Dani Arribas-Bel

Martin Fleischmann

Afternoon of 6th June. The academic year is coming to an end, and we have arranged a meeting with Martin Fleischmann and Dani Arribas-Bel, despite their many teaching responsibilities. Both are members of the Geographic Data Science Lab at the University of Liverpool. Dani, who was born in Zaragoza and trained as an economist, has developed a research career in modern computation and new forms of data to understand cities. Martin is an architectural researcher with strong links to urban morphology and a promising young researcher at the International Seminar on Urban Form (ISUF)[1]. Their prior experience and their work together in Liverpool prompted us to discuss, from an academic perspective, data applied to the analysis of the form and behaviour of our cities.

Pablo Martínez-Díez (PM):

Dani, Martin, thank you so much for agreeing to do this interview. We are very excited to include your reflections in this special issue. We have prepared some notes that might guide the conversation. We would like to discuss three main aspects: the opportunities of working with data, to understand where we come from and where we are going; your innovative *Urban Grammar* project, which applies artificial intelligence to the study of the form and behaviour of our urban fabrics; and your sense of responsibility in your work.

Dani Arribas-Bel (DA):

Thank you all. We look forward to discussing the form and behaviour of cities. This call for papers is really engaging.

Sergio García-Pérez, Pablo Martínez-Díez and Mar Santamaría-Varas (SGP, PM, MS):

Dani, it's been nearly a decade since you wrote 'Accidental, Open and Everywhere: Emerging Data. Sources for the Understanding of Cities'[2], an accurate text on the opportunity that the digital footprint offered for studying human behaviour. It's an inspiring vision about becoming an 'urban scientist'. However, the debate about the scientificity of the discipline of urban planning is long standing. To what extent has the digital era improved the processes and scope of urban planners' work?

DA:

I was in shock when I realised that 10 years had passed. I wrote it 10 years ago, although it was published a little later. So yes, it's 10 years old. I think it came at a very optimistic time, with the peak in new forms of data of the early 2010s,

1 Readers interested in more detailed CVs, please visit the following links: https://darribas.org/ https://martinfleischmann.net/

2 Daniel Arribas-Bel, 'Accidental, Open and Everywhere: Emerging Data Sources for the Understanding of Cities', *Applied Geography* 49 (May 2014): 45–53, https://doi.org/10.1016/j.apgeog.2013.09.012.

Granularity, consistency and scalability in morphological studies. On some recent Works by Dani Arribas-Bel and Martin Fleischmann

SERGIO GARCÍA-PÉREZ
PABLO MARTÍNEZ-DÍEZ
MAR SANTAMARÍA-VARAS

Sergio García-Pérez, Pablo Martínez-Díez, Mar Santamaría-Varas, "Granularity, consistency and scalability in morphological studies. On some recent Works by Dani Arribas-Bel and Martin Fleischmann", *ZARCH* 19 (December 2022): 186-197.
ISSN versión impresa: 2341-0531 / ISSN versión digital: 2387-0346.
https://doi.org/ 10.26754/ojs_zarch/zarch.2022197412

which was clearly a force for good. That's when Google was using the slogan 'don't be evil', when using Facebook was still fun, when a little thing called Airbnb appeared. It was almost a different age. I think it would be difficult to write such a paper today in the current climate and, in a way, it's good that this is the case. That's because we've already come a long way, so it's probably not so novel to say that so many new sources of data are available.

At the same time, I think we're starting to realise it wasn't all that great and there were aspects of these technologies that at least I was overlooking. I think now we're aware that it's not such a fun place to be but, ironically, it may be more useful. These 10 years have been a unique opportunity to learn about dealing with data. It's something I said in the paper, but probably, it didn't have enough relevance. So, I think we're now beginning to take a more sober and, therefore, more useful view of how these sources can and should be used.

Disciplines such as planning or geography—perhaps most social sciences—have been rather reticent to introduce these sources and techniques that we might discuss later. At some point, they have realised that it's hard to carry on with business as usual, when business is not so usual. When everything in our lives has changed so much in 15 years, it's really hard to argue that your day-to-day should remain the same. I think that, little by little, the most traditional parts of the disciplines are waking up to the fact that we live in a different world; a world that certainly has an abundance of challenges. We might as well try to make the most of the benefits that the world presents and be optimistic. Are we really going to be here in 5 or 10 years? Who knows? But I think there's definitely been a change in academia—I don't know so much in practice. That article took two years to be published. Now, it's a topic of growing interest, maybe too much. But, as I was saying, we're starting to take a more sober perspective.

SGP, PM, MSV:

Changes introduce advantages and, although we must be optimistic, we must also consider the disadvantages. In recent years, we have seen the influence that data management is having on the real estate sector, for example, with a clear impact on our cities. This dark side has led to more privacy-conscious users, and even legislation seems to have moved in that direction.

DA:

On the one hand, in writing the paper I tried to capture what people were feeling at the time. Privacy was part of the list of things to consider, although it wasn't at the top. Now, privacy should be put at the forefront and actually help lead innovation of more privacy-conscious technology.

