eIUS: Neo-Geography Experience Report 1

In the text: Some barriers in '{{…}}' to provide more contextual information.

Interviewee profile

Researcher at the Centre for Advanced Spatial Analysis (CASA) at University College London, in part collaborating with the GeoVUE NCeSS node; lecturer at Birkbeck College

Time spent in research

"**10% right now** in this term because I have to work on teaching, but **before that it was like 90% [period including projects mentioned in this interview]**, the other ten percent was administrative stuff that you have to do for projects and other things, research intended as a broader term I would say (..)".

Research area and Research question(s)

"I have a PhD in geography and I would say that I am sort of a **quantitative geographer** rather than a human geographer so I like to **analyse data, I like to crunch data to see the results, [use] modelling and (..) GIS, Geographic Information Systems as a tool for me to do analysis**; so I'm not interested in the computer science side of it. I have to know something of course, but I **use GIS and geovisualisation tools** as the tools they are, and the data I crunch more is health data because we have a **partnership with the Primary Care Trust** [PCT] **in London and I am actually charge of analysing different health data coming from them**, and my particular role in this is a transversal interdisciplinary because I have to do **analysis on data that rely on the epistemological and the medical geography domain**, so **analyse the data, build a model or find out targets or areas where a particular disease or a particular outcome is clustered or is it different geographically from all the other areas around**, and then I also have to **deploy the data in a meaningful way** and in an effective way and **in a way that public health analysis can understand** and sometimes **also public users can understand**, and these imply some complicated **thoughts on interface, on how I display the data**, which **kind of media** decide to use because in the health sector there are different priorities for deploying data, most of the time it's **paper**, the media, so you need to think about how to **format the data in an effective way**, so it's a long process that starts from **knowing how to model the core of geographic analysis techniques to this is a new area**, a change for me, I'm going, let's say **geographic visualisation; geovisualisation**. Although geovisualisation relies more on the technique and on the **technology** rather than on the concept, you've got to think the **concepts that are coming from different domains**, but this is a different topic."

Research Lifecycle

Literature Review – Start of the research process – already includes implications on data collection

"(..) literature review is always important for my analysis of course, but I don't collect data, **I have data that has been collected before** [e.g. by PCTs] and in a way this is a bit of a pain for a researcher because researchers want to have control of everything and in a delicate part like data collection (..)". {{"Some researchers are unfortunately constrained by the fact they take the data without knowing sometimes the specific purpose or how a researcher should know how the data had been collected, so it's the first struggle of getting the data in my case that has been formatted and collected before."}}

"**Working in partnership with the Primary Care Trust** means that you have to **read all the NHS documents** in my case, that means that I have to **browse into this deep, big documentation of the NHS**, the sales target and everything, so basically I would say from the outset my main sources are my **Primary Care Trust that publish documents and these are the guidelines that I have to adhere to**, and then of course all the **other case studies** and all the **other material** and work that all the **other PCTs and all the other researchers are doing in the Health area**; for example, there are some points of reference in London that are probably the **London Health Observatory** just to say one, (..) and in terms of scientific research from my side is the **ordinary academic literature search**, so whatever I can get; scientific libraries and **being at UCL I'm pretty lucky because the library is pretty big**, **electronic research**, **electronic sources** and whatever I can get via **books**, whatever."

Data collection process

"So I would say probably the **data collection is something that we skip**. I also have **consultations with an analyst** that is something that is, as a researcher, is good, especially in a domain that is really hard to approach that is the **health domain that has different semantics and different agents** playing in it."

EXAMPLE of a concrete project/research activity (summary of research process, also includes **data analysis implications):** "I'm actually finishing some work for this **Primary Care Trust in London** and they have a deadline of a **report**, and this work was about teenage pregnancy, in particular **this part of London has the highest rate of teenage pregnancy in all of the UK**, so they're trying to address this, and it's a delicate problem isn't it? Because teenagers and pregnancy; you have a particular delicate age and whatever, so last year we tried **to propose an alternative method to try to reduce this**; it was targeting people and trying to reach these people, these cases, so we tried to **follow three different ways of targeting people by looking at their life**, basically; what does a teenager do? Basically he or she goes to school, they are probably registered to general practices and then they stay home in our small model, so what did we do? We basically **analyse the overall frequency of teenage pregnancy** in Surrey and **see where they were located** and see how much of these we can **target choosing different ways, geographical areas**, so we find out that if we want to target houses, so where they live, we need to use a lot of Unis to reach the eighty percent of these teenage pregnancies, if we use a school we can reach a higher percentage of this, if we use the general practice, the registration, we can reach a different hundred percentage, these are actually **all guidelines for health analysts, health officers to choose particular geographical areas to target**, and then the **analyses were also a basic statistical epidemiological analysis on the location of all this teenage pregnancy**, the overall **geographical clusters** that are

