you cannot as a human. You can't draw

these patterns out of just reading the records. It's impossible. We don't have the brain power for it but computers allow us to do the big scale mapping, we just have to be aware that we lose out on the accuracy, we lose out on the details when we do it.  And so they both have important things like this following the individual, but also mapping the overview. They have important things to tell us. At different ends of the scale, and they're both. You can't understand the records just by doing this, but you also can't understand it just by going, you know, record for record for record.

**H\*6**

Smart is actually pulling in lots of different resources rather than just a list of names.

Smart is helpful if you're looking for a specific individual. But if you were trying to do statistics, You know how many people came from the specific country or how many students were there in a particular year, It's very difficult to pull that information out from an alphabetical register. Which you could potentially do with visualization.

**H2**

So I mean, I would say that in terms of that book particularly, I mean what the

way that record is kept is by class. So the 1st and 2nd and 3rd and fourth years would

all register and matriculate into their year each year and they would each send

the registry each year. So what you've got there is ability to see who was in class with each

other, who are the peers, who was actually sitting next to Andrew Lang when he was a

student in the 1860s? That kind of thing. And you don't necessarily easily manage to get

that information when you're doing it through your database. You would have to reconstruct

that by doing a search, and that isn't something that you'd have to do because it's all on the

same page, so it's very easy to go out to the original register and see that. So by dealing with each individual and they abstracted information in the biographical register, you have

lost the context. You've lost some contextualization of the original record. Now, I don't

know how anybody who doesn't know the records would answer that question. I think it would be quite difficult for them to go back and think about the difference between what you've given. And the raw data in its original format. And I mean there are, for example Bob uses things like college minutes and Senate minutes to show that people have got degrees or that they were bursars, all that kind of thing. So he's abstracted the information about the individual, about their rewards, or about their particular position. but you lose all sense of historical context about what that means. If you go back to the original record of the

college minutes or the Senate minutes, you can read about what else was being discussed at the same time, and how many other students were also being nominated for the same bursary  or scholarship, whether there been a competition, and whether the person that you are actually checking because they had a bursary, whether they had had to write an essay or take an exam, or whether they were the son of some professor and therefore got it because it was privilege, you know, so you're losing, you're losing a sense of the nitty gritty detail. So I would say that. One thing I would hope your study does is to encourage people to go back to the original sources because the very fact that you've kept the references in, I mean I'm  just looking at this now, you know for these people, you've got your references, you've got united college minutes there from 1897. That actually tells you what that where this person was listed. She was also in the university or register. She's mentioned in the citizen. So you've got the ability to follow up. And it's brilliant that that reference material is there in the in all these entries because it does show A that it's authoritative, that B, that there is potential for the places that you could go to and check this to validate it to make sure that if you wanted to you could do more research yourself. You could get into the university and say can I have a look at UYUC400 because I'm really interested in seeing what the college said about this woman Gertrude von Petzold, you know.

**H4**

It removes the human aspect. Well, it just removes that original source, doesn't it?

And there’s the family information and what they do afterwards with it. How would you have found that? [explaining] Getting all that original information and everything else. Yeah. And digitization kind of makes it more accessible doesn’t it?

**H\*1**

So I guess what gets removed is the sort of a particular context to sort of particular material reality, the form changes overtime, you may lose sense of the original labor, so perhaps if

you're using the database you don't have a good sense of what Bob Smart did a particular piece of work that maybe not not be evident to you.

Yeah. I'm just saying. So I think potentially what gets taken away is the yeah, the,

the, the site of the Labour, so A database drawing of Bob smarts work may not be immediately obvious that as Bob Smarts work in the way that it would be if you were holding bobs Smart’s volume in your hand.

In terms of what gets added. I guess it's a sort of a layering of functionalities, possibilities

and a possibility for changing context and changing kinds of use and new access points.

That's it. Shifting meaning over time essentially loss of meaning if  kind of atomize the

data in different ways and rearrange it.

**H3**

I need to go back and have a look at each one. And you've got the original record, haven't you? And I think it's very, very important to recognize that that is the thing that is most

essential. That's on what everything is based and that is the thing that we have to go

back and check because we've got to reckon. It's possible that at every stage we've added an additional errors. That's the possibility. So manual transcription the stage we've essentially got is hopefully getting into a form where it's kind of readable. He publishes it, doesn't he? So he shares it. So what you've suddenly got is something. It goes from being one document held in the archive. At this stage [transcription] you've gone. OK, you've got something that can be distributed publicly where it can be consulted without damage to the original materials and where it's just much easier to read. It's much more approachable.

So that is a massive, massive step forward. And we continue that with more manual transcription. And that's, I mean, that's so, so important. And the moment you

publish that I, for example, this precise book [1413 transcription], I was able to

use in some of my doctoral research. And since then, when I've wanted to check things up.

Which is really helpful. So that I think it's got an index at the back, hasn't it? So you can look people up and it's. Yeah, that's that's a big, big step forward and I think the moment you've made that step, you've got a resource that will be used by scholars and that was probably

the most important step. The alphabetization I think was quite important and the efforts

to link together different periods and link together later careers, because I think one of the

issues is that the early transcriptions are, you have no context, so you can use them to say ohh I want to look, as for example I do remember doing, at in the 1550s, whether there were

particular sort of spikes in number of people matriculating and there are and you could see

it and you can also see it in various points, you can see the negative impact of the Reformation and 1559 and 1560. You can see that in the matriculation records. You can

also see the impact of plague in the 16th century in the matriculation records. So I was

able to do that with just the transcribed and published versions. But you don't really have any context, you just have a name and it's up to you to do the sort of stitching together. So this is a massive step forward in that it tells you ohh this person goes on to do this and I think that's amazing. I think the drawback here is that it looks definitive. But it almost certainly isn’t as Small himself apparently said to you. There was a lot of judgment calls that had to be

made, and I think it is quite possible that this is the point where a lot of errors might potentially be introduced, even though can I just say smart was an amazing scholar and a paleographer on a scale that I will never ever be. But there is room in the process for the introduction of a lot of errors and assumptions at this point. So I think this is a really important and useful stuff. I think the resource that you've got here will probably be

of great use for genealogists, scholars that want to look things up quickly. Absolutely.

But there's a possibility of going wrong. Digitization quick text research with records yeah

Absolutely. I think the ability to get the data out quickly and doing that overall

analysis, which you had to digitise to do that quickly because it would have taken like

years to like go through and count it from the physical records, I think that was a

massive step forward and enabling you to see patterns and relationships. I think that's a

big, big step forwards. And I think the extent to which it is useful to the wider

community depends on the extent to which those resources are published and my recollection was that they certainly relatively recently weren't. It was something that, that special collections and people could use, but that wasn't certainly quite recently. It might be now. Like just available online. But I think if you do, you've got an amazing resource which I think again, particularly as a lot of family history people want to search things without

necessarily buying books and things like that. I think you might well be able to share this

data with a group who probably aren't actually going and looking at the earlier published

Matriculation records. So I mean I think something else that I'm conscious of is just quite a lot has been done in recent years. And I think it's interesting.

**CS\*6**

I feel like information gets added and personality gets removed. Or there are other things that get added, I guess like. So like in the later data ones you know that there's a lot of

data enrichment there, which is exciting and interesting. You can see more, more things going on. I guess the ability to access the data sort of gets added in a way with some of

those, right, they become more accessible. More queryable. But you don't get the context.

Context I guess the context sort of disappears a bit as you go through the transformations. You know when you look at this beautiful page with. Whatever lovely Latin is going on here.

Very nice slightly flex nib pen like it has a personality that is missing from this lovely block of Json. Sure. I mean the visualization lets you be more curious, which I think is the fond bit of it right it it and I guess it brings back some of that context. It it lets you I think especially there's something about looking at maps that makes you feel more connected because they are explicable and relatable in a way that like seeing that someone's from Shetland doesn't have the same feeling of remoteness that's seeing it on a map does.

So you get back some of that especially when you can follow people around afterwards, you

got a little bit of their story and in some ways that's a story that I feel like you couldn't

get from text. You do still miss the beautiful calligraphy and could you not put all of the

data into the visualization like one of the things I was enjoying reading on some of the previous things was like there was someone who sounded super interesting in the LLA data.

Here. Ellen Darling studied astronomy. How exciting is that and just because a lot of these are like, there's a lot of church stuff, a lot of just languages and there are ones that stand out like this lady did maths. This lady did astronomy and like, I just wanna see the cool people who did.

**CS2**

I should know that can I went and looked at all the screenshots but I genuinely don't know.

If you mean actual pieces of data, I don't know. I wasn't paying enough attention. I guess what gets added is accessibility and availability for each transformation. It's much easier to work with the database than with manually  transcribed records, even if they're very elegantly transcribed.

**CS5**

So I think when you're going, depends on the formats you're going between. So going from the original print records, I think there's a lot that can be learned about even the wealth of the university and the time periods just from the material or physical records of students.

And then I think there's always a possibility when you go from one format to the next that I suppose it like record is gonna be missed and some information might might be left out accidentally. And then I think there are advantages to having things in digital form because they can be searched more gradually and it's easier to build connections to other information. But at the same time, It's removed from that original kind of like physicality that it

had that I think brings you back to like, that emphasizes how many years ago, something was in a sense that, like students, were enrolling in the university by hand and like signing their name by hand. And there's a sort of decontextualization that happens as things get digitized. But of course having the the photos online of what those older records looked like helps with with understanding that too.

**CS\*1**

So OK, so what’s added or lost. OK. So I mean I I think the part of my answers in

terms of added is what I said previously, you're adding possibility, readability just accessibility of the data because in this form [original records] If you are finding a certain

Person you would have to manually search for them in the right year range, and even

then even then it's not very  readable. You struggle to identify writing, I don't think much

is lost at this point. I don't know. I guess these are ages. So I don't think much was lost

at this point because I think date of births were are all the way to the end. But I think there was something. So yeah, previous most of my previous comment, but the with the LLA is I think it was. There was a bit more information you said about their grades and things like that. That data when it was carried forward was completely dropped off. Admittedly, you

only have it for the LLA students.

[explaining that we were focusing on th geographical information in the content vis]

OK, so this was that data was carried forward, it just wasn't in this visualization, OK.

Yeah, it doesn't really fit in this kind of visualization, because you’re visualizing different data,

just wasn't sure if the data being carried forward. So I felt it meant lost because it was represented at one point in the process. But yeah, you're right, it's it's all about data about

locations and things like that. So it's not, it doesn't really fit in. So yeah, nothing's really been

lost that's relevant to the to the general story of the visualization.

**CS\*4**

All right. Let’s go forward a bit. With these ones [smart], you've got

the name and subject and date And then they're their career and their bits of life history.

Has anything been removed? I don’t know.