Some things have happened in this time that might have made you change this list of priorities. Currently, we're aware that, in fact, the most granular data are not the most useful for many applications. This seems at odds with the idea that data is a privacy-breaker. It's a false dichotomy.

I think we can do better. Rather than saying that we either give up privacy, or we give up new and exciting sources of data... why don't we think how to combine datasets—or aggregate them—so that they remain useful while preserving privacy? In this sense, we're only scratching the surface of what we can really do.

On the other hand, granular data is really hard to work with. It requires specific skills, time and resources that not everyone has access to. It's important to consider the level of professionalism that granular data requires. So, how can we generate synthetic datasets that retain statistical properties but do not reveal any private information? How can we analyse and learn without seeing the real data? The dichotomy of data and privacy is no longer valid.

MF:

If I move away a bit from privacy, there is another issue that's very similar to what we're talking about. We have so many datasets that come in so many different forms (accesses, structures, etc.) that only a data engineer can sift through them and turn them into something useful. It's a great challenge, because one person doesn't necessarily have the skills to complete all those processes, even if they are open. If we find a group of people to spend their time 'chewing' to ensure the process of releasing and anonymising the data, we will simplify some of the work. It would be very useful for the rest of the community.

SG, PD, MS:

The recently published work 'Open Data Products-A Framework for Creating Valuable Analysis Ready Data'[3] tackles this idea; a call for the need to think about processes that make data more accessible. A way to simplify the abstraction of data (whether public or private) and transform it through transparent methodologies into open services that are accessible to people without advanced computer skills.

DA:

That work was jointly developed with some colleagues from the Geographic Data Science Lab[4], but of course, the wider community shares the idea. It's a natural evolution after seeing how the amount of data we have has increased in these 10 years. It's a bit like the Wild West though... we have more data, which has a value but is not ready for analysis—or at least not for the kind of analysis we need. It's important to consider what to do with it and how to make it useful. Many of these datasets are now accessible, but not fully finished. They cannot be treated as the traditional data we were used to working with. Rental listings, GPS traces and satellite images contain potentially useful information, but it's not easily usable. They are not spreadsheets that we can download from the census office to make the map we have in mind. They require more work from us.

But, at the same time, these datasets unlock countless possibilities. Going back to the example of the census—which is updated every 10 years—its low frequency reduces its granularity. All the interesting activities occurring in the 10 years between updates are not recorded in the database. These new data sources—many of them real-time feeds—offer extraordinary possibilities but are not ready to go. We're moving towards a world where there will be no canonical dataset for analysis. Until recently, the canonical datasets in urban planning were the official cadastre and local planning. We thought we were using that because it was the best resource we could use, but we were actually using it because we had nothing else. We could have intellectual debates about what would be the best means to measure income, but the only means we had was the census. So, in a way, it was an intellectual experiment.

In the world we currently inhabit, that thought is no longer intellectual. It's a real empirical question: What data do we use to measure certain things that we care about? And the challenge is how to create descriptions that are more accurate than before with the new forms of data? The starting point is a full, 24-hour picture of the planet, or a real-time snapshot of the places people shop with their credit cards, and so on. The challenge is how to translate these big datasets into analytics, into useful, ready-to-use data. This path raises very interesting questions about the design of these datasets, about their governance, their reliability, their sustainability. Frankly, most companies that generate them don't last more than five years. The people who were part of this ecosystem in 2015 are out of business or have been successful and bought up. It's a really important question about sustainability that we might have not asked ourselves before. That's why governance is so important. In this new world, realising that is just the first step.

3 Dani Arribas-Bel et al., 'Open Data Products-A Framework for Creating Valuable Analysis Ready Data', *Journal of Geographical Systems* 23, no. 4 (1 October 2021): 497–514, https://doi.org/10.1007/s10109-021-00363-5. infrastructure, analytics or a combination of all of them, where each step of development is designed to promote open principles. Open data products are born out of a (data

4 https://www.liverpool.ac.uk/geographic-data-science/

MF:

Also, what will this huge amount of data bring about? I would expect certain level of standardisation. There have been countless discussions about how to store and catalogue data. Releasing data should involve following certain standards, certain common objectives focused on constructing datasets so that they are consistent, comparable. A few years ago, everybody did it their way. The EU is making progress, thanks to the INSPIRE directive[5]. Before, it was unthinkable that all EU cadastres could be consistent. Today, we're getting there. I hope these experiences will reach other aspects of the data world: being able to process in a standardised way from mobile phones, from GPS navigation... I see this process generally happening from the bottom up. Surely in 10 years we will rely on standards to help us.

DA:

Standardisation is a controversial topic, but of necessary interest. I think standardisation is a good thing, in certain aspects, but I don't think it has to be the top priority. Firstly, because we could end up with a new canonical dataset. It's interesting to think conceptually about the right way to measure something and then have the means to do it. A new canonical dataset might not contribute in that direction, by forcing us to perform a certain type of analysis, because the data is collected in a specific way. Secondly, because once things are standardised, they don't change any more. INSPIRE is a good initiative, because all countries are doing similar things, but they are all using a 20-year-old format. Trying to open standards is good but letting us approach it in different ways is also good. We definitely need more diversity.

