UNIVERSITY OF SOUTHAMPTON
Faculty of Physical Sciences and Engineering
School of Electronics and Computer Science
Web and Internet Science Group

Software Evaluation Focus Group Transcripts

by Samantha Kanza (sk11g08@ecs.soton.ac.uk)
October 2017 - November 2017

A Software Evaluation Study, conducted as part of a PhD
in Web Science: Computer Science & Chemistry

Contents

1 Trial Focus Group

This trial focus group was conduced on 31/10/2017 and comprised of 3 scientists (Participants A. B & R). The following colour key is used for the Participants and Focus Group Leader:

Focus Group Leader
Participant A
Participant B
Participant R

Focus Group Leader: This is a trial focus group to figure out if this is the best way to run these software evaluation focus groups and to see if everything makes sense and whether anything has been missed.

Focus Group Leader: Do you think it would be better to have this up on a projector and have everyone look at it together, so bear in mind all of your documents are going to get put in a test google account which this hooks up to, but do you think it would be better to have this up on a projector, we can all look at it together and people can take over the mouse and keyboard etc, or make everyone bring laptops and have everyone load the software individually.

Participant A: First option I reckon

Participant B: Yeah

Participant A: getting everyone to bring laptops would be tricky, there's always 20 minutes of oh can you load it? Oh I can't load it. are you on the network?....no I'm not on the network..why isn't it not working..etc

Participant R: Not to mention I don't have a laptop I only have my desktop

Participant A: You could get a smart board and just pass stuff around.

Participant R: Yeah.

Focus Group Leader: When I run through this we're going to talk about the general evaluation of the software and then talk about some ELN stuff after that. The first question I'm going to be asking is: What type of document have you bought and how typical a representation of your working document is it. That's just generally for you guys to say what it is and how confident you are about your ability to tell me about it. Then I'm just going to ask some general questions like how much of your documents are usually made out of words and pictures and diagrams, and what formats are they usually in. So that's because as you can see at the moment this is quite text based and it is obviously harder to pull through documents that have images etc, you lose some of that information so I'm trying to get out of you guys how to typical maybe these documents are and how likely I am to have to deal with images etc and then how that might affect the usefulness of the system. Do you think that sounds reasonable or do you think there's anything else in that area that I might have missed?

Participant A: I'm very much a, if I'm doing something like jogging my memory if it's like a setup or something I'm writing about or if this is for taking notes on day to day, taking pictures and having

something that can deal with a heavy amount of pictures, is important.

Participant R: I'm the opposite, in that I don't need the images I just need the protocol and I need to save what factors have changed in the day. So as long as it can do micrograms and various other symbols we're good to go.

Participant B: Yeah I think I'm more like Participant A, it's drawings, diagrams, graphs, etc

Focus Group Leader: Out of interest when you guys create your documents, typically stuff you would actually write up on a computer because I am looking at this in the environment of when you would actually put this on a computer, what would you actually produce documents in.

Participant A: LaTeX

Focus Group Leader: So then PDF

Participant B: Yeah, LaTeX or Jupiter notebook so HTML or PDF

Participant R: Either word or prism or excel or powerpoint if I need a diagram realistically

Focus Group Leader: What formats do you use for your data roughly?

Participant B: CSV

Participant R: Prism, matlab

Participant A: CSV

Participant B: also tab separated, anything that would be delimited data

Focus Group Leader: Then these two tag things, this one shows the different services that we need, these are four different tagging services. Open calais goes through and tags based on wikipedia articles, it goes through those documents and your documents and looks at how other people have tagged those documents to figure out how to tag your documents. These tags aren't actually words in the documents, these are tags that it thinks this document is about. Chem Tagger and gate, these go through and actually pick out chemicals or things that they think are chemical compounds or elements out of the document, and then the ontology, so an ontology is basically a hierarchical structure of concepts for this purpose, this one has chemistry, biology and physics ontologies in there, so it can pull things out, and again its things that are actually in the text. So again the chemical elements are more weighted towards chemistry but equally stuff out of the ontologies and open calais will bring in more biology and physics if it was about that. So I've got them split up into services here and they're in the order that they are ranked. And then here this is because these come through with different rankings I can't really put them on the same set (open calais vs the rest of the tag) but these are the other ones, and again these are ranked in orders of the tags. I'm going to be going through and asking people which of the tagging services do you think have the most appropriate tags for your work. Because I'm interested in what both chemists and non chemists think about that. As two of these are quite heavily weighted towards chemistry.

Participant A: Is there an online service type thing where you can download tags that are more

relevant to the niche of the subject you do. If I had this could I go somewhere and find semi conductor optics tags and then somehow put them in that in the notebook and it would recognise those more.

Focus Group Leader: Yes, at the moment potentially we could add in most ontologies to this, at the moment it's got a selection but if there were any ontologies out there that had more detailed terms we could add it in. Also as long as things came back in a similar format we could use other APIs. At the moment these three push to different APIs. So I'm also going to be asking are there any of the tagging services that you don't think look useful. There's a chance the non chemists will think the chemistry ones aren't as useful. And how would you have tagged your document if you'd done it yourself.

Participant A: Date, and year and project, or if it's relating to a certain project.

Focus Group Leader: Would you have gone beyond that?

Participant A: No

Participant R: The methods that you'd have used as well, what experiments, what setup it was, you kinda need to know that as well, as there's going to be different things like ml and hour that have changed.

Participant A: Mine would be super basic

Participant B: My one would pretty much just be project, project number and project name because I mostly use physical notebooks at this point, it has project number and name at the top of the page, and then it's got page number continued from and continued on every single different project so you can flick through going through each one. That seems to get most things done but I'd probably use ontology tags, I might not use the chemTagger tags it doesn't seem particularly useful for me. If I'm naming a chemical by name it might be useful so that that project will always refer to it so it will always pick up those pages but other than that it wouldn't be very handy.

Participant A: It would depend how much of your other work uses it, so I would imagine this would just be I did duhduhdh in the lab and for me a risk assessment would probably end up being outreach stuff that we did so that would be useful to have.

Focus Group Leader: What about for your electronic work because bear in mind this isn't just your paper lab notebook stuff I'm thinking about I'm thinking about.actually if you think I didn't make that clear then I should probably say that at the beginning.

Participant A: Yeah I thought this was just for your day to day everyday work.

Focus Group Leader: No, okay hang on that's good I should definitely clear that up. So, the premise of this is not just for your paper stuff and it's not necessarily looking to replace your paper stuff either, this is to be ideally to be the bit of software that when you go to your computers to do stuff, be it type up your reports, type up your papers whatever, this would be the bit of software you'd go to. Obviously there is the hope that this would further encourage digitisation of your work, but this isn't just assuming your paper stuff but it's assuming your electronic stuff as well. So with that in mind if you were tagging your electronic documents would that be the same answer:

Participant A: No

Participant B: No

Participant A: If I was doing a COSHH form or wanted to look at stuff pertaining to that I would use the tags more

Participant R: Also if you're trying to do a methodology or for whatever reason you wanna use an abbreviation for the first time then the chemTagger tags would come in really helpful, because I use EDTA all the time and it's not something that rolls off the tongue what it stands for

Participant B: Say if I were writing things up then I'd want it to also, apart from everything that's tagged there, also say is this a methodology, is this a set of results, is this a paper, is this a thesis, be able to work out what kind of document has gone in there.

Focus Group Leader: Okay that's interesting.

Participant A: Can this do more than one notebook, like a notebook for lab work and a notebook for admin or would it all be on the same system.

Participant R: Can you have subtabs.

Focus Group Leader: This is showing you everything because that's how it's been coded up for the prototype, but the end goal of the system would be that whatever of this was decided to be useful would go back into a notebook, would be done on a per document basis, so then your tags would be done per document so that the stuff that goes away and computes the tags and sends off to the services to figure out what it's about would all be done per document, so you could have whatever setup you wanted, and then the searches could be done per folder, per lab book, whatever you could restrict it down to as much as you wanted. To actually bring it back if you guys thought we were just talking about paper, for electronic stuff what formats are your documents and how much of them are made up of words pictures etc, is that the same or does it differ.

Participant A: More words, equations if they count as words. So pretty much all text. Much fewer in terms of pictures.

Participant B: A lot of pictures, lots of methods.

Participant A: Even in a paper?

Participant B: do you mean journal paper or?

Focus Group Leader: I mean anything you might write up to do with your work, your thesis, your pre thesis reports, papers, reports for your supervisors, literature reviews, presentations, anything that associates with your lab work.

Participant B: A fair amount of pictures, but a lot of words, mostly describes the pictures

Participant R: I'm still the same, always a nice mix

Focus Group Leader: Would you guys make use of an option to add or remove your own tags.

Group: Yes

Focus Group Leader: So search, at the moment it's a fairly basic search, you can do a regular search and you can type in and have a look (does search examples for generic search). You can search as you'd expect, also you use a slightly advanced search, you can just search in the title and just search in the text .

Participant R: can you do 2 words at once

Participant B: can you do boolean

Focus Group Leader: Not currently, but if you think it would be useful?

Group: Yes

Focus Group Leader: At the moment you can do multiple terms, because the way it works is it takes your multiple words and splits it up and then you search through the array of words and depending on whether you're doing title or text or all of the above, it then computes the term frequency for the total for the text and then adds them both together and then it orders them by that.

Participant R: Do you have to put and between the words or can you just put them in without it.

Focus Group Leader: You can just put them in without and.

Participant R: The slower pieces of software involve brackets and ands etc

Participant B: Basically gmail search, email search, the way that it's able to do two terms as in let's say nitrogen and water, not this. From date, and then you input the date. If it can do that kind of stuff then it would be very useful.

Participant A: And use a * if you can't remember what it ends in.

Focus Group Leader: So basically regex search as well then?

Group: Yes.

Participant A: Can I ask about if you already have a text document you want to put in there. Do you just like you would in drive drop and drag or like how does this interface with other stuff.

Focus Group Leader: At the moment this interfaces with google drive, when you first login to this application you get a list of all your documents in google drive and you get to choose which ones you care about putting in, as obviously you might have some you don't want to come in here, and then it will do it. It persists the data so it basically goes has this file been updated since I logged all the data on it, if so go and check it again if not keep the data, and then that way obviously if you had a document you could add it in. Again this is currently how the proof of concept works, for the final bit of software it would be something like google drive where it would actually interact in like that.

Participant B: Have you considered whether it's possible to implement with Microsoft's...what the university uses for online files. Online file syncing it's the same thing as dropbox and google drive. The university has vetted it for various privacy things that need to be adhered to because google drive doesn't do european economic region.

Focus Group Leader: this is kind of working in two ways, at the moment it's hooked up to google drive and the initial prowess was doing it off google drive as a proof of concept, because of the APIs and the software affordances you have. However it has currently been written in a modular way in that everything it does is separated out from google drive, so you essentially get a bunch of files out of google drive, give the files to my software and then my software does everything else, so in that sense anything that would allow you to integrate it would you could, the problem with any sort of proprietary software unless they've given you really good access to your APIs you're going to struggle to get your stuff back in. Google was picked because it was the best way to get stuff back in, because you can use the APIs and do things externally but you can also do things inside the document with Google App scripts and you can make applications that sit inside google drive. That's why it's been chosen for the proof of concept but realistically the way it's been written and if we ever got funding to do it properly then the look would be to integrate it properly, and that could be with a number of different bits of software. Me; Do you think just being able to search by the title or the text is worthwhile?

Group: Yes

Participant R: Sometimes you do a keyword that I would never put in a title

Participant B: You might remember that you said that particularly set of words in that sentence and you know that's the page you're after rather than the tags.

Participant A: The more search terms the better

Focus Group Leader: I've said would you expect to see any other advanced type of search, you guys obviously said boolean but is there anything else you'd expect to see

Participant B: Not really

Participant A: maybe, I can't imagine it doing an image search very well?

Participant B: if you could implement something that could work out whether an image is a graph, like if it could search whether there is an exis on PNG

Focus Group Leader: For future work it could be integrated with google image tagging and maybe get a basic idea of it. Additionally, if you could add and remove your own tags, you can add in image tags.

Participant R: I suppose realistically as long as you give a legend to the results part that you put in diagram wise then you should be able to find it that way.

Focus Group Leader: Equally if you tagged something as an image then you could do image: searchTerm

Focus Group Leader: I've also looked at markup (shows markup in system). You know you talked about wanting to know what kind of document it was Participant B, this has actions in it, like this is a synthesise action for example and this is a remove action. It's currently got it marked up with actions from chemTagger, it's got it marked up with chemicals, so this is a chemical element, and it's got it marked up with things from the ontology which then say what these words mean from the ontology (showing tooltips) and you can turn the markups on and off. At the moment everything but the actions have top markup priority so if you turn those off you'll see any actions underneath, but all the tooltips work. I've partially put the markup in to show where things have come from in the document so that people can get a better idea of it, and particularly the ontology tooltips are there because you can actually see what the word means. But in terms of an actual system do you think any part of this is interesting or useful, the actions for example or having the tooltips for the ontologies, or is it just something you wouldn't expect to see?

Participant B: So if the ontology is particularly reliable then I would say that it is useful for passing on notebooks to people. Because when people leave and they've written stuff in a particular manner, if you can add markup yourself, so if someone keeps on referring to a particular process all the way through and you don't know what it is but it highlights it and gives you the definition that they wrote, then obviously thats a lot better than having to look it up or send them an email or whatever.

Focus Group Leader: So you'd like the option to add your own markup as well.

Participant B: If a word, such as bromochlorothienopyrimidine (example on system), it gives you a decent description but it just says chemical element, so if you could edit that to say whatever is relevant to you.

Participant R: Maybe something relevant to its risk assessment like whether it's toxic etc.

Participant A: It would be good to put your own tag around some text and then call it definition and then you could search through your own definitions in some form of glossary.

Participant R: I don't think the definitions are really useful, the more you specialise the basic things you tend to forget the definitions of them because you move so far away from it.

Participant B: So for example if you had something like this chemical that is always used in a particular concentration or a particular method and you can't be bothered to write it out every time then you could add it to the definition saying typically this unless otherwise stated.

Participant R: Yeah this percentage etc.

Focus Group Leader: Also, when you were searching, just to dial back to search for a minute, when you were searching for a word, if you were just doing your general search, would you expect it to prioritise the results by it's existence in the title, term frequency or tag weighting.

Participant R: Frequency, if you've mentioned it more than once. Kind of.

Participant A: Half the time if I'm writing something I'll give it a title that might not mean something amazing and then I'll write the document and kind of forget about it and leave a working title, so I'd go with frequency.

Participant B: Yeah, my titles tend to be pretty vague, because things evolve and then you don't want to change the title.

Participant A: Or you just leave it because at the end of this paper I'll have a different title anyway

Participant R: Would also like to be ordered by date as well as frequency.

Participant A: If you can tag stuff by date and then get a chronological bar of all the documents.

Focus Group Leader: When you say date do you mean created date, so when you order the search you'd expect it to be by most recent document or then by term frequency?

Participant B: I'd probably say toggle between relevance or sort by date.

Focus Group Leader: Interesting okay.

Participant R: Yeah the option for either or.

Focus Group Leader: When you're looking for date unless you're doing specific date searches you're talking created date.

Participant R: if you know you've done something in the last week but you don't know where the hell you saved it or whatever you called it but you remember a keyword, but knowing that you wrote it months ago that has it 5 or 6 times is not helpful at all. The option either or would be nice.

Focus Group Leader: *shows term frequency graphs / ontology graphs* If this was to go into an actual document system, so remember you've got your tags, your searches, your ability to have tooltips and markup etc. What would you expect to see in an actual system. So imagine this was in Google Drive and you're opening up your document what would you actually expect to see?

Participant R: Depends on what it is, if it's a protocol, see everything time and date and everything bulleted.

Focus Group Leader: No, in terms of everything here, how would you expect to see the tags represented or would you expect to see them, would you expect to see them somewhere close to the document or would you expect to be able to hover over them and see it or would you expect to be able to just click a thing in the document so you can see the tags, or add and remove them.

Participant A: Hovering over sounds good

Participant R: Yeah hover. They don't have to be highlighted necessarily, if you can then click the option to show the tags that are available so you know whether you need to add or remove them then that's helpful.

Focus Group Leader: So when you say option to show tags or not do you mean kind of more like this markup thing.

Participant R: More like the highlighting yeah. Purely so it looks like a plain text document.

Participant A: Then you can toggle something on and off.

Participant R: For whatever laziness sense, if you copied and pasted that into a word document or LaTeX then that's going to be bold, different fonts, green etc so it would be nice to be able to change that format. Because I think realistically if you're doing everything online anyway then you're going to be copying and pasting things across into things like thesis.

Focus Group Leader: So you'd want it in normal format and then to be able to turn the hover on and off.

Participant R: Realistically the hover is amazing and probably needed at all times but I think the actual being able to see which words, it would be something that you'd need at the time of editing to add the markups for if you need any extra or if you want to change any, but otherwise it would be nice to just have it as plain I think.

Participant A: In terms for the notebook for this type of column (column of all the documents name) I would like the ability to close it / hide it and then reopen it. As if I was typing in this bit (document bit) I'd want the option of more real estate. But I like the side tabs, they're useful and fun.

Focus Group Leader: So you'd like the option to be able to view the documents by tags?

Participant A: I think I prefer that more as something that I couldn't get rid of and then I could occasionally expand and see all of the things (documents).

Participant B: I would probably say that I would like a UI somewhat like mendele, if you're familiar. I mean that one does a fairly dumb tagging system and you kind of have to do it yourself most of the time so it's not semantic at all. But it has a lot of the searching features and a lot of the space that you might want, the only issue is that if you're loading in PDFs, because I would probably load in the PDF you can't interact with it you'd have to go back to LaTeX, so the notebook itself would just be something to keep track of outputs rather than help you actually do anything.

Participant A: Is this supposed to be more of a document edit tool or more of a document management tool because I'm not 100% sure.

Focus Group Leader: This is showing you everything that can be done with the documents to figure out what should be done in the actual system. What would happen in the actual system would depend to an extent on what you were integrating it with. Imagine, say we open up Google Drive (does this), so say because this is currently integrated with Google Drive, imagine we got the funding/the code such that it was integrated properly into this, if I then go into one of the documents in Google Drive, so this is the document we've got out from Google Drive. In this document, what would you ideally want to see from what's feasibly shown in my system. For example, would you, because Participant R you were saying about the markup, would you want something in here to say turn the markup on and off.

Group: Yes.

Focus Group Leader: Would you want to be able to hover at all times and see what these words were or would you want to be able to turn those on and off, would you expect to see something here

about the tags, like would you expect to be able to dropdown a tag menu or have something come up here.

Participant A: These all sound good, turning things on and off and having all of the options.

Participant B: I think the options to turn it on and off, I'd probably opt out of the hover / display.

Focus Group Leader: So you'd want to be able to configure all of these?

Participant A: If it wasn't your document

Participant R: Or you were sending to somebody else

Participant A: If someone sent you a paper you'd probably want to have those options

Participant B: But I want the ontology stuff to appear on the side rather than hovering over

Focus Group Leader: You'd like to be able to see the tags but not have them attached to the document, but presumably you'd also like the ability to hide them.

Participant B: Yeah being able to hide them would be useful, if you're going through it and this system is throwing up tags and boxes every time you try and highlight or edit a bit of text then that would probably get on my nerves very quickly.

Focus Group Leader: Presumably you want an overall ability to say default don't show x y z and turn it on per document.

Participant B: Yeah

Participant A: Like a tab

Participant R: If you had default settings that you predefine and save as your preference and then obviously you can fiddle depending on the way that you've set it up independently. As I think it's very user dependant because I've obviously looked at it for certain things but you guys would use it very differently. And some of the things I'd prefer to have all of the time others definitely wouldn't.

Focus Group Leader: More on the ELN side. Previous studies I've done and other people have done suggest that scientists have a greater tendency to use software for writing up reports/thesis/papers etc rather than during their lab experiments to take notes. What do you guys think?

Group: Resounding agreement

Participant R: they also hand you a notebook at the university that is your lab book

Focus Group Leader: Even I got a paper log book and I'm in computer science

Participant R: Your supervisor reads your paper lab book. Default is write in book.

Participant B: Another concern is the lab book that I have has locations for me to sign a reviewer to sign and countersignature and a reviewer signature and I don't know because I've never looked into

it, but if you had results that were patentable because you need to show the roadmap towards it, are electronic signatures accepted at all.

Participant A: Oh Yeah.

Participant B: Because if you had an ELN and it showed the entire road map but you couldn't counter sign it or reviewer sign it then would that be enough

Participant A: Or prove that it was their signature

Participant B: Exactly

Participant R: You'd have the date on it wouldn't you, date stamp of when it's first done

Participant B: yeah but it's whether that's submissible

Focus Group Leader: Electronic signatures can be a thing, but bear in mind this is aimed at people when they come to use software to interact with their work rather than to replace paper, a lot of my big conclusions so far is that we can't replace paper, and it's better to build a system to work with it and improve the current software offerings that you guys have rather than try to replace that paper lab notebook entirely. Would you guys be more inclined to use an ELN if it was aimed at the process we just discussed? So writing reports, data analysis etc rather than replace your paper lab notebook.

Participant A: I think my main desire for it would be for it to backup my lab notebook so there's always a safety net. In terms of analysing data I wouldn't use it for that, I'd be using matlab for that. Or something equivalent.