**CS3**

Yeah, that's a really interesting thought experiment. So objectively, like

mathematically very little gets  taken away and you gain an awful lot, like I say, in terms of

your, you're adding structure and potential for different organizations and views on the

data. So from like an accountants perspective, like it's continuous improvement, but

there is, you know, you're looking at history and you're looking at something which is

a piece of like handwritten evidence. And so there's the qualitative aspects of the like

the records themselves, like the physical books. And then there's also like. I don't know. I mean, because these are physical artifacts that message about the data representing women's presence in academia, it's very easy to think that when you have a nice,

neat database that you have all of the data and that there aren't things missing and that

when you're presented with this, it's as it is. But having that book and being able to see that

there are two separate books and that there are holes in the information or differences in

the way it was recorded gives you more of feel for just what was left out and what might not be accurate and what might be like, you know, the more human aspects of the data. So I think that's lost in the transcription. But but it's not quantitative, right? It's not. It's really

to put a  value or a variable on what's lost there, but it's really important.

**What stands out when you compare LLA’s with regular students?**

**H\*5**

I mean here, obviously the numbers are a lot, and that suggests that there is a massive trove of unmined information here of possibly on. I'm not sure how well documented the LLA history is. And also as I mentioned before, the pure bureaucracy, you need to let women both you know, this widely spread in the UK, but also in the colonies complete like the subjects like. How did that work practically like, what are the logistics there it opens for so

many questions that I mean with the men like or the women who got degrees at Saint Andrews, you know, they show up, they take a boat or a train or a carriage or whatever and they show up and they're here and then they leave or the, I guess the doctors, a lot of them don't necessarily show up, but you get some of the doctors. They just get the paperwork. But for the most part they they do that. Whereas with these women It's not like a Doctor Who has

you know, the doctors are apprentices, and then they're like mentors. Signs off on them and the paperwork goes at Saint Andrews. And they get their fancy degree in the mail. That's a rough sketch, but you know you get the picture, whereas with these women, like someone is educating them, someone is taking the time to teach them. I'm assuming it's not all book learning. But maybe it is. Maybe they're just sitting at home at their desks, studying in the dark, like with candles. I I don't know. Like there's so much history here that hasn't Been mined and that hasn't been as far as I'm aware anyway. looked at by the university by.

social historians, historians of of women's rights and I suspect I mean, as soon

assuming Saint Andrews isn't the only university who did this. I haven't read a lot of

late 19th century women's history. So I have no idea what work has and hasn't been done

But that is probably, I would assume based on the fact that no one at the university seemed to really have a good overview over the numbers that no one has actually been aware of the scale of education that was actually going like organized education that was going on for women in this period, particularly when compared to their male counterparts. Because this is organized education on a large scale. And in the pandemic era.

**H2**

In terms of this thing [content vis], I mean there's some things we've

got very, very little data and there are other visualizations where you've got loads of data.

So it just depends on what you're looking at, I guess. What strikes me is that what I

would have expected to see, which is the difference between female students in LLA students, so that that's very clear, partly because it's a longer period that the LLA was

active for  so yeah, I mean that the difference in detail, I suppose I mean if you look at that one as well, you know the further education, we don't have anything for the LLA’s there.

But what was impressive in terms of your circle was the way that the peaks hugely grew

because of the number of LLA students who were part of the universe community but weren't in Saint Andrews. So that's what really stood out for me. I think that the amazing extension of that peak between those two red arrows, that really was powerful.

**H4**

Well, it's not as much information on this. So no matter how sophisticated you're

gonna have your search engine, it just does not gonna have enough on them. Because you can only visualise the information that you've got.

**H3**

How many of them there were. I had. I had totally underestimated how many of them

there were, and I think you suddenly go. This is a massive part of the story of women's

education and of the role that this university played in women's education. And that just stands out for miles and I had since the time I started postgrad, I had been aware of the existence of the LLA. I certainly heard about it when I started doing stuff in my masters year

But it's very, very interesting. I would totally admit that I hadn't regard it as part of the

port core core purpose of the university and yet actually it was changing far more women's

lives than the women who came in person. I think that's interesting and something I've learned from this.

**H\*1**

Partly just the sheer volume. So clearly, at various points the majority of students were LLA students. The fact that there's less data. Obviously whoever’s keeping the records seems less interested in where these students are coming from and where they're

going to. So yeah, it's noticeable there's a large scale of students who otherwise weren’t

really included in the record or the work to communicate the record.

**H\*6**

Obviously, it's a much shorter time frame in comparison to a 600 year history of the university. And then the massive spike. Yeah, just such a short span. I suppose because they're not in the main register, we don't have as much detailed knowledge or we

do, but it's not captured in the same way. It's in a different series of records.

**CS\*4**

The numbers. I wonder how much time the academic spent on these sort of external students Did they try and give all the students equal amounts of time, or did they thought that  the LLA degree was a waste of time. And so they didn't spend much time with

The LLA students. Did they have to study by correspondence?

Does St Andrews have some kind of remote degree programs for people in other countries? I can’t remember.

**CS\*1**

So let's see. So the LLA’s, this turns out actually looking at this that they come from much further afield. And obviously that's a benefit even now with remote learning that you don't have to be on site. It emphasizes that it opens up [education]. Opportunities for females and to the point where actually for the males, these are the birth locations of

the males. You're actually more likely to study at Saint Andrews if you're abroad and female,

then if you're abroad and male during these time periods, So that's interesting. At least based on the data we have here, that's an interesting thought, which is not, not something I would have expected, that something that makes sense  because there's actually benefits to the remote learning process. I thought it was quite interesting this point here. [single LLA in careers] as you said about this person here that actually the studying during the

Transition from a remote learning to in person learning. It's quite interesting to see that that one person just transitioned to cross between the two. In the way that the data is

visualized without your explanation, I would be looking at this graph and thinking. Why? Why is there only one LLA student? How did they know her career? What what in the LLA records said their career? Because that doesn't seem to be the right records. But now I understand that. Because their data is carried across from here, from the other parts. So yeah, it's. It's interesting. But also at the first glance and without an explanation, it's confusing. But at the same time, it's not good to just drop the data, so it would be interesting to sort of say with this combination have one data point and this is because there's this particular student was this interesting case? Because otherwise you're presenting presenting a very, very interesting data point. But without the explanation as to why. Why it is that way.

I mean, it's how much data do you want, how much information you want to give in a visualization as well.

**CS5**

Well, it's a whole different register. But it's a whole different it's a whole different format I

So there's this sense that it's just like this other thing that's not as official. I don't like to

sort of kept on the side. I think to me is indicated by the fact that it's like a whole separate book. It's interesting too, because I feel like in the looking back at the the two different enrollment record books that have been in the visualization, the women's one looks so sparse in the way that it's put out that really like the numbers in your graph show that

there's this huge influx of women who wanted to attend, but there's, I think, just something

about how the text looks more crammed in in the universities, like so-called official records, and so it's kind of interesting from that standpoint as well, like. Yeah, I guess the two books to me at least the pictures of them here don't really seem to do justice to like this huge increase in the number of students. I mean is it looks like the LLA's took were like more than 50% of enrollments if they were counted as enrollments, is that correct?

**CS\*6**

OK, sure. Well, the big thing for me is just like I wouldn't have expected there to be as many as there were [number of LLA’s]. Especially for something that you didn't know about that

wasn't in any of the original transcripts. It's a big number of women who are doing, it's a big, big number. If you look at the number of men, there's a lot. Yeah, it's full because like,

these are all women like ohh, it makes sense that the the number of males is lowish anyway because Saint Andrews is a small place and there are only so many people you can really fit into it, but honestly, one of the things that I find most amazing is that it's hard studying not a university now and doing all online, and this was in 1877. When did the post  office start right?  Things that are amazing. There's bloody a lot of them. They're not all in the UK as well. Like, look at this little doll, way out here, and I think Austria, Switzerland.

Were other universities doing it? They must have been.

Back to the comparison, I mean I'll it it. It's the obvious ones, but I guess I'll mention there are obviously lacks of data and where they [LLA] went afterwards compared to to the rest of the students, they very much drop off the map afterwards. Literally drop off the map.

**CS2**

Just how big the number was? Like I assume there would be a jump from the leader, but

That's such a big jump.

**CS3**

So. I think two things. Firstly is the number of them. And secondly, is the diversity and the spread, right? So like the number of them, I'm struggling to get honestly and objective idea of this, but it looks like there are significantly more and then this number at the top suggests

that there are about four times as many LLA students as there were like male registers

students, right? So that's a huge huge number of students who were studying remotely.

So then the other thing like I said is that spread. So it seems like it's probably true that

geographical location was quite a limiting factor because it was expensive and time-consuming to travel, and yet it looks like that wasn't such a restriction for the LLA students, because this seems to be quite accurately distributed by population, like big points of population, right? Like you have a lot around the major cities, you have quite a few rural ones, but it doesn't seem to be geographically concentrated around St Andrews, so it seems to be a real enabler for both women and like people who live further away from the

institute.

**What did you think about the textual descriptions? [Did you need more/less information in these descriptions?]**

**H\*1**

I thought they were useful. Yeah, I think I felt like I got it. I guess I have a bit more

contextual knowledge about the records that probably helps, but I think I got it, yeah.

**H3**

I mean, they were absolutely essential. Definitely not less. I think you could probably in some cases maybe a couple more lines would be helpful. I'm also conscious that I came to this resource with a pretty high level of familiarity with Saint Andrews Records, and

having used some of these records, and I suspect if you didn't have that degree of

familiarity, you would have needed more context.

**H4**

They're good. I mean, yeah. I am quite happy with the snippets. Because it can get in the way of everything else to you know, you get bogged down with some of the the sort of how they did it, but it's enough who did it when they did it and what they did is perfectly fine.

**H\*6**

I think it was okay, don't think I'd want less.

**H\*5**

I am a words person. I will always want more words. I have colleagues who go ohh. No.

She sent me an e-mail again. So don't, don't rely on me for this. I'm known with the IT

department as the person who sends crazy long emails that no one wants to read.

I I go to, I go to museum exhibits and get upset because it doesn't have the information

I want.  I always take more information but I I understand that not everyone likes words.

**H2**

I read them very carefully. I found a few spelling mistakes. I'm such a nitpicky person, and I think I. Yeah, I I think they were essential in that you need to have them to be able to understand what's going on. But if you hadn't been here to help me, I would have had to

have read them extremely carefully, and I might have had some questions. So they're a bit

like a caption in an exhibition. The other thing is that an exhibition caption has to be written really carefully so that it's it is understandable by anybody who's using the system,

and I know that Wardlaw Museum, for example, write their captions for a 14 year old because that's kind of the average able to being able to interpret whatever they've got

So sometimes you've got quite technical language here. Sometimes you've unpacked it

really well. You have explained what the difference between the LLA scheme and the and the physical attendances, but I think perhaps some of this might have needed a bit more

contextualization, a bit more background knowledge about why these dates are important.

I know that I don't. I don't think that's a problem. You know, for me, but I don't think you say anywhere that the United College was formally 1747, and that's why the record begins then.