SG, PM, MS:

This reflection reminds us of Richard Sennett and his concept of 'open'[6]. Open is adaptable, it has the capacity to evolve. Standardisation can run the risk of moving away from this concept of openness.

DA:

Yes, Sennett is very inspiring. His views on technology often help us to think about how to do more, rather than less. Standardisation plays that role; it helps us to connect more datasets, to use open tools in a more interoperable manner. Ironically, standardisation, rather than being about having things in a single format, is about having an ecosystem of open tools and standards that allow us to plug and play.

SGP, PM, MS:

The idea that data not only helps to answer new questions (climate adaptation, social segregation or gender equality), but also to approach 'classic' studies—such as Jan Gehl's analyses of the quality of the ground plan, Kevin Lynch's perception of the city, or Jane Jacobs' urban vitality)'[7]—is also implicit in these texts.

MF:

The foundation for all these works by architects, urban designers and town planners was always a solid intuition, although they were driven by empirical evidence. When one tries to translate that intuition into a script, it loses its essence—especially if one tries to apply it to urban design in order to establish rules about what a shop should look like, its position or its façade. It's hard to rely solely on numerical analysis, because eventually we could end up with the same cities everywhere. Can you imagine something fully parametric? It's true that we have enough data and tools to do that. But it's not clear to me that something fully parametric is good enough. And that is coming from me, who is pushing a lot for quantitative data! I think that, in the end, it would be a step backwards in architecture.

DA:

I suppose it can be understood as a back and forth between big-picture thinking and fine-grained data. Like a pendulum. And it's hard to find a balance at any given time, because we're always products of the time we live in. In a way, in the world of Jan Gehl, Jane Jacobs or Kevin Lynch, as we said before, these questions were intellectual. They didn't have opportunities to think with numbers, with computers. And they got a result that paved the way for at least fifty more

5 The INSPIRE Directive aims to create a spatial data infrastructure in the EU for environmental policies or activities that may have an impact on the environment. https://inspire.ec.europa.eu

6 Richard Sennett, *Building and Dwelling: Ethics for the City* (London: Allen Lane, 2018).

7 Such as Xavier Delclòs-Alió and Carme Miralles-Guasch, 'Looking at Barcelona through Jane Jacobs's Eyes: Mapping the Basic Conditions for Urban Vitality in a Mediterranean Conurbation', *Land Use Policy* 75 (June 2018): 505–17, https://doi.org/10.1016/j.landusepol.2018.04.026; Gabriele Filomena, Judith A. Verstegen, and Ed Manley, 'A Computational Approach to "The Image of the City"', *Cities* 89 (June 2019): 14–25, https://doi.org/10.1016/j.cities.2019.01.006; Damiano Cerrone et al., 'How Many, Who, Where, What, and How Long: Public Space in Russian Monotowns through Jan Gehl's Theory', *Sustainability (Switzerland)* 13, no. 9 (1 May 2021), https://doi.org/10.3390/su13095105.

years of urban work. It's like the theory of relativity, which took about a hundred years to be proven. With cities it's kind of the same thing. I think we're at the other end of the pendulum now. I'm not going to say this time is easier, but now we're trying to integrate and embed new data sources and flows into thinking about cities. This is precisely what I would like historians to remember about this part of urbanism.

However, there is a risk that we completely forget to think about the big picture, acting as if these ideas are facts and these notions are laws. As Martin says, I think we definitely need to consider the big-picture thinking part of the world because, ultimately, data does not come from the sky. We create data and the way we create it reflects what we want to do with it. The last 10 years we have reacted to the new data landscape, but in the next 10 years we're probably going to get closer to what this data landscape wants to be. And, if we forget that data by itself will not solve our problems, we will probably waste the next 10 years making very fancy maps that tell us very little about what we want to know.

SG, PM, MS:

Thinking with computers is not the same as thinking without them. Now, you are starting a new line of research, the project *Urban Grammar*[8]. You think with computers to try to describe and understand the form and behaviour of the city, of the territory. Can you tell us a little about how the project is progressing and its impact?

MF:

You just mentioned that we're trying to think with computers to do *Urban Grammar*. I'm not sure if that's really the approach. I would probably say that we're capturing many different aspects that describe the city: from streets, layouts, buildings, their footprint, their size, their density, the location of shops on the streets or the distance to the nearest body of water. The things we can know, from the architecture to the population. The challenge is to combine these variables and develop a classification of space that helps us to understand these urban environments. The method covers the whole space, and it combines form and function into one classification. So, it not only assesses the morphology of the place, but also how the place works. By combining and simplifying these hundreds of variables, we aim to find 15 or 20 different types of urban (or non-urban) environments. And we can then measure all urban environments and compare them. We're looking at similar units across different cities.

DA:

In a way, we're thinking with computers. That classification is not something that we pull out of a hat and say "these are the 16 classes that we think we should have". That classification actually comes from a data-driven approach that's effectively done by a computer. But, at the same time, we're not quite giving the "raw" data to the computer so that it can tell us what we're looking for. We're trying to inform it with theory. We feed the algorithm with considerable experience and knowledge that have been developed in recent decades around urban morphology and urban environments. And what *Urban Grammar* tries to blend it all. It's about "letting the data talk to the computer" about what we already know of cities. We don't want the computer to spend a lot of time reinventing the wheel; we want the computer to do what is really hard for us: sift through a myriad of numbers and values and synthesise them into more digestible classes.