around and some **other basic statistics**, so last year, the model itself or the analysis itself was trying to point out that it's better to actually target schools so you can reach a higher percentage of teenage pregnancy; why targeting and why this? Because teenage pregnancy are a cost that is mainly imputed to people and to the Primary Care Trust; I know it's sad, it's all about the money, and also because being the highest rate of teenage pregnancy is also a constraint for the Primary Care Trust itself because it needs to target this NHS set target and the key issues (..); unfortunately [teenage pregnancy in this area] being the highest one in the UK it is actually one of the targets, so the **next analysis goes back to the new data collected and sees something has happened**, so it's sort of a **sequel of the other analysis that was suggested two years ago**, and this is actually a real case; I've just sent the analysis, and for this case, we now go to the **deployment of the data side, it's a paper report**, so what they **sent to us was basically an Excel spreadsheet with a lot of graphs and tables** but **when I did the analysis I used something different**, but **you need to choose files and software that are actually known by anyone**; not only geeks and geographical professionals like me or other people."

More on sources for data: "For the **health domain** I use websites that have information that can be deployed online, publicly, and I can download it, but, for the data that I have to use for my analysis, for example, teenage pregnancy in this case, individual level; patient to registration to a GP, unfortunately there's nothing I can do and I have to rely on the **NHS system and to browse this system I have to be in a place that is into the intranet**; the NHS system is pretty complicated and closed to the outside world as it has to be, and sometimes to broach some data you need to use a 'smarter terminal', a **VT hundred terminal** that is something that was deployed twenty years ago".

"(..) **geographic data** I used to basically get the data, the ordinary sources like **Edina, Digimap** and all these **web internet sources [UK Boarders, DAFPD files=old field postcode directory files for linking postcode with coordinates]** that I use to download the data, the **Census** and all these kind of sources."

Data analysis – modelling – geovisualisation

"On the other side, the **modelling and the calibration of the model, the analysis of the results are less biased**. Sometimes they are biased because in my domain that is **medical geography, health geography**, there are some **directives** and we have a big boss that is the **NHS** that sometimes **sets priorities and targets sometimes without knowing geography**, so you have to struggle with that, but in the scientific process that we were talking about before, we probably miss the data collection part and sometimes I miss the real outcome of my analysis because what I do at the end of my particular research, I pack up my analyses and I give them to the public health [experts] finally that should make some decision about that, and then target these areas and then reach these people."

{{"Then it means that it is a really long, long process that sometimes is diluted in times or diluted in the bureaucracy of the NHS and so you don't actually see if what you did earlier was effective or not;"}}

"(..) sometimes we have **feedback** and so far so good;"

Use of tools/technology:

"I use **GIS total software** that can help and improve the speed of certain analyses that were **done before on paper or calculators**, so **I rely a lot on software** and I rely on that I became some sort of a **power user rather than a developer**;" {{"it is a big difference in computer science and this is what I see, especially for software between the **users and the developers**, it's like these two universes that don't speak to each other and one has to work for the other and they should have some synergies."}}

"I was actually appointed as a lecturer and I am **teaching programming for the GIS** but I don't want to feel too much like a developer because I want to feel and listen to what other users need."

"GeoVUE (..), the computation and technology in this are really important because when you **deploy the results of my analysis**, sometimes there are **graphs and charts** and it's okay, you can find some **effective ways of displaying data in charts and tables following some simple guidelines** and rules that are set not by the software but by the researcher themselves": Tools developed by GeoVUE make the process easier.