I mean, a little bit of explanation like that might help, you know, the university was founded in 1413. That's why the records begin then. Or it might be founded in 1410. But the records begin in 1413. 1579 was chosen as a break in the date because of these, these thirds are actually really nice. They are thirds, aren't they? This division is really interesting, but the first third is the period. That's the Middle Ages up to 1579, that's up to the period of the implementation of the post reformation reform of the university. So really that first third represents the medieval university. And a bit later, but basically medieval university. From 1580 onwards that's the period at which the Reformation of the college is under the new

foundation, really set up Saint Mary's as a College of Divinity and the two other colleges of of

UM Arts and Philosophy, and then that runs through the period right through 1747, when the

next major change happens. And you could say that the next major change really happens in

1889 because it's the 1889 Scottish Education Act, which then allows the admission of women in 1892. That's really the next major change around here. So I would say that this last little bit you need is as big a division here as you've got. You know in there so that you can show that third massive change when the university really became basically what it is today.

**CS\*6**

I like them. I thought there are about right. I mean, I guess it depends what kind of context you want this to be in, but if you're talking about someone just scrolling through quickly in the museum, I think this is nice. The the the average stay time at any museum and exhibit is

something horrifying like 3 seconds so. I think that a lot of this visualization ohh I'm about to

pay you a compliment does speak for itself. Like you can see like especially being able to

see the transformation a little bit at the side that gives you a good idea of what is going on in

Each step and it's nice to have a little bit more background information, especially the people.

**CS2**

They were very helpful. I don't think I need less information. I mean, obviously they had a whole mystery about who'd actually built this page. No. I think in general I think I ask you a few questions so I think I guess maybe I need a little bit more information because I wasn't clear on a couple of points. But I don't think I needed much more. I might not have worked out what the people at the bottom were about without being told, and I I thought they were very cool, but I would have quite liked to have been told what they were about and I might have worked it out, but I'm not sure if that makes sense.

**CS\*4**

I read most of them. It was nice to see my name there with all these other people.

I can imagine some people might want more information if there's sort of sort of

they have sort of professional interest in the data and the story it's telling. They might want more of it. Yeah, and I found it interesting, but if I wasn’t a programmer at the time I wouldn’t necessarily be interested wouldn't interested or know anything about any of this.

No, I think that's quite good. If you can condense decades of work into a couple of

sentences, then I think you’ve done a good job.

**CS5**

Yeah, I think there were helpful. I think it could be nice for someone visiting on their own to have maybe like another paragraph or like something hovering over the visualization that just indicates what the different visual encodings mean. Because I think it's pretty

intuitive, but I feel like with with anything that's not just like a bar graph or scatter

plot, I always want to make sure I'm understanding. Like, yeah, kind of what was envisioned and with that encoding. Yeah, but yeah, I definitely liked having the caption on the site.

**CS\*1**

It it was, it was plenty. There was times when, because I was sort of in my my mind, I was

sort of used to looking at graphs and visualizations, sort of like small amounts of data, sort of like numbers and things like that when I looked at this I thought, oh, something else to read. But at the same time it's valuable. I don't know if you could, you couldn't really extract anything else of that without losing information, and you couldn't sort highlight certain information. I mean you could highlight periods. But I think I think as it is, it's probably

exactly what you need.

**CS3**

So I'm a very visual person. I sometimes struggle to understand paragraph written text like that. I do have to read it either very carefully or multiple times to understand it in context. So personally I think it's super useful information. It explains a lot. I think if I haven't been talking about the visualization with you as I go through, I would have figured it out from that text. So it is useful. And it could be enhanced more by being closely linked to the context that it refers to, right? So like having a little bit more of a either it's position on the screen or like a some kind of color coding that matches it with the the things that are changing or what it describes, and even having more of it and shorter dotted around the screen. Like just to explain kind of guide me through the different aspects of what's happening that would be great too. But yeah, the text is important. I think it adds to the storytelling narrative a lot as well.

**What do you understand about the transformation stages? [Which elements helped you most?]**

**H\*5**

For me, the center one, the the circle is probably the most useful one. Now, this is twofold.

Umm, one of it is because I have a relatively good knowledge of, you know this. I you know, I worked with the code I saw smarts books. Right, so this isn't completely unfamiliar

information to me. The other reason is. I'm I have a very poor recall for images.

And so for me. This is interesting, but my brain can't call up the previous one you showed.

To compare them. Which is why I asked you that question at one point about how

is this different from the other one? Because I can't recall it enough like I can recognize

patterns if I have both of them in front of me. But looking at this one having the other one gone means I've kind of lost it.  Which isn't necessarily criticism, and it's just explaining. That's how my brain works. And so for these, I would have liked like the

ability to compare it to something. kind of would have worked with the way my brain works. Which is why this one was nice because I can compare, I can compare the Andersons to the smart to, you know, the work that was done. Uh, I don't have to leave one

behind to get to the next.

**H\*6**

Yeah, this [screenshots] is what I'm used to, which is possibly why it's more

comfortable for me. I think with the central and visualization, I would need your

assistance to understand it. So I don't know if you wanted to add the description.

to the description what you're looking at here. Or maybe for other people its

more obvious.

**H2**

I could imagine that somebody that wasn't me might have been happy with the caption, the, heading and the central thing that they might have thought that the right hand stuff [screenshot] is just extra that that might not be necessary. But when you get to the visualization at the end, it absolutely is crucial that is essential. And I really like the diagram. I like the the presentation of this. As I said in the circle the the way of bringing it all together, I think that really helps to demonstrate the continuity of the of the whole project from 1413 as a conscious Interrogation of the universities community. I think that's a really, really helpful

thing to have. I think the paragraphs are essential because they give you an explanation of what's happening in the in this picture, you wouldn't know otherwise, so I would basically

say you need them all, but in this slide, for example, this bit is just a gobbledygook to

me [database screenshot]. I'm not a computer geek, you can have it there because it gives me the background and it's great, it's excellent and they need to know it's there, but I'm

not gonna sit and read all that code. And now other people might be really interested in how you've coded it. They might be interested in looking at the underlying stuff. It does nothing for me, but then that's just me. Whereas I would really like to go back to even this kind of

information I find fascinating because it means that you can actually see what data is there and what you're basing it on? And then the handwritten I love, because that's just what I

regard as, you know, the basic source material so I basically think you need them all.

But some of the right hand stuff is more interesting to me than other bits, and when you get to the end, when you're actually doing that visualization stuff. You absolutely need this bit as

Well. Because I think this was really impactful because this was the 1st place where we

saw both this table at the bottom and the dots [first vis screenshot]. And I think that that has a massive impact because that shows you that the work that you've done in creating the underlying databases enables you to ask questions of the database and to present the results in an interesting way. So I think this is really great. I mean, yes, it's static, but

nevertheless I think I mean this was quite early, wasn't it when you did this? That was the 1st. That was the first kind of wow you can tell us that from this data, you know that's fantastic.

**H4**

The visualization is too complicated for me. So this is fine [text]. And then this is great [screenshot]. And you can see what Bob Smart did and the information there.

**H\*1**

Uh, so I guess the text I'm sort of getting who did the work and more precisely, what what it

involved, and perhaps why it happened. From the visualization I'm getting when the work happened how long it took and what section of the records it's being worked on and

perhaps how it relates to previous work. And from the the image I'm getting some idea of what it sort of might have looked like. The sort of manifestation of the transformation.

I certainly found the visualization in the middle the most useful. I think that information is very

hard to keep track of just in text form. So I yeah, I found this the most useful.

**H3**

I think they were all essential actually, because I think the thing that I focused on most and

that remembered was probably this central visualization on the screen, but I would not have

been able to make sense of it without the text on the left. I did need that text to understand

what was going on. I thought that the screenshots were also pretty essential and I thought I

actually found even though what my instinctive wish was was to like go and explore the data further and I do think that like if he could have that other stuff [whole book] that would be like a way to get people into those resources, and I I would like to look at those and but I do

actually think that having the screenshots, even just the screenshots, was really important in getting you to see the different way the data looked. I do think that mattered. I mean, it's interesting seeing how this is laid out [LLA records]. I for example I'd never seen this document in the original. So that's fascinating. It's telling me something that the way they've

bothered to lay it out, that they are obviously assuming that these women will be picking up

multiple subjects if that makes sense. I not everyone does. There's obviously some who just did one thing, But it's it's an interesting way of laying it out and the assumptions that are

being made. And so I I think what I think all three are necessary and I think had you lost any one of those three elements, it would have been less easy to follow and I've got I've got less from it.

**CS\*4**

I guess this line showing the student numbers catches my eye most. The other components are sort of slightly more abstract. and seeing how the the source data changed over the

centuries that was nice to see.  This fit with the 25 little segments [Smart] if you straighten it out, it could be books on a shelf. A bit too abstract but not everything lends itself so

easily, so visualization. So, yeah, here's another something if you straighten it out,

then it's just the line chart of numbers. It looks a bit like St Andrews skyline that you see

in some logos. You put them all together and look at all three then you get an idea of what's happening. The thing in the middle gives you feel for the numbers, students involved, time scales and watching the data on the right become more and more structured

**CS\*1**

Yeah. So I'm looking at this. [circles vis] I was initially quite confused. I see the time periods here then. Then there's another scale that's also time periods. Which is obviously these are the years of the students. Steering the periods and these are the date dates they were transcribed but. At the beginning I was a bit confused just by there being 2 date axes.

Because there's two day date axis but height here I I assume this is number of data points

that really long one. Thre was a lot of students recorded during that time. But when it's up against the state date access, it feels like it has some duration attached to it. I feel like we're recording stuff for this, this long to write people for this year, which doesn't make any sense, but because the sort of like vertical datum, there's sort of meaning on the vertical axis or this axis,  whatever you call this axis that comes out from the center. Umm, so vertical meaning which is dates and so my first glance, I mean these employ range the in this case in

this range here these are the right, right the ranges they were transcribed. Here these are the Rangers. These ones were transcribed. So these are the periods of all these happened.

And for this that's not the case [the long official squiggle]. And that's quite confusing.

And it's just, I assume this ties to this. This is just a graph version of this. The extent.

So I think the same about this [first visualization] It's sort of like the fact that

you're showing a vertical data visualization here. And also showing the range of

years here on the same axis is a bit misleading. But at the same time, I feel that this is a very strong visualization in terms of it very clearly, once you understand it, summarizes what happened and when it happened. And I think that I think that's really good.

They were definitely helpful, I mean, obviously when it got to visualization stage [content vis] this was very helpful for understanding the data. But up until that point it was helpful but just to glance at sort of say OK yeah, that's what they were working with, things like that. I gained a lot less information out of that until it until it came to the visualizations. So I just thought, yeah, that's XML. It wasn't really as information rich because I also cause it's just a single screenshot. At you couldn't possibly show all the data. And the text was daunting when you've been parsing graphs. It was hard to get the information out of because

there was sometimes dates. When you see a date date range, you have to check. OK. The

student. Yeah. OK. Or is it the scale? Because it could be talking about the transcription dates. So you can't quickly glance at this text and say, OK, that's a transcription date or.

OK, that's a student attendance range date.  So the two different timelines and then within the actual paragraph of text, it's hard to at a glance know which is which when

you're reading it. I don't know how you fix that, unless unless you color-coded the different the dates, date ranges on each axis or this was read for instance and this was blue.