Participant B: It might be useful to keep track of results, but I tend to agree. You'd want it for daily use kind of thing because there's a lot of data you go through, a lot of data and methods.

Participant A: And a lot of things come out as useless.

Participant B: If you can tag things as Nah then that gets rid of things immediately and you don't have to rifle through an entire notebook of null results.

Participant A: I think it would be very good for legacy, so if you have a student coming in behind you and they're doing you work, then I think all the things you described with the tags would be useful. Really really useful.

Participant B: By the point you get to the journal paper and the thesis. Usually you have to hand do the tags anyway. Particularly with papers they get you to put at least 4 tags on a paper, at least for the ones I've published. So the fact that you have to do that anyway precludes the fact that this can do it automatically for you. Which might be handy if you want to tag it with 20 things to hit it with as many things as possible.

Focus Group Leader: So almost this would be more of an interim stage, so a stage beyond the paper lab book but a stage before all of your final writeup pieces.

Group: Yeah

Participant R: Definitely, it would realistically be, say my lab book is purely used for the day I did it, the experimental settings that I used, but the actual data and the results are always on an excel file elsewhere so this would actually be a way of, I'm probably a bit less organised than some people but this would be a way of combining everything together rather than my colleague who cuts and sticks everything into a paper lab book to see how that data went etc. So I mean you can see it would be helpful to then type in this protocol, see the different changes and you can see the data and what worked and what didn't.

Participant A: That kind of important intermediate where you know the stuff you did that day, and you know what was important and what was useless, and you add the important stuff in.

Participant R: Yeah and add THAT DIDN'T WORK in capital letters by things, so never look at again.

Focus Group Leader: So would having the semantic tags and something to auto mark things up and auto tag them encourage you to use a bit of software to put in this particular place.

Participant R: Yeah

Participant A: Yes if they were good.

Participant B: What I've found with ELNS is that it's clunky, at least for me, I spend too much time scribbling stuff and it's way faster to scribble than it is to type anyway, and that slows me down so I don't use my laptop to take notes I use a paper lab book. For a little lab book I could feasibly cut all the pages out and ram them into a scanner and then all of them will be uploaded but that's a lot of effort. Either that or I have a binder instead of a notebook but that's not really going to work in a lab environment.

Participant R: And that nice moment when the binder breaks because there's too much stuff in it.

Participant B: Or it's not in order.

Participant B: If there was something intuitive and fast to use that gave you drawing and scribbling and writing abilities that you could interface easily to this, like an e-ink / kindle but with a stylus of some description.

Participant A: It would be good to recognise a partial differential and a load of greek symbols.

Participant B: Although to be fair there's that app that does that, you take a picture of equations and it tries to differentiate it for you. It's mostly an interface problem, because the stage at which this would be useful for me, is the stage where an interface that I like doesn't exist or at least I haven't found one. But it would be very useful to arrange all of that data that I'm producing automatically and well. As opposed to what I do which is what's the title?

Participant A: It would be better for others to understand.

Participant R: Is this aimed at every single level possible because obviously you guys would go into a postdoc knowing everything you do. Whereas if you started at the beginning from there this would

be like a golden tablet kind of thing and you'd be really excited about it and it would literally save your ass from day 1.

Participant A: I reckon this would be more suited to undergrads.

Participant R: Who are you aiming at or are you aiming for everyone

Focus Group Leader: It's currently aimed at scientists full stop, people that do work in the lab. And part of this is also trying to ascertain where it would actually get used. Or where there might actually be a market for it. This is part of the investigation.

Participant R: I think that some of the tagging would be really useful at earlier stages of scientific career. Like **Participant A** said with the undergrads. PhD students if you have that from day 1, you'd never look back. But then the more prestigious that you get, and the more knowledge that you know, you don't want as many tags in your face as you already know what something is.

Participant A: We're all so used to something else as well.

Participant R: Because you've got the options then you could give user settings and it could be aimed at different people depending on what they wanted. I think the older a researcher gets realistically the less they want flashing at them as they know what everything is with their eyes closed. I was just thinking what **Participant B** said about not needing things, but then going Biology way from day 1, I'm at a single cell level and originally I learnt about everything, but the basic stuff has now gone out of the window. If it was constantly flashing at me telling me what a basophil was I'd throw out of the window. **Participant B** and **Participant A:** Agreement.

Focus Group Leader: But do you think if you were a first year undergraduate student

Participant R: And you didn't even know what a cell was

Participant B: Then you'd feasibly want to see all of this stuff

Focus Group Leader: On that logic. Maybe you guys might have been more prone to LaTeX your stuff from year 1, but then it if was somewhere you stashed all of your uni work, because you are going to be producing stuff that is more likely to be in a straight PDF or word document for essays and cw handins etc

Participant R: The earlier in that this is put. The earlier in years that this is given., the more organised they'll become.

Participant A: they'll be better at using it.

Participant R: They will be better for it realistically the quality and then losing everything and transfer thesis going out of the window and all the stuff I've gone through wouldn't happen, it would organise them at an early stage which would then give them a much better work ethic later on.

Participant B: Particularly would be a help for modules that rely on previous modules, because

Participant R: They forget

Participant A: Oh yeah

Participant B: you'll forget stuff exists that you've already done, alright I need to learn everything about lasers today and you've already done 2 modules but you don't know exactly what you've learnt about it

Participant R: And then it would bring up your notes about it

Focus Group Leader: that could go through your notes and even if you put the slides in that you got in a good form it could go through the lecture notes too.

Participant A: Exactly

Focus Group Leader: Because this can do, because you know in Google Drive you can have powerpoints / slides in there too.

Participant R: So say your key cards are in there too then you could literally hover over the word later on when you're writing your dissertation then it could bring up useful info on it.

Participant B: On that note if you're able to, so at the moment your ontology definition comes up for a particular word, if a button comes up next to it that just goes straight to all documents containing that

Participant R: Like a search within, if you really want extra information and you know that it's there

Participant B: If it's somewhere then that would be useful

Participant A: Maybe level appropriate packages, like undergrads definitions package

Participant B: Yeah

Participant A: Because that would probably be too basic for a PhD to want

Participant B: At an undergrad level you probably don't want custom definitions you probably want dictionary definitions

Focus Group Leader: So different levels

Participant A: Yeah as a postgrad you wouldn't want to be told what isotope was again.

Focus Group Leader: So maybe a configurable setting in the beginning that's also how much detail you get, might still be interesting to tag your documents with isotope but you don't want to highlight it

Participant R: So everyone could get given the same package and then say for undergrads we recommend these settings and for postgrads who don't want to see these then press this button. So there's a basis for them to start with rather than having to go through it all themselves to figure out what settings they want.

Focus Group Leader: You said it would be interesting to click on something or hover over something and say show me documents containing that particular chemical. On that note, obviously you can just search for terms, but would you want to be able to do a specific chemical search or a specific

Participant R: i think it would be more word than anything else, it wouldn't be the actual chemical signature as you'd have been more likely to write the word, if you're doing glucose you'd actually write glucose. **Participant A** and **Participant B:** Agreement.

Focus Group Leader: Would it be interesting or in any way useful to have a document and say find me things similar to that or would that not be useful. As this has come out of previous studies.

Participant A: What do you mean similar.

Focus Group Leader: It came up as a potential feature people might want, and it hasn't currently been implemented as it was hard to see how to do that in the proof of concept but it's something I wanted to ask in these evaluations.

Participant R: An example I can see like if you could put words under a subheading, so in biology for example if you're talking about anesthetics, that is a group as a whole that these chemicals you know belong to that group, when you would say show similar it would come up everything that was a general anesthetic rather than something that has a similar chemical pathway because that could be most things.

Participant B: Like all of organic chemistry

Participant R: If you tried to get that and it came up with that it would be a bit scary but if you were trying to do like a sub category then it might be useful.

Participant B: Physics example: There are two courses in 2nd year physics and one is quantum and the other is wave physics but both of them have similar equations of the same form, and if you know one you can apply one to the other. But it wouldn't tag them immediately based on subject because it would just say this is an equation for a wave, and this is an equation for a wave for quantum physics. So being able to recognise equations and then apply that to saying these two things are basically the same so that for a student would be semantically quite good because then you can say I've already done this course and I did well so I know I understand that so I can understand this. But that might not be immediately obvious to a tag if this says these are wave equations and these are quantum.

Focus Group Leader: Would having a more accurate search across your work, being able to have everything tagged and marked up, would that encourage you to use a system or put your data into a system more or not.

Participant A: it would encourage me as i think if I had this or I was going to use it it would be in the situation that I'm in a research group and this is how we preserve our data, so the easier it is to search for me, to search someone else's work and for them to search mine the better it's going to be.

Participant B: It's mostly ease of searching other people's work. I know how to look through my stuff but push comes to shove I doubt I'd be able to find things in other peoples log books.

Participant A: Yeah.

Participant R: On some days depending on how I'm feeling things don't necessarily get saved under the right name and then trying to do a nice little apple search hoping it will come up with the right thing and it never does.

Participant A: Or you save it in a folder and you didn't realise it was going to that folder

Participant R: By pure accident of pressing enter too quickly

Focus Group Leader: So would it be really useful, for example google drive is collaborative, if you were able to pull through if it was marking up everything on the collaborative folder or everyone in the research group folder and everyone was using stuff like this then you could search through everything.

Participant R: Would there be a privacy level issue. If you had access to their stuff and suddenly got an amazing idea and you changed their experiment oh so slightly.

Focus Group Leader: It would more be you would have as much access as you'd always have.

Participant R: as much as they'd allow.

Focus Group Leader: I know that you (Participant B) have shared stuff within your research group (other previous participants), It's not suggesting additional access, if you already had access.

Participant R: Yeah that's fine.

Participant A: You might all have your own research group language for how you tag things.

Participant R: I ask because my lab is in the middle of lab wars, and if certain people found my lab book online...that would be really bad

Focus Group Leader: No definitely not for stealing other work, it's just assuming you did have access or if you were leaving your stuff for legacy purposes.

Participant A: Surely it would make it easier to know if someone nicked your work as you'd be logged in to see it. I'd be really annoyed if someone modified my tags without me knowing.

Participant B: On that note, how is the ontology kept, do you download it and then would user made ones be held locally or is this something that would update across a network.

Participant A: user made ones would be very good.

Participant B: Basically is there going to be a master file somewhere on your computer that says this element, definition: this, user added: no. Or yes.

Focus Group Leader: The stuff that is generally out there would be networked (or is currently), all the ontologies that have been made by other people that we're using that contain the different chemical terms for example, that's all coming from another central ontology, however if you choose to either overwrite or add you would get your own personal file of your tags which you can go in and

change.

Participant R: it would be good to personalise it.

Participant B: what happens in the event of no network connection.

Participant A: How would it work online

Participant B: Because the university occasionally goes down, not all the time but sometimes.

Participant R: We often have network issues

Focus Group Leader: things can work offline, there definitely might be points where you'd want to connect and check for an update potentially because things do get updated but you could definitely opt to have local copies if you wanted

Participant A: how often would things backup or autosave

Participant B: Yeah

Focus Group Leader: The tags or document system itself

Participant A: The whole thing

Focus Group Leader: the document system itself would depend on what platform we were using, Google drive however it does normally. Bear in mind this whole system, this wouldn't be what you got given,. You'd be given the document system you were already using and you would have this additional functionality on top of it. So the documents / everything else would depend on the document platform. The tags and everything else could be configured or saved when the document saves.

Focus Group Leader: Last few questions, do you think that semantically tagging your documents would have any impact on the efficiency of your work?

Participant A: I reckon it would have impact on the efficiency of other people going through my work, maybe my work.

Participant R: It depends how specific it is. And the reason. If this is an interim and you have a particular data set that is of a particular importance. Every day experiments and it's the same thing so you constantly ended up with the same few set of words that wouldn't be that useful. However, being able to change it to like the settings / various other tags like that that would be useful would be great.

Participant B: I think that if your career happens to be particularly long lived, by that I mean my supervisor occasionally says oh I don't know where that is and will disappear for a couple of hours whereas this would probably take 5 minutes to find something. In terms of my efficiency I'd agree with Participant A, that it wouldn't really impact me but other people's stuff would be faster to look through and if somebody wanted to look through my stuff then it would be faster for them.

Participant A: I'm guessing if you could look through other people's stuff better then that would

improve your work.

Participant B: Yeah. It's a good collaborative tool but it doesn't impact me specifically it would impact other people when they're looking for stuff. But if most people who do that could talk to me directly anyway they could just ask.

Participant A: Unless someone prompted you for like what is this data trend, and you might think it's this but then you could be like actually no it's that.

Participant R: Then again it's the level of user that you are **Participant B** and **Participant A:** Agree

Participant A: the more involved you are.

Focus Group Leader: Would the following features encourage you at all to further digitise your work or add your own tags to your work.

Participant A: Yes, but again i think I would be doing it more for someone else than for me, but then for me doing things to teach other then helps me learn things and structure things so that's not a bad thing

Participant B: I think yes it would be useful long term. My main issue is still an interface. At the stage I would like to use this the interface still isn't good but it would be useful.

Participant R: I'd say I would use it, I think that it would be very useful, there's a lot of very generalised terms that after specialising in something so small I've forgotten the basics and I think that's quite an easy thing to say. Early on pathways of general allergies have kind of gone out of the window. Whereas everything this particular cell does I can tell you with my eyes shut. With that it would be useful.

Participant A: I think it would be good if you were forced to do it from the off.

Participant R: it would organise you so much.

Participant A: In first year everyone is less organised because they don't know what they're doing or how to organise

Participant R: Or how they learn

Participant A: As you go on you get into a rhythm but if you get everything tagged from the first bit of chaos.

Participant R: It encourages you to write as well. I think I personally wrote different subjects into different books, but if you could tag papers in it that you know are relevant or key bits of extra reading in each one then you'd know what you were doing. I wrote everything in black but extra reading was in blue. That wouldn't be necessarily counter intuitive to someone who's just come up from college and they've only just turned 18 and they haven't got into studying yet.

Participant A: Then you wouldn't have to worry about keeping track of physical things it's all in

one place.

Participant R: Yeah if you lose it or it gets wet walking to and from campus's or ripped or burnt or stolen or whatever, you lose it in moving from place to place.

Participant B: One other thing. Just to ask. What kind of files does this interface with. And could you do a file type white list, so say everything relevant that you want tagged is .PDF can you tell it only do .PDF or can you tell it do .PDF do .tex do .doc and .docx

Participant A: You don't want it coughing up a temp file

Participant B: If you went into a LaTeX folder you know how many files that has, and you know that most of them are irrelevant and then would you want it to necessarily interact with your .tex file or would you only want it to interact with the final product

Focus Group Leader: At the moment it tags .txt and .PDF, but again I think configurable would be the way forward.

Participant R: then realistically you could save your diagrams in PDF anyway

Focus Group Leader: Presumably then would you guys want, as part of your configuration in the beginning would you want options for what types of files you care about tagging

Participant A: maybe

Focus Group Leader: Anything that was written would be sensible, and also remember you;d got to be able to read it, it's not likely to be able to read .log and .aux, it would be set to sensible things with the right extensions, which is currently .txt and .pdf and then if it was extended it would probably have powerpoints and presentation files. But you'd expect regular file formats.

Participant B: So I never use word documents, so if there was a word document in my drive it's most likely a COSHH form that I don't care about.

Focus Group Leader: At the moment you chose exactly which documents you tag anyway so you would only take what you cared about. But then if it was working on a per document basis rather than an external system it would have your standard formats that you could configure.

Focus Group Leader: Any other comments?

Participant A: Re similar documents, It would be good to hand link documents/projects rather than expect that to be done automatically. A computer might mess it up. In terms of the software actually one question, when it's importing stuff can you tell it not to do it.

Focus Group Leader: Yes.

Participant A: So it can be done on a document by document basis.

Focus Group Leader: Yes

Participant A: thats good. I reckon it would need to come with a lot of things you could toggle on and off and also a really comprehensive manual for someone to get the best out of it. You'd need to sit people down and teach them how to use it or people would lose their temper with it quickly or not use it very well, and then it would fall flat on it's face.

Participant B: I think my comments for the prototype are mostly on the privacy elements for grants that require privacy.

Focus Group Leader: What about if you had this ability in a private document system. This isn't the thing you would get. This is a proof of concept project showing you what can be done by tagging etc, it's currently done like this because it was quicker and simpler to use external services using the Google Drive API to pull out documents rather than doing it via app scripts. But it's also been done to write it in a modular sense so all the functionality is separate from Google Drive and currently you can switch around where you get the files from. The overall idea would be if this was actually integrated, it would be integrated into Google Docs or Evernote etc. Hence asking what would you expect to see in the document, as you wouldn't get this you would get Google Drive with some tagging features. It would have enhanced search and the tooltips/highlighting etc could be done inside the actual document.

Participant B: As long as implementing it for something like that is feasible. The other thing is the issue of ELNs and signatories. Those are the things that aren't really in the scope of this project but are worth considering. These might dissuade people from using it.

Focus Group Leader: Do you think that when I start this next time I need to give a better explanation of how the proof of concept works and that this is what the project is trying to do.

Group: Yes.

2 Chemistry Focus Group

This focus group was conduced on 15/11/2017 and comprised of 6 chemists (Participants L, J, S, AP, AJ & AK). The following colour key is used for the Participants and Focus Group Leader:

Focus Group Leader
Participant L
Participant J
Participant S
Participant AP
Participant AJ
Participant AK

2.1 Introduction - Explaining the purpose of the system

Focus Group Leader: I have been looking at electronic lab notebooks for my PhD thesis, I did a bunch of initial studies, I looked at what was out there in the market, I observed people in the chemistry labs and I did some initial focus groups with physicists chemists and biologists, my interim conclusions were that there's a lot of good reasons to digitise your lab data, there's still a lot of reasons why we should be looking to do that, and that people do actually go and use quite a lot of software for what they're doing. But ultimately there's still this kind of barrier to actually using electronic lab notebooks in the lab, and there's quite a lot of barriers trying to replace paper, so I'm not really looking to try and create a system to replace paper because I don't think that's particularly feasible right now. But what I am looking at is trying to improve the current software offerings of the software that you would use if you were actually writing things up on a computer, and hopefully by improving those we might encourage people to digitise their work further. So, my initial thoughts were when you look at everything, ELNs aren't very popular even though there are loads of them, but electronic notebook software like Google Drive, OneNote, Evernote etc is a lot more popular, and if you break down the features that you would need from an ELN and a EN there's actually a hell of a lot of overlap. For an ELN you need everything you'd have in something like Google Drive and OneNote and then you'd have a domain layer with domain applicable tools, and then a further semantic layer on top that handles metadata and tagging.

***shows diagram ***

This is a diagram of the breakdown of the three different layers, being everything here is generic notebook functionality, this would be domain based stuff, and then at the top is the semantic layer which looks at tagging your notes, doing a better search, linking things together, recognising chemicals etc. The bit I've focused on making was the semantic layer. It's all well and good for people to say that they like their documents to be tagged and made better to search, and to have more information about them, but no-ones actually given a great deal of information about what they actually want and how they want it to be achieved. This is a proof of concept system that I've designed. I'm going to show you this in two parts. There is a backend part of the system that shows a representation of what can be done. It shows what tags have been pulled out for each document, which bits of the documents have been marked up, what ontology terms have been recognised, and where different chemicals have been identified. The front end of this system would be in Google Docs for this proof of concept idea,

and so you will be shown a mockup of how this functionality could be implemented on the front end. This focus group is to look at both aspects. To investigate how well have the documents been tagged, how things have been marked up etc and also what you guys think of how the front end looks and how you think that would work, and to talk a bit about ELNs.

2.2 Question 1 - Documents

2.2.1 Question 1a - What type of document have you brought and how typical a representation of your working documents is it?

Participant J: This was a blog post and it has an image grab on it and a bit of text and stuff because I mainly work with computing stuff so I don't really do chemistry in terms of get chemicals mix together see what happens...so this had some actual structures and stuff in it which I thought may be useful for identifying. It's probably not particularly typical of what I usually do because most of the time my stuff probably doesn't weirdly refer to chemicals as such.

Participant S: This is basically the only type of document that I really ever produce that has words in it. If you went on my work PC it would just have folders and folders of random spreadsheet which I wouldn't really know how to decode. So this is just there, it's electronic supplementary information that will go along with the paper that I'm currently writing so when I've got a load of data together or I've done an experiment I will then formally write up my lab notebook and get an ESI document like this started and slowly add to it as I know what is going into the paper and make sure anything that isn't going in the main body of the text will go in here so it will have all my experimental write up and loads of loads of figures and digital data that will have gone in there. That's pretty much it

Participant AP: This was basically the appendix from my transfer report with the materials and characterisation methods, so all of this is a list of my materials for all of the chemicals I used and those are the characterisation methods about them.

Participant L: Mine is like a part of my thesis, it's the experimental procedure for my thesis, so I basically put three different types of synthesis, I work with gold metal particles, I did three different synthesis of different sized particles and it just described the synthesis of them basically and that's some molecules that I functionalize my particle with.

