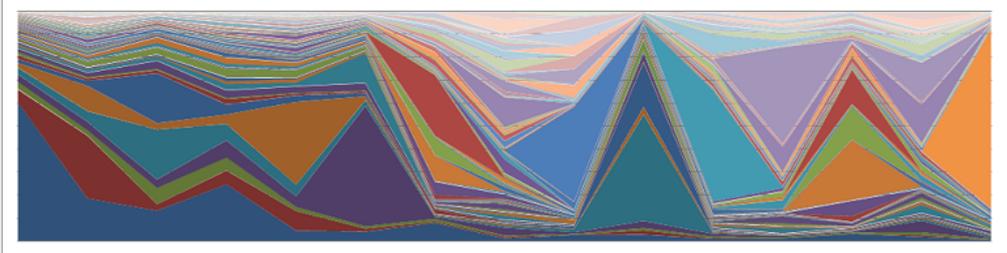


# NameSorts

NamSor Applied Onomastics – blog about naming and data mining of names.



## What's in a Twitter name? A glance at the Irish digital Diaspora

To jump directly to the interactive map, [click here](http://cdb.io/1beWaVB) : <http://cdb.io/1beWaVB>

([onomastics.co.uk](http://onomastics.co.uk) reblog)

It's been a while since I published a first 'Feature of the Month' in [onomastics.co.uk](http://onomastics.co.uk) and I can measure the progress made. The article, published in March 2013, showed maps of [French and English investments in Africa](#), established by recognizing the names of Company Directors, instead of the traditional measurement of capital flows (FDI).

At the time, ***NamSor Applied Onomastics*** software was new and I was still exploring how such data mining tool, which recognizes personal names, could be useful. I was uncertain whether the social benefits would exceed the risks inherent to such powerful technology.

Names are a Code and contain a lot of information about an individual, but there is no determinism. Human groups of different levels can be recognized through names, but human societies are fractals. Each group can be broken down again and again, from different angles. A first name, a last name, a Twitter handle are part of a person's identity and may indicate a social intent, the belonging to an ethnic/linguistic group, a geographic origin, beliefs, ... however at the finest grain level, every individual is unique and an exception to the group.

Genetic code, at one point, was thought to contain all the information needed to 'build' an individual from the physical point of view. After years of research, it seems that part of the information and the 'algorithm' are elsewhere... Still there is huge interest in applied research such as [23andMe](#) that 'decrypt' the genetic code to provide insights into a person's ancestry, as well as hints about potential health issues.

The Name Code and the Genetic Code share the same ability to fascinate : each

can somehow statistically be recognized to have an influence on your life, social status, average income, career... both relate to a family history. Each Code can be misleading and yet insightful. Fleur Pellerin, the French SME & ICT Minister, was born Kim Jong-suk in South Korea. She is both truly French and truly Korean, one name indicating a culture, the other a phenotype and genetic heritage. Considering only the Genetic Code would be denying a part of our humanity, which comes from being a child, a teenager, experiencing life, interacting socially, being part of a country and a culture, making one's choices.

Twin studies would tell a lot about the links between those two codes (Name, Genetic) – if only there were more twins. *Even though identical twins possess the same genetic makeup, they may go through different experiences throughout their lives that shape their personality, behaviour, and psychopathology in ways that make them unique relative to each other (Hughes et al., 2005).* Twins will have a different first name. Twins might also have a different last name, if -hypothetically- one twin was raised in Russia and the other twin was adopted and raised in the United States. In that case, what would the Name Code and the Genetic Code tell about potential Health issues (smoking or alcohol addiction, obesity & diabetes, life expectancy, etc.) ?

An article published last month caught my eye '[Scientists seek volunteers willing to have genetic code published on internet](#)': the hunt is on for 100,000 British volunteers to post their genetic information online in the name of science, as a North American open-access DNA project arrives in Europe. Personal Genome Project UK's mission is '*to make a wide spectrum of data about humans accessible to increase biological literacy and improve human health*'. The organization recognizes that '*Even if a person's name, home address or facial photograph is specifically excluded, a dataset like the one we are building is far from anonymous. It is simply too easy for someone to connect the dots and reveal a person's identity.*' Genetic Code is a very personal data. Would you like to see yours published along with your Name Code and Identity? Yet if the identity of participants can be protected, I can see huge scientific value in such Open Data.