**CS5**

I think they worked nicely together. I think having the visualization as the kind of focus and

having the images of different materials or the the data as something that

you kind of saw hints of had to click to see, I think to me that put emphasis on the process

rather than on the output of the transformation and which seems like that's what you

wanted to. At least that's what I'm getting from this whole visual and the way

that you have the captions and stuff on the side explaining the process but I think at the

same time having those photos is super helpful, because I don't think the process would be as easy to understand the significance of without seeing that very first format of the records all the way to the end but yeah, but I think the way the timeline is centered puts a emphasis on in the process of the transformations.

**CS2**

I think the circular visualization was the most helpful. In understanding what kind of transformation was happening to the data. I found the images really helpful for just

sort of get a visual hook or what was going on. That was really nice. And I like the map.

Map is kinda doing something different from the rest, but it's sort of it's not telling

you very much about the transformation of the data, it's telling you about the data, if

that makes sense, I think. So in terms of transformation of the data, the circular visualization was definitely most useful. Well, given that I clicked through, I think if you just showed me the end results, I wouldn't have any idea what was going on, but they're kind of building up very well

**CS\*6**

Depends on the transformation that happened I think for some of them the bit on the right is the most useful. Just because the difference between them is not all that obvious. This gives you a bit of a hint about what's going on [vis], but it's kind of useful too actually see

the difference in the structure of the data. The move from the scratchy to the like, more ordered I think is a really nice reflection of what's going on over here. And I do like this sort of way of trying to visually categorize what is going on between the different transformations, where the value added is in a particular transformation. I think for some of it, it's a bit of a

combination of a story, but as in between this and this, you can see that there are differences. I didn't really give you a very good binary answer. The question was which bit is the most useful. I mean like obviously the text tells me what happened and so is the most explicit. That said, I'm much more likely to look at the right and evaluate the changes between the structures of the data than to read the text first like it's a a bit

more of an overview.

**CS3**

Yeah, I think they were all really important to me, the use of the build up over time as another dimension was also really important, so like like if you presented this final visualization and all of the text at all of the examples all at once, I would be like what is going on right there is

something really valuable to the storytelling process and all of the elements contribute to it.

Seeing an example of the digitization is pretty cool, being able to explore specific records and see them as a point on the like you maybe not even all of them, but just like you know you give an example of a record as a as a like an XML. Seeing that as a point somewhere

in the visualization would really help connect the two, I think. But there is still, there is

still a strong sense of connection in terms of the the graphics which are used, but I didn't

feel the same sense of connection in terms of like. How to say like the there's a lot of like the number of records is quite a significant on this wheel and I didn't get a feel the the number of records anywhere else. Right? Even when I got to the final visualization I was like

struggling to understand like over time how does a number of records reflect the transcription process so that.

**What did you think about the juxtaposition of screenshots of the transformations (instead of say text boxes) with the visualization of these transformations?**

**H\*5**

I think it's a very good thing to have because the thing with the visualization in the middle, particularly for someone who hasn't, you know, been as in depth with it as we have is you don't you get the overview, but you don't understand what that means in practice.  But with with the pictures, let's go back a couple of layers. With pictures like this, where you can actually see in practice that we start with. When I get all the way back to the beginning that we start with like this handwritten thing. And then you know, we get the transcription by the Andersons and a couple of layers and Cant And then you get smarts which looks different again. That will help people who haven't worked with it as in depth to kind of understand what these lines mean in practice. That this isn't just he's doing what they did all over again. You know, it's actually they did an important job getting it from this to this format and then he added built onto that with this amazing, you know, information for research project and then and then you go from pure kind of old fashioned research into technology and digital help and digital tools. And I I think that's. Without the images for a lot of people, that wouldn't make any sense.

**H\*6**

It was helpful. My inclination is whatever I want to go back, I want to click over here [the button]. But no, it's good to have the two and that you can slide it back and forth.

**H2**

I like the circles in the middle. I think it's great. I yeah. I don't know what these little people at the bottom are. I had to ask you to explain then. So I'm not quite sure whether they are necessary. And I I think if they are necessary, they need to be clearly explained why they're

there and what they're doing. I think the look and feel of this is very black and white.

And I mean, it might be that because it's on a black background, that's nothing you

could do about it. But you know, you could put make your headings bright blue or something, or turquoise or yellow or, you know, you. You could make it a bit more. I mean, the colour

comes in when you go onto here [first vis screenshot]. So what about? Is this accessibility compliant? So in terms of what you're presenting here, I mean this, if you were, if you were to just as you say, you know, if you have problems seeing red or green or whatever, I mean these numbers are extremely small. So I I don't know, I'm just asking the question you're right.

**H4**

Well, for example, at the very beginning I got the three phases [periods]. That was OK, but the more of the designs came in it gets more confusing for me.

**H\*1**

I guess because they're on two separate screens it is hard to think of them as being two parts of the same thing. I think probably this image these screenshots added less

than the other two parts. And of course they're just screenshots, so you're not

really getting a sense of how they created or used. So I'm not sure I got that much

from the image in the context of the other two parts.

**H3**

I liked that [vis] and incidentally, I have to say I found it pretty easy to navigate between the central visualization of the sort of like with the lines changing over time and the stuff on the

right. I actually quite liked the fact that, It sounds really silly, but you've got the a little taster [blured screenshot snippet] of the one on the right, and I think that's actually quite important because I think otherwise just a little arrow, I would probably have overlooked a better even though I technically knew that the resource was there. And think if you've got, this probably less so [XML image], but with some of them where I've got something like this [trancript] or even more where I've got one of the actual manuscripts that is just definitely, slightly tantalizing and you're going, Ooh, yeah, that's really cool. I want to look at that. I

have to admit, you may have seen what I did just then. I did instinctively want to just click

on it. Basically I like having that showing on the right hand side. It's it's helpful. And it was a it it grounds you in which record you're looking at.

**CS2**

I think the circular visualization was the most helpful. In understanding what kind of transformation will happening to the data. I found the images really helpful for just

sort of get a visual hook or what was going on. That was really nice. And I like the map.

Map is kinda doing something different from the rest, but it's sort of it's not telling

you very much about the transformation of the data, it's telling you about the data, if

that makes sense, I think. So in terms of transformation of the data, the circular visualization was definitely most useful. Well, given that I clicked through, I think if you just showed me the end results, I wouldn't have any idea what was going on, but they're kind of building up very well

**CS\*6**

I think it gives a nice sort of addition to the story. Having this [vis] here and like you've clearly had a bit of a thing about the categorisation of what different transformations added.

And I think the visualization gives you a bit more of that high level categorization than

looking on the right and trying to figure out for yourself what's changed, and sometimes it's very obvious that sometimes it's less obvious.

**CS5**

Yeah, I thought they were nice compliments. And so yeah, kind of similar to what I said before that I think the visualization puts the emphasis on their process, but that

process becomes more meaningful and the significance of that process becomes easier to grasp when you have the the digitized photos and screenshots on the right side that you can see.

**CS\*1**

I think that's good that there are cause because this information is just information

you glance at. You can see the side. You can see it slightly. And you're aware it's there.

But it's not. You're not going to get much information out of that. So you just want to look at that, say, yeah, I understand that. And then I understand what data was at that stage

and then go back to the main visualization, which is where the information is.

**CS\*4**

Did you think of making it more data dynamic?

Yeah. Yeah, seeing how the data was recorded and transformed, it tells you a lot about

How the university operated and what they thought was important. For the the two periods that weren't sort of digitalized or visualized as their numbers here just…[original circle]

Placeholder?

**CS3**

I yeah, I like that juxtaposition. I wouldn't always call it a juxtaposition though. I think the complexity and the digital, the the amount of digital in the visualization is kind of respective or or respects the the stage of organization in the data, right so starting with these little sketchy lines and then going into more structured but not necessarily differentiable

lines. And then you have these kind of blocks and then you have markers that show the rest of the transcription process there. I mean it shows an improvement in complexity of the data in line with the transcription efforts so. I wouldn't necessarily always call it a juxtaposition. It's it's more of an enhancement in many ways.

**Did you want to dig deeper into certain aspects? Which? And why?**

**H\*5**

Yeah. This is just scratching the surface of what you could do with this material and this kind of analysis. You can expand it to a bigger map, you can expand it to different aspects of their lives. You could, I mean, heck, you could. Let's get ourselves to the to

the very end. You could build this map, but for careers. You could look at the doctors.

You could look at the lawyers. You could look at the teachers. Because like the

doctors because a lot of them didn't necessarily go to Saint Andrews. They got their papers

from St Andrews, right. Are they more widespread than, for example, the teachers who might

more naturally be Scottish, who then go home to teach? Or priests? You know, do we see

any of those lines? Career wise. Some of those, like oddball careers where you know, you only have a small handful of people who end up doing them. Do they have any, you know, connections? There's so much analysis that you could do with this. And with this[provenance]. And or, you know less visual mapping analysis and more kind of statistical analysis. This is just a tiny piece of and I'm sure there are many more ways of

doing it that I haven't, you know, touched upon now. But this is I mean, honestly this is kind of what I did my masters on just for sieges and the 10th and 11th century. So like this kind of visualization is something that I find very interesting and there's so much you can do.

**H2**

You mean as I was going through this? Does the data that comes up about the individuals, is is there more than that data available for each individual?  What I would want to have happen would be to be able to click on this guy and go to the link to the database so that it links out so that I can then click on here and then find all the information about Alexander Murden. So I want to get to the underlying data. That's the thing I would like to be able to do.

**H4**

Yeah. I would have liked to of course the LLA’s. The LLA’s are of interest to me, yes.

**H\*6**

I think I wanted to play around more with the final. Just because this offers opportunities that we haven't had before in terms of the way that the data has been structured. Because I've never read Bob's book from start to finish.

**H\*1**

I sort of wanted to click on this and manipulate it. I don't really know why, it just

sort of looked like something you might be able to sort of do. I thought I wasn't able to

see the dates in the middle. That at certain point that kind of became perhaps on the smaller side. So I sort of wanted to be able to move this around as if zooming in it I I thought I sort

of can zoom in a bit but. The other thing I sort of wanted to know that I was finding how

to get was who had done the work. I only just noticed these guys down in the corner.

And sometimes I actually wanted to be able to see, like what were the bits that they did?

What were the bits that Bob Smart did. If I was to look at it now, I sort of know because I can work it out but like that was the only thing that I want to sort of dig deeper.

**H3**

I absolutely want to dig deeper into the primary sources themselves, there's no doubt

about it that that was really, really sort of tantalizing. I would have loved to see the date represented for the earlier periods. But I think I probably do have some questions about the actual processes and decisions which were taken at each point. I think I probably would do, but I don't know. I suspect that quite a lot of people probably wouldn’t have done. It's probably me coming at it having tried and to an extent failed to deal with some

historical data at times, so I was curious about their methodologies, I suppose. So

yeah, I mean, I think in that sense for me, and probably about one in 100 of the people who will ultimately access it. I think there is a a niche group who probably would be really

interested in that, sort of extra document you could click to which would explain the

methodology at each stage, but that probably isn't what most people want.