Participant AK: This is a pretty typical finished write up for an experiment, what I would normally have is to start with is the materials that I plan on using, I often have a brief outline of what I'm gonna do so it's more like a checklist, add this chemical now, wait 10 seconds, take measurement etc. I try and write up what materials I'm going to use, what I;m gonna do and what I'm aiming to do and then after I've done the experiment I'll take what I've written in my actual lab gbook and put that in with the figures and the tables and whatever else, whatever calculations that I've done.

Participant AJ: This is a typical page from my actual lab workbook that I use when I'm running experiment sin the lab, so it's got some brief explanation of my kind of reference of what kind of experiment it is. And then the parameters that I'm going to need to refer to later when I write it up. And then the so I'm in electrochemistry so a lot of the data that I collect is compiled by a

computer program, either a custom computer program or something that has been developed. So then it's feasible to note down all the filenames of the data that I've recorded and saved and then I also use those filenames to refer back to later if I want to look at say what temperature I was running the experiment at and I can see what time I did the experiment and refer that to our own data log of lab temperatures on a daily basis.

2.2.2 Question 1b - How much of your documents are usually made up of words / pictures / diagrams? and Question 1c - What formats are your documents usually in when you create them?

Participant J: I've done a bit of different usage of different systems, so some of my stuff is done in books, so on paper, and some of that will be words and will also be some coding, notes on stuff like that, probably not much to do with the actual chemical structures on the paper stuff. But then I also have our notebooks, two different blogsites and OneNote. So lots of different formats, also alongside some word documents and things like that especially when it comes to reports, although actually most of my reports are in PDF format. Those will have a reasonable amount of words in them but also some coding and then notes on coding as well as Excel spreadsheets for like data and numerical stuff that's come out because a lot of the...some of my stuff is to do with obtaining descriptors and that usually comes out Excel or CSV formats so that it's able to be shared.

Participant S: Pretty much all of my stuff will be, aside from Office, so almost all of my documents will be in Word, in whatever templates I have to be using for whatever paper I'm writing, there are a few in Excel, few powerpoints and then stacks of things like ChemDraws, ACD Labs data, raw NMR data off the spectrometers, and then normally lots of origin fits as well. That pretty much covers a lot of it. I use words very very little, this document is quite a good example in that it maybe has two or three pages of actual writing and all of that is pretty procedural and then the rest of the document is figure with caption. This also will have a...there's a paper that goes with this that is a lot more wordy with images and probably has a lot more words that will describe what's actually going on. This is very much functional to get across the information it gets across and that's it. Pretty much other than writing on paper or putting together an ESI with some procedures in it I'll have nothing with words in it aside from in my lab book. Paper things are about the only things that exist with any kind of writing.

Participant AP: Actually most of my stuff is on paper, I'd rather do writing but for software I normally use word and origin for my data. And some imaging processing for key images. I usually have a lot of words normally, my documents are mostly words and there are figures. So it makes sense to have the tags.

Participant L: The document i have here is just words with just one figure with some molecules. I use a lot of numbers for concentration volumes because this is an experimental procedure it's just words. In my lab book it's the same mainly words, some scheme that I just draw when I make some gels for particles so I will just draw quickly by hand the scheme of it whilst I put in what went well. But mainly words.

Participant AK: I've got a mixture of words and graphs and tables. If you counted the pages the graphs would take up more space because they're massive, I like to be able to see them so...I do make

them take up a page. But if you scaled them down it would probably be about 50/50 because I like to make sure I've written out what has happened and what I think has happened because otherwise you looked at the graph you might draw a different conclusion, which isn't necessarily a bad thing but I like to make sure that I put captions so that I know exactly what the data is showing, and what the experimental conditions were.

Focus Group Leader: Participant L what about your formats

Participant L: For me I use mainly word and spectrum with origin but I always export as JPEG so I just copy everything in JPEG into word and after I transfer to PDF.

Participant AK: I commit a deadly sin in science and I use excel to plot up.everyone's like urgh why would you do that but I just don't get on with origin. So I use excel to do my data processing and also to plot the graphs and tables and then the rest of it is in word.

Participant AJ: When I am working with the lab (this document was originally handwritten) I have a copy book so it makes two copies of all my lab notes whilst I'm working in the lab. And then periodically I'll write up sort of bits of methods and analysis of data that I've taken so all the.the majority of the data is comma delimited text files and then get exported into Origin and I produce graphs in Origin and export them in JPEG. Often when I'm doing data analysis I'll export them all into a publisher or a powerpoint file with captions and then paste those across into a word document when I'm writing up a report.

Focus Group Leader: Moving onto the tags. These are the tags that have come out of different systems (NB: Each participant was given a copy of their personal documents and the tags as well as being able to see it on the system). OpenCalais looks at wikipedia articles that have been tagged and looks at what content has been tagged to try and infer how other articles with similar content should be tagged. These are categories of your documents rather than things that have been pulled out from the text. ChemTagger and Gate are the automatic chemical recognition pieces of software so they have pulled out what they think are chemicals in your pieces of work. ChemTagger also identifies actions (gives example showing the marked up actions in backend). As you can see Gate typically seems to pick up less than ChemTagger. For the ontologies, ontologies are basically hierarchies of relationships and concepts that we use in the semantic web so I've currently just got 6 ontologies in here, some chemistry physics and biology ones. This piece will pull out the matching ontology terms. Your documents have been run through all of these systems and currently pulled out everything it can find, and part of this software evaluation is looking to refine this process to figure out what things we should actually be pulling out. So on that note if you all have a brief look at the tags for your categories.

2.3 Question 2 - Tags

2.3.1 Question 2a - Which of the tagging services (if any) do you think has the most appropriate tags for your work and Question 2b - From the perspective of automatically recognising the chemicals you have used, do you think ChemTagger / Gate have done a good job? Question 2c - Are there any of the tagging services you don't think look useful?

Participant J: Across my one, so I don't have any gate tags at all and so I have the different tags and the different ones and none of them are the same. So each group has different tags. I guess I'd say the open calais ones are potentially the most useful, so my ones they're not...these chemical structures that I have, they're not named so they don't have any...I mean there's a family class that it has identified, so they don't have specific names so they haven't picked out specific names. And for chemTagger it's sort of bits of the molecules which isn't particularly useful really at all. I guess the OpenCalais tags are the most useful because they're kind of the grouped ones. I'd still say it's missing a number of things but um...yeah the open calais ones are probably the most useful ones.

Participant J: Ontology tags, I mean it says ring and graph and fair enough there is a graph in it and there are rings but equally it's not particularly useful but potentially if it was a longer document it might have more.but I guess the gate tags it hasn't picked up anything at all so from my point of view that would be the least useful one. The ChemTagger one's I think it's probably picked out parts of the image perhaps, for instance it's got an NH group but when that's on the left hand side on the molecule it's sometimes written as HN and sometimes it's written as NH depending on which side of the ring it is and it's identified tags as both HN and NH, not realising that that are both the same thing. There are some other tags that I don't even know what it's meant to be referring to. There are only 2 of them...it's got CI which I don't even think is even written on here. I don't know if it's meant to take the chlorine as CL and has just misdone it. The other thing that annoys me from a chemical view is that it's all lower case, it's not taken in the fact that when it says ME, that's a methyl group so that's meant to be an uppercase M and lowercase E, or O is oxygen so that's meant to be a capital O. So from a chemical point of view from reading it you kinda glance at it and go that looks wrong, even though I guess technically the letters are still the same.

Participant S: To me that's simple and it should get it.

Participant J: Yeah...your ones look like actual chemical names whereas all of mine are just elements or groups and looking at that.it looks wrong.

Participant AK: IS that the Gate tags or Chem Tagger tags.

Participant J: I don't have any Gate tags. So ChemTagger

Participant AK: I do have some capitals in mine.

Participant J: Maybe it's recognised mine from the image?

Participant AK: So some of mine. Like 4-Platinum is normally written as PT-4 because it's the labelling of one of my electrodes, but it's pulled it out and given it it's full name and given the platinum

a capital P.

Focus Group Leader: These services all bring back documents in different ways so there is a chance that some of the bits that have been marked up may have got converted into lowercase whilst trying to do comparisons. As ChemTagger gives you entire XML, Gate gives you partial XML and the Ontology tags markup around the terms it's picked out as tags.

*** Gives demo of ChemicalTagger***

As you can see this basically marks up everything apart from and's and full stops, because it has things like noun phrase and verb phrase etc. I took this and stripped out the actions and chemicals.

Participant J: I imagine it's potentially due to...actually I dunno some of the things are referred to in the text and they are capitalised whereas a lot of my stuff is off an image so I guess it had to take that text from the image. I mean I'm also..in terms of stuff I'm also missing a fair number of tags that I would expect if I was going to tag it.

Focus Group Leader: One quick question, as this is an additional feature of ChemicalTagger but they don't make it available in their code library that other people can use. They provide structure highlighting on chemicals, do you think this is useful?

Participant S: So it finds that structure?

Focus Group Leader: Yeah it will give you a diagram of the structure when you hover over it

Participant S: That's vastly more useful than a name

Participant J: Yeah, as long as it does it correctly

Participant S: This is interesting in ChemTagger as it has pulled out the UPAC names of the compounds that I've synthesised. I don't even know the UPAC names of the compounds that I've synthesized, it's an entire line long because my structures are complicated. If I could search through structure through my list...if I've got stacks of these documents of the synthesis of all of my different compounds then that would be vastly more useful. It's interesting that it pulls out my UPAC names although I wouldn't know how to search on that but if it could pull out the structure

Focus Group Leader: When you say pull out the structure, what do you mean?

Participant S: So on ChemicalTagger it's doing it for common things but it's interesting to see if it would do it for something novel.

Focus Group Leader: So you'd want to see the diagram of your structure? Or be able to search based on a diagram or words about your structure?

Participant S: If I wanted to go back to from a compound that I've made, that's what I'll care about, I don't want to go through and look for generic generic structures so if the example up there of DMF, that's a common structure that would be everywhere.

Focus Group Leader: I mean how would you see searching that going, if you had a system and you

were typing something in.

Participant S: What I'd want to do is draw a picture of it or give it a SMILES or an inchikey of a structure and then I want it to pull out the document that has the synthesis for that compound in it.

Focus Group Leader: Can you clarify SMILES and Inchikey.

Participant J: SMILES or an Inchikey are computerised representation of a compound, but the thing is, they have all their issues, so basically SMILES aren't unique and they depend on how you start in the molecule and where you draw it from, inchis are meant to be better because they are meant to be every compound has one inchi because it describes the way...but you can still do it differently because you can add extra layers to talk about chargers.but it goes this is the chemical formula of the compound, this is how it's all connected together...structure...so basically it creates something that a computer can understand and make into a picture or search for structures that are similar across two molecules.

Participant S: What I'm thinking here.it's picked out the name of a compound that I have synthesized, no-one else has ever made this ***says name of the compound*** I got this name by drawing my structure and drawing my structure in ChemDraw and copying and pasting what it told me the UPAC was, and I've given it my own code for how I refer to it. What would be useful if I had a load of these files, if I could draw that structure and copy the SMILES from ChemDraw or the Inchi and put that in the search box and pick out the document where I've had this written up. It's those novel compounds, the complex structures that are interesting to look for, as opposed to what mines done. If you have a look at the Chemtagger it's picked out every single common chemical. That's totally irrelevant to what it's actually about, m not going to need to go and search for the solvent, ChemTagger has obviously picked up everything that vaguely sounds like a chemical, but every single one of those chemicals apart from the three new chemicals I've reported here will be in every document I've ever produced. It's obviously potentially useful for it to be marked up like that, and it's interesting that it's picked up some common commercial things, so I use a dye in some of my experiments called HBTS and it's picked up that and you can imagine if someone's thinking if you're putting it into an online database of these things and people want to find out how people have used HBTS in different experiments then they could find that so it's probably useful to have that marked up but in terms of being useful to me looking through my own work...it's a lot of data that I don't want and it's not picking up the stuff that would be useful for me to find my own things.

Focus Group Leader: So would it be good to have it there but it would only be useful to you to have the unique stuff?

Participant S: Yes

Focus Group Leader: So for your tags you wouldn't want common things in terms of what you can search on but still have that underlying metadata

Participant S: Yeah exactly, if I put common chemicals into the search box I've probably written them in every document.

Participant AK: I've got the same thing actually, and mine's actually picked up if I've spelt alumina

with a lower case A or an uppercase A, it's picked that up as two separate tags. Nearly every experiment I'm going to run will probably use alumina as we use it to polish the electrodes as you can't use them unless they are clean so I would never search for that.

Focus Group Leader: So any more comments on the most appropriate tags for your work?

Participant S: So...basically Open Calais, ChemTagger, so...Open Calais has picked up nothing, it's picked up general words...one of them is chemistry.wonderful well done. It's picked up two words that I don't even know what they are.

Participant AK: It's probably more useful if its multidisciplinary, if you've got several projects and some are biological ones and some are more physical.

Participant S: The thing is what it's actually tagged is stuff that is incredibly generic and isn't actually about what I'm working on. It's interesting that I've picked up these two things that i don't even know what they are. I would be interested to know where it got these from. It's like...it's picked up various

Participant AK: It's picked up abbreviations and then assumed it knows what it is

Participant S: It's picked up NMR spectroscopy stuff which is great because it does have NMRs in it but it's not what my thing is about, and anything where if you've written up an organic compound it'll have an NMR in it.

Participant J: Yeah

Participant S: So it's not picked up anything that identifies what my thing is. The ChemTagger is interesting, although it's picked up a load of nonsense as it's picked up some measurements, parts of people's initials, postcodes, units...it has picked up the three UPAC names of my novel compounds which is interesting, so if you could take that and convert it to something that is searchable then that would be very useful. Gate Tags is hilarious, carbon nitrogen sodium...wonderful.the ontology tags again have picked up very generic things, so it's obviously picked out everypretty much every method or technique that I've used.

Focus Group Leader: Probably because it's using a chemical method ontology.

Participant S: which is great but going from that to if I was going to go through and find every experiment I'd done then potentially it would be useful but I can't see it searching that..I'd want to know what compounds I'd done work on, or which paper it was to do with. The only words it's actually picked up which have anything to do with what I do is transport and binding. Which is interesting because it has kind of picked those as themes, but then everything I do is transport and binding, so there's nothing I could identify what this document is again. It's also picked up self-assembly which is quite fun, which is related to what I do but it isn't this and I don't think I would have used the word self-assembly in there.

Focus Group Leader: It will pick out the superclasses of things it's recognised in the ontology terms.

Participant S: It's clearly gotten close to a theme but It's not quite there but again I'm not going to search by theme, I know what I do and I want to know which experiment it is.

Participant AP: Funnily enough I do self assembly and it's not on here. Self assembly is one of the most important tags and it's not present in any of these, and the other one is DNA. There's no DNA like in ChemTagger but in OpenCalais I get New York because I wrote somewhere that I did the experiment in new york. As you were saying the Calais tags are so generic they don't tell me anything like...solution.

Participant AK: It's the sort of thing you might put into a literature search if you wanted some really generic stuff, if you weren't looking for anything specific just general knowledge. So if I wanted to know about cyclic voltammetry then I might type it into a search engine to see if i could find anything very broad, but if I'm looking through my own stuff

Participant S: yeah.again if you're doing something on my stuff which doesn't have a name really

Participant J: It might be quite good for people who...it's not necessarily that useful for us, if it was other people looking into a much larger system then it could be a lot more useful

Participant S: Yeah

Participant J: For instance it's tagged mine as organic chemistry, which it kind of is more, obviously is't modelling and stuff which isn't really referred to in this post, but it's definitely that compared to say electrochemistry.

Participant S: Yeah

Participant AK: Mine has definitely come up with electrochemistry tags, but if someone was to search through my experiments then that would come up with all of my experiments.

Participant J: Yeah

Participant AK: But then if all of my experiments were on an online database it might bring up just mine and Participant AJ's experiments but it wouldn't be any more specific than that/

Participant S: If someone is looking for someone who has done an experiment or researching NMR spectroscopy for example then they're not going to be interested in someone who's just obviously run a spectrum because they've used it which is the key difference if you're actually searching for one of those things, you want to know about the technique not just find every paper that's ever used it because it's too ubiquitous and I assume it's the same with voltammetry and things like that. There's a difference in a paper about it or a paper using it.

Participant AK: Yeah.

Participant AP: For me most useful tags were the ontology tags, because they were a mixture of the ChemTagger but also like general things like spectroscopy and there is assembly actually, no self assembly tho. These are the most interesting even though some of them are still very general like process. Which could be in everything like mixture. But in general they are the most useful instead of

say the ChemTagger which okay apart from some tags it's just be a list of the chemicals used. Can be interesting but there are some like NPS that for me is nanoparticles and nanoparticles I wouldn't put them under a ChemTag that I'm looking for. The Gate tags are just the elements. But I think the ontology are like the most...the ones that makes it better. The main things that I do. Even though it is a bit too broad when they do the research.

Participant AK: I found the ontology tags are also not that useful, they're just singular words. For me although the Open Calais ones are really broad tags and not particularly great for searching, they are better than the ontology tags because they pulled up things like cyclic voltammetry, not sure where it got ultramicroelectrode from...but it's pulled out the name of an equation that I've used. Whereas the ontology tags are too general.

Focus Group Leader: The ontologies that I used that I could get my hands on were a certain amount of available chemistry ones that were small enough to run through my system. So these may have been more aligned to other types of chemistry, they didn't look particularly specific to electrochemistry.

Participant AK: It's really hard to tell I'm not sure to be honest. But I'm...the most important part of my document i feel is the aims of what I'm trying to do and then the second part is the summary of what actually happened. But it's only pulled out...it looks like for Open Calais it's only pulled out one and for the others it's maybe pulled out one or two.it would be better if it was the phrase or phrases from the aims.

Focus Group Leader: ***Demonstrates the actions in semanti-cat*** would these have been more useful?

Participant AK: Yeah the actions would have been a lot more useful maybe...would these action tags have said just come back as labelled as heat or pull out the bits that it's tagged? Currently just the words but that could be changed. ***look at the different actions*** none of these actions are right however. It's not accurate but it would be useful if the tag was the phrase. These would potentially be useful.

Participant J: But then if it did that for all of them you'd get everything...like washed electrodes

Participant AK: It would be hard to differentiate between what's useful and what's not. The last thing about mine is I've got a couple of numbers in my ChemTagger but they're not the useful numbers. They are numbers I've written in the text. It hasn't pulled out any numbers from my figure captions or the table and really the stuff in the table, again that's the summary of the important data so that would have been more useful to have. If I knew that I wanted information on an electrode that was 480 microns, then I might tag that in and I'd find the experiment where I calculated the size of the electrode, in which case that would be quite useful. But it hasn't tagged any of the tables or the captions.

Focus Group Leader to **Participant L:** Which did you feel were the most and least useful for your stuff?

Participant L: Ontology definitely.Open Calais had extrapolated out the wrong things, it is talking about chemotherapy...I've never worked with that, the application for my particles are optics, definitely

not that. Even quite far is...so I work only with gold nanoparticles, it went to platinum, silver, can be useful in terms of literature reviews and things like that. ChemTagger got most of my chemicals so it's quite good. Gate got almost nothing. Ontology got...from these ones it can understand it's an experimental procedure that the document is so could be quite good. Just a few random words but not too bad. But definitely the last one.

Focus Group Leader: Can we just have a clarification of the type of chemistry you do...I know **Participant AJ** and **Participant AK** do electrochemistry

Participant J: The document here is chemical modelling. This is referring to quantitative structure, activity relationships, but the actual stuff in here is kind of organic.it's chemical structures but it's to do with transport.

Participant S: Super molecular and transport? That's what I do

Participant AP: Inorganic Chemistry & Nanoparticles

Participant L: Same

Participant AJ (about **Participant AK**): Our stuff is fairly similar

Participant AK: I'm looking at electrodes and how small we can make them, how sensitive they are and what interesting properties we can find.

Participant AJ: I've been developing different ways of performing electrochemical techniques and improving the speed at which you can perform the techniques. So increasing the rate of data recording, high frequency data. So in terms of the tags that it's pulled out...the open calais and the ontology tags are both quite like everyone else they're quite broad, and quite general. It's pulled out things that so for example it's pulled out frequency. Well everything is going to be recorded at some frequency so that tag is going to be in every single page of this lab book, and this lab book is sort of used as a reference document when I'm writing up so that as a tag isn't particularly useful as it's just going to tag up every single page with that. What would be more useful is if it were to tag say the frequency at which it was captured. From that point of view some of the ChemTagger tags are more useful, as it's picked out some numbers and units and things. So when I'm tagging I might want to look at what data did I capture at a frequency of say 500KHz so that would be a useful tag for me to then refer back to. The ChemTagger is also inconsistent so it's picked out some of them but not other ones, and it's not tagged some of the ones that would actually be more useful for me as a tag. So for example it would be useful to tag say the size of electrode that I'm using so I could refer back to when was I using aluminum or when was I using platinum or when was I using titanium and what size were the electrodes that I was using, that's the sort of thing I would want to tag.

2.3.2 Question 2d - How would you have tagged your document if you'd done it yourself? and Question 2e - Are there any obvious tags missing?