I should add that we're not the first to try to do this. Our contribution is more about being able to combine granularity, consistency and scalability at the same time. We have developed a very granular classification. The current example for Great Britain is based on 14 million spatial units. It's consistent—because we developed it in the same way for the whole of Great Britain—and scalable, because we're prepared to adapt the methodology to other contexts. In our experience, it's difficult, if not impossible, to find these three characteristics at the same time in the literature. We often joke from those three, you can choose any two of them and you will find a large number of papers: those that are granular and scalable, or granular and consistent, or consistent and scalable. Achieving these three characteristics at the same time is the great value of your project.

SG, PM, MS:

You talk about granular, consistent and scalable 'spatial units' of analysis. For a second, let's take a deeper look at this point. How many spatial units did you say? How did you define them? Is there any geometrical particularity of these spatial units?

8 For more detailed information on the Urban Grammar project, please visit: https://urbangrammarai.xyz. Some research papers on the project have also been published recently: Krasen Samardzhiev et al., 'Functional Signatures in Great Britain: A Dataset', *Data in Brief*, May 2022, 108335, https://doi.org/10.1016/j.dib.2022.108335; Martin Fleischmann and Daniel Arribas-Bel, 'Geographical Characterisation of British Urban Form and Function Using the Spatial Signatures Framework', *Scientific Data* 9, no. 1 (7 September 2022): 546, https://doi.org/10.1038/s41597-022-01640-8.

Figure 1. Example of enclosed tessellation.

MF:

More than 14 million, more than you can imagine. The spatial unit is defined by the enclosed tessellation, a measure of the morphology of the place[9]. It's not the plot, it's the equitable distribution of space between buildings and it makes it possible to compare urban fabrics where the plot (as an administrative, fiscal or geometric boundary) is not consistent throughout the territory. We have areas of historic centres where the tessellation is very small, representing approximately a single building. However, if we look at a modernism fabric—where the plot is usually associated with the building and not with the open space—the enclosed tessellation distributes that space among the buildings. Therefore, the tessellation covers the space, whether it is urban or not. In the countryside, tessellation units are often delimited by elements such as rivers, roads or railways. These units can be very large. Tessellation can adapt to the context. This is one of the key aspects of this spatial signature. Morphological tessellation is adaptable to the urban environment and always collects information at the same level, not necessarily from the same area.

DA:

I would like to say that this is an example of how to think conceptually about what would be ideal, and actually do something that comes very close to it. In this spirit, we spent a lot of time reading, thinking about the right spatial unit to work with in terms of form and function. And then we got as close as we could to a consistent, scalable unit. The result is the enclosed tessellation. We have to compromise; we always have to compromise. But, compared to any other option we evaluated—and we really thought about many of them—this was the closest to having a conceptually meaningful spatial unit, applicable to other contexts.

9 For more information on the spatial unit 'enclosed tessellation' see the article by Daniel Arribas-Bel and Martin Fleischmann, 'Spatial Signatures - Understanding (Urban) Spaces through Form and Function', *Habitat International* 128 (October 2022): 102641, https://doi.org/10.1016/j.habitatint.2022.102641.

SG, PM, MS:

Form and behaviour have a different temporal dimension. In contrast to ecosystems, land uses or buildings tend to be more static (or have a long-term transformation time); function is more dynamic. A regular home can be transformed into an office in a short time, or one economic activity quickly replaces another. How does *Urban Grammar* deal with this temporal dimension?

DA:

There are two answers to that question. The first is more conceptual, about how form and function feed into each other and what their timescales are. I think you are absolutely right. Form is a slow process and function is a bit faster. But, at the same time, function doesn't move erratically. It always happens in the context of the form and, in many ways, it's constrained by it. So, I think the value of *Urban Grammar* is not necessarily knowing the relationship between both in real time, it's rather knowing how form and function fit together, what are the constraints of form on function, through a study of the slow process.

The other answer is related to the second part of the project, on which we're working now and about which we have not yet published too much. We have been trying to bring the static classification of forms into faster ways of updating. To do this, we're using satellite imagery. The vision (not the product yet) is to be able to take a satellite image, extract the footprints of buildings, and, maybe, extract other data for the form and function domain and apply the *Urban Grammar*. And then get much more granular data on how cities change. I don't think we'll get to the point of giving a real-time view of how function changes. I don't think that's what we're trying to do either. Function is fast, volatile... high frequency, and we could get lost in looking at what's happening every second when maybe what we need to do is look at the longer-term trend. *Urban Grammar* fits more into this longer-term trend.

PM:

We would like to congratulate you because when reading the project documentation, the effort on the replicability and traceability of your work is evident. For us, there is a highly complex moment that perhaps concerns some of the conceptual issues we have been discussing. It's the moment of clustering the results; the moment when, from all the variables analysed, 16 morphologically similar regions are classified. How is the description assigned? In one way or another, we believe that we're returning to a traditional methodology: Can we identify with a historical centre, with a residential periphery or with a business centre what the computer has grouped together?