**CS5**

Yeah, I think. For me, especially when I got to the LLA's, I think that's where I wanted to dig deeper. And you mentioned as well like not everyone that enrolled got degrees. And so like that whole kind of narrative around the LLA students and how they compared and contrasted

with the official records is very interesting to me because again, I'm doing work on gender bias myself. So I think those are the sorts of things that catch my eye these days.

**CS2**

I don't know if the actual transformations of the data and actually has much more

dig in. I'm sure there is more to dig into, but I do not have the right expertise to know what

that would be. I'm. This visualization [content] made me want to kind of explore it more.

This is very cool. Sort of interested in kind of where people ended up. Well, that's interesting.

What does it mean when the lines are highlighted like this? And I hover over Edinburgh?

**CS\*6**

If we go right back to the start. Some interesting spikes in this data [original circle] Especially around here, like what happened in 1820 that was a sudden very large spike. These are all pretty uniform and I'm guessing there was a set number of students that they were taking through that period. There's like a cool little crowny pattern going on, you can see ups and downs in there and it's cool. And seeing the LLA’s these are two interesting spikes here as well. The other thing is the obvious thing that I've sort of mentioned of me getting nerdy

about particular people like look at this astronomy person. I wanna know everything about

them, but we don't have the data for that.

**CS\*1**

I was wondering when I saw that [single career lady] data point, why we had that

information for that person and nobody else. So yeah, I don't know if there was

anything else.  During the process where I've thought, is that what's going on

Here?

**CS\*4**

It's a nice sort of looking Web page and it would be interesting, if it was slightly more.

interactive with some of the bits. Because it’s fun to interact with a nice looking web page.

The bigger stories like the LLA that I didn't really know much about that would be something

That I would go off and find some other book about it which takes you beyond the scope of this. Playing with the map was interesting. I think yeah, there are different ways which make me want to just go deeper.

**CS3**

Yeah. Yeah, I did a couple of things specifically. So as as I mentioned, having more context

by being able to just say OK, here is here is a record somewhere on this wheel, I would

like to see what that record looks like. Like it's hard to make that connection between

the example you present and its position on the wheel, I think that would really that would be

cool. Again, not all of them. It's just to help understanding of like the transformation

process. The other thing I might like to see is the life of a single record from start to finish. So if I had like you know the original book and this is how it became transformed here and then

here and then here and here. I wonder whether that would allow me also to better understand what was lost right? Or the decisions that were made during these transcription

processes. And that's again kind of interesting qualitative information. As for digging deeper, I mean I think the number of students and generally like the sex of the students is.

like the key aspects that are visualized here and I think it does that pretty well.

**Did the series of visualizations trigger any questions for you? [about particular records, patterns, absences in the records, the record keeping, the University itself]?**

**H\*5**

We've basically got on detours because I had questions. Yep. So. And also and I and also the questions about how cause, I mean this is just a taste, right? This is just a taste of

someone's massive piece of work and and so that's always interesting. So both the work that they did, so the historiography of that, but also what information is in that work and

I love this. This kind of thing and. And wondering about how you know like this visualization. Actually this one particularly because it visualizes the subjects. [lla scan]

Just like.  What made them choose? Why did Kate only do French and German?

UM Jordina did French, but she also did Physiology, Physiology, and Helen. She was, you know, she did philosophy in English and Physiology and French and German. Are there are there some of these that none of the women do? Does you know? I mean, right now

we got no one in Latin and Greek, right? Does does no one's do classics. Do some of them do classics? Hygiene. That's a subject. Astronomy, right, which I mean, cause Latin, Greek. I I recall enough of the 19 century can see women doing that, but like astronomy. Like, do they do this? Do do they actually take these subjects or is it more of a they could have taken the subjects, but obviously everyone's gonna go French. German like the fancy, you know the useful subjects for a 19 late 19 century woman. Like that kind of information would be super interesting. Oh, oh, and one more thing. These guys are clearly writing themselves into this book. Which is a fascinating concept. Because you have each well one, you have each person signature. But two, you also know. How they choose to present themselves to the university. because with like immigration records, usually that's, you know, some poor

bureaucrat trying to spell people's names and where they come from and has no idea.

Whereas you know David Ramsey you know, he knows where he comes from he probably knows his age. This isn't someone else guessing if there is guessing involved. David is the one doing the guessing. And so that's a really interesting. It's a very different take on it

From. All of the other biographical sources that I've I've done some family research and stuff, and, you know, church books, immigration records, all of that's a someone else is writing

down the information, which means there is a room for error. Whereas if Jacobs writes his

name as Jacobus, that's how he chooses to present himself to the university with the Latin

form of his, you know, English name, Scottish name, Scottish name probably given he was from Fife. And also I mean, I'm no script analysis but an analyst, but the differences in in writing you see here George is much older than everyone else And also writes in a very

different way. Also what happened to this guy who got his name like spelled over?

Cause he's spelled it badly. He basically wrote so badly someone wrote with pencil over top just to correct him. Also, Robert was so eager to write his age he wrote it over there.  Because and so you get all these questions, I mean none of them are answered. Most of these aren't answerable questions, but they're human questions. About the people who this Book documents or this these sources document be they smart, be they the graduate rolls The LLA register. And I think that, yeah, there's lots of lots of questions and no answers.

**H\*6**

It's maybe not relevant for your PhD but any potential plans for extending it? And we could go up to 1920s?

**H2**

Uh, yeah, I think there's. I think there's more data about further where they went afterwards. As I said, I think the records of the general counsel might be useful in plugging gap. You know, there might be more information available about careers postgraduation from both the

alumnus records and the registered General Council. But it is really difficult because

the university doesn't exist to gather information about where students go after they finished

being a student. Umm, it's not systematic apart from I think there are. There are a couple of points in history in the 20th century where they did a survey of former students to see where

they'd gone, but basically information about former students is almost accidental in where it's kept because it's dependent on the students keeping in touch with the university and telling them about where they've gone or what they're doing or whatever, in which case they would amend, you know, change of address and all that kind of stuff. That's the biggest gap.

**H4**

I mean, have you found all the records? I mean have all the records been looked through? So is there anything else like I was telling you about Ronald Cant’s catalogue that got lost and we found it when I was here. So is this as much as you can get now? And this database, it can be expanded even more? Like for example, say Mary Imandt, her articles that she wrote. Can that be added to her record? And then you could see where she’d gone and stuff. That’s really exciting! Not only for her but for everyone. So a woman that was at the university in the 70’s her dissertation could be added to her record?

**H\*1**

I guess the reason that particular work is done is not always transparent. So like

clearly there's more attention paid to this sector [1747] from that sector [1413]. And I can sort of infer that might be because there's more information and there's more detail and there's more records. But that's the sort of question that is sort of the first one that came to my mind.

I guess it's hard to visualize what's not included, so I'm sort of assuming that there's more data in this segment than that segment, like sort of per record that may not be true, but.

The fact that the records are not the same over time probably could have been is

something you sort of want to know more about. What the original record original sources

look like and what's in them what's not in them? Were you understanding your

expressing margins of error? Are there any other any data points that are questionable, where maybe we can’t get someone's name. We aren't sure about someone's date or

Someone's matriculation status or whatever. That sort of hard to get across. So I guess when you kind of get to the final visualization I'm just sort of introduced the to the idea that some students don't have data. That's sort of interesting, but you only really see that at the

End. So this is I think this is the first time you're sort of introduced to the idea that there may not be data for some so students and in that sort of raises the question of How does that fit into all the previous things I've seen? So I think those are the main things I'd want to dig deep into.

**H3**

I think I've I think one question I was left with was to what extent is their quality

control ensuring that at each stage you don't get additional errors introduced? Because in my experience popping off an in various ways often introduces errors. And these texts have now gone through multiple people's hands and I think there is a potential for by the end of it, you to have ended up with the repetition of quite a few different errors. Now I'm sure people worked really hard, but I think the question in my mind would be. How often is the end result checked against the original manuscript record, I suppose. And that would then affect as a

historian how I would feel about using the ultimate data, if that makes sense.

**CS5**

I think. so seeing the maps and kind of where people would come from, it made me curious about I guess whether there were differences in the students themselves beyond just where they came from. So like their economic class, for example, like where the people that were coming from further distances did they tend to be wealthier or was it the opposite? Or is there like no correlation at all? Yeah, that was something I started to think about

when I was looking at the maps.

**CS2**

Yeah, I I had a lot of questions about this [career lla] that you answered. I was very curious as to why anyone person had a career after the LLA, but it turned out to be a record keeping issue. The issue I guess I was curious as to whether or not you could use other record sources to trace people who did the LLA so this could be filled in. But I imagine that's a lot of

work and I don't know if anyone is interested enough. Yeah, I was just wondering whether or not you could kind of go and find sort of find Alice Martin from Sheffield and see what

she ended up doing, yeah. Find more information. And say if this graph can get kind of filled in.

**CS\*6**

One of the things that I was wondering about especially when you have when you go through these first few, they are pretty data light. I'm assuming these are the

things that they studied. It doesn't exist for everyone. Largely, it's just a list of names. Presumably that data existed somewhere else question mark, it seems odd to

Me. If you go far enough out this is but this is quite nice because they say where they got

their sources from, I don't know how to read these sources but I assume these things in brackets are the sources. So yeah, I guess it made me wonder a bit about what the the

rest of the university records must have really looked like cause this is very slim list [original records]. I don't know what this last column is [comitatus]. So what is this Fifensis and Angusiens. See, you lose that when you go through and modernize it all, don't you?

But English ones are the same, Stafford and Cumberland. Perth is different again.

There you go. This answers the question or or your question right? These are super weird and interesting. These lines are how long it took to make each of the things right, the transformations? So that took like 30 years. That’s basically an entire career.

Ohh and then it gets super short when you're just taking a digital format. Nice, although I'm a bit surprised by that one [smart right]. I guess there's probably was also in a computer somewhere, just because I'm looking at the text doesn't mean it didn't exist as a digital document somewhere, because that's pretty cool. It's nice how short this suddenly gets

when it's digital.

**CS\*1**

Well, it's the same thing again.That last data point with the person that that was in

the LLA records and we had her career information, which I thought was strange.

As we didn't for anyone else. I did wonder where the grading information went.

But as you said yourself, it wasn't. It's just because all the other information, I

think all the other information, the information made it to the end and was represented in the data, but that wasn't.

**CS\*4**

There were a few where I wasn’t quite sure what some of the things meant. The little people down at the bottom left, I wouldn't have guessed what that meant. If you read the text. You can see that it's a guy with two daughters and then someone else comes along and.

But then you know with a lot of the stuff on the right I’ve seen before.

**CS3**

Yeah, the biggest outstanding question I had was really like what was lost throughout this process. And I think those I think the methodologies that I suggested might help answer that. I think as well it's hard to understand how much work is actually involved in all of these processes in relation to each other, right? And like there's a lot of effort that goes into the transcriptions, but just the number of records alone doesn't always tell you what's involved, like maybe an indication of the time or or something like time taken to resources involved. I'm not sure that that might be questions there on like, OK, how much effort actually went into these transformations.