Participant J: That's a very good question...in terms of some of the things, I mentioned earlier it's pulled out a whole bunch of elements so I guess parts of it, so on my drawings they are done where they've got a common backbone and they've got an R group which is used when you have multiple

structures and they're pretty much all the same but with different bits on the end and it goes 1 is this and 2 is that. It's basically pulled out a lot of those, I guess that's kind of useful but it's not really helpful because it wouldn't help you search it in any way. What it's kind of missed out on completely is the family of molecules which is in the text, they're called tambjamine and it's not picked that up at all. But I don't know if that's because I guess it's a niche area. I don't know if it doesn't recognise it...erm it's picked up on a couple of the structures in the things but I don't know because it's picked up pentyl group as a tag which I haven't referred to as pentyl group but it does have pentyl written in the text but it hasn't picked up on sort of the benzene ring in it and the other ring systems which potentially would also be useful.

Focus Group Leader: So those are the things that are obviously missing?

Participant J: Yeah, in terms of it tagging the chemical structure, it's pulled out some of the things from the text but for instance, it's got obn but it's just written it out as that, it hasn't kind of expanded on what it means. The other thing it's really missing, what I would have thought is that there are two things...logP which is one of the descriptors and that's probably quite a key thing and EC50 which is another measurement, which are quite crucial in the measurement stuff and they're not exactly unusual measurements in terms of doing transport studies. They're widely used so I'm a little surprised that they weren't there because they're actually written in the text rather than on a plot or anything but I think those are most of the ones, I guess potentially like simulation, as I've said simulating a couple of times. I mean I guess that it's not got a lot of text on it so without kind of inferring what it's about those are kind of most of the things that are obviously in the text that are missing.

Participant S: It's interesting you saying it's missing out on things like logP..How up to date can an ontology be for stuff thats only been published in the couple of years

Participant J: I mean tambjamine fair enough it's quite new. But EC50 and logP they've been around for ages.

Participant S: 10 years though, think about the volume of literature there is and the volume of literatureI don't actually know how relevant the ontologies can be to stuff that is novel

Focus Group Leader: That's quite difficult, the problem is firstly we don't have ontologies covering everything. There are definitely some chemistry ontologies but there are presumably a lot of areas that aren't represented currently in terms of different areas of chemistry, and you've also got the problem that people tend to often make ontologies and then think oh I don't want to use existing ones I'll make my own so you can then end up with conflicts and you can update ontologies and you end up with new and deprecated properties and if you've linked to your ontologies via the web then you should be pulling out the latest version but if you've got a saved copy and you are accessing it offline you won't get the latest version. Mine currently pulls it from the web but that wouldn't work if we were doing this somewhere without internet access so then a local copy would have to saved.

Participant AK: Could you get a feedback loop of people using the system, suggesting tags and suggesting oh I can't find this word...obviously you'd need internet access and then that would go to whoever was managing it and they would go oh we don't have this particular family of chemicals we

should try and add these.

Focus Group Leader: There are collaborative ontology endeavours where people work together and suggest new additions to ontologies based on their needs but it does kind of depend

Participant AK: On how people engage with it and you could have someone being like oh it must be like that and someone else disagreeing

Participant J: I guess to different people things can mean different things, like the way that people use the stuff in the different areas.

Focus Group Leader: Participant S you've kind of already answered this when talking about how you would tag your document, but is there anything obviously missing from this?

Participant S: It depends what I'm tagging it for.

Focus Group Leader: For searching.

Participant S: Who's searching? So if this is a system for me to look through my work then there's a hell of a lot of things I'd put in here.I would need it to recognise what it is to start with, so it's supporting information, which project it's part of, it would need to pick up the compounds that are relevant for me to search by compound, by my experiment number and my name for it.

Focus Group Leader: What do you mean by compounds that are relevant?

Participant S: Literally just the three that I've made, because that's how I'm gonna look for it. All of the starting materials...it's interesting the tagger has picked them up and I wouldn't get rid of that because having thought about it again if you found a contamination or something in the lab and you had to look up every time you used it and search for it that way. But most importantly would be looking up where you've written up that compound that you've made and how you've done it. A list of things that are in it, potentially even just pulling out the headings so I can see it's got synthetic procedure, The NMR spectra...if it just pulled out those i could search for them. That's probably most of it. Which may not actually...these kind of things won't be pulled out by something that is looking for chemistry related words, because it's just going to pick up the things. What the ontology has picked up is that the themes and the areas but I need the practical detail. I'm really impressed now that I know what the ontology tags does is actually that it's picked out that it's binding, transport, self-assembly, it's not quite there but for putting that in a database of documents about certain subjects so you could search on subjects that way it's very impressive, but in terms of actual

Participant AK: I think the ontology tags would be more useful if you were searching and it came back so say like I'm looking for something and I type in electro platinum and glass, if those three tags were all in those documents it would narrow down the number of documents rather than saying any documents with electrode or any documents with platinum or any with glass. If you set it so you had to have at least two or three words for the search to come back that would potentially be more useful.

Participant S: I think it's just, it's obviously picked out tags that are most occurring but it's the really practical detail it's missed. Which is what first and foremost you would be going for is when did I write the synthesis out for this compound, where did I paste this, which plot is it in. And

without the ability to find the information that way...its use as a notebook and a database of your own work is kind of limited. It's good that it's picked out compounds and everything but it's just missing that level of extra practicality which I can't just get from what I currently have, I've got the folder structure where I go Project -¿ Papers -¿ ESI and it's there.

Participant AP: Of course I would add DNA because that's what I use and self assembly, so ontology tags were quite good, in putting together some words like it understood x-ray spectroscopy so like a single tag like electron microscope could be two different words but it took it as one phrase

Focus Group Leader: That's because they directly appear as terms in the ontology so it sees that phrase as one specific tag.

Participant AP: This could be extended when I look in my lab book I normally go and I look for synthesis of some particles. And I may have three or four different synthesis for each and then I have to find all of them and compare them for instance. So I guess if I picked the word synthesis and gold nanoparticles I could find it but it is useful. The thing I look for most often is characterisation or I want to see how have I done it in all my experiments to compare them to look for the parameters. Other than single words maybe the actions maybe would simulate this kind of behaviour.

Participant L: For me it's a bit similar, what's missing is DNA, I wonder if it's because it's in upper case? Or maybe if I will have put better information it would have picked it up I don't know. This is not something I will search in because I do this synthesis now I don't even need the protocol so it would be more for someone new who is coming who I need to teach the synthesis to and if they can search as I tend to use gold nanoparticles but different sizes. You didn't pick anything up about the sizes. Also probably something about temperature as the only difference is the temperature, after I make observations I talk about color so maybe if you could tell me something about this size goes with this colour, and after you realise that yeah that in the end I just finish with how I characterise them so if I can see the technique I use then I assume that would be useful.

Participant AK: I've got the,...it would be really useful if I could have the measurements with the descriptor and the units, it's pulled out a couple of numbers, it's very hard, I looked at them...at one point it's got 10 to the minus 8, and I looked back through and that's not particularly useful, whereas if I could had the descriptor with it and the unit then I could immediately identify what that was rather than scrolling through the whole experiment so if it could have more of that that would be useful. It would be nice if the key aims and key conclusions turned up more keywords, so I don't know if...I've used an ELN program in the past where every section is in a different box so the aims is in a separate text box to the summary and conclusions and that was in a separate text box to the method and I think that changed the way that you searched through the experiments. I'm not entirely sure on that but if you could make the aims and summary separately searchable

Focus Group Leader: I have done some templating prototyping and things can be split into separate sections so that could change how things are searched.

Participant AK: I want most of my tags, apart from my numbers most of my tags would be in my summary because I don't write it up formally so I would potentially put the key findings in the summary including the numbers. But the aims and the summary would have the most useful

information for differentiating between experiments so I'd like those to have the most tags coming back ideally.

Focus Group Leader: Would people find it useful to restrict or specify the bits that are tagged?

Participant J: It depends, cus if it could identify say a procedure like thing or this is like an experiment report it would depend on what's present in it then it might be useful, but where it's more generic notes vs a formally written up thing the structure is completely different

Participant AK: You wouldn't really want to restrict it but if it could show you where it found the tag that might be more useful. Any keyword you'd brought it up and it showed you that it found that tag in the bibliography oh good that's the paper I'm looking for, got the reference now, or that's no good I'm not looking for a reference paper I'm looking for the actual number and carry on going through.

Participant J: I think section identification might be quite useful

Participant S: Yeah

Participant J: Because I do put headings in, this one doesn't really have it because it's just notes but actually writing it up it would have a lot of section headings.

Participant S: Essentially I've got a few compounds in here that it's tagged from ChemTagger that just comes up from in my in the short intro where I've said I adapted this procedure from this paper...in fact that even may only be in the bibliography*searches* it is. So it's picked it out of the bibliography and tagged it as if I'm making those compounds which I'm not so if it could filter stuff like that.

2.3.3 Question 2f - Would you make use of an option to add/remove your own tags/chemicals?

Group: Yes

Participant S: I feel like you'd have to

Participant AK: I'd like to be able to add tags, because sometimes the system isn't going to pick it up but you wanna make sure that it does for future knowledge, on here I have put the date on the experiment but it hasn't tagged it. That would be really useful, as sometimes in my lab book I use the date to cross reference things and also it seems like breadcrumbs.what did I do last year.ok I've found it now.

Participant S: Again, in a formal write up I won't put my experiment number or my shorthand names for my compounds in it so if I want to match up the experimental write up from a lab book to wherever I've then written it up in a document I'd need to add that manually as I won't put the shorthand in.

Participant AK: I've got the date. The experiment number can be really useful if you're trying to search for something.the other...the last thing I've written down here is that so like I said with the ontology tags, you need to know if I search for something, I'm not ever going to search for one word and I'm probably going to put multiple words in and I want all of those words to be in the document

for it to come back as being found at least if it's the first one on the list I definitely want it to contain all of those words otherwise I would get way too many things to look through, I might as well just look by hand.

Participant AJ: I think for this sort of document. This is my own kind of shorthand notes that I'm writing whilst I'm in the lab, it would be useful to tag it by type of experiment and then by materials that I've used to do the experiment or...whatever...so it would have a sort of like...so I'd be able to search it by, when did I last do that particular technique or that particular technique, and then within that, what materials did I use and what did I apply to my electrode, what was my electrode material. I think I would definitely need to add some of my own tags to be able to do that, because you develop your own kind of shorthand when you're writing that kind of thing,. And some of them are standard electrochemistry techniques which it would be able to pick up itself but where you're derived your own techniques they don't; necessarily have a name that you;d be able to search for in published literature so you'd need to be able to put in your own tag to work out what that is.

Participant AK: I suppose you could have an overarching system and then have local definitions and then it wouldn't affect anyone else's versions of their tags

Focus Group Leader: Yeah when we look at the design side of things later this can be made to be configurable per document, and I don't know how much you guys share within groups but then things could be configured on a shared set of documents too.

Participant AK: It would be good to be able to share more easily. I don't know what it's like in other groups, but we don't really have much sharing but I have in the past used software when I did my year in industry where it was very shareable, it was everybody used the same software it was called ELN which isn't particularly specific so I'm sorry about that. But it basically meant that parts of it were tagged so you could search for it, you could make things shareable or not shareable and you could share specifically with other users if you had their username which was quite a useful system. Of course sometimes it crashed then no-one could do any work for the day, everybody went for coffee..

Focus Group Leader: I do have a list of ELNs that I identified *gives list to Participant AK*

Participant S: I will often have to send a plot to other group members, the other thing that's quite interesting in terms of if you could tag things with experiment numbers, when you get to the point where you have the write up and I'm EPSRC funded so all of my data has to go online, so if I can go back and if I've got the list of experiment numbers to go with that and just bring up all of the data that went to make that document and being able to work back from raw horrible numbers through to notes to the write up at the end would be useful.

Participant J: I guess some of that, potentially, you say your stuff you save it, you save it in a hierarchy by file, so if I guess you were taking the document off your computer and it was saved in molecular/project/. It could potentially pick out like some of the structure from that rather than you having to manually do it all yourself, but obviously that's not in the actual text of the document.

Focus Group Leader: One of the logic of doing it on top of an existing cloud based notebook was to facilitate sharing.

Participant AK: Looking at this list (my list of ELNs) the Accelrys and Contour are both sticking out in my mind so it could be one of those.

Focus Group Leader: Contour was iLabber and then was acquired by Accelrys, and now it;s by Dassault Systemes. So maybe it's this one?

Participant AK: Maybe that does look familiar.

2.4 Question 3 - Search

2.4.1 Question 3a - When searching for a word in the search bar here, would you expect the search to prioritise results by: Existence in the title, pure term frequency or whether it was tagged as an important term for that document? and Question 3b - Do you think these advanced options (by title or by text) are actually useful? Would you use them and are there any other advanced options you'd make use of? Or would you prefer to only use the search bar with no customisations? and Question 3c - How would you expect this search to work?

Participant AP: I think I would like it to be in the title first, because if I write synthesis probably in my document I have it 300 times but I don't...but I care if it's in the title because maybe I am looking for the synthesis of something.

Participant S: Yeah

Focus Group Leader: That's interesting I've had conflicting opinions on the title being important

Participant J: I suppose if it has appeared in the title then it is something that you care about

Participant AK: I would say the opposite because I haven't got a title I've got an experiment number and all of my subheadings are generic

Participant J: But if it didn't appear in the title it wouldn't appear at the top of your search, so if you search for glass plated electrode, if it was in your title it wouldn't return any results then it would go down to the next section.

Participant S: Yeah Participant AKL I suppose it comes down to what are you searching, are you searching through writeups for a paper or are you searching through write ups of your lab book

Participant S: If I'm searching for one compound synthesis I'd want it to pick up everything that has mentioned that compound and then find the thing that's got synthesis in the title

Participant J: i'd probably say potentially like title but what would be quite useful but you know when you search and there's like filter, A-Z, highest to lower, lowest to highest, if you could be like filter by frequency.

Participant AK: But sometimes it depends, you know what you're looking for when you're looking for something most of the time or at least you have a vague idea of what you're looking for so if you

could decide during a search what you were prioritising.

Participant S: The thing is I wouldn't be searching for one type of thing I'd be searching for a specific thing.

Focus Group Leader: Currently this facilitates just searching the title or the text

Participant AK: Could you put prioritise like I prioritise title then text, rather than one or the other

Focus Group Leader: It could definitely work like that, part of this group is to understand what users expectations of the search function is and how they would want it to work.

Participant AK: Date is really important to me I often track things on when it happened

Participant S: I've never written a date in my lab book

Participant AK: Sometimes I call my experiments...the last part of the filename will have the date in it because I can then go back to that day in my lab book and say this was the day that I collected all of this data with these conditions

Participant J: Does it actually tag it in terms of if it were like tagging it so obviously it's got like these ontology tags, but would it tag it if it was doing dates it would have the tag type would be a date

Focus Group Leader: It could do

Participant J: Because that way you could be like, instead of searching for a date in th text somewhere, you could be like find me all the documents that are tagged with this date which would probably bring you back all the experiments, I don't know but then there's another search thats like if you're referring back to the experient like on the 14th October I measured whatever

Participant AK: As long as you didn't tag it with the date that the document is created, as I refer everything to the day that the experiment was done rather than the day I write it up and it might be a week later that I write it up or I may have created the document in advance and started filling it out in advance in preferpartion for a day next week when I'm doing the experiment, the important thing for me is the experiment day.

Participant S: I think thats key tho is that everyone works different so for me.experiment number is key and that might be four or five days...it doesn't matter what day...might be three or four days, and I'll have one page of one things that I did in the experiment. It might be a three day synthesis and some people will start a new page every day and do different experiment numbers, but for me it's one experiment for one thing so you should be able to set the dates for how you work.

Participant AK: that made me think of another thing, if I'm doing things over several days in my experimental it will say the 14th September to the 21st because that is the date range so it would need to bring up potentially more than one date. It would be good if you could accept and reject tags as you go along in case they get it wrong.

Participant S: Yeah, it's when it's tagging your stuff in the title, like if I search for synthesis I want it to identify that are a synthesis as opposed to where the word synthesis is in the text, so that if I then put compound plus synthesis the first thing to come up is the thing that is the synthesis of that compound

Focus Group Leader: It sounds like you guys would want to have multiple combined searches

Participant J: I guess it would be things you'd want it to be able to do like must have all of these words, or this and this

Focus Group Leader: So boolean operators?

Participant J: Yeah the stuff you can do in google if you really know how to use it

Participant S: Yeah, the thing is you aren't going to be looking at something just with a keyword and filtering what comes up, you're going to be looking for something very specific. This process with this compound it should be able to find those.

Participant AK: WHen it's your own work you're never searching broadly you're always searching specifically.

Participant J: I guess it's how you search potentially on sciencedirect when you're searching for papers, you can put author and category and if you really go into it you tick boxes and stuff and then get it to search but it does the same sort of thing but if does it off of that

Focus Group Leader: What would be the most useful thing for you to search on?

Participant S: Compound. I'd want a library of my compounds that I've made, if i could click on the pictures of my compound and it brings up the synthesis of it, the characterisation things, then the...every experiment that I've run on it, all in one place that would be...I would more likely use that than a search I think, I would drill down through all of the stuff that I've done with it.

Participant AK: I can imagine that's quite representative of most organic chemists, most of the times they are looking to synthesiseis that fair? But the pictures, normally

Participant S: It's very specific to me though...I make a compound and then do experiments on it. Other people will be just making lots of different compounds.

Participant AK: In which case the picture with the functional groups is really important spo for organic chemists that would be really great. For us I don't know what we'd draw...

Participant S: If you'd manufactured an electrode could you then go...in the same way if you search for I did it with this electrode of this size, would you be able to drill down.

Participant AK: thats words though, you wouldn't need a picture for that, you'd just be doing a straight word search

Participant S: You want to find the work you did on a specific thing.

Participant AP: Procedure, like normally I look for different procedures, so my procedures could be synthesis of compound x or cyclic voltammetry

Focus Group Leader: Also would you guys ever find it useful to search for stuff in your images? Which is hard but Google has a level of image recognition, although I think it's quite generic.

Participant J: It would potentially be useful so depending on how good it was you could almost have an include images text box and you could search it and see if it comes up in the text and then go include images and see if it comes up. It would be lovely if it worked but I remember to find the issue that every paper publishes with a structure that can't be read by a computer.

Participant AP: Could you not look in the captions

Participant J: I assume they do that potentially anywhere depending on how the captions are done.

Participant AK: It hasn't picked up stuff in my captions I don't think.I guess a lot of the stuff in my text, if it's in the caption then it's normally referred to in the text.

Focus Group Leader: It has pulled out a couple of things in the text, but at the moment it's looking at the .txt file version that it downloads from Google Drive it will only take things for the captions if they are included as separate text. So it has pulled out a few things from the caption of the image here but for example wouldn't get this title here as it's clearly part of the image.

Participant AK: So the legend and parts of the title might be quite useful but the shape of the graph isn't useful. This is a very generic experiment anyway and it's not shown anything novel.

Participant J: Whereas mine has, my top image has got structures and it's picked up a whole bunch of those which I assume is potentially because they're quite easy...it's text in the image so OCR for that would be quite decent.

Participant S: If it could be ChemDraw intelligent that would be great.

2.5 Question 4 - Markup

2.5.1 Question 4a - What do you think of the markup? and Question 4b - Are the tooltips useful? - most specifically to see the ontology descriptions? and Question 4c - Would you make use of the ability to edit the tag / chemical descriptions yourself? and Question 4d - Do you have any suggestions for a better way to markup the documents to make it more useful?

Participant L: the information about the compound would be nice

Participant AK: If you searched for separating molecules it would be good if it picked up on the tag of centrifugation.

Focus Group Leader: If you wanted to search through the markup of the document rather than the initial text you were given it could include things from the descriptions that are associated with

these tags.

Participant AK: It depends

Participant S: I know what centrifugation is if I've written it

Participant AP: But what if someone else is looking at it

Participant J: I suppose you could potentially choose whether to show them or not

Participant AK: I like seeing the markup because then it feels like it means that you can see what's been tagged and I think that can only help you...sometimes when you're searching for something the problem isn't the computer the problem is you not knowing which what words are the most important to type in so to search more smartly that might help you

Focus Group Leader: *** demos how you can change tag descriptions in the document***

Participant J: For instance the first four of those are quantities and measurements and some of rest are elements

Participant AK: I would change those and give them descriptions like maximum observed current or something like that and then later on I would rename the 2-Platinum and 1-Titanium etc as electrode.

Focus Group Leader: So these could be changed entirely if you wanted, and you can turn on and off what you don't want to see

Participant J: Something came to mind about how Mendele does automatic processes of papers and stuff and it comes up with all of the tags and sometimes it tags it as needing review and stuff and you can go in and be like oh no actually it's not really this or change the stuff down the side. Especially if you were doing experiment number and stuff and if you could go in and see it or potentially edit it like if you were like oh experiment number I think that would be super useful. Or if you could tag it like...later on I need to come back and have a look at this potentially. Then because obviously if it pulls it out if you have the time you might want to go through and check it but if you don't.

Participant L: The tip thing is not for me, like I know what centrifugation us, but if I had to teach someone so in my PhD I teach physics students to do chemistry so this would be super useful because they have no idea what some of these things are so they could just go in and have a quick definition. So that would be really good.

Participant S: There are vastly more useful things than that if I'm being brutally honest, just why...if it picked out my compound names and then do a tooltip that linked to all of the documents that had it in that would be amazing.