DA:

There is a technical answer about how we actually chose 16 regions instead of 17 or 15 and then there is a more conceptual answer that brings us to the point we discussed earlier. Our work is truly hybrid: it's not a data-driven classification and it's not a totally theory-driven classification. At some point, some people have a very hard time with this and say "well, but doesn't the data we get come from data?". And it's partly true. We're trying to be open about how we do that selection. Ultimately, I mean on the technical side, we're doing unsupervised learning. If we already knew the classes beforehand we wouldn't be doing this exercise. This process is repeated when we decide what features we show to the algorithm, and in what form. That's totally subjective and we're convinced that being subjective is fine. That's the point where we can embed everything that the not necessarily data-driven literature has given us during this time.

There is another point in the chain in which computers are much better than humans: in trying to group what is similar. What we did in the project was to try to be as open as possible about how we made that choice. Because, ultimately, we're taking 14 million data points and summarising them into 16 classes. It's a trade-off between detail and abstraction. And no single point is optimal. The idea of optimal doesn't really work in this case. Ultimately, it's our decision, we try to make the most sensible one, making our decision open and being transparent about how we make it.

SG, PM, MS:

After these first results you have obtained for the UK, with the classification of more than 14 million spatial units in 16 morphological regions, what are the next steps for the project? Even if it's not yet published, although we're sure it will be soon, what can we expect to see in the coming year?

DA:

Maybe one year is an ambitious timeline. The next immediate step is the discussion about satellite imagery. Getting a computer to recognise these classes, these signatures, in satellite images. We have almost finished most of this work, although

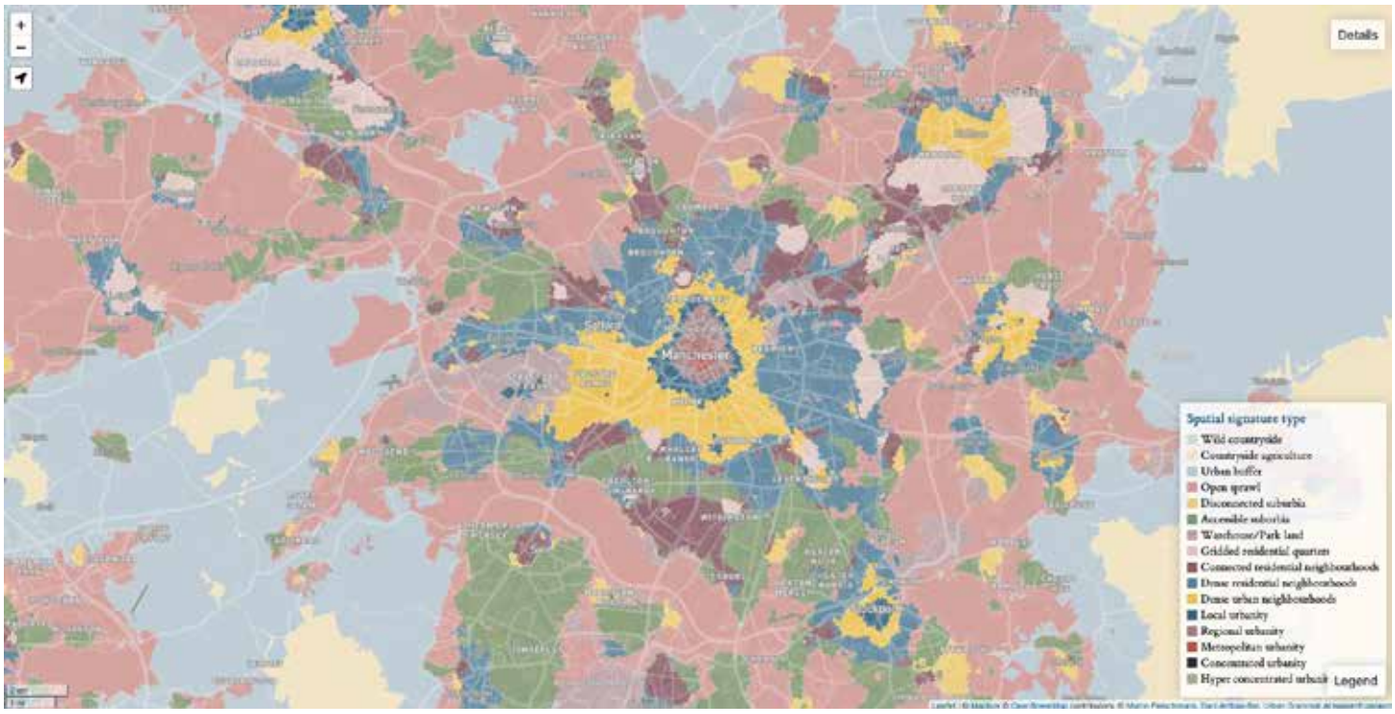


Figure 2. Sample of the results obtained from the classification into morphological regions of the UK. For an interactive version of the findings please visit https: https://urbangrammarai.xyz/great-britain/#

I don't know if we can have it written immediately. Next, we will check if what we have achieved is good, or at least good enough. We will wrestle with the question of what does good enough mean, and when can such predictions be considered valid, and whether they are useful enough. When we manage to answer these questions, we can apply that training to a catalogue of satellite images, to then start playing with the idea that spatial signatures capture how form and function evolving. The idea of *Urban Grammar* was to first develop these signatures, the building blocks of meaning. Then, the grammar looks at how those signatures change over time and helps us to identify the rules of behaviour or change. Right now I think this is closer to science fiction than to science. It's good to have an ambitious goal to wake up to in the morning.