**What values if any do you see in visualizing both the processes that historical documents go/have gone through and what’s within (content) these records?**

**H\*5**

Yeah, I think and we've kind of touched on this already, it's it's the difference between

historiography and history. And it's the importance of knowing both to be able to work

with the material in a source. Because here's the thing. If we didn't know. If you didn't have this graph For the LLA's if all we were working with was smart, like if it stopped there, or if it

stopped with what I was doing in 2014 This part of the information Wouldn't be there. It's not in smart. It's it's not actually documented. They're only documented when they fit into

his category of student at the university. And they would be lost. And so, and it also highlights a very important thing with any analysis of any information is that you have holes.

There are areas that you aren't covering. There are biases, there are, you know, difficulties.

There are transmission errors, It's very important because this this looks very true [database screenshot]. Right, this looks like fact to a human mind. But actually, even if it's statistical, even if it's, you know, the big picture, It's still an approximation It's not fact. There are biases

and holes and errors and I like to connect things with what I do. And I've been working with some AI. We've like discussions about it, not actual AI, but like talking about the issues with AI and the whole thing with biases and how that skews your data and skews your AI is relevant here as well because like without the LLA, you've skewed the whole thing and you don't even know it. Because it looks like a complete register. But we're missing we're we're

missing these women. And it's funny how these women are taken out of it. But the doctors who did exactly the same thing they didn't study in Saint Andrews, right? They just sent it off like paperwork that their mentor wrote. He's a good doctor, basically. And, they get their papers. It's the exact same thing. Right? But because they're doctors, they get to be included or the OR the honorary degrees they didn't even do anything and they get to be

included, but then these women don't get to be included because they don't fit the bias

of what a student should or what a university graduate should look like. And I so I think that's

The historiography is not the important thing, but it's an important. The metadata and the context of the source is as important as the source. To make a complete analysis of

something.

**H\*6**

OK, well I suppose for the specific information, it's helpful for research inquiries. If you're

looking for specific individuals, but these visualizations are looking at a wider picture.

So if you're looking for, you know trends or when there's been a massive spike in the number of students or what caused that rather than looking at the specific information.

**H2**

I think the transformation of the record is really interesting and I think that's quite that's

a novel concept and that wasn't mean when you were asking all those questions about the record keeping. I thought that was really interesting that you were asking those questions. I hadn't realized that you were going to be presenting it in this way. And I think it is very

interesting seeing this. I think that's fascinating and I think. This is this is not just you

addressing the content and enabling the engagement with the content. This is you thinking about the Universities conscious act of keeping record and I think that's a really, really interesting perspective, especially as somebody who's responsible for the university's records. It's fascinating to be able to see that continuity and chronology of record keeping presented in this way. I love the visualization of the data [content], that’s great. And yes it’s brilliant and it’s absolutely essential and it’s a really good thing. We want to be able to access that after you go please. But this bit [provenance visualization] is really exciting for me because this is the first time I have actually seen any attempt to tell that picture in anything non-verbal and I think that’s brilliant. So I think this is really good.

So you passed your PhD for me.

**H4**

Well, it depends what you want out of it. I mean if If someone researching particular people or particular subject matter, then they might just want to get into the database and just find out about it. Whereas people like yourselves, if you're really into, you know, computer science and and you want to see the processes, but for me as an end user I mean, it's fascinating, but I don't think I would need it.

**H\*1**

I think it's really useful for demonstrating labor. The value of Labor, the cost of Labor. Not that you've been able to attach a cost to this, but one could start doing that. And how that relates to the outputs. There was a lot of hidden labour, little confident costs, hidden histories. So I think visualizing the process helps to bring out something that is not normally.

obvious via, say, our website or via our reading. So yeah, it helps justify existence of labours of all kinds I guess.

**H3**

I think there is an absolutely massive benefit to visualizing the records, the data, the

records themselves hold, and I think that's like that would have a similar value to when

people do things like sort of visualizing mortality data or stuff like that. I think there's

a massive value to doing that. It shows different patterns and cycles and the life of the

university. And I think the general public and I think historians will be interested to

see how patterns shift over time. Now I’d be really interested, this sort of circle I've got on

screen here. I'd love to see how that worked going back to the 1420s and frankly, I'd love to

see how it worked in relation to modern times, though my guess is that you'll just get this

massive late 20th early 21st century spike, but I'd be interested to see what happened.

Did it just break the system? It went so completely massive. I don't think it was any question that there's a value in visualizing the makeup of the university and I think would be really

interesting to see visualizations of what places people come from change over time. I did like

those maps of where people were born, and I think I'd be really interested, more maps of that type for more periods. I'd love to see that. I'd like to see breakdowns or subjects and things like that, but that isn't always applicable to the way people studied in the past and the way they're recorded in the way our degree structure works. So that's probably not relevant.

Regarding how the data was processed. I think it's one of those really interesting things where if you had visualization of the actual historical data itself, I think I'll go back to that time and time again. I think I'll probably consult it quite frequently. I think a visualization of

what happens to the historical records, the journey they go on, so the transcription, the

digitization and things. I think looking at this [provenance vis] has made a very powerful point. I think it makes a point, amongst other things, for the purpose of what's special collections does and the fact that what special collections is doing now is building on the work of people several decades, even centuries before. But I suspect I wouldn’t consult that

very often. I think having seen it once or twice, it's like that's a cool thing that made the point.

But I've got the point now if that makes sense. So I would be interested in going back and looking at OK that data on the LLA’s and like where they came from and all that I'd go

back and look at that. But I don't think I particularly, I mean, I might someday in the

few years time, but most of the time I wouldn’t be regularly consulting the story of the transcription of the records and the publication of the Records, unless I suppose you had a lot of information on the methodology, in which case I might possibly consult it for my

own as a model for my own research projects. But I think this is something that could be like wheeled out to lots of people across the university and lots of times when you like go into schools and all sorts of things where if you're trying to explain, OK, what does a big

transcription and digitization project do? I think this is really interesting story of that.

**CS2**

I think it helps add context and meaning to the records. It's much easier to understand

how some of them haven't been transcribed when you sort of note that somebody sent 52 years transcribing and only got through sort of 100 years of the sort of 600. And you know, it's a particularly thankless task, transcribing historical records and it's important and it's nice to actually see how the final visualization has com about and the effort that has gone into it. And at this point, you know, it's historical information in its own right. I think going along. Yeah. I mean, I think there’s as much value in visualizing the process as there is in visualizing the results quite often. Depending on the results, obviously.

**CS\*6**

I mean so like I think some of the conversations that we literally just had like being

able to talk about the length of time that it took. I think there's a lot of value

in knowing the journey of the data that has got on you to this point. Because this data has been on a, just from the first digitization to this visualization being produced, it has been 120

ish years. And it's gone through quite a lot of changes of that time, and this data is the combination of different data points and different sources. And also has gone through a lot

of people who have different lenses on how to treat the data and deciding the structure that it

will be recorded in and like think all of those people who are part of the history of that data

And all of that history has informed what has ended up creating this visualization. It's not, you know it didn't just happen overnight and it's very easy to look at visualization and feel like it just happened overnight and not over 120 years.

**CS5**

I think so yeah, this is something I've thought about with my own thesis work as well, and I think there's a lot of of decontextualization that happens with data these days because

things are so easy to just grab if they're online, and especially if it's large amounts

of data for things like machine learning. That's like the prime data set and the kind of the

quality or, like representativeness for the end goal of a model isn't always thought about very

well, often isn't. It seems like. And so I think what's nice about this is it it forces you to slow

down and think about formats and in that that process of transformation and that time

that it takes and and then also like the number of people that are involved and. Yeah. And thinking about the person that even added more information to the records that have been in them before. You know, even just choosing what data points to add is something that often can show like a certain set of values or or biases or kind of whatever you wanna put it. But there's I find like with my work looking at archives at the University of Edinburgh, the

descriptions talk a lot about degrees and honors that people received, and they tend to be

men because historically more men were enrolled in the university and they were teaching at the University. But then when you get to the volunteer work and social related engagements

there isn't a lot of detail. It's really skimmed over. And there's just sort of the more you read is that again and again and again more you feel like ohh, that just like wasn't valued. Someone just decided that wasn't important. Then why did they decide to do so?

**CS\*1**

Well, there is value in the sense that it it helps you because this gives you such a rich understanding of how the process of the data came together, how the LLA's got added

in and things like that. It explains a lot about the density of the data and the

amount of information we have and how it differed. So if you have the raw data and no

explanation as to the process, you might wonder why certain information is missing at

certain points. Why suddenly at certain points like the LLA period we have additional data.

The the the the is suddenly appears at a certain time. I I think that. This visualizing the process is valuable. It's very valuable because of the fact that it helps understand.

The difference in the data that was collected during the period that's being represented.

**CS\*4**

I guess at some point people will start studying the history of the source of data

transformations and things that will become data itself. Yeah it shows the amount of work that’s gone into working with all this data. It shows the benefit of using computers over writing stuff in books and the way it can change from XML to json so quickly and then the geotagging stuff. I don't know if historians are all interested in the transformation of.

of the data. If someone was very interested in the history of the university or something, the

student numbers then visualization of the history of the transformations of the data wouldn't be their main concern. If there's someone who’s got vague interest in lots of things so I guess for me seeing both is interesting.

**CS3**

Yeah, I think there's value in both. They tell different stories. Yeah. For me like.

I think the biggest the biggest impact is just recognizing how much work has to go into

creating data useful for visualizations? That's pretty valuable, I think. Yeah, It's a

tough one. I think I've seen other kind of big projects like this as well that just give you.

a sense of like. It's a very qualitative sense of, UM, opportunity or excitement. That comes with the digitization of records like this. So simply saying, hey, look, we've got all of these books and we digitize them. As like. OK, great. Who's gonna do work with that data. But then when you start telling a story like this and you say these are the process that they have gone through and you show it alongside the records and you bring some kind of gravity to an understanding like why these records are important and you allow someone to have a kind of a peek of that data and then you say all of this data has now been transcribed and these formats and this is what they've gone through. You have the opportunity to kind of come up with new questions and become involved and invested as you're engaging with this and

the potential for visualization and understanding and working with that data at the end is

much greater. So I think the value comes in presenting the final product as as much like

that final product being the transcribes like searchable database, right. This kind of storytelling visualization is really valuable for attracting interest and then doing something useful with the data going forward, right? It's almost like a sales pitch for like, why should you care about this data set? It works very well.

**Do you think that visualization is an effective way of showing, and emphasizing provenance?**

**H\*5**

Yeah, I think it's a good way of illustrating the work that's been done. And the provenance, up to a point. This is a tricky source in that it has, you know you've got,  the original material and then you've got a transcript of the original material, and then you add material. Right. And then and then you add more material up here when we get to that one, so it's. So it's less about the provenance of the thing than it is about the evolution of how it became the data that you have today, which I guess is provenance in a different way than I'm just used to archives where provenance is like it's one and done. This is the provenance end of story. And then that file just kind of moves. So but it is. But it does a very good representation of the the way information is added and the way the information is changed, even if it isn't technically changed, it's just transcribed or, you know, it is still changed and that it's gone from handwritten material to printed material to code.