Participant AK: Or if it picked up the name and gave you a picture of what the structure looked like

Participant S: Vaguely yes

Participant J: It could be useful, it could also pull up...I know obviously for a structure if it exists

it depends if it's tagged with the name or the structure, because if it's tagged with the name then the name needs to be converted to the structure and that's not necessarily good conversion but then that goes away and say searches with ChemSpider.which may be what ChemicalTagger does. And then it could get properties and do like basically how much it weighs

Participant S: Yes

Participant J: And calculated stuff, it could bring back the name or other synonyms that it's known as

Participant S: if it could go into my inventory and tell me what cupboard it's in, that's what I want

Participant AK: That would be so useful

Group: Loud agreement

Participant S: The amount of times I've come out on a postit note with the name of a compound and some data and where I can find it

Participant AK: What would be quite cool is if you could somehow link your batch numbers to your experiment because sometimes you go through your experiments and you find that this is exactly the same but it's come out with a different result and you find out that the batches aren't exactly the same chemical because there's some change in impurities in the chemical so it's really important to track batch numbers when doing experimental work so it would be really cool if we could link that in somewhere as well. On here when I write out my materials used, I also write down a bunch of numbers, one of the tags did come back with those numbers which was quite good. But that's to do with the chemical and the batch so if you could link with it's CAS number?

Participant S: Yeah

Participant AK: That would be good

Participant AJ: We've got a lab inventory list that has all the chemicals that we have in our lab so if we could link it to that

Participant J: There is a system called quartzy

Participant S: Yeah...if there's one thing that makes me grumpy is having to take my lab coat off, throw my gloves in the bin, log into quartzy...and again if I was using an ELN at the time and I could just hover over something I wanted to use

---

2.6 Question 5 - What do you think of the current design of the system? Does it look like the content is well represented, does it seem useful, how would you have expected or wanted to see the content from the backend represented in the backend?

*** shows the different front end sections ***

Participant J: Can you click on the tag and go to where it is in the document?

Focus Group Leader: Not currently but you could

Participant J: Because if you could say click on cyclic voltammetry or it depends on the type of tag so for instance if it was argon and you could click on it and it would say 1 of 4 and then it would go to the first instance of it if it's actually tagged something.whereas if it's tagged something like cyclic voltammetry it might just be related like it thinks this document

Participant AK: that might be going a bit far as you could use ctrl f

Participant J: Or if you could right click it and find it in the text. If it was highlighted in the text you'd expect the text to be hovered over or clicked, potentially for something to happen. But if you wanted to see everything then having the bar on the side is good.

Participant AK: I quite like it how you can see what's been tagged in different colours, I would like a little more difference in colours as I think it's really difficult to differentiate some of the blues and purplesbut by the side is really useful because it means that you can still look at the document, it doesn't get in the way of doing that.

Participant S: I'd want a tooltip, bearing in mind comments about how useful these tags actually are, I'd want to see it with experiment numbers, compounds used, stuff that's actually useful information that you can go straight to

Participant AK: Would you want to know which ones have been made by the ontology and which ones have been picked up by the others because obviously the ontology could be several degrees away from the original word

Participant S: I want the key specific details, the practical details

Participant J: I guess the alternative is if you can turn them on and off, and if you have them together and sorted and if at the bottom you have an option to turn different types of tags on and off. So if you did add hierarchy like experiment number, project type, those would be a different class of tag

Focus Group Leader: Here are the options that allow you to turn off the different tag/chemical channels, so you could turn off everything, you could turn off the hovers, specific tagging areas

Participant J: Then if you could say you only wanted to search by experiment details then if that was a class then you could just be like thats the only one

Focus Group Leader: So you'd want to turn them on and off in relation to searching and what's displayed?

Participant J: Yeah, I think so, I guess...depends how easy it is to do but obviously if they are classed in certain ways and that is in the underlying structure. I was wondering when you were saying about doing it for different documents like if you could have different classes of documents, like almost apply a tag template to it. Say this is an experimental write up. Because then if you personalised it all it

would be like excellent this is exactly how I want it for my procedures you wouldn't want to have to reset it when you made a new document.

Participant AK: going back to what you were saying about practical details, would it be possible to have another pane where it separated the tags into numerical value, experimental tags, that kind of thing so that you could almost...if I knew I was searching for I dunno the mean diameter of the electrodes I was using then I know where to look in the tagging sections. But still also have the open calais, chemTagger, gateTags, ontology.

Focus Group Leader: You could add as many panes / separators as you wanted and that in itself could be customisable.

Participant J: I guess you'd want to separate into stuff like quantities, measurements, chemicals, processes, like procedures, and then kind of more document tags potentially I dunno about any other ones

Participant S: Is' about separating those things that some outsider would want.

Participant AK: So it's not like you're removing any of the tags you're just arranging them in a more user friendly way so when I came to look at them I could see what important ones were there.

Focus Group Leader: So does the current toolbar / side pane system work?

Participant J: Yeah

Participant AK: It's unobtrusive which is good

Participant S: It's an interesting point in that I would need the complexity of a full working office system. If I'm in word and I'm producing a formal report for all of my origins or all of my ChemDraws will stay embedded until the absolute last moment so that I can still edit them and that would be very annoying if I couldn't do that because getting up the files is annoying.

Participant AK: You could still do it in word and then upload it and convert it into docs for your backups

Focus Group Leader: This does facilitate drag and drop. The idea behind the domain layer of this software is to try and include some chemistry specific packages where there are open source versions available.

Participant J: Does google docs do...so you know how LaTeX pulls in images and if you edit the source image it edits it in the document, does Google Docs have the capability to not put a JPEG in but put a linked image in as if it did you could say if you wanted to edit it do like right click and open in origin and then it opened it up in the program that you have installed on your computer, and then you can save it and the new updated version would appear.

Participant AP: You can do it in microsoft with powerpoint.

Participant J: I know you can do it in word if it has the special things specially for chemical structures

Participant AP: for powerpoint sometimes what I do, I copy paste the graph from origin to powerpoint and then I if you click on the graph in powerpoint it will open it in origin

Participant J: Does it open it the graph within that powerpoint rather than opening where you've got it which are two very different things

Participant AK: but you do wanna keep them separate because you don't want to accidentally edit your original and realise you've accidentally deleted everything.

Focus Group Leader: Do you guys have any other comments about the front end design

Participant S: For the tag descriptions, It's kinda hard to say given that the tags themselves aren't very useful

Participant J: Some of the tags, there are loads that are all marked chemical elements whereas not all of those are actually chemical elements, some of them are chemicals, but for example chloride is tagged as chloride rather than being tagged as an element whereas alumina is actually a compound.but it's useful to see them all like that I think, obviously when you get a really long document though this could get quite difficult

Participant AJ: Maybe this would be better ordered in a hierarchy so you have a group and then the tags in that hierarchy

Participant AK: Like you have all the chemical elements at once and then have all the other tags in different groups

Participant J: but I assume if it's all there in the underlying stuff then you could have different settings for how you wanted it to look

Participant S: yeah or you could turn off the elements.

2.7 Question 6 - Previous studies have suggested that scientists have a greater tendency to use software for writing up papers/reports/thesis rather than during their lab experiments to take notes, what do you think of this?

Group: Yeah

Participant AK: I think it's starting to move that way though, a lot of people are having that conversation of how do we move to store our data electronically, and how do we move so that we can do writeups in the lab, I know that our group has talked about having a tablet with handwriting recognition.

Participant AP: like a tablet with electronic ink?

Group: Yeah

Participant AP: I would find that useful

Participant S: I don't take anything into the lab where I would be upset if I threw it in the bin at the end of the day,.

Participant AK: I don't like things that go between office and lab I like everything that's going to be used in the lab stays in the lab and anything that's going to be used in the office stays in the office. I don't even take my pen in.

Participant S: I have lab pens and desk pens. I don't take my phone in...I have sometimes taken it in my pocket and if I do I have to take my gloves off.

Participant AK: I try not to though because you don't know what you've touched, you don't know what you touch in general life but in the lab you know there is a high risk of something, and the last thing you want to do is put something irritated next to your face.

Participant J: So I did my placement I did the implementation of an ELN and they had...it was a pharmaceutical company so firstly it was highly regulated which is good for doing ELNs, so they always have to use the same template whereas untemplated work is much harder to put onto ELNs because it doesn't recognise the stuff. They were able to access it from multiple things, and with Google Docs you can obviously do things via your phone or mobile device, and they had waterproof tablets you could wash and disinfect that only stayed in the lab and when you were outside the lab you could access it on the machines.

Participant AK: We talked about putting ours in a ziplock bag

Participant S: The way I sort of see ELNs being useful is more as a filing system I think.

2.8 Question 7 - Would you be more inclined to consider using an ELN if:

- It was aimed at that stage of the lab process (e.g for writing up reports / data analysis etc).
- It semantically tagged your work?
- Facilitated a more accurate search across your work?

Participant S: Yeah, when I'm working in the lab I just need to be able to scribble things down. But if it could be the overriding way of sorting things so again you'd go under an experiment number, you could scan in the scribbled notes from the lab and you'd be able to find all of the things that related to that experiment, the data sheets, the characterisation, whatever you produce and be able to sort it in one lace that would be something much more useful. The one I used before it was just a replacement for paper and it's taking something that works already and making it harder.

Focus Group Leader: So I'm more interested in improving the current software offerings and hopefully trying to increase some of the digitisation but I'm not looking to entirely remove paper as that doesn't seem that feasible right now

Participant AK: For me an ELN should have three parts, a section for me to write what I would

like to have happen in the lab and for me to write up almost like a recipe for what is going to happen in the lab because I like to go into the lab and follow what past me has told future me to do, so future me or present me knows exactly what I need to do as I'm often half asleep.and the second bit is any supporting notes, anything that comes up along the way, any quick observations that come up along the way that I just want to note down and go back to later, and the last part is the bit where I write up properly including maybe central literature that I've used to help me decipher what's happened and that would be more like what this final thing looks like. For me I'd need three separate parts and they all need to be linked so that they don't get lost.

Participant J: I think for my stuff, a lot of the most important things...there are so many different document types and everything it would be how it was tagged so if I could pick out a compound or project type and see the experimental stuff, the background information, the reports, the presentations that type of stuff, even if they were in all different formats but you could see how they were tagged and pulled together especially for trying to disseminate information...if you could pull it all together rather than having to manually trawl through all these folders, that's probably what I think would be most useful. And if it could tag it automatically it would save you a lot of time.

Focus Group Leader: So would any the features here encourage you to use this system

Participant J: It would potentially be good as a starting point, as it if brings out tags that some of them are useful it saves you some of the time, but obviously you'd have to use it to know which kind of ones were useful, because then you might have a whole section of the ChemTagger ones that aren't great but you might want to add other things

Participant AK: If it said it had no ways of tagging or searching then I definitely wouldn't use it, so if it had any sort of searchable function then it would make it more likely to use it but also does depend on how good it is.

2.9 Question 8 - Do you think that semantically tagging your documents would have any impact on the efficiency of your work?

Participant J: It would make writing things up a hell of a lot easier I think as if you're trying to search for stuff or pull data together, or potentially if you could link between like say you wrote on the 14th october I did this and say click on that to take you back to the bits to the previous experiment document rather than having to be in your lab book and be like...14th...where is that.

Participant L: Or if you're in the library and you don't have your lab book with you having this would be useful

Participant S: I wouldn't underestimate how quick it is to flip back through a lab book until you see the structure. 5 seconds

Participant AK: Yeah but your structures are big and obvious, a single word or a single number, would be much harder.

Participant J: Or if you're trying to find multiple instances to compare between them.

Participant S: On tags, these tags, no

Focus Group Leader: And if it was the tags you wanted?

Participant S: Yes absolutely

Participant AK: It needs to have tags, you can't get around that

Participant S: You can't rely on general words

Participant J: It needs to be able to learn...sort of thing.

Participant AK: It would be cool if it could do that, especially if you choose your own, you'd put in your own tags but then it could pick up on that and start doing that, I don't know how plausible that is, just don't make it too smart

---

2.10 Question 9 - Would the following features encourage you to further digitise your work or add tags of your own?

- Semantically tagged documents
- Stronger search

Participant S: it would have to become that all encompassing project management thing, so if it did have everything in one place where at the base level you had every experiment and everything associated with everything to do with that experiment and it was findable and searchable and stored in the same place and you could go bigger and look at the project and the reports to do that and it would link back to below and it could link in with inventory and things like that then yeah I would bite your hand off. Being able to save time searching through different folders in my computer for different things.

Focus Group Leader: Part of what I'm trying to understand here is is it just a case of we haven't nailed the right application and functionality yet or whether it's a case of even if we did that wouldn't make a difference and people still wouldn't use it.

Participant J: No

Participant S: No Participant AK; I'm definitely looking to keep more stuff digitally and less stuff in a lab book, just because it's digital, it's so easy to store multiple copies, I pray every day that the chemistry building doesn't burn down but if it does this book would be lost but the ELN wouldn't be and I would definitely consider using this probably more to back stuff up because I do...as well as the filestore I use space on my OneDrive and also space on various Google Drive accounts that I have so I would definitely put it into Google Drive and if it had this tagging I'd say oh great, upload and convert to docs cus then I can search through docs much more easily

Participant J: I'd definitely use it for the tagging part as I already store most of my stuff in Google Drive or Dropbox, as it's so easy to share and it's backed up and stuff and you can access it anywhere,

and I've got so many different document types that even just the ability to have the metadata tagged about that and searching because especially if it didn't require you to change the method with which you input stuff so if you were writing your stuff in largely the same format as you were, then what have you got to lose, you haven't got to change anything of what you're doing it just gives you that extra

Participant AJ: Definitely having it searchable by tags would be really useful, if I'm say writing up and thinking about a particular experiment I did 18 months ago, trying to trawl through all my data...I've got a sorting system but all of my data is stored in different files and things and it's obviously got its own sorting system, but if I could search by tags in an electronic lab book it would be a lot faster to find that data. And as I say at the moment all the notes I take in the lab are all written in a duplicate book and then I've got to manually make sure that the copies are kept somewhere else and having it online it would be much easier to make those copies. So that would be really useful

---

2.11 Question 10 - Do you have any further comments, and is there anything that could persuade you to use an ELN?

Participant S: Please put quartzy in it

Participant AK: That would be a really good idea

Focus Group Leader: Remind me what quartzy is

Participant AK: it's an inventory thing, you can order through it and you order chemicals through quartzy and then see and it will automatically save stuff about that chemical

Participant J: I think it's aimed at chemists and stuff but it's not prescriptive

Participant AK: we use it for our inventory

Focus Group Leader: Any further comments?

Group: No

3 Physics Focus Group

This focus group was conduced on 21/11/2017 and comprised of 4 physicists (Participants A, B, AL & AM). The following colour key is used for the Participants and Focus Group Leader:

Focus Group Leader
Participant A
Participant B
Participant AL
Participant AM
Group

3.1 Introduction - Explaining the purpose of the system

Focus Group Leader: I have been looking at electronic lab notebooks for my PhD thesis, I did a bunch of initial studies, I looked at what was out there in the market, I observed people in the chemistry labs and I did some initial focus groups with physicists chemists and biologists, my interim conclusions were that there's a lot of good reasons to digitise your lab data, there's still a lot of reasons why we should be looking to do that, and that people do actually go and use quite a lot of software for what they're doing. But ultimately there's still this kind of barrier to actually using electronic lab notebooks in the lab, and there's quite a lot of barriers trying to replace paper, so I'm not really looking to try and create a system to replace paper because I don't think that's particularly feasible right now. But what I am looking at is trying to improve the current software offerings of the software that you would use if you were actually writing things up on a computer, and hopefully by improving those we might encourage people to digitise their work further. So, my initial thoughts were when you look at everything, ELNs aren't very popular even though there are loads of them, but electronic notebook software like Google Drive, OneNote, Evernote etc is a lot more popular, and if you break down the features that you would need from an ELN and a EN there's actually a hell of a lot of overlap. For an ELN you need everything you'd have in something like Google Drive and OneNote and then you'd have a domain layer with domain applicable tools, and then a further semantic layer on top that handles metadata and tagging.

***shows diagram ***

This is a diagram of the breakdown of the three different layers, being everything here is generic notebook functionality, this would be domain based stuff, and then at the top is the semantic layer which looks at tagging your notes, doing a better search, linking things together, recognising chemicals etc. The bit I've focused on making was the semantic layer. It's all well and good for people to say that they like their documents to be tagged and made better to search, and to have more information about them, but no-ones actually given a great deal of information about what they actually want and how they want it to be achieved. This is a proof of concept system that I've designed. I'm going to show you this in two parts. There is a backend part of the system that shows a representation of what can be done. It shows what tags have been pulled out for each document, which bits of the documents have been marked up, what ontology terms have been recognised, and where different chemicals have been identified. The front end of this system would be in Google Docs for this proof of concept idea, and so you will be shown a mockup of how this functionality could be implemented on the front end.

This focus group is to look at both aspects. To investigate how well have the documents been tagged, how things have been marked up etc and also what you guys think of how the front end looks and how you think that would work, and to talk a bit about ELNs.

3.2 Question 1 - Documents

3.2.1 Question 1a - What type of document have you brought and how typical a representation of your working documents is it?

Participant AL: its just a page from my log book, its very standard of me, its just a sketch and a quick description of what i was doing that day, nothing really worked, there wasnt anything interesting so its quite basic.

Participant AM: I got a...what seemed like an undergraduate lab notebook discussion and conclusion about ohms law, but it seems to be the kind of discussion that you would expect to write in a lab book, I did similar experiments, I found these results, these were the errors and this is what I would do when I do it...that kind of thing.

Participant A: mine is the biggest paper Ive worked on during my PhD and its very...I picked it because its big and its very specialised and its both a experimental and Participant Bretical so I was wondering how your software would deal with that with regards to the tags...and I gave you...I was quite interested to see what it would do with a pdf / LaTeX file.

Participant B: So mine was a sample report so its not something that would be produced on a daily basis but after a run of experimentation youd make one of these per sample and its...its not particularly obvious terminology that is useful for this type of tagging to flag up so words that it...words that havent been flagged up necessarily by any of these or by the combination of all of them, would have been interesting to see if it managed to pick them out but maybe thats something to think more about in future.

3.2.2 Question 1b - How much of your documents are usually made up of words / pictures / diagrams?

Participant AL: Probably in terms of space 50/50 text and diagrams. I dont sketch data images I just keep them on my computer and reference them in my log book, I dont like writing that much so its just shorthand

Focus Group Leader: out of interest would what got written in your lab book go into some software, like recording your daily work in Google Docs for example?

Participant AL: No

Focus Group Leader: what from your lab book does get recorded on the computer?

Participant AL: Pretty much nothing, this is a record of what I did and it references the data that I save and I just collaborate between the two when I need to write something more official

Participant AM: Mine was 100% text, but generally it would be a combination of text and mainly scribbles and doodles and proposed experiments and stuff and tables of results.

Participant A: I would probably be similar to Participant AM, mostly text, explanations, writing stuff down, but in terms of records I would keep lots of folders with raw data sets that have been saved for a rainy day but just kind of stand alone and then all the information is in the title. In terms of meshing them together when I write stuff up in my lab book Im really lazy Ill just write it down and be like look at file duh duh duh, I wont put them together

Participant B: sort of 50/50 but the caveat there is that on a daily basis in my log book I can have weeks of every single entry is just images with very sparse description and then I can have another set of weeks where its pages and pages of just text: instructions, how I did this and how someone can replicate it. There is usually a middle ground where its a 50/50 document of here is a bunch of images, detailed explanations...its more of a here is a bunch of images, next page here is a whole bunch of text

Participant A: yeah, ad hoc

Focus Group Leader: when you write stuff up, presumably you caption your images? But Is a lot of your info associated with the image in the image or is it in the figure legend? Yeah...this one that Ive brought is pretty written up, its the kind of thing that I might put in a thesis rather than in something that I wrote in my lab book or even had written up for the same purpose. This sample report is quite detailed to the point where I really wouldnt have gone to this much effort, it really would have been the same plots, the same kind of quality of data but really not adding quite as much text around it. It would have had captions to explain it and then some more details, but yeah definitely not the three pages that this has ended up as, it would have been more like 1 and a half.

3.2.3 Question 1c - What formats are your documents usually in when you create them?

Participant AL: If ever Im writing I would always write in LaTeX as a practice, if its on paper its just a stream of consciousness, here is what I did and a sketch of what happened and what it means.

Participant AM: for me it depends on the type of experiment Im doing. If Im writing a more efficient report say a write up to give to someone else Id do it up in LaTeX and make all the graphs and figures captioned well and do the tables correctly but most of that would be transcribed from notes that Ive made in a log book, but Ive started to use OneNote a lot to try and convert...its difficult because sometimes you do an experiment you wont necessarily have a computer with you and you scribble notes down with you and maybe write those up into a OneNote file and they are quite good for collaborative work as well. As you can share them and they sort of do drawings and stuff in them, but if I ever want to draw a diagram or write an equation but Ill do maths and equations and diagrams with pen and paper and if I can remember later to transfer them over to OneNote then I will.