"(..) other outcomes are also **maps**, and maps are pictures, and pictures are worth ten thousand words as they say, so I **use GIS a lot to prepare my maps** and then **image editing software to prepare my maps** as well, so again, technology is involved a lot and with the invention of **Google**, things changed a little bit because they actually **provide some sort of browsing of geographic data in an effective way**; **base reference maps, basic geographic information** but if you think about it, after the deployment of **Google Maps and then the Google Maps API and Google Earth**, there's been a **big interest by the public users** about browsing geographic information and generally your public user wants to know about where they are, what's around and all these things, and this, for geographic analysts; I'm talking about analysts in this case, not about the developer; means **new ways of actually visualising our data using some pretty good base reference maps**, and from my point of view again geovisualisation, you know how important this is, so all this work entangled in a big scenario of trying to **deploy geographical data in an effective way** and tried to do it in a way that is **easy**; web **internet browser, no plug-in, nothing**; we're talking about for example Google Maps, so not even the plug-in of Google Earth that you need to download; formats that are easy to use; KML [=Keyhole Markup Language is an XML-based language schema for expressing geographic annotation and visualization] with all the problems that are related to KML and an easier way to program web pages using the API's, so all these things together brought me the interest into using Google Maps to try to explain my maps a little bit more (..); [in a] **GeoVUE project** with people that were more interested in the development side of it, (..) [including interviewee 2; see experience report 2] (..) a tool, a software [was developed, the **GMapCreator** (http://www.casa.ucl.ac.uk/software/gmapcreator.asp)] to help me and other people to work with Google Maps and to try to put their data on top of Google Maps, so you now see how all these things get back together; I'm really interested and proud of being part of this (..) because actually one of my outputs that I then developed, the **London Profiler** [http://www.londonprofiler.org/], is used as one of these **use case**s as we were saying before."

"(..) my two big outcomes of my analyses in terms of deployable and deliverable are actually two websites, so for me, it's easy to say that **the web is important** mainly

because in the world we live in; in the city we live in, the internet is pretty well accessible to anyone; (..) even if you don't have a computer you find a library, you can go and you can browse, so let's say **that is one of the main media to reach people**. You don't need to buy any software because the **internet software is free** and if they work or not work is another problem. **In my case you don't need to download any plug-in or anything for my website and my website is also very, very light to load** so you don't even need a very fast connection and you can reach a lot of users. (..) you need to focus maybe on what are your users, but if you focus your website to a particular kind of user and you **allow them to interact**, that is one of the other key points, you know, all the **key points of the web 2.0 is participation and social networking** (..); for **analyses as well, using methodology that had been documented and improved** [openly, through the web]";

Tool development, exemplifying adoption and use: "**GMap Creator, London Profiler and Maptube** [http://www.maptube.org/] are related and are actually ordered chronologically" in the timeline they have been developed at CASA: "I was involved partly at the end for some cartographic hints and other hints on its use, so we developed this software [**GMap Creator**]; it is a software that allows you to load a shared file for the none-geek, [i.e. not] GIS people, the shared file is (..) the de-facto standard in the geographic information deployment and is a form of file for mainly vector data. You can have them in points, polygons and lines so mainly it is a file that contains boundaries and statistical data related to some sort of geographical units, so from that, loading that into GMap Creator doesn't require you to know what a shared file is or whatever so it's for the end users, again, **GMap Creator allows you to load a shared file and then to create a thematic map into this software and then to have as output, a Google Map mash-up**, straightforward, but this Google Map mash-up should be so good that the output itself is like an HTML page, so something you can click, you double click and you see in your web browser, so it's targeted at the general user; (..) so my idea was why don't we deploy a website (..) to basically build a sort of portfolio of mash-ups containing data that we at UCL and CASA deployed, so the **London Profiler** (.) [is] basically **a website that shows socioeconomic data for London, and this socioeconomic data comes from different sources** (.) [e.g.] **the census or the organised statistical sources** but **some of them are actually created by us, by the researchers**, so there's the **added value of a new way of visualising the data** that this GMap Creator applies to something, [the] London context, and also the data themselves. And it is actually a successful website; we have a lot of hits and it's **browsed by different kinds of users and different people that I would never expect**, but you focus it to someone and someone else comes along the way. **Maptube is the natural sequel of London Profiler** and it is more web 2.0 oriented and more than the London Profiler (..) you can load whatever part of the world you want and it allows you to see **different layers, one on top of the other, playing with transparency and everything**, and it also allows you to load **different types [of data] coming from different services**; (..) Maptube is more open in terms of things that you can do and has been watched and used a lot where media have been involved a lot in this **London Profiler and Maptube websites, they have been used for the BBC** I think, Look East programme, they have been featured into the BBC radio a lot of times, they've been featured in Channel 4 so we had **a lot of interest by media** and general public into

this because they are probably a website that is effective in what they want to do that is **displaying geographic and other contextual information in a very easy way**."