**H\*6**

I suppose from a professional point of view it's the validation of progress. I quite like seeing the narrative of how things have progressed over time. I don't think you necessarily see the connections. I think it's helpful to have a visual representation of the things and

how they're connected. Having you know the script here from the XML next to just an

entry from a calendar [Smart’s records] or something you wouldn’t necessarily make those

Connections. I think having this here [visualization of transformations] helps you compare them.

**H2**

I think so, yes. I think it helps to be able to. Well, how else would you do it, I mean you've got a load of different sources that you've got to interpret and your visualization has enabled you to pull together a composite picture which allows you to build a story whereas. Otherwise you just got to go look at all the different sources and there's no way of comparing it or bringing it together. So I think it does work, yeah.

**H4**

I think it is. Well, some people learn things by reading. Some people learn things by hearing and listening to, you know, and then other people, Images help them understand things so as well. Again, it depends on the kind of person that you are. It's very intriguing. But I'm the kind of. I like this [screenshots]

**H\*1**

I think this particular visualization is useful for showing the labor and the  history has gone into. I think the screenshots of different forms, I guess in a way you want those to kind of show the transformation in a way a bit more less static, so as all before and after. So you see what processes, what the difference between the beginning of the transformation and the end of the transformation is. So maybe that's [images] less successful in showing the transformation.

**H3**

Yes, I do. And I think the visualized version is more powerful than just a paragraph or text would have been and an actual fact I didn't know the full story, but I knew about 3/4 of this story and I think it's still made a point to me seeing it in the visualized version and made me see things that hasn't really picked up despite actually having quite a lot of the background

information.

**CS2**

Yeah, I think so.I think it was useful, I think otherwise I'd have lost track of the dates and which bits you were talking about. Very early in the process.This meant that I could kind of

keep track of what we were talking about and kind of which bit we're talking about much

more easily. Because otherwise I would have just been going sort of yes, 1877 to 1897. I don't remember when that was or who transcribed that, where we've got to. But this worked quite well. I do keep trying to zoom in on it. You can't see or hear me, but my scroll wheel is just kind of  going. I think the animations really help. I think the only transformation I struggled with was the line being drawn above the dots. All the rest was very clear.

**So for example, if you had just the text instead of the visualization of the provenance?**

No, I’d have lost track and got confused. This gives me context of the dates and you know I can see that this chunk was digitized. Whereas if you've just given me the dates It would require a lot more concentration. I'd probably end up sketching a timeline if it was important enough and I needed to care. If that makes sense. Otherwise, I just nodded and smiled, and not kind of kept track of which bit was digitized.

**CS\*6**

Yeah, I actually thinks that the visualization was even more stark difference in some ways than the actual transformations. Especially when you're looking at like what looks like slight variations in the layout of text, understanding the difference that like this it's me hopping on about this again. And I think it helps structure the way you think about how those transformations have happened.

**CS5**

Yeah, I think so. Yeah. I mean, I think the kind of circular representation of time. I think it was nice to show  for this group of records in particular, because it has gone through like this, the same information has gone through so many different changes. And I think there's

also this sort of ongoing never ending aspect to a circle that I think kind of comes

through with this as well now that I'm like looking back on it after stepping through it 1 by 1

and that like there's always more information that could be added or new transformations that can happen as technology evolves and the challenges are just like constantly trying to keep up with the the changes in new technology so that if you want to make sure

that like datasets or information can be accessible, you have to make sure that the

formats are going to work with what the latest computers are tools that are out

there and so this I think the circle circular representation of the records through time

indicates that this is just something ongoing. Something that people who work in

like an archiver library always have to be thinking about and in order to make sure that the all the information they're managing is discoverable.

**CS\*4**

Yeah. So some of the visualization bits are a bit abstract, you know this sort of dashed line and dotted line when you’re flipping through the bits you think ah, something's changed. Then you can read what changed on the left and top feet, see an example but looking at the bit in the middle. You wouldn't sort of necessarily know what all the different rings meant.

**CS\*1**

I think so, yeah. This graph, as I say, I think this graph shows it well and as you watch the graph being constructed you gain the ability to read the graph. I mean, if you were given

this graph at the beginning, you have no idea what it meant. But you the fact that you see

the graph being constructed I think that it's a valuable way to represent the data because afterwards, once you've been through the process and seen the graph growing, you could then look at it and get a quick summary of how it all worked without looking back through the stages. You can go back. You you can go back if you want and read the text samples here. But it gives you it gives you a good amount of information,

What I didn't quite get the need for personally was this. The count of people. It may be that some people would find that useful, but for me it was for me to seemed I I'm not. I'm not sure if I'm sure what it was trying to get across other than the people happen to participate in

this particular part of the environment the process.

**CS3**

Yeah, I think it can. I think there are a lot of different ways of doing it. I haven't seen

this in particular encoding before, but it's very effective at presenting the layers of work that transcription has gone on and and you know this timeline over the last 100 years of how they have transformed, there are loads of loads of other ways that you could show transcription processes and I'm sure they would open up different questions each way would answer and and invite different questions.

**Do you think that visualization is an effective way of showing and emphasizing what’s within the records?**

**H\*5**

Yeah, I think that the thing with the LLA's is important and I think this is an important visualization, it would also be very interesting to do the similar thing with the honorary degrees with the doctors who never came. It basically pull out all of these. These are the people who actually came to Saint Andrews. These are the people who, for

various reasons, weren't actually students. And then here's the overlap of the people

who kind of did both, you know, the like, honorary degrees. You'll have some students who

get honorary degrees and that kind of thing. You have doctors who did undergrads here. And so there is some overlap, but. And you have the two women who did LLA and then degrees.

So, but I think this is a very important visualization to show. There are holes. The data may not look like it because it's fancy data, but there are holes. And they are clearly gaping.

**H\*6**

Yeah, because you still got the same information, but it's being represented in a way that's possibly easier to understand than just an endless text.

**H2**

Yes, absolutely. Yeah. It allows you to interrogate it. But I think that the questions you're

asking of the data at the moment, are relatively limited. But the visualization is only as good as the data you put in so you're always going to be limited by in what you can do because they'll always be gaps and you may not be able to get the answer that you want.

But I think the fact that the data is in databases and is therefore searchable and you can

play with it opens up all sorts of new research questions and enables you to to tell stories

about the university of different way and in a very appealing way, which is great.

**H4**

Yeah, that's pretty cool. I mean, this is just you kind of you see the the scale of it. When you see it like this. Yeah, you know, I mean, you see a page with names and it's a page of names since you have no idea how many there are unless you had the physical book in front of you. But this is really this is very good.

**H\*1**

Yes. So I think, I mean, I guess the only thing is that it gets very busy on this screen. So

it's less obvious. I mean, it makes macro point that all of these people are in Scotland and in the Central Belt Scotland, but you sort of want to be able to scroll in a little bit. But yeah, I do think it's the most effective visualization of this or explanation of this data that I've seen so.

I would say so.

**H3**

100% yeah, we should do more with it. We should look at more options and ways of showing that information. Absolutely.

**CS2**

Yeah. I think with the concept of visualization and this visualization in particular are effective for conveying information in the records. I think it's a hard to visualize that because it's very busy. Yeah, I think visualization in general is effective as a technique. Equally, I'd be very surprised if anyone with any visualization experience would say no it’s not an effective way.

**CS\*6**

Yeah, I think. I hadn't even noticed these numbers up here, and it's useful to have them, but I think that even just having these points is very visually striking of the differences between how many people were there, how many men were there? How many women were there, how many were studying from abroad like it's much more visually impactful and especially things like if I saw the name Heinberg and a Donna I would not know where that was. I would not think of it as being as far away from Saint Andrews as it is. I think having it here

is quite powerful.

**CS5**

I think so. I haven't spent a lot of time looking at enrollment records before. My sense is that a lot of it is like the students name and then where they came from. So I feel like if that's the case with these records that a map makes a lot of sense because, yeah, the information is mainly about where they came from.

**CS\*1**

I do think it is because the location data is if without visualizing it, it is really hard to

understand and you can get a lot of information out of visual out of location data.

But just looking at the raw data is absolutely meaningless, but you could say how many records did the person come from Brentford or something like that. But you might want to know how many people came from around the London area and with here you can see it because you can actually see the map itself where in this, whereas otherwise you'd have to sort of find out how many people came from each Individual location around London.

Uh, so I think I think that seeing on the map is also very good also shows hotspots and outliers and things like that. If you mean you have a single place single person here that came from and that's just one data point. If you were looking through the

data, you probably wouldn't even  know about that. So because it's just what the single data point amongst many. So I think visualization is very good. Visualization is very valuable

for map data. Location data and also the filters. On top of that is very useful to be able to filter and without that I wouldn't have found that here we had their career. I think very

valuable.

**CS\*4**

Yeah, this is very interesting and very attractive. It's a nice way to explore the data

And it could be hard to do something similar with sort of occupations rather than

geography, because geography has such an obvious way to represent it. I guess grouping people by occupation would be more of a challenge. Yeah, it's very good. I want to zoom in on some of the busier bits, but I can't.

**CS3**

Yes, definitely. Yeah, there are. There are elements of this visualization which are

not present in the in the transcription visualization. Yeah, this is definitely one of

the most effective ways to present kind of the final transcription and some of the key messages there, so yeah. I can't think of another way of doing it honestly like you would

you would have to visualize this to have any sense of of a data set with like 5000

data points in it.

**Do you think that the way we combined visualization of provenance with visualization of what’s within the records was effective or not? Why?**

**H\*5**

Particularly for the LLA's. I I like the LLA’s partially, cause it's women history partially because you know, it's stuff I didn't work on. It's new stuff to me. But but I think like even if you didn't have the LLA's, even if this was just a students, Being able to see this [the overview of provenance] and having started with that first slide of like the roll of you know these students went to here, this birthplace that they wrote themselves and then being able to see that based on all of these people's work, all of these almost 150 years of work.

And we've been able to build this. Is very, very powerful. And then you add the fact that

you're suddenly introducing people that most people would have never considered university

Students.  And it's even more powerful because you have both. Both the visualization of their their lives in a rough way, I mean, birthplace isn't necessarily representative where someone lives, but you also have the representation of how a group that is that big [lla] and that prominent in terms of the universities, not necessarily teaching power because they weren't actually here, but in terms of, you know, people who are connected to the university are completely silent within the kind of official register of the university Both in terms of the the graduate rolls and in terms of the work done by the Andersons, by smart by, you myself, and and Alice and and like, they [lla] weren't there. That's incredibly important Because it it gives them Not a voice, but it gives them representation where they otherwise wouldn't have it.

**H\*6**

This [content vis] would probably be more useful. I'm not saying that this isn't useful.

But I don't know if they necessarily need to go together.  I suppose it depends on what

you're trying to achieve. It's just what questions is this trying to answer in comparison to what you'd be using this for. I don't think there's any reason that they shouldn't be together.

Yeah, it just it you may be using it in different ways, but there's no reason that they

couldn't be side by side.