Focus Group Leader: Thats interesting as some of the later work in ELNs has been to trial OneNote as an ELN, this contributed to looking at all of the overlapping features of ELNs and ENs as over the last 5 years or so generic notebooking software has been trialled as an ELN and equally a lot of the features that one would want from an ELN appear in ENs. However one of the things that did come out of this was that scientists liked the collaborative side of things and they liked that it was a decent interface but they were missing the domain knowledge side of things, I guess in your case that would

be the ability to do the graphs and handle equations:

Participant AM: Yeah Ive been looking at buying a I think it hooks up with OneNote, essentially you write on a piece of paper with a special pen and it reads your handwriting and translates everything youve written into OneNote or a PNG file so that you can stick it in there. I find OneNote really good, screen clipping tools, you can copy and paste images over. Im always a bit hesitant especially with the nature of my work now using Google Drive as that's stored on the cloud

Participant B: In breach of European economic region

Participant AM: Yeah

Participant B: problem I have

Participant AM: People doing PhD thesis and stuff. From what I understand the legalities of that, because its out of the UK and EU, its somewhere in America. Its the same reason that you cant save your results to dropbox technically, that would be a hesitation of using an online platform at least with OneNote you can save a copy directly to your desktop and send it through a collaboration on shared server which is better.

Participant A: Throughout my whole PhD I cannot think of a single time that I used anything other than LaTeX, or NotePad or my lab book, problem with using something like word is if I want to write up any equations I use them pretty often...it would take me ages so I tend to do them in LaTeX. Ive been using zotero for all the papers I have and they were all PDF and there are well over 300 of them, so for me one of the most useful things for tagging would be tagging PDFs so any system would need to work well with PDFs as I use them so much.

Focus Group Leader: Its an interesting subject divide as typically all the physicists I talk to only use LaTeX whereas the biologists only seem to use word and the chemists are mixed but more word.

Participant A: Well can you imagine writing something like that in word.

Participant B: Things of this format...I . Throughout the PhD Ive mostly used word because whenever Ive had anything heavy duty it was LaTeX but that was just for me, but whenever it was a notebook or something like that it was gonna be collaborative and most people dont struggle with Word so I was kind of bending to the slowest member. Yeah..more recently Ive been using Jupyter notebook. Jupyter notebook is pretty damn decent, particularly because its not pretty much all the features youd want from LaTeX plus python plus HTML plus will just generate a PDF from LaTeX, but I find it a little bit clunky and it still doesnt do on the fly diagrams so thats the main issue I have with it and why I dont use it all that much but Id certainly consider it better than trying to write..writing up something like this document I guess I would use Jupyter notebook over LaTex and definitely both of those over word.

Participant A: Just to add...because youve just reminded me. This paper I had to write in Overleaf which is an online LaTex collaborative environment. Its so useful because the other person writing this was in Russia and it wasnt useful to just send stuff across.

Focus Group Leader: Moving onto the tags. These are the tags that have come out of different

systems (NB: Each participant was given a copy of their personal documents and the tags as well as being able to see it on the system). OpenCalais looks at wikipedia articles that have been tagged and looks at what content has been tagged to try and infer how other articles with similar content should be tagged. These are categories of your documents rather than things that have been pulled out from the text. ChemTagger and Gate are the automatic chemical recognition pieces of software so they have pulled out what they think are chemicals in your pieces of work. ChemTagger also identifies actions (gives example showing the marked up actions in backend). As you can see Gate typically seems to pick up less than ChemTagger. For the ontologies, ontologies are basically hierarchies of relationships and concepts that we use in the semantic web so Ive currently just got 6 ontologies in here, some chemistry physics and biology ones. This piece will pull out the matching ontology terms. Your documents have been run through all of these systems and currently pulled out everything it can find, and part of this software evaluation is looking to refine this process to figure out what things we should actually be pulling out. So on that note if you all have a brief look at the tags for your categories.

3.3 Question 2 - Tags

3.3.1 Question 2a - Which of the tagging services (if any) do you think has the most appropriate tags for your work

Participant AL: the tag glass is not great because that will appear on every page of my logbook. The chemistry tags have just tagged a lot of molecules which I guess could be useful but I only work with a limited range of them so...yeah...I dunno if Id ever want to search for silicon dioxide because again it would be on most pages.

Participant AM: The open calais ones describe it quite well, they sort of bring out everything that I would think about wanting to search for or you know..associate with it. The ontology tags, for it being a discussion and conclusion of the experiment theyre quite non specific and vague but on the other hand I cant think of anything off the top of my head in the text that would better describe it.

Participant A: Its tough...because the ChemTagger got the abbreviations which I would have wanted that everything else missed. OpenCalais got two really fundamental ones and Ontology was just generally good. There are some hyphens Im not sure where theyve come from.

Participant B: Can it not detect new lines

Participant A: Maybe

Focus Group Leader: I think this is more because I ran it through a PDF reader so it may have hyphenated words at the end of lines. I converted the PDF and then ran this on the .txt file based on what types of files I could send through to the APIs.

Participant A: fair enough.

Participant B: the most useful seem to be ChemTagger. Open Calais is very generic, as is Ontology.

Open Calais reckons this is an astronomy paper.

Participant A: Oh yeah me too

Participant B: This is odd...

3.3.2 Question 2b - From the perspective of automatically recognising the chemicals you have used, do you think ChemTagger / Gate have done a good job?

Participant A: Gate tags were a bit rubbish, very much the least useful. In the Chem tags it has...its got loads of bits of the equations which isnt very useful, so ideally if you could tell it to ignore the equations that would be very useful. ChemTagger got more of what I wanted.

Participant B: Gate tag didnt catch anything. ChemTagger was more useful. My first instinct on ChemTagger is that its pretty specific and its got some unusual tags considering the sample name that this whole document refers to is V629, and they managed to catch that V629 is something of interest...but the only problem is that it didnt . V629 is a game sample which appears in the title, and I dont think its referenced anywhere else in the document, but if you go onto page 2 it switches to another sample because for the second part of characterisation you need V629 which it didnt recognise which is a bit weird considering V269 was recognised. And the main problem there is because V629 is a game sample but V628 is a seesam and its put seesam as one of the tags. You might interpret this document that V629 is a seesam. Obviously thats fairly drastic...if somebody was completely ignoring what the document said then that might be a problem.

3.3.3 Question 2c - Are there any of the tagging services you don't think look useful?

Participant A: Gate. Even though ontology tags got a lot of results that are relevant I wouldnt say they are necessarily useful, they are a bit too generic, even though I guess they might not sound generic to someone, even physics just doing a different field, like photoluminescence doesnt sound like that generic but it kind of is for this paper which is super specialised.

Participant B: Gate.

Participant AL: Something of interest on mine for the ChemTagger it hasnt actually picked up GST as a chemical, probably because the chemical name is GESPET but everyone calls it GST, so I dont know if it knows acronyms.

3.3.4 Question 2d - How would you have tagged your document if youd done it yourself?

Participant AL: If I wanted tags it would be for the SEM as thats what I did that day, using the SEM in ECS

Focus Group Leader: Whats the SEM?

Participant AL: Scanning electro-microscope, so I might want to look at what are the results of the SEM or did I use SEM on this. As well as something like the GST there, thats what my PhD is around so this is characterising that specific layer on that day. That's the only useful bit of info on this whole page really.

Participant AM: I think for the OpenCalais ones its what's there mostly, I would have added the kirchoff's law...like the physical...the well known physical effects...the physics behind it I would have tagged. This other one here...internal resistance I would have tagged that too. For the ontology I dont know...its hard to think about what I would have wanted to say, its like the key results but I dont know how to back name that to an ontology or a particular naming convention to be able to isolate that.

Participant A: Kind of a mixture of the ontology, open calais and chem ones. Chem ones has gotten the abbreviations that I guess Id like, quantum, parabolic that type of stuff. I would have definitely put those in. Then I guess all the keywords but OpenCalais and ontology have a couple of...so Ive highlighted the chunk that ontology gave me about two maybe I would definitely want in there and for Open Calais probably about 3 of those. Id do it more sparsely to be honest.

Participant B: I guess the main problem are the things that are of interest to me that are input with a particular unit, so wherever it says nanometers, the number before that and the unit should be a tag. It caught 500 milliwatt.

Participant A: Maybe it caught it because there was a space between the number and the units.

Participant B: Maybe...although here 500 meters watts...but yeah so where it says 500 miliwatts thats very helpful because we want to know how much power does this laser output so having that is useful but it seems that it doesnt tag it if its written correctlythis like that so.knowing what wavelength it emits, knowing what pulse duration is and all of those are going to be number space units, and having those flash up would be ideal.

Participant AL: Yeah I agree

Participant B: It alsoOpen Calais caught mode blocking, and it seems that some of them it ignores the abbreviation but it got the long written out version. Ideally getting both or working out that one means the other or if you could address that later and say this tag is the same as that...so here PL and photoluminescence are meant to be the same thing but unless you tell it that I dont see how it would be able to do it, but once you told it that surely you should be able to know all of those entries are just the same thing. Use the same tag and dont generate two tags. Otherwise its okay, its a bit sparse, it probably got about 50% of the tags Id want to identify this document.

Participant A: Thats an interesting metric. I think for mine out of all the stuff that it spat out I would say its 75% of the things I would want in there did appear somewhere.

3.3.5 Question 2e - Are there any obvious tags missing?

Participant A: Yes, this whole paper is about a bosonic cascade laser and none of those have put that in it...to be perfectly honest the chemistry got BCL but thats like...what the paper is about like if you had to pick one chunk of keywords and it didnt get that so thats like a flag.

Participant B: Abbreviations.

3.3.6 Question 2f - Would you make use of an option to add/remove your own tags/chemicals?

Group: Yes

3.4 Question 3 - Search

3.4.1 Question 3a - When searching for a word in the search bar here, would you expect the search to prioritise results by: Existence in the title, pure term frequency or whether it was tagged as an important term for that document?

Participant A: weighted tags would be a good search metric but I wouldnt trust software to do it for me, so in that sense I would go with times that something had cropped up.

Participant AM: I think it really depends on what youre using it for so if I was to look back through my lab notebook and the electronic version I would be able to quickly skim through if this is a frequency thing or even a weighting thing but one of the things Ive found is for knowledge transfer when you leave. One day your supervisor might want to look back and see if you wrote down some instructions of how to use the SEM for the next student to do that really quickly so I would imagine they wouldnt have that benefit of hindsight of knowing what you did so might just type in SEM or how to use SEM or methodology for SEM or something like that so I guess the choice in that regard depends on the application and I dont know which one would be best for which application.

Participant A: What kind of thing do you search and do you tell it specifically search in title, search in authors

Participant B: That would be quite useful

Focus Group Leader: for here at the moment you can search specifically by title and by text but more advanced options could be added.

Participant A: Figure captions would be good to search on

3.4.2 Question 3b - Do you think these advanced options (by title or by text) are actually useful? Would you use them and are there any other advanced options you'd make use of? Or would you prefer to only use the search bar with no customisations? and Question 3c - How would you expect this search to work?

Participant AM: I would personally rather have one that basically tries to read my mindnot the advanced sort of boolean ones and wild cards and you might not know that information anyway because usually you look by title and by author and everything but in this one for example.the Open Calais tags. Ammeter and multimeter and electromagnetism but these arent terms that I can search through a document necessarily unless I want that term to come up a load of times and if its a collection of different lab books and notebooks together it would be like heres 50 different results from 50 different experiments that all use the term because you measured current or something.

Participant AL: I at least...in paper form I do a lot of just searching for units, if Im trying to find

out the power of a laser on one day I just scan looking for miliwatts because I know it was a miliwatt measurement so I dont read the text. So if I could type in and it would pull up all the milliwatts so i could get all the laser power measurements that would be useful.

Participant A: It would be useful to narrow it down to a date range too.

Participant AM: Date tagging would be good actually, if you knew the vague time...I guess...text bars where you run get data and save it from like data lockers and stuff should have a time code and date associated with it so that data should be able to be pulled out and know it was between July and September...heres all the raw data in that time and then if you search for parameters within it would narrow that down. So if I knew it was a 500 milliwatt laser or something that would narrow it down.

Participant A: What format does it come back to you like...when you search in the bar or whatever. When it comes back on the screen how can you refine it.

Focus Group Leader: At the moment the search is quite simple as one of the aims of these focus groups was to figure out what users wanted from a search.

*** Demos search, shows searching general terms and also specifying by title and text ***

Participant A: So when we get back this list of documents can you then reorder that, would there be a way to look at these by date or most relevant. Me; Yes that's functionality that could definitely be integrated.

Participant A: Im surprised Open Calais didnt do a better job given that thats the one looking out for other things.

3.5 Question 4 - Markup

3.5.1 Question 4a - What do you think of the markup?

Participant AM: Looking through, especially the actions where the actions are sort of like commanding actions or even actions as in like conclusive statements so this one here says find the initial modelocking results so I would now know that I could look through the document and find the results for it, whereas the chemistry one its sort of like stir this add 9ml of brown fluid and then heat...both are useful but the actions are something that I would be interested in. Marking up in the document I think it useful because it makes my eyes follow things and then I would see which bits have been picked out. I will get a rough idea what its about. But this is me thinking about this on another side. This is me thinking about things Ive not read, so putting myself in the shoes of a supervisor who will want to go and find out that data later, say years from now and go yes thats worth me reading.

Participant AL: I think it depends what youre searching for, if I search my whole log book and it was giving out all the words for each page then it wouldnt be terribly useful, as then Id have to check each page and that's no faster than using the paper book. If youre searching a large body of work then you have to isolate the bits that you would care to check, which is what the tags are essentially doing.

Focus Group Leader: So do you care about seeing them marked up / coloured or do you just want them separate and not shown in the document?

Participant AL: It would be nice to be able to turn them on and off.

Focus Group Leader: So this is a feature, you can turn each of these tagging/chemical recognising services on and off. *** shows tick boxes that do that *** and there are mockups of settings to turn them off per document or per entire notebook in the front end Google Drive

*** demos that mockup ***

Participant A: I would probably...at least if I had it I would turn off the actions because I dont know why I just find them, less useful to look at. Id rather look at highlighted keywords like I guess that pleases my eyes better and if I was looking at a document to do something then it would probably be the whole document

Participant AM: When youre reading it.whilst youre writing it I guess you dont need it to highlight those actions and maybe the next day you dont.

Participant A: But yeah Id know...I dont know why I feel like I dont like them.

3.5.2 Question 4b - Are the tooltips useful? - most specifically to see the ontology descriptions?

Participant A: In terms of hovering over and seeing a definition I do like that idea but I would want to be able to edit it.

Focus Group Leader: That is a feature that would be implemented. This tab here in Google Drive shows the tags and their descriptions and they could be edited to change what shows up in the hover.

Participant A: Some of what I do is really specialised and Id probably have to put in the definition myself for someone using it a year later.

Participant AM: It would be good to be able to hover over things and get more information.

3.5.3 Question 4c - Would you make use of the ability to edit the tag / chemical descriptions yourself?

Participant A: When its looking for definitions can you tell it where to look. If I gave it my documents and to be honest I wouldnt write my own definitions Id go to RP photonics and copy and paste what they had written there, could I tell it to look there for my definitions.

Focus Group Leader: It could potentially be hooked up to other 3rd party services, although specifying direct tag definitions for certain documents in certain places might be a bit tricky to implement.

3.5.4 Question 4d - Do you have any suggestions for a better way to markup the documents to make it more useful?

Participant AL: This is an aside but for me the biggest barrier would be actually digitising the work. I wouldnt be prepared to

Participant A: Old habits...yeah

Participant AL: The speed and ease of paper data entry

Participant A: Yeah

Focus Group Leader: One of my main conclusions thus far is that I dont think right now we can totally eradicate paper until we have much cheaper better durable hardware that you can write on as well as a piece of paper it isnt going to work. But I am looking to encourage further digitisation and improve the current digitisation process. Typing your stuff up **Participant AL** definitely took me a while to figure out what youd written (NB: **Participant AL** sent a copy of one of his handbook lab book pages to be tagged in this system).

Participant AL: If you were going to go through and give meanings and definitions for specific keywords youd want to be able to put things in a subgroup of your own, you wouldnt want to do it for everything that would be too time consuming. But for specific samples you might want specific descriptions.

Focus Group Leader: So youd want settings for specific tags across the whole notebook and you might want more specific ones per document sometimes?

Participant AL: Yeah...Or even...this is a bit abstract but if you could give the keywords their own tags, so you could tag that as sample and date and what material it is so if you wanted to see all of your silicone samples then in your subset of tags you could search for silicone.

Focus Group Leader: So tagging the tags?

Participant AL: So if I tag all the keywords in my document and if each tag has some tags to go with that then I could easily find all of my samples.

3.6 Question 5 - What do you think of the current design of the system? Does it look like the content is well represented, does it seem useful, how would you have expected or wanted to see the content from the backend represented in the backend?

*** shows the different front end sections ***

Participant A: I like it because it doesnt take up much space of what Im used to seeing if you know what I mean. Like its not too intrusive, I like that you can minimise the sidebar and tell it to go away. One of the things I would definitely is colour coding that I can do myself for various things if I

can give certain families of keywords colours that would probably be useful instead of just being stuck with a set design but other than that I like it.

Participant AM: Once its been used I guess I might get used to it but I think my confusion at the start would be how to use it and how it looks different from an ordinary google docs. So I would look at that and if there was something about tagging and I didnt know how to look for something I would just go ctrl f and I probably would miss that there are these semantic tags and chemistry tools up there.

Focus Group Leader: Is there any way you would expect this to interact in a more intuitive manner?

Participant AL: A mouseover function would be good, for the toolbars or the tags you could mouse over them and it would be like we think its this for the assigned tags. But again youd want to be able to disable that.

Participant A: I would like a button in the corner that was like, disable all the functions so I could turn it off without having to turn it off per document.

Participant AL: I assume in that case you could just open the initial document without running it through

Participant A: Im assuming it runs by itself always

Participant AM: Yeah Id have thought it would always be there

Focus Group Leader: The idea is its always there in the background but unobtrusively hence you could turn it all off. But yes a turn of all function would be a good idea. Could you just re-explain what you meant by the hover thing:

Participant AL: If youve gone through the effort of tagging words in the document you might as well have it so that it displays the information concerning those tags when you mouse over. So if you mouse over V629 and it thinks that its picked up on it being a sample

Focus Group Leader: So youd want the tooltips but with your custom information?

Participant AL: Yeah

Participant A: I almost think im imagining this as something I would put a document into and it does all of this at the end but how does this deal with it when youre live writing, when youve finished a sentence go oh I think this is a tag? As if I have to keep stopping this every 2 seconds to make it do it at the end Id get quite annoyed.

Focus Group Leader: sensibly this could either be set to run on closing the document or when the document had been inactive for a certain amount of time. Obviously google drive saves as you go along so it would need to be set somehow.

Participant A: Id want a manual do it now.

Participant AL: Yeah when youd finished.

Participant AM: I in a way would prefer your semanti-cat thing

Focus Group Leader: You prefer how it looks?

Participant AM: Its something that looks like it would be a thing that I could put a document through and then it would do all of the tagging and everything for me and then I can search on that later.

Focus Group Leader: So you like having a separate system?

Participant AM: Unless the functionality can just be turned off and when Ive finished I can do do all the tagging now

Focus Group Leader: That could be set to be like that

Participant AM: Yeah its one of those things that could get a bit awkward as one of the biggest turn offs for lab books are the amount of different ones that are available and theres only one type of pen and paper.

Participant A: That is not true *laughs*

Participant AM: You know what I mean, pen and paper is always the same but youve got OneNote and Google Drive etc theres so many different ways of doing it, if the software can be independent and be applied to whatever method youre using, whereas in this case it looks like its forcing me to use Google Docs.

Focus Group Leader: Its showing a prototype of using Google Docs in the front end because that was the cloud notebook that met the most of the requirements that we had elicited from users. However the functionality is very decoupled from Google Drive and most of the functionality is in the backend and just communicates with Google to get the files and its been written in a way that it could be used with many different types of system.

Participant AM: So it could be run as a type of plugin.

Focus Group Leader: Yeah

Participant AM: I suppose the question you really have to ask yourself is, what is your main purpose to use this for. The way I see it as its just a really neat advanced search function and I could tag and search specific things.

Focus Group Leader: This links in with the last set of discussion questions.

3.7 Question 6 - Previous studies have suggested that scientists have a greater tendency to use software for writing up papers/reports/thesis rather than during their lab experiments to take notes, what do you think of this?

Group: yes **Participant AL:** 100% agreed

3.8 Question 7 - Would you be more inclined to consider using an ELN if:

- It was aimed at that stage of the lab process (e.g for writing up reports / data analysis etc).
- It semantically tagged your work?
- Facilitated a more accurate search across your work?

Participant AM: Theres two really good applications here, I could feed in a load of downloaded PDFs and documents and papers and stuff and semantically tag all of those things to find key pieces of information that I want out of the open literature in the world and then I can also apply it to my own personal lab book or notebook if I was to use an ELN and search that later for some results for some paper or particular methodology that I used. I guess for me I would use that for PDF / Open literature sort of searching... for my own lab book Im trying to use OneNote but the real problem that I have with it is I cant scribble things down and if I had a piece of paper...and they do exist there are little bars and handwriting gesture recognition things...I think theyre 100 or something

Focus Group Leader: Ive seen a new one called remarkably come up and that was about 500

Participant B: Ive been looking at that with one of my colleagues and saw the price and were like noo.