Concrete example of use (GMap Creator, London Profiler, Maptube): "(..) example of other users that use the website; basically our website itself shows different socioeconomic data for London, so you can browse, you can zoom using the Google Maps API, you can watch at different stages and this website was actually **used a lot by real estate agents and companies**, so far away from what I was thinking about; why? Because basically when you sell a house, you need to also provide contextual information about the area itself; (..) so this website that contains information about ethnicity, information about involvement of Information Technology by people that contains data about higher education, that contains data about health, that contains data about crime, transport and other house prices was actually used to sort of foster information for the prospective buyers to buy a house, so I think this is actually really effective in a way; it's not an academic use but it's a really effective use, it's been used a lot in terms of data and support, now, **if we want to go more into the academic and deployment of the results domain**, I can tell you that I used this **to actually visualise some health data** that has been deployed again at London-wide level; it's actually an interesting reserve, this kind of data coming from four years of hospital admissions for the whole of London, **we calculate the risk of hospitalisation for every single postcode but in this case we deployed it I think at output area level for thirty different long term illnesses**, so if you type in your postcode you can see a map that is showing your risk to be hospitalised; (..) [looking at user generated maps sometimes] I find out the strangest correlations ever that people submit to me, but this is what you have when you open information on a media like the internet."

On coupling data from different sources: The "process is what is called 'loose coupling', meaning you download the data and you use the data and you couple with the data that you have using a join or whatever in here so it's a sort of loose coupling, it's not like completely integrating".

Collaboration (and issues of sharing confidential data)

"(..) the **health domain** of my research does actually contain people that are not even from academia; different sectors, probably different backgrounds as well, so different languages as we say, and data sharing is a bit complicated because of health; we have some restraints, also for deploying the data themselves": confidentiality and security issues of health related data: "we are **not constrained by the location** so we can do our research (..) [in the] office and **send the analyses straightforward by email**; the coordination in the term of exchanging data, like **physically exchanging data**, is mainly unfortunately for health data (..), so **I have to go to the Primary Care Trust, get the data, encrypt it in a 128pgp encryption** because people are really concerned about data themselves and I understand because sometimes I have to analyse data that are a bit **delicate, like those teenage pregnancies for example**"; "(..) whenever it is possible that means it doesn't breach any confidentiality issue, the **data can be sent by email, encrypted** again." This kind of data usually lies on a "NHS server, protected (..), chequered with any kind of firewalls and encryption that you can think of".

"In the **academic domain** with my **colleagues, data sharing is really high**, especially in **CASA**, there's a **good flow of information and data** between us that is mainly **managed by emails and [f2f] meetings every week**".

Dissemination

The interviewee does "target the main publication in my research field that are mainly on GIS and applied research, so **GIS, geographic analysis domain** on one side (..) and sometimes also sort of **special planning journals before the outcomes**, like the websites and the ways of visualisation of this data and **also all the other health publications; health journals and medical geography journals**." {{"The problem is that the health domain of publication is pretty wide and the health geography is a small bit of it and sometimes it's really hard to find the right journal to submit your papers to because most journals are interested in the policy, in the social implications of your analysis and when you actually submit some papers you need to provide that contextual information so all the health policy environment where these analysis is embedded in (..)"}}; "(..) and of course the **geovisualisation** one, and **book chapters** as well." The recent book chapter is "about this geovisualisation tool I deployed [the London Profiler], (.) another paper came out on a **journal of applied spatial planning** (..)."

Combining publications with source data: "(..) my own perspective, and this is also typical of **geographical analysis** (..) [which is] **interested in the tool itself**, you know, deploying the mathematical algorithms for the tool. For **my research I can release data and publish data at the level that I can**, especially for the health [area], so I **publish maps, tables, whatever I can provide** and whatever they can have because unfortunately, again, we have a lot of rules that try to **preserve confidentiality**".