Another challenge is to take the signatures out for a test drive. Trying to intersect them with other phenomena that we think are related. This is what I imagine urban planners and architects are interested in; seeing how different signatures relate to different levels of emissions, walkability, productivity, inequality and so on. There are many theories in the literature about how these things relate to each other but, as we don't have the data, empirical study at a granular scale is less frequent.

Then, there is another, more ambitious challenge. Maybe not for next year, maybe for the next decade. We have shown that you can do this study in a data-rich landscape like the UK. Could we have planetary signatures? In less data-rich contexts like Africa? Could we do a similar classification for the whole of Europe or the United States? What could we learn by comparing them? It's a really interesting challenge on the horizon, and it's precisely on the horizon because we need to solve all the other stuff first. There are quite compelling questions to ask about how these rankings relate—how do they differ and how are they similar? I think that's tremendously interesting and, you know, it's an empirical question as they say.

SG, PM, MS:

Sometimes topics such as spatial tessellation, data clustering or algorithm training are not easily understandable to the stakeholders interested in the project. However, you are doing a good job of disseminating and even generating graphic tools that help people who are not familiar with these concepts to understand them (e.g. clustergram).

MF:

We create different outputs of the project, depending on the target audience. We have academic papers, which help us to disseminate our results; we have code, to ensure the replicability and transparency of our working method; and we also

have outputs that have nothing to do with the above, which are, simply, clear and beautiful representations, the presentation of the 16 classes obtained, a good description of each of them, with good references (toponyms, common places) so that people can understand them[10]. We try to adapt the output depending on the audience. We like to choose the way we show things. For example, we can't probably tell urban planners the details about how we selected the 16 classes, or how we dealt with 14 and a half million spatial units. They are not interested in the total number; they are interested in the 16 classes, their definition and how they can use them. So we adapt the message and the output so that all stakeholders can understand us well.

DA:

I always think this is a bit like climate change. It's an extremely difficult and complex issue that relatively few people devote their entire careers to. However, it's important enough for the rest of the world to know about it as well. So there is a huge amount of work that goes into translating, summarising and repurposing the results from the more technical end of the spectrum to other parts that are not as technical or detailed but allow people to relate to them, to make them useful for some of their contexts. Otherwise, they wouldn't be able to access any of those works.

And something similar happens in our cases (perhaps not as relevant as climate change!). As much as I would like everybody to master Python and know about enclosed tessellation and supervised learning, I have to admit that this is not the case. People have problems that need solving and we believe that some of the stuff we do is useful for them. So the challenge is how to reframe it in a way that can connect with what they want, what they can understand, what can be useful to them. And I always think this is a bit like teaching. When you prepare a class you have to think about students' background as it determines what they will be able to learn, regardless of what you tell them. You could tell them everything, but if they only have a grounding in A, they will only be able to access B and C, even if you turn all the letters of the ABCs around.

It's really difficult for people like us, from the academic world, because it's so far out of our comfort zone. We're very used to writing in a concise, obscure way. You know, obscure is currency in academia. You want to sound dark and difficult and here you have to do exactly the opposite. It's a challenge, but it's also fun and, above all, a good ending to make our work useful for more than just 25 people in the world.

PM, SG, MS:

In addition to publishing the code notebooks that make your work transparent and replicable, you both collaborate as volunteers in the development of open source code (GeoPandas, PySAL). You are committed not only to making the code less abstract to improve its applicability for the user, but also to its generation in open format. What is the added value of working in open source?

MF:

We're asked this question time and time again. If you think about the standard academic life—where the only thing that matters is publishing—you don't really see the motivation. And maybe we shouldn't do it that way because, quite often, we're not rewarded by the institution. For me, it's always difficult to answer this question because I've never asked myself 'why should I do it?', because it came naturally to me. This is the right way to work. We're working on an investment, financed by public money, so everything we do is for the public, it should be done publicly and it should be done in a way that the public can use. If we're writing code, what's the point of leaving it on your hard drive? No one will be able to understand how it was done. Or build on it.

SG, PM, MS:

Do you find journals that recognise this open work? Is it easy to change the inertia of the journals?