Focus Group Leader: So what are the rest of your thoughts

Participant B: Strongly the issue isnt the software its hardware, writing stuff down or jotting stuff down is way easier on paper, having all of these extra features (like **Participant A** mentioned for Zotero) I use mendele but if you can semantically tag PDFs because I bulk throw a bunch of documents are mendele and then have to sort of work out what they are but if something can go through them and I just say right modelocking ones and I want them to tagged with results and I want them to be this and it just spits them out without me having to lift a finger then thats very very useful if it can handle PDFs. In general I wouldnt use this with things that I have digitised it would be stuff that Im aiming to give to someone else or stuff that I am using which is someone elses because usually I have a good handle on what Ive written down...another thing I wanted to bring up back to sample names, so in my group weve got a sample database and that has row by row each sample plus column by column wavelengths, powers, etc. Pulling the wavelength or whatever for a particular sample and using those as tags for a document would be incredibly handy because if the document hasnt worked out that this is one micron which it does say in the title but it hasnt got it. It didnt catch that but if it worked out that V269 and spoke to my database and said right so heres all the tags that you would want for that sample then I dont have to go to my database, work out which are all othe 1 micron ones and go back to where my documents are held and work out where the sample document is held and compare those. If I can just go I want all of my 1 micron documents then I can do that. But it does involve talking to the database to generate tags first which I dont know if thats a feature that you can easily add. And ideally it would work on an excel spreadsheet or CSV too. Yeahhh user generated tags, quasi auto generated tags that I know are correct and it can then work out based on the document where in my tags it has to look, thats the kind of feature that would be useful. But its

not the kind of thing that would sway me towards ELNs but its something I would add to what we have already just to make it easier to search through.

3.9 Question 8 - Do you think that semantically tagging your documents would have any impact on the efficiency of your work?

Participant AL: Similar to what you said Participant B, this seems like a really useful piece of software but to search other people's work. If I could run this through mendele and it could say here are all the papers about this specific thing then that would be fantastic. About my work I dont digitise any of my lab books its not worth it, everything I do digitise is quite a formal report thats written up but I know whats in them because I spend a month writing them or whatever so I struggle to see it within my work but with work that I havent written I think it would be very useful.

Participant A: There is nothing you could do to an ELN for me that would make me use it on a day to day basis. Ive been in the lab all day today doing stuff and I love writing in my lab book (ish) but even I have to like stop myself and be like write the thing down and even that feels like it takes too long. I feel like its something I would use as a knowledge transfer kind of between members of the group, like I would put...I would never use it live, I would drop in important documents and highlight them up for whatever poor schmuck is going to get the legacy of my work.

Participant AM: Like a knowledge management database tool. Weve got a lot of legacy crap written down in lab books and Im sure some person will at some point if we had this sort of thing would get tasked to transfer everything to electronic format. Or even if it could read handwriting. In principle if you could do that and get old stuff and make that electronic...I guess my opinion of ELNs are its a slow changing thing, I think that it probably one day will happen.

Focus Group Leader: The only paperless labs Ive heard about so far are industry.

Participant AM: When trees go extinct we have no excuse

3.10 Question 9 - Would the following features encourage you to further digitise your work or add tags of your own?

- Semantically tagged documents
- Stronger search

Participant AL: (to Participant A) I think that you put it very well when you said it feels like an imposition

Participant A: You dont feel like you have time I feel like it;s one of those things that youd have to sit down and do well and then it would be really useful but its something youd have to be forced to do.

Participant AM: Yeah, its one of those things when it comes to writing your PhD thesis youll say I wish Id done that but during it you were like no this is fine Ill remember this forever so its fine...

Participant A: I think its probably if you didnt do it like every week though as youd be like this is really important and maybe this is really important. If you did ELNify your work every couple of months then you could still kind of find the important documents or whatever that would be useful for you or someone else later but they wouldnt just be like this big database of crap that would pull up loads of stuff you dont need in the end.

Participant B: The closest to needing something like this that I have actually is my log book has

Participant A: A contents page?

Participant B: Essentially yes it has a contents page and each page in the log book has project number, project title, continued from, continued on. When Im running 5 projects simultaneously that's pretty much all I need. All I need is to be able to go to the previous page is there and the previous page is here and the next page is there. I dont really need to be able to work out exactly what the meaning of everything Im doing is, just page by page, working out where projects are continuing.

Participant AL: I think if you want people to use this youd have to teach them how to use it before they start working in a lab. If this was in undergrad labs and knew how to use it and realised it was efficient and quick I would continue it but now.theres things Id do differently maybe if I started again but what I do works so why would I spend weeks/months trying **Participant A:** You already feel like you dont have enough time anyway

---

3.11 Question 10 - Do you have any further comments, and is there anything that could persuade you to use an ELN?

Participant A: I think the most useful thing you could get behind it if you wanted students to do it would be their supervisors being like look this is going to be really annoying but you have to use it, and then people would have no choice but to use it.

Participant AM: Im super for the psychology part, forcing somebody to use it makes somebody want to use it less but if you can highlight the if somebody would give me a compelling argument at the very beginning and give me a persuasive argument of the advantages it would give me then I would consider it but if somebody said do it because I said so

Participant A: But remember being a first year PhD student and someones like this would be good if you could do it but up to you

Participant AM: But that's not a compelling argument, if you gave me like actual statistical evidence of something or youve got to be sort of influencing and persuasive with it.

Participant B: I started using Jupyter notebook not because its been particularly convenient as its been a steep learning curve but because I thought it was cool.

Participant A: I think youd need to have the supervisor making you do it and the whole group has to do it, if its just you using it by yourself.that wouldnt work

Participant AL: I can see a scenario where every lab has a touch screen in it and whenever i wanted to enter something into my log book I wrote it on there, scribbled away, commit and there it is under my name with my time, date, whatever and I go back to my experiment and then carry on and whenever I want I can just compile it

Participant AM: As I said Ive been using this OneNote thing and I actually recently worked on a little project and I went to the effort of setting up a shared OneNote thing and gave everybody little tags and in the meeting I said oh Ive sent you all a OneNote thing and they were like what...and I was like its really cool you can write things down sort of like an ELN, and the first 2 weeks nobody used it and I could see on the thing that it was just me, person x, person y, person z and I was writing notes and making subtabs and nobody used it and then 2 weeks later people started to slowly use it and then eventually its now got to the point where it looks like a big fully functional project notebook

Participant A: Thats really cool

Participant AM: Some people have gone like oh yeah thanks for sending me that because they can actually use it and see what other people have done and add to it and they can go and question what does this mean to people. So it has been useful so I guess a team effort is how do you train a set of monkeys to keep their ELNs

Participant A: If you have a slow afternoon when you have time then you might be more willing to do it. We tried to use OneNote but we dropped it because no-one liked it.

Focus Group Leader: Any further comments

Group: No

4 Biology Focus Group

This focus group was conduced on 23/11/2017 and comprised of 5 biologists (Participants Q, R, AN, AO & AQ). The following colour key is used for the Participants and Focus Group Leader:

Focus Group Leader
Participant Q
Participant R
Participant AN
Participant AO
Participant AQ
Group

4.1 Introduction - Explaining the purpose of the system

Focus Group Leader: I have been looking at electronic lab notebooks for my PhD thesis, I did a bunch of initial studies, I looked at what was out there in the market, I observed people in the chemistry labs and I did some initial focus groups with physicists chemists and biologists, my interim conclusions were that theres a lot of good reasons to digitise your lab data, theres still a lot of reasons why we should be looking to do that, and that people do actually go and use quite a lot of software for what theyre doing. But ultimately theres still this kind of barrier to actually using electronic lab notebooks in the lab, and there's quite a lot of barriers trying to replace paper, so Im not really looking to try and create a system to replace paper because I dont think that's particularly feasible right now. But what I am looking at is trying to improve the current software offerings of the software that you would use if you were actually writing things up on a computer, and hopefully by improving those we might encourage people to digitise their work further. So, my initial thoughts were when you look at everything, ELNs arent very popular even though there are loads of them, but electronic notebook software like Google Drive, OneNote, Evernote etc is a lot more popular, and if you break down the features that you would need from an ELN and a EN theres actually a hell of a lot of overlap. For an ELN you need everything youd have in something like Google Drive and OneNote and then youd have a domain layer with domain applicable tools, and then a further semantic layer on top that handles metadata and tagging.

***shows diagram ***

This is a diagram of the breakdown of the three different layers, being everything here is generic notebook functionality, this would be domain based stuff, and then at the top is the semantic layer which looks at tagging your notes, doing a better search, linking things together, recognising chemicals etc. The bit Ive focused on making was the semantic layer. Its all well and good for people to say that they like their documents to be tagged and made better to search, and to have more information about them, but no-ones actually given a great deal of information about what they actually want and how they want it to be achieved. This is a proof of concept system that Ive designed. Im going to show you this in two parts. There is a backend part of the system that shows a representation of what can be done. It shows what tags have been pulled out for each document, which bits of the documents have been marked up, what ontology terms have been recognised, and where different chemicals have been identified. The front end of this system would be in Google Docs for this proof of concept idea,

and so you will be shown a mockup of how this functionality could be implemented on the front end. This focus group is to look at both aspects. To investigate how well have the documents been tagged, how things have been marked up etc and also what you guys think of how the front end looks and how you think that would work, and to talk a bit about ELNs.

4.2 Question 1 - Documents

4.3 Question 1a - What type of document have you brought and how typical a representation of your working documents is it?

Participant R: Mine is a section of my thesis for my PhD and this is the kind of thing anyone within all years of their PhD will be doing. Theyll be writing up and doing bits of sections and so the tagging is very relevant

Participant AN: Mine is a second year larger piece of work that was marked, this would be quite typical of a student and might be good for other younger students to be able to look at work if theyre looking for specific things within bits of work.

Participant AQ: Mine is a second year lab write up which is what people would do weekly, so its quite a short piece of work. When it comes to going back and looking through, when it comes to revision and stuff like that I think a tagging system would be very useful.

Participant AO: Mine is a library project from second year, basically kind of like a mini review. It would be good to have a tagging system to see the relevance and what focus I had when I wrote it.

Participant Q: this is a sample of one of the results chapters of my thesis and in terms of where I got my data this is how everything is usually presented, so an introduction to the background of the experiment, figures of the experiments data and what the experiment is telling us that, and that is typical for what I would put in my thesis.

4.3.1 Question 1b - How much of your documents are usually made up of words / pictures / diagrams?

Participant R: Its mostly words and data, however theres a lot of it which gets transferred onto the computer. With the databases you would be using your lab book obviously but when you have a spare minute youd be straight on a computer doing it. Diagrams would be used more for a poster presentation or something along those lines.

Participant AN: We tend to use more words and data than figures and pictures. But they are used for more as a summary or go alongside what youve said. I feel like that the tagging system would work even if there were quite a lot of pictures, youd still be describing, they compliment each other.

Participant AQ: Same as, mostly data and writing and every now and then there would be a diagram, more for presenting work rather than actually just doing it day in day out. But you will always have a figure legend if youre going to have a diagram. There will always be wordy stuff in there.

Participant AO: I would use, most of it is data and writing but the odd diagram. I think it kind of acts like a word sync. If the tagging system would be good for that, it goes into more detail which you dont have to write into the main body but is still relevant.

Participant Q: The majority of it is words so for each figure I guess I have maybe a page / half a page minimum of text that goes with it and that's including figure legends so Ill have an extensive figure legend plus a body of text that goes with it, and Ive got a lot of figures. A lot of figures, a lot of text.

4.3.2 Question 1c - What formats are your documents usually in when you create them?

Participant R: Id say this is going to be a really big generalisation but Biologists dont tend to use LaTeX. Unless were made to write out PhD in it in which case its specifically in engineering or computer science. Were very specific in that we like word, word is good, word is easy, word is straightforward.

Participant AN: word is safe.

Participant R: not that were not open to change.

Participant AQ: except when you move a figure

Participant R: and everything breaks...theres an easy way to do this, make a table and make it invisible and put the figure in there.

Focus Group Leader: Or I could teach you how to use LaTeX?

Group: No.

Participant R: But anyway, what things do we use?

Focus Group Leader: Yes the formats you create documents in.

Participant R: Mostly word documents, excel initially into large pieces of data, and then it will get transferred into Prism for proper diagrams. And then powerpoint to make pretty pictures.

Participant AN: Same, GraphPad, Prism to maybe make some of the figures and all figure legends to go with it and to handle a lot of the data, but mainly word and powerpoint to present what data youve got and theorise about it.

Participant AQ: Yeah, Word, but everything you do in Excel always ends up in a word document when you come to presenting it.

Participant AO: Sometimes i draw my diagrams depending on how complex they are.

Participant Q: Everything is in word, I cant code or do LaTeX or anything. I wish I could as word constantly messes up my images but I cant.

Focus Group Leader: Moving onto the tags. These are the tags that have come out of different

systems (NB: Each participant was given a copy of their personal documents and the tags as well as being able to see it on the system). OpenCalais looks at wikipedia articles that have been tagged and looks at what content has been tagged to try and infer how other articles with similar content should be tagged. These are categories of your documents rather than things that have been pulled out from the text. ChemTagger and Gate are the automatic chemical recognition pieces of software so they have pulled out what they think are chemicals in your pieces of work. ChemTagger also identifies actions (gives example showing the marked up actions in backend). As you can see Gate typically seems to pick up less than ChemTagger. For the ontologies, ontologies are basically hierarchies of relationships and concepts that we use in the semantic web so Ive currently just got 6 ontologies in here, some chemistry physics and biology ones. This piece will pull out the matching ontology terms. Your documents have been run through all of these systems and currently pulled out everything it can find, and part of this software evaluation is looking to refine this process to figure out what things we should actually be pulling out. So on that note if you all have a brief look at the tags for your categories.

4.4 Question 2 - Tags

4.4.1 Question 2a - Which of the tagging services (if any) do you think has the most appropriate tags for your work

Participant R: I think with mine, which is looking at asthma and basophil activation. All of them have picked that up which is great apart from Gate. But, Id probably say that if I was looking for something specific the OpenCalais tags may be a little bit too generic on this one, just because everything that I do is to do with basophil activation and allergy / histamines / biology. However the ChemTagger has pulled out some of the specific chemicals that Ive mentioned. That way if I was looking for a certain antibody then I could just look it up and I would get the right receptor which is helpful because I cant remember the last time I mentioned CD4, whereas C63M I mention loads.

Participant AN: Mine is actually about ketamine and its picked that up in more than one place. Not sure about the ontology though but that would be very specific. ChemTagger has picked up a specific longer version of a chemical that Ive used and it seems to have picked up that.

Participant AQ: I think with mine, the ChemTagger was the best, and I agree that the OpenCalais is too generic. Its picked my work up as chemistry, and its got too generic terms. So yeah I think ChemTagger is the best one for mine.

Participant AO: Yeah, so OpenCalais is a bit broad, ChemTagger has all the different types of chemicals that Ive mentioned. Ontology tags, its good to some extent its just some of the terms arent really relevant. Gate Tags havent given me anything.

Participant Q: Id say the Open Calais because with the exception of one / two words...so neuroglia and anatomy I wouldnt say fit, but loosely fit, but the rest of them are how we would classify this research. Venus is a fluorescent protein and in terms of the research itself we use it in anatomy. Loosely anatomy is in here but its not anything specific. The ChemTagger has pulled out, each of the P3 P6, these are postnatal days, so these are measurements of time not a chemical, same with the

E14. Some of these are vaguely chemicals...as its got proteins we use, loosely it would fit. And the lego vectors would I guess fit in chemicals. But its not as nice as the Open Calais. The ontology tags are probably too broad, because its more pulled out things that are in there but not specific what the topics are. But the Open Calais fits really nicely actually, really really nicely.

4.4.2 Question 2b - From the perspective of automatically recognising the chemicals you have used, do you think ChemTagger / Gate have done a good job?

Participant R: ChemTagger yeah. **Participant AQ** / **Participant AN** / **Participant AO:** Yeah.

Participant Q: Not really but thats because we dont really use chemicals as such in my area of neuroscience / biology.

4.4.3 Question 2c - Are there any of the tagging services you don't think look useful?

Participant R: Gate Ontology is okay but a bit and there are so many tags that it kinds of flutters between really specific and useful and a tiny bit random. Ive got some tags that are just too generic. But then other tags are very relevant. I think it kind of flutters between the two.

Participant AN: Gate / Ontology

Participant AQ: Gate

Participant AO: Gate

Participant Q: ChemTagger probably not as well. The ontology maybe but it would be very broad overviews rather than anything specific.

Focus Group Leader: At the moment all of the science ontologies are used on all of the documents and there isnt currently any distinction between which ontologies are used for which document as I was looking to see if having more of them there would make things better or worse and it seems like its potentially made a worse impact in that its added too much stuff.

Participant R: Its made it a bit too general. Shows the group the chord diagrams and the term diagrams per document to show which specific terms have come from each ontology. General consensus that the biology based ontologies were the most useful.

4.4.4 Question 2d - How would you have tagged your document if youd done it yourself?

Participant AN: It would be good if it could give you tag options based on the general terms so you could narrow it down. Although its picked up most of the things I would have tagged it with.

Participant AQ: I dont really know how Id do it differently, but just following on from Participant ANs point if you had some sort of tagging led system so when you click antidepressant then after that it skims down tag options to whats relevant to that word.

Participant AN: Like filtering?

Participant AQ: Yeah, like a hierarchy for each word. If you look at the words open calais and

ontology have picked up, theyre all well and good but if there could then be a filter for other words related to that specific word that would be quite cool. I dont know how else I would tag this document.

Participant AO: Id say Id have put more on immunotherapy, as its given me things on immunology, neuroglia and immune system. Only ChemTagger has vaccine in it so OpenCalais and Ontology tags, Im surprised it didnt pick them up. ChemTagger has quite a lot I dont recognise, it seems to have renamed some of them or changed them. Otherwise ontology tags are a bit broad but everything else seems fine apart from Gate. Maybe a bit about neurodegenerative disease as well.

Participant Q: I suppose that for...I guess it would be different to be thesis specific for me more than anything so on a broad scale you could tag this as immunology, neuroscience that kind of thing..but Id probably break it down to experiment specific things, for example PCRs for different techniques that we use, what tissue is used, what age the tissue is...that kind of thing because my thesis looks at development right through to aging so I have different time points and different tissues that I look at. And, its useful sometimes to group things by experimental technique because then you can see how or what if you changed any of the parameters of the experiments then you could flip back and be like this date is different, what did I change and pick that up..something like that...but I guess it depends how specific you would want the tags to be. For me...Im usually quite vague with these things so I would group them into more broad categories rather than getting really specifically into it.

4.4.5 Question 2e - Are there any obvious tags missing?

Participant R: Theres one Im a bit surprised about, which is that the text itself is about asthma and allergic rhinitis and those are weirdly the two diseases not mentioned. Its got the broader concept but I think because Open Calais its got allergic rhinitis therefore allergy its kind of missed them. Which I think I would have liked them in there just as like a broad thing. Allergy has gone too broad.

Participant AN: Mine are missing a few I would like to put in, Open calais is along the right lines, its got some of the basics like antidepressants and ketamine but it hasnt picked up some of the more specific terms.

Participant Q: In the context of the tagging system you have, I would say no. Because for I guess a broad tagging system that you could apply to lots of different disciplines this would fit but if this was coming from me, experiments, techniques is something that would be a useful way of tagging stuff. I think thats probably the only thing I would really want. And I guess if you have this type of software you could personalise it.

4.4.6 Question 2f - Would you make use of an option to add/remove your own tags/chemicals?

Group: Yes

Participant Q: It would be easier...you get through things and be like oh these things should be labeled like this. And I wouldnt necessarily if I had a set of software Id be like these are definitely the things id want, Id rather have the option to play with it I think.

Focus Group Leader: One of the reasons for marking up your data like this and getting the tags is to improve your search. So just a few questions about searching.

4.5 Question 3 - Search

4.5.1 Question 3a - When searching for a word in the search bar here, would you expect the search to prioritise results by: Existence in the title, pure term frequency or whether it was tagged as an important term for that document?

Participant AN: Whether it had been tagged as something important in the document, and then prioritise if its in the title or not second.

Participant R: I dont know if Id want the title though.

Participant AN: It depends if youre looking between documents or

Participant R: If you know youve definitely written something with that in the title then yeah, but if youve got a document where you write about a specific drug for example and then all of a sudden youve definitely not put it in the title because youre going on about the disease but you know you mention that drug youd want the frequency personally.

Participant AQ: Probably frequency, I always find it quite frustrating, you know on word when you do find and it brings up..

Participant R: parts of the word?

Participant AQ: Yeah, or it will bring up a word that is in one of your references and it will always bring that up so it will take you to the bottom of your document and that's not what you want. Youd want priority but not with the title necessarily and be able to prioritise it in different ways like headings in the document or words that are constantly in the document, but then have a filter system so youre not ending up with it searching in your references, or your ability to pick and choose.

Participant AN: Or the ability to just search the references?

Focus Group Leader: You can search by just title or just text and you could add different options.