**H2**

I think this worked very well. I think you gave primacy to the record transformation.

Because of its position within your presentation. I think I understand that the building up to the visualization of the data it is necessary, but I think you could probably also do a different presentation which went straight into the visualization of the data. Without, I mean the record

keeping stuff might not be as important to other to some people. And they may say they we they would just want you go to the end to do the the visualization to tell the stories. The record keeping story is interesting, but I know from experience some people just want.

You know they want the answers and people's questions might be still related to the individuals rather than to the trends. And and so you have to choose your tool according to the question that you're asking, I guess.

**H4**

I mean it's difficult when you see something for the first time and you're trying to grasp something that's completely new to you. So it's kind of difficult to then to be able to describe how you think about it and how you understand it. I would have to sit in the and really play around with this all night to be really sort of answer your question.

**H\*1**

They seemed like 2 distinct things to me. Rather than one thing that sort of fairly related.

I'm not sure why. I guess just because they look so different, it just seems like you're going from one quite distinct project to another project, and of course they're actually very closely related. So I don't know if there was some way of making it somehow look more similar. One sort of transposing onto the other somehow I don't know. That was just my gut feeling was

it just seemed like 2 separate things.

**H3**

I mean, you've got two stories in that one visualization in the centre, and they're actually two.

time periods because you've got the the circle in the middle with the years of like the

students, and then you got the outer one. I actually. I think they work OK together.

It's not that. Like I was sitting there feeling that was a massive problem, but I also have

to say that I would be probably just as interested in the inner circle with all the data,

shall we say. Going back to 1420 and I'd find that really, really interesting. And then I think I

probably would be interested in the outer circles as their own thing probably. So I probably wouldn't have an issue with the two being split necessarily. I mean, I think there is almost

an interesting story in. Because which of course you can't get from this because the

matriculation data you've been looking at only goes up to 1897. The process of starting tracking that with transcribing that starts just before that but overlaps teensy bit and then really is a later period, and so in a sense,  they're both 2 separate stories about what's going on, but also two separate stories about the time periods in which they're happening. And I think that's one of the reasons why I’m not anti them sitting on the same chart, but I also think they could work perfectly well as two separate charts. It might be interesting if you

frankly had a view unlike the size of the university, you actually have the modern data and you also could see when people are doing stuff with the transcription and that would be actually kind of interesting to see how the universities archival management sort of maps onto the university's own growth, but I don't think that's a massive story to tell.

So I'm OK with them being on the same screen together, but I also wouldn't have issues with them being on separate screens.So I I have to say that that particular little graphic with

the LLA over the the conventional seasons was really interesting for me. Actually, that was one which was to be fair. I probably also would have found that interesting if I had the two, the two lines together, I'd be like just actually seeing the two superimposed on each other. I was just like, oh, that's a bit.

I love this. I love this [content vis]. I think you have to reckon that it does get a bit confusing, but I reckon if you had, yeah, you consume in quite a bit, can't you? And if you had? I love this. I love this, and if you had just that little bit more flexibility with. Yeah, I really like it.

So many cool places. Saint Andrews. There we go. Yeah. OK. You can have so much fun

exploring this.

**CS2**

They were helpful in a sense that by the time I looked at this, I had some idea where the data came from and what the significance of like what was and which bits were being Visualized. I take it back, it was effective because it told me which bits were being

visualized and what's significant for people. So yeah, I think it was effective.

**CS\*6**

I liked it, I think it lends itself to the way I've been talking about this as the transformation is being part of a story of the data. You don't wanna do it for every visualization you ever create, presumably, but I think it's a particular lens to look through the final visualization

Through. And it's a lens that doesn't exist unless you do something like this. It's very easy for

that stuff to get missed, and it is all part of the history, especially, you know, lots of the people who've been involved in this transformation have been in Saint Andrews and Saint Andrews is a very historic traditional place. Being in St Andrews as a place is part of the identity for a lot of people and so this sort of like large transformation of data that is handled largely within Saint Andrews.

**CS5**

Yeah, I think it's effective because it shows the records at kind of like an overview level with more like the process visualization and then the maps seem a bit more detailed where like you can really drill into specific people. And but because like you said, there are so many people I think rather than just like alphabetical which isn't really that meaningful to people, it

seems like place is good because people more likely to have connections with a particular place that we will particular letter of the alphabet and so that makes it more interesting way to just scroll through the information.

**CS\*1**

I think it’s very effective. Because, I mean, if you go, if you go through the first time you want to look here[screenshot], see what's going on. You wanna you wanna understand what

data was being created at that point? You you definitely want to understand that and then

then later on when you're going over it again. You can sort of see it on the side. The small preview of it in the side and you understand just by looking at that because you've seen it

before. You're like, OK, that's what it looked like at the side. There was times when I didn't

look at this heading initially. The first thing I did was look at the data here[screenshot]. Even the first time around, first time I think I looked that was this image and thought OK, that's what the data was there, especially when it came to XML because when I came to the first

XML page I said OHH it's XML now and I hadn't even looked at the heading here or the text  I just saw the XML. I think it's valuable, I think it gives a lot of information in this much quicker reference. If you're flicking through it. If you're looking for the points at

which it was XML, you could see it. You could see it, see it in the image a lot quicker than and react to that a lot quicker than you can search for the text I think.

I think that was very effective being able to go into here [content] and play around with the data and things like that from here. Because I mean it's showing you all this [provenance], all this information, all the transformation, how the data went through this, they all this process but that last step being able to transfer then move across to here and then explore the data.

Emphasizes, even beyond the just allowing you to explore the data yourself, emphasizes the actual value of the whole process and what what this whole process has enabled, we couldn't have done this. Done this with the original data or or any of the earlier stages, but because the data is now transcribed digitized in a database. Because of all this. Just being able to go across the here and say this is the. This is the data now explore it. Emphasizes why all this whole process was so valuable.

**CS\*4**

Yeah There's a nice progression of sort of things start off with some ancient leisures And it's sort of dreadful handwriting in them and through years and years of work you end up with a cool map that you can interact with. You have thousands of people's lives here and

 you can find out about them or certain things about them and you have all the work that the people have done over the years with the data and it brought these people into the light.

They are not buried in some dusty old book.

**CS3**

Yes, I think it was effective. I yeah. It it gives you it gives you a more tangible sense of

the transformation at each stage in the process. And example is a really good, a

really good way to fix that. I feel like a little, a little more connection between the two of

them would be beneficial because it it is so effective. But essentially, yes it is.

**Any comments?**

**H\*5**

It's very well done.

Congratulations. I am thrilled to see what you've done with it and it's a fantastic stepping stone, whether it's you or someone else To to to keep researching, to  keep looking into this, to find new trends and patterns, and the the work that could be done here is many lifetimes probably.

**H\*6**

Can't quite remember the last one, but I think this is easier to use.

**H2**

I think it's an amazing piece of work. I think it's really great. And my main my main comment is that I would encourage you in some way within all the restraints that you have in terms of

intellectual property and everything else to share the research that you've done and

the underlying data with the library so that we can continue to benefit from your work

because your research has for the first time pulled all this stuff out of the of the record and it will be a great shame if after your PhD the the data is no longer there and accessible, so I know it's not something to be dealt with right now, but if they give you 3 months to make minor corrections, use the three months to get the data out and make it accessible.

But the other thing is that you know you can put you, you will be able to put your data into

the into the online research portal, you know through Jackie Province outfit within

the library so as as an adjunct to the E thesis. I don't know how you're going to

capture this and submit it as a PhD because so much of it is interactive. But I do think that if you can possibly manage to enable us to have continuing and sustained access to the software that enables you to do this kind of stuff as part of your research output, that would be brilliant. So that's my main, my main hope.

**H4**

I think it's intriguing that you're doing the LLA though. I think that's fantastic as I was doing my Research, when I find out you're she [mary] did that up. What is that? And I just thought it was just amazing. I thought I was so proud to be in St Andrews. I I think Mary was in the

first year [of LLA] she was really early in the LLA. Her family was hugely educated and

And she was a socialist because where she came from and everything but she would

have fought for everything that she got I would think you know. I'll send a book for you.

I am included in lots of people’s academic work because women studies are so important

now. I mean, they're so popular now. Yeah. So I did this years ago and people are writing about women there's two Dundee women who have done so much way back in 1892 and they were pioneering so it is quite cool.

**H\*1**

Just like I thought it was. Yeah. I mean this. This is hands down the best sort of storytelling of around archival labor that I've seen so I hope we can use this in some way or even archives in general can use in some way to show where the sort of end products of libraries and archives come from and how long it takes and why they look the way they do. So I'd be interested if this could be reused in other contexts. And certainly this kind, I know

you've done this kind of work for a long time, this kind of geotagging of data which again I think is really effective. So just to say, yeah, I feel like we're somehow not using this as well as we could. Either at Saint Andrews or as a sector that is sort of trying to describe it's funding needs, its resourcing needs and what we mean by sort of long term curation?

**H3**

I think it's very, very good resource I think. I think it's a really interesting resource which I

enjoyed using and I learnt stuff from it. I think I suppose I'd have question as to what's the intended audience was? Because I suspect the public would be much more interested in exploring the data and having a gateway. The data for the university and having a

gateway into how to research individuals within the university, which is quite possibly what they are interested in, and maybe I'm making judgments about the general public, but there's

certainly is a huge sector whose interest in genealogy. I think you a lot of alumni will find that

interesting. And I also think there's a lot of people out in the world who would be interested in the history of their families, and I think you'd also find things like scholars who wanted

to research things like, shall we say, women education or like ministers, who sort of like, get

sort of maybe things like relationships between people who go on it into various careers and sort of their relationship with Saint Andrews and things I think people might be interested in that. I said my question is, who is the audience? Who is to be interested in the story of

What happens to the records beyond People working, the archives and museums sector and to an extent beyond maybe people, university management who have to listen to presentations from the Archives and museums. I think regards that's an advertizing tool for what the archives do is is is really good. But I also actually do think that probably I'm not sure what the intended audience is.

**CS2**

I really like how effective visualization you've achieved using black and white. It's very clean.

**CS\*6**

I like it. I think the more I spend time thinking about it, the more I like it. I think it lends itself

To exploration and sitting and thinking about it. And I think it's quite easy to not do that when you first interact with the visualization. And I think there is a lot to get out of it.

**CS5**

No, I think so. It was really fun though.

**CS\*1**

I don't think so. Not that I haven't already been brought up some a few times.

**CS\*4**

More of a question than a comment. For the periods that have been less extensively

worked on, will you get to do any of that or is that sort of for someone else to do later

Or?

**CS3**

Ah, I feel like more use of color could have been used to increase the differentiation between processes or to draw my attention, even right to specific parts of the process. I might be tempted as well to say like because this is circular visualization I sometimes found myself a little confused by the 12:00 o'clock. You know you've got a period of history like 100

years, 200 years ago, there's 400 years difference, like immediately next to each other, so there's that. By any other comments, I mean it's a really cool visualization. It very

effectively tells a story in a short period of time and an interactive fashion. Yeah, I like it.