MF:

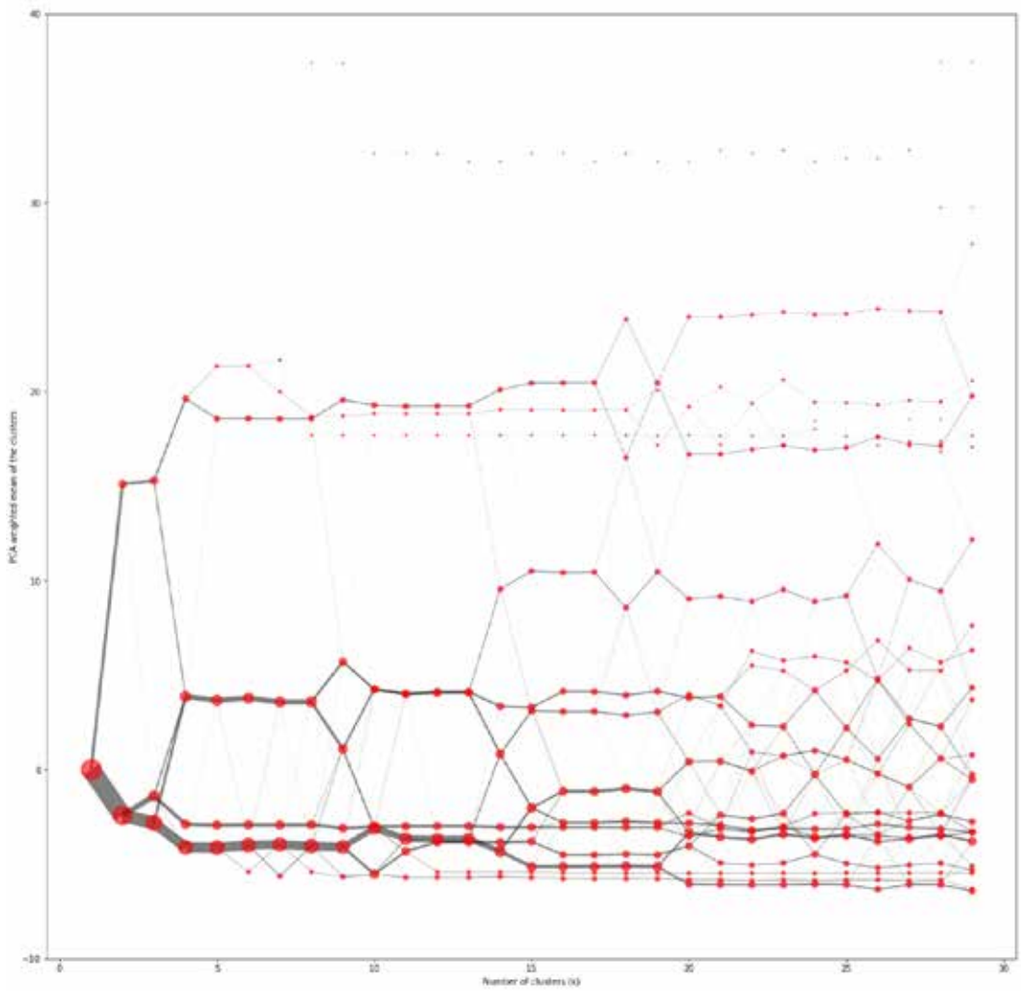
Some journals are open to publishing computational notebooks. Most journals encourage you to share the code although they don't really care how it's done. It's more like providing a link to your repository. At the same time, those journals that encourage you to share the code generally don't penalise you if you don't do it. So that motivation isn't there either. I have the feeling that it's changing but as happens in academia, the change is very slow.

DA:

I think it's changing but, as Martin says, it's changing very slowly. I've changed a lot myself over the years and my views on it have also changed. I don't think the punishment

10 https://urbangrammarai.xyz/story/

Figure 3. Example of a clustergram.



approach works—the idea that if you want to submit an article you have to submit code—, because then, people end up not submitting to your journal and go to others. Journals should do more with the current approach and say that we can publish articles as usual but, if we do these other things, they will recognise the value of the intrinsic work and will try to give us credit. In Environment and Planning B that's what we have done. We have opened a new section—Urban Data/Code[11]—that tries to unearth and rediscover some of these developments, to publish not only the article, but also these artefacts as open data products, or open source code. The idea is that an article can have a valuable impact for academia, the development of a server package or a data product that others can use do too.

I think we also have to enter into a more cultural debate about what "open" means. We should do it because it's good, and not just because it gets more recognition. Although it may seem the system is not based on traditional economic incentives, it's not that different from other widespread systems in academia (like peer review). I mean, the whole peer review system, if you told someone who didn't know about it, they'd say it sounds great, but it's quite naïve. You'll never get people do it and yet it's the only currency in research. So, I think we have to make it part of the culture. That this is how things are done and this is what is expected. And, in the same way that we ask people to review for free, we should ask people to do things openly. If it comes to that cultural change at some point, the price of not doing so will be lack of trust in our work, or no credit or reuse of it. Getting to that point will be enough of an incentive for everyone else to do it. But how do we get there? I'm working on that, but I'm also pragmatic in the sense that, as things stand now, we should give incentives.

11 Dani Arribas-Bel et al., 'Urban Data/Code: A New EP-B Section', *Environment and Planning B: Urban Analytics and City Science* 48, no. 9 (30 November 2021): 2517–19, https://doi.org/10.1177/23998083211059670.

SG, PM, MS:

It has been really interesting to see how much the data landscape has changed in the last 10 years. Your reflections from academia are a great guide for our work. Thank you for your sense of collectivity and open work. These reflections on what to do and how to do it will definitely be very useful to the readers of Zarch and will help to reflect on the role that educational institutions should play regarding these new identified challenges. Thank you for your time.