Participant R: Its good to have the option

Participant AQ: Yeah

Participant Q: I would probably expect it to be by titles more than anything else, I would never think...I have no idea how many times I might have written a word in a document and the way Ive organised my data and my thesis is by title so I would search based on that, and it would be better to search on the title to find out where things are and go from there. I dont know how often I use certain words, probably more than Id like to think about. I guess the only downside to doing it by title, if I do this for example in endnote, looking at papers, it brings things up that have it in the title and the text, and you cant necessarily tell where things are in the text. So I think that if you could show me exactly where things are, if you could search by title and then it could map it back, and if you do it by text map it back to where in the document that text is.

Focus Group Leader: So its in this title here but in this paragraph here? Maybe with a preview?

Participant Q: That would be useful otherwise you have to trawl through it to find it, and you just know it has a word in there but not where. At least for me as I dont like spending time hunting I like things to be there.

4.5.2 Question 3b - Do you think these advanced options (by title or by text) are actually useful? Would you use them and are there any other advanced options you'd make use of? Or would you prefer to only use the search bar with no customisations? and Question 3c - How would you expect this search to work?

Participant AN: Depends what I was looking for, if it was something specific and I knew there were a few specific papers out there then I might look for the title or keywords, but if its something generic that Im just looking up then I might just hit and hope that something good comes up.

Participant R: I think it depends on your mindset, if youre having a bad day and trying to look for something youre just going to hit and hope, all specificity kind of goes out of the window. If youve had your first cup of coffee and youre in a good mood youd be more likely to go for customised options.

Participant AQ: It depends what youre looking at as well, if I was trying to scroll through a thesis then Id want something more specific than say a masters student.

Participant AO: Id prefer to search by the text in the document because I remember last year for my dissertation we had to look at a lot of the things we were looking for it was the details within the actual files, that gave us what you needed instead of the title because the titles might be so irrelevant for you to look for if youre looking for different things and different pathways.

Focus Group Leader: In the backend Ive marked up where things have been tagged or recognised as chemicals in different ways, Ive marked up the actions in purple *shows actions* and then here is marked up to say that these are the chemical elements recognised by ChemTagger or Gate, and then these are the ontology tags. You can turn all of these on and off, and if you hover over the highlighted words it will either say that it is a chemical element or give the description from the ontology. This markup idea was partly derived from how ChemTagger works and partly to show which bits have been tagged or recognised as chemicals by which services.

Participant Q: searching by date could be interesting if youve got multiple experiments and you were looking for a particular one, I mean it depends...I organise my files like that on my computer so Im always looking, and in my paper lab book I always track things back to the date and even my analysis has all the linked data files whatever.as in the stats program or whatever. Date would definitely be useful, I dont know how else I would search the things, other than the odd word or date, that's usually my go to! Or I just look endlessly through the lab book, I vaguely did this...in this range of pages...where is it now? So yeah I guess being able to search that kind of thing would make life a lot easier.

4.6 Question 4 - Markup

4.6.1 Question 4a - What do you think of the markup?

Participant AQ: As in the highlighting etc?

Participant AN: yeah because it gives you the context of what youre talking about.

Group: nods

Participant Q: Yes, I think it kind of links back to the search function in that you can see whats its brought out, rather than it returning things like you have this many tags in this document and not telling you where it is. I like knowing where things are. I think thats useful.

4.6.2 Question 4b - Are the tooltips useful? - most specifically to see the ontology descriptions?

Participant AN: Yeah particularly for younger students, for first / second / third years who may be reading primary papers for the first time or the beginning of reading primary papers may not actually understand what theyre reading all the time, and if you can just hover your mouse over it and see what it is there and then instead of having to pause your reading and go away and search it and then come back to your reading it would make a difference.

Participant AQ: I still dont understand all the papers.

Participant AN: I think it would be exceptionally useful particularly for students.

Participant R: I think its still handy for other people as well, as the more that you focus down on a particular subject you forget the broader terms and its really stupid but its so so true. Theres a lot of basics that you completely throw out of the window, so even with some basics just reminding you what that cell does is really helpful. Just as a silly little reminder when youre like okay, absolute brain freeze, what is this?

Participant Q: I think if there were students after me who were looking at my notes and didnt know what things meant it would be useful, most of the things that I write...because Ive written them I know what they mean.

Focus Group Leader: What if this was collaborative and you were looking at other peoples stuff?

Participant Q: This would probably be useful, as if you look through other peoples lab books and its not how you would lay it out so its hard to know what they were getting at sometimes so if theres something that would streamline that and say what they meant and where things fit in that would be very useful. For collaborations it would definitely be a useful tool.

4.6.3 Question 4c - Would you make use of the ability to edit the tag / chemical descriptions yourself?

Participant AN: I think it would be nice particularly to be able to update them if things changed or if its not quite right, particularly in the wider context of the reading. Group Agreement

4.6.4 Question 4d - Do you have any suggestions for a better way to markup the documents to make it more useful?

Participant Q: I guess what I would expect to see would be slightly different because the ontology terms for example are quite broad so they would bring up things as stuff I wouldn't necessarily flag as things I would count as important. I guess that comes under subject specificity. I dont know what I would expect to flat up in my work if someone else was looking through it or what they would expect to pull out of it as you know what is and isnt important in your own stuff.

Participant R: I think sometimes having terms in your own words to your own understanding is much easier, everyone does revision notes.

Participant AQ: Your own descriptions for things?

Participant R: Its like seeing things in your own handwriting, like if you describe things in your own words rather than writing things like how a paper should sound, like just having this is a cell responsible for allergies would be a lot easier to understand for me.

Participant AN: Like personalised tagging basically?

Participant R: Yeah

Participant AQ: That would be pretty cool, or I dont know if this is possible because i dont know anything about computers but if you had the ability of it being connected to the internet and it brought up links which you could go to that would have papers with the relevant terms.

Participant AN: That had the right tags in the paper?

Participant AQ: Yeah that would be mental.

Focus Group Leader: You can do some stuff in google docs anyway.

Participant R: Like could you copy one of the ontologies into google scholar and see what it hit?

Focus Group Leader: Not exactly.

demonstrates explore toolbar in Google Drive

you can search the web via this bar and search for things relating to this document and bring up images and web results and insert things from this bar into your documents and you can search throughout your other documents in Google Drive here. You can also insert references from papers here.

Participant Q: And could you link to other file locations, as I currently write them all out by hand in my lab book.and it takes me ages as I put things into millions of folders

Focus Group Leader: Yes

Participant AO: So you know how weve talked about the hovers and being able to edit them yourself to suit you, could you also kind of maybe draw diagrams to go with that.

Focus Group Leader: Its not currently integrated but its very possible.

*** Gives demo of ChemTagger ***

I used the api from this project (ChemTagger) and their actual example shows highlights of structure diagrams, although they dont currently make that info available in their api, but it could be done using ChemSpider or something. You could also set custom pictures as highlights if you wanted.

Participant AN: For the tagging, could it not link to a document through Google Scholar to bring you straight to that document, so if you think it has anything that might benefit your piece of work then take you to that.

Focus Group Leader: So search the web for similar documents?

Participant AN: Yeah that would be cool.

Focus Group Leader: So far weve looked at the backend and whats possible to do with the different tagging and chemical services. Now Im going to show you the front end and how this could be implemented inside Google Docs.

4.7 Question 5 - What do you think of the current design of the system? Does it look like the content is well represented, does it seem useful, how would you have expected or wanted to see the content from the backend represented in the backend?

*** shows the different front end sections ***

Participant AQ: For the tag descriptions, could there be links so that when you click on them it would take you to where those tags are in the document?

Participant AN: I like being able to tag actions as if you were looking for a protocol in another paper that would really speed things up.

Participant Q: This would work for me, because it is familiar, I tried downloading an online electronic lab book in my first year and it took me so long to figure out what I was doing I saw it as a waste of time, I would rather have something where I know what I'm doing and how to deal with it and the additional tools are all you have to learn about it and i dont have to learn everything, and I guess you can put things straight into it and it will back it up in the cloud and that would be beautiful. I think this is a nice way to do it, I dont like learning new things. Its easier not to have to learn something totally new if youre just adding new things into something familiar.

Participant AQ: I think its nice in that its not frightening looking, like for computer phobes. It just looks like a normal word document and it just highlights a few words, its just a small thing on the side. So its very user friendly.

Participant AN: I think it needs to be more specific for chemistry biology and physics independently,

but with that it would be exceptionally useful as it seems to tag it fairly nicely and its user friendly so would be good to be able to use this for searching through documents. It seems easy to use but also to have more options if you are looking for something more specific.

Focus Group Leader: The logic of doing it on top of the existing notebook was that there was no point building the notebook part from scratch and also that people use these pieces of software already and for a reason, GDocs is really similar to Word and people use Word and Google Docs and Evernote etc because they are useful to them, so I figured if you built stuff on top of that you might have a better chance of people using the software that they already like with additions rather than having to learn a whole new platform.

Participant AQ: Id be much more willing to learn an extension or plugin than a whole new set of software.

Participant Q: the main thing for me with my paper lab book is that I like sticking stuff in, basically like a scrapbook, what I would want...I assume I can do it with Google Drive but is so I can have figures, I can have text, I can annotate notes, not necessarily like nice neat notes, just all over the place...I like doing that in my lab book because thats how my brain works. As long as it can do that then for me thats fine.

Focus Group Leader: That concludes the system side of the questions. Ive got a few more questions about ELNs in general.

4.8 Question 6 - Previous studies have suggested that scientists have a greater tendency to use software for writing up papers/reports/thesis rather than during their lab experiments to take notes, what do you think of this?

Group: Mass agreement of very true

Participant Q: 100%, everything I do thats not writing my experiment of the day is on a computer. All of our data is put in excel files, Graphpad files, word documents, all of our images are stored on computers from the microscopes, because when weve finished all of that has to go to the supervisors, so everything that is in your lab book should be in digital form somewhere.

4.9 Question 7 - Would you be more inclined to consider using an ELN if:

- It was aimed at that stage of the lab process (e.g for writing up reports / data analysis etc).
- It semantically tagged your work?
- Facilitated a more accurate search across your work?

Participant Q: It would be useful because I find that I collect data in my lab book, I do stats, I print off graphs and stick them in my lab book so I have everything in one place but it is cutting and

sticking nonsense. The only time things are really combined properly is when I put them in writing like for my transfer, as thats the biggest combination of all of my work because all of my figures are in there. But even in that I have like four folders of draft figures. Data figures, stats things, figures, if I could just have...ideally what I need is something that integrates my data tables and my stats program but I think that maybe I just need a better stats program or do my data differently I dont know. Because thats the main one that I find I jump between. I jump between excel and GraphPad if Im processing data. Thats the main thing, putting things into word is also a pain for writing things up properly, so if you had a platform that would integrate everything without it spazzing out that would be great.

Participant AN: I think it would be nice to have a combination of the paper lab notebook and the software for this. If youre just running into the lab to go and run a last minute experiment youve got your paper notebook in your hand whereas when you come out and youre running a series of experiments you can keep writing things up gradually and putting in data and adding the observations and they could be tagged to observations other people have found, which would be nice to continue the research.

Participant AQ: I think you need to marry the paper notebook and the software for tagging. Practically in the lab setting the lab notebook is your best friend.

Participant R: Equally its nice to have...youll come back with your lab book and youll make an excel document, prism file whatever but the main thing is that you can put in that diagram from that day and then it will have tags and figure legends where youve written what you did and where it came from. Its good to have that electronically.

Participant AO: I prefer my notebook in the lab because everything is everywhere in the lab, and I dont want something I have to look after something to make sure nothing gets spilt on it. I dont want to have to worry about that. I dont 100% agree with the use of computers, Ive written up my lab protocol etc and then having it laid out with what this chemical does or having explanations of what things are would be helpful though.

Participant AQ: I think to be perfectly honest with myself I would not make the effort to go and put my past work in there unless I have a free afternoon as you dont think to do it. But if it becomes a habit and becomes ingrained into your routine then the software works really well but I dont think its something Id be constantly willing to do.

Participant AN: I think it would probably be once every week or so that I might be willing to chuck things from my paper lab book into this and take any research Id found and put it in there too. I do think it would be useful to tag it and relate it back to stuff and you can see what youve done and where this relates to other people's work, but my paper lab book is just the best.

Participant R: Im very attached to mine, I think an issue with everything when it comes to projects is that you name things completely unconventionally but you think its relevant at the time trying to find a particular document and its not there which is where this would come in very handy because youd then know which experiment it was and you can see the dates etc which it would be useful to search on and get date order and then youd know what kind of experiment it was and you could look

back on the actual data, equally I think that realistically no-one analyses the data until the day before the morning of the lab meeting, which is true unfortunately, and doing those little last minute jobs. Thats when it would realistically get added.

Focus Group Leader: So thats your paper lab notebook but what about the things you already have on computers? SO say this software existed when you guys were writing you your documents, would having the semantic tags make any impact on whether you would add this into google docs etc? Or would you just still continue to use word? My system is more aimed at this particular stage of the lab process rather than replacing the paper lab book.

Participant Q: Yeah, I mean this would make my life easier if I dont have to copy and paste things and I can just search for them in software and find it in my documents and paste it in like that it would make my life so much easier. Particularly if the formatting didnt go squiffy every time I add something. I would really love that, it would save time as I spend so much time faffing around with figures and then realise that its messed everything up. I would be happy to use this kind of thing I think it would be a lot better and a lot easier. I dont know if it would be a lab book use for me or more of like a tool for writing. If I think about lab books and the way I think about them and this the two dont really go together as the lab book is very much scrappy notes just so I know what Ive been doing during the day whereas this would allow me to process all of my data, write it up and visualise everything in the way that I currently do but without all the hassle of jumping through folders.

Participant AN: For big pieces of work where Im looking for multiple documents and trying to see what Ive previously written to see if it flows nicely I probably would use the software like that if I knew it was available and I knew some more of the functions. If you were to put it out there it would be nice to have an instruction manual / tutorial for it as well to say this is all the functions that you can do with this because I think that there are things that are in word that Im only just discovering many years into university. But yeah I would probably use this for writing up large pieces of work.

Participant R: I have lots of random word documents that are saved completely where I know but they have stupid titles like questions I need to ask about my thesis or why is this important so actually being able to go through these documents and figure out which random paragraph of the question I decided to answer that day when I had a spare 5 minutes would be really useful. Then it would be like you type it in and it suddenly comes up with a nice paragraph that I wrote when i had time so I could copy and paste it into my thesis. And Im more organised than most students i know.

Participant AQ: I think I dont really have anything else to add to that

Participant AO: With me, I dunno I can imagine myself going onto a tangent if there were tags and Id get really interested in what the different buttons and functions did and then if it had links between tags and they linked to research around that area or anything related to that area I can imagine it expanding my knowledge, which is good if youre still writing things up but bad if youve already written it?

Participant Q: I dont know if the tagging would make an impact. I dont think I spend an awful lot of time searching for things because I usually know what Im looking for and I have a vague idea of where it is. The only time I really search for stuff is, one if its like a specific something so if I have

an excel spreadsheet of loads of gene names and I need one name and I dont want to trawl through 6,000 genes to find it, or if Im looking at stuff in EndNote...occasionally I search for things to find it if Ive really lost it, so that...it would be more of a searching thing than a tagging thing that would probably be useful for me. Just because I dont think where Im at at the moment, it would be useful to start tagging stuff...maybe if I started a new project it might be helpful but at this point I dont think of things in terms of tags.theyre just...I know where they are in folders, I know how Ive got all my documents organised and thats where I kind of look for stuff. The search would encourage me but it would depend how much depth Id have to go to for adding tags. If it was manually adding tags myself probably not.

Focus Group Leader: The idea behind this is that it would tag most things and you would have the option to tweak those afterwards.

Participant Q: Yes probably, but if it was me trying to set all of that stuff up myself probably not.

4.10 Question 8 - Do you think that semantically tagging your documents would have any impact on the efficiency of your work?

Participant AO: Definitely

Participant AQ: Im not efficient

Participant R: I think initially its going to take some time but thats more because you need to personalise it to what works for you and then once you know that then I think that from then on you would be good to go, but obviously there will be certain things in the initial stages of uploading documents and there are going to b a few things that you want to tweak and change around and once its done it would be fine.

4.11 Question 9 - Would the following features encourage you to further digitise your work or add tags of your own?

- Semantically tagged documents
- Stronger search

Participant AQ: Sorry no

Participant Q: Probably not.just because at least to the extent of what I know, my work is as digital as I think I can get it. All of my data is digitised so the thing that I would take from this as a benefit is that I would be able to streamline combining all of these things not necessarily tagging them and searching it. I like the idea of your software being able to integrate all of the different bits of functionality more than the tagging.

Participant AO: What if your work is already digitised, I have some sequencing data that I have to

sequence and search through so what if this could do it for me.basically we get sequencing data and then we have to look at it against the wild types to see if there are any mutations...even some of the results from glass arent great.

Focus Group Leader: I think that would probably fall into the category of a specific domain based add on.

Participant R: I would, I dont think it would make me do it more because I think that the balance is there if that make sense, everything that is data goes in the lab book and I write out the protocol every day, whereas everything that is on the computer is always going to be my data, my words, my everything so its something that would be used obviously because google docs you can throw in word documents and be done. So thats not stopping me from doing something Im already doing its just a transferral across and that could be actually really helpful when youre searching for documents that are stored in random places. I think thats it.

Participant AQ: The other benefit of it being on Google Docs is that the girls Im in the lab with weve got a Google Drive and its got everything on there, what needs ordering in stock wise and what weve done and its a really good way of actually communicating in the lab, not just for your own personal lab. Definitely more useful in a collaborative environment.

Participant R: Also encourages you to save it in places other than your C Drive, as if the University gives you laptops...Ive gotten through 6 and theyve wiped the hard drive on every single one and all of my work has gone 6 times and if I didnt have dropbox or an external server it would be gone, so by having this there it would be the biggest lifesaver in the universe if all hell broke use.

4.12 Question 10 - Do you have any further comments, and is there anything that could persuade you to use an ELN?

Participant AQ: My issue with going away from paper is that a few things...writing for me is quicker and if youve got gloves on then the last thing I want to be doing is tapping away at something because youre worried about what youve got on your gloves. I dont think twice about picking up a pen, which I know is highly unhygienic because I put it in my mouth afterwards

Focus Group Leader: Do you have special lab pens?

Participant AQ: No...its the same pen, ive probably eaten a lot of bad things. In my mind Im much more carefree, and I think most biologists are about picking up a pen and writing it down, its just a quicker instinctive natural way to do it. I think eventually things probably will end up electric but its just going to take a big transition period and it would be a slow thing but its something that will happen but it wouldnt be an immediate change.

Participant AO: If I had more time during the week I would, I like having the paper thing just because its rough and its quick but I would like to also write it up just so that I know I have a cleaner more legible version of what Ive done and my results so I can easily sift through it instead of my lab book where I have to manually look through it, whereas here I could just use the search function or

try and find things that are related.

Participant AN: I dont reckon I would use it in a day to day basis for data collection but for the final writing up and any interim larger pieces of work, not just 500 words of just spitballing ideas onto paper, or onto a computer I probably wouldnt use it for that but bigger pieces I probably would use something like that, but more use it to record what Ive done but also possibly even use it as a research tool so to expand what Im looking at and see what other people have done but paper initially and then computer may be secondary.

Participant R: After actually starting in industry and starting to use OneNote Ive seen how useful having that electronic document can actually be, to the point where I kinda wish that I had OneNote throughout my PhD rather than discovering it near the end as you can have tabs and put your dates down the side and see this experiment, that day, here is the data for it. And I think encouraging an ELN would be particularly useful, even if students, depending on their level. This would possibly be better to have tags for different levels of people so kind of if youre a first year student then you have everything tagged to death and then the more advanced you get you filter upwards so with that respect I think realistically any undergrad who would write up their lecture notes would get such a world out of that and it would be so incredibly useful and as far as lab books obviously arent going anywhere but the actual writing up process is so ridiculously useful but I think that it would be really good, especially if you threw it in front of a PhD student I think theyd kiss you.

Participant Q: I think that to get...the one major concern I would have is I think that you would have to almost choose a discipline to really get the tags right, and even in terms of biology theres a lot of sub disciplines. The biggest sticking point would be to get it to be specific enough by tailoring it, as even under biology.to me biology is plants and animals and my work is more towards biomedical science. In terms of encouraging me to use ELNs, I think at this point it would be unlikely that I would just because everything for my PhD is paper based...for postdocs I dont know if I would, its more that Im so used to sitting down and scribbling things and I always think when I get to the computer it's the final bit, its the formal side of things, Im processing data and writing, and I wouldnt take a computer into the lab I dont think ever because I am messy and my students are messy and I dont want chemicals on my computer, whereas lab books...whatever. If my supervisor had made me use one I would have just got on with it, if I didnt have the option Im sure I could manage, it would be an adjustment but it wouldnt be the end of the world, but given the option I quite like paper, and I think even if I had an electronic lab notebook i Would still have paper notes that I would put into it. All my notes are on paper even though all my writing is on the computer. I have sticky notes and lots of paper notes. Even if I dont have paper I will get a bit of tissue and write on that and take it out with me instead of putting it in my phone or whatever because its still easier. I know I will always have things like that around and some of the stuff I work with I dont want to be touching a phone or a computer as I work with some relatively nasty stuff, maybe it shouldnt make a difference to getting them on paper or a phone but I use my phone outside the lab whereas my lab book doesnt get used outside the lab. Im quite fond of my lab books.

Focus Group Leader: Any further comments?

Group: No