Bibliography

Arribas-Bel, Dani, Seraphim Alvanides, Michael Batty, Andrew Crooks, Linda See, and Levi Wolf. 2021. 'Urban Data/Code: A New EP-B Section'. *Environment and Planning B: Urban Analytics and City Science* 48, no. 9 (30 November): 2517–19. https://doi.org/10.1177/23998083211059670.

Arribas-Bel, Dani, Mark Green, Francisco Rowe, and Alex Singleton. 2021. 'Open Data Products-A Framework for Creating Valuable Analysis Ready Data'. *Journal of Geographical Systems* 23, no. 4 (1 October): 497–514. https://doi.org/10.1007/s10109-021-00363-5.

Arribas-Bel, Daniel. 'Accidental, Open and Everywhere: Emerging Data Sources for the Understanding of Cities'. 2014. *Applied Geography* 49 (May): 45–53. https://doi.org/10.1016/j.apgeog.2013.09.012.

Arribas-Bel, Daniel, and Martin Fleischmann. 2022. 'Spatial Signatures - Understanding (Urban) Spaces through Form and Function'. *Habitat International* 128 (October): 102641. https://doi.org/10.1016/j.habitatint.2022.102641.

Cerrone, Damiano, Jesús López Baeza, Panu Lehtovuori, Daniele Quercia, Rossano Schifanella, and Luca Aiello. 2021. 'How Many, Who, Where, What, and How Long: Public Space in Russian Monotowns through Jan Gehl's Theory'. *Sustainability (Switzerland)* 13, no. 9 (1 May). https://doi.org/10.3390/su13095105.

Delclòs-Alió, Xavier, and Carme Miralles-Guasch. 2018. 'Looking at Barcelona through Jane Jacobs's Eyes: Mapping the Basic Conditions for Urban Vitality in a Mediterranean Conurbation'. *Land Use Policy* 75 (June): 505–17. https://doi.org/10.1016/j.landusepol.2018.04.026.

Filomena, Gabriele, Judith A. Verstegen, and Ed Manley. 2022. 'A Computational Approach to "The lmage of the City"'. *Cities* 89 (June 2019): 14–25. https://doi.org/10.1016/j.cities.2019.01.006.

Fleischmann, Martin, and Daniel Arribas-Bel. 'Geographical Characterisation of British Urban Form and Function Using the Spatial Signatures Framework'. *Scientific Data* 9, no. 1 (7 September): 546. https://doi.org/10.1038/s41597-022-01640-8.

Samardzhiev, Krasen, Martin Fleischmann, Daniel Arribas-Bel, Alessia Calafiore, and Francisco Rowe. 2022. 'Functional Signatures in Great Britain: A Dataset'. *Data in Brief*, (May): 108335. https://doi.org/10.1016/j.dib.2022.108335.

Sennett, Richard. 2018. *Building and Dwelling: Ethics for the City*. London: Allen Lane.

Sergio García-Pérez (MArch, PhD in Architecture) is Assistant Professor of Urban Design and Planning at the University of Zaragoza, Spain. Funded by the Ministry of Economy and Competitiveness (2015-19) and Ministry of Universities (2022), his research focuses on urban morphology and the analysis of its quality, considering the regeneration possibilities of different urban tissues at different scales**.** As practitioner, he has worked for the urban planning office 300.000 Km/s (National Prize for Urban Planning awarded by the CSCAE in 2019), collaborating on projects related to urban analysis, housing and urban regeneration (2020-22). Research stays at *Universidade do Porto* (2018, 2022) and *Université Paris I-Sorbonne* (2019). His research has been published in journals such as *Cities*, *Sustainable Cities and Society*, *ACE*, *Territorio* or *Zarch*.

Pablo Martínez-Díez is an architect graduated from the School of Architecture of Barcelona. At present, he is the co-founder of 300.000 Km/s. As co-founder of 300.000 Km/s, he has an strong background in using new forms of information (open, social and big data) to address contemporary urban and territorial challenges through diagnosis, reports, planning and participation. He is an expert in data processes from gathering to publication. He is a pioneer in Spain in the use of open data for urban policies. He has lectured at prestigious academic institutions across Europe (TU Braunschweig, Welsh School of Architecture, Institute of Advanced Architecture, UPC Foundation, among others), as well as participating regularly in debates and conferences of significance in the sector.

Mar Santamaria-Varas is an architect graduated from the School of Architecture of Barcelona. At present, she is the co-founder of 300.000 Km/s. Her professional activity has been related in the last 15 years to digital technologies, spatial analysis, urban planning and public policies at an international level. As the co-founder of 300.000 Km/s, she has elaborated strategic reports, diagnoses, guidelines, data observatories and master plans in the field of environmental health, regulation of economic activity, new forms of mobility, housing vulnerability, data and digital divide, among others.She is a pioneer in Spain in the use of open data for urban policies (Associate from the Open Data Institute) and the fostering of digital/data skills among citizenship. She has also a lectured at several international faculties (Welsh School of Architecture, TU Braunschweig, School of Architecture of Barcelona, IAAC, among others) and public institutions.