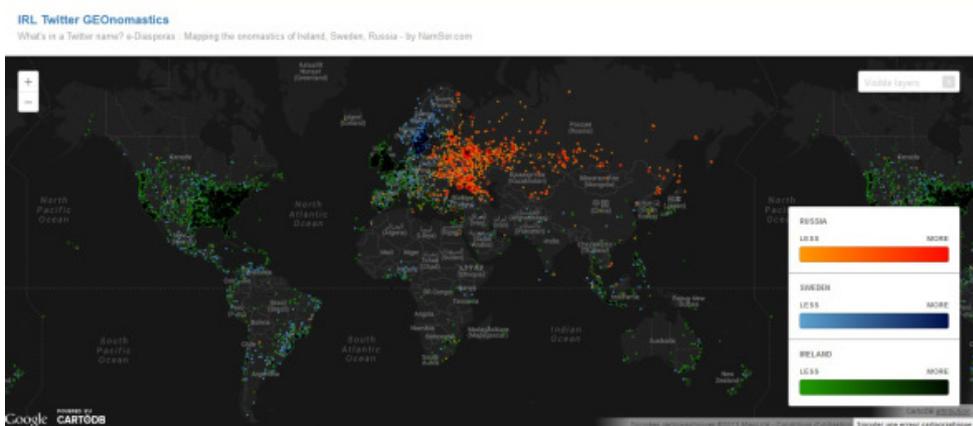
The Name Code, as such, is not personal data. Personal data is all information about yourself, that you should be allowed to keep confidential. A name is given to you as a communication tool, to interact with the World. There is a social intent in giving a child a common name, or a rare name that will more immediately identify a person – though I believe that one should be allowed to change names, just as Casanova did (who named himself Chevalier de Seingalt). There are legitimate reasons to keep one's name and identity secret sometimes: you should be free to do so, unless that freedom infringes on someone else's rights. A personal name (except *possibly* when it becomes a trademark) doesn't belong to anyone : it's been used before, it'll be used again, it's often shared by several people, it's found in the press, it's made up for fiction books ... Could a democracy work without the citizen knowing their politicians' names? How could historians do their research if we were to erase all personal names from the archives?

We see potential social benefits in applied onomastics and name data mining,

that clearly exceed the risks of misuse : not just in social sciences research, but also in economic development, tourism, marketing, health, urban planning ... We've helped one EU country reach out to its Diaspora in the US to [originate foreign direct investments](#) (FDI) and create jobs. We're currently helping a BioTech scientific cluster raise its game through better understanding where the talents lay in that field, and [where the brain juice flows](#) internationally. We're trying to find local partners to launch [AgroDiaspora](#), an economic development initiative in Africa to foster stronger links between Sustainable Agriculture Transformation Projects and top-level BioTech scientists of African heritage, who could help make local plants climate-change resistant, among other benefits. We are also very excited and enthusiastic about a paper we submitted to [ICOS 2014, the XXV International Congress of Onomastic Sciences](#), which will take place in Glasgow in August – as we foresee very positive outcome from that research.

In last month onomastics.co.uk feature '[The Impact of Diasporas on the Making of Britain](#)', Eleanor Rye mentions a very interesting research into what surname-based sampling can reveal about historic male migrations in the UK and Ireland.

We are currently conducting similar applied research on Twitter. I love Twitter. The freedom to choose one's handle and name. The limited amount of structured information that goes with an account : a location, a language, a short profile, a few pictures. What's in a Twitter name or handle? Anything : real names, company names, fancy names, pictograms, ... the amount of information produced through Twitter is enormous, but it's possible to filter this 'bigdata' in a way to make sense of it. We created geographic maps of e-Diasporas, by recognizing the Twitter names of geotagged tweets: Irish, Swedish, Russian, etc. We call this Twitter GEOnomastics, borrowing a term from [Dr. Evgeny Shokhenmayer](#). Below is the map of the Irish e-Diaspora, along with Swedish and Russian.



Irish Twitter GEOnomastics

[Click here](#) to access the interactive map:  
<http://cdb.io/1beWaVB>

How does it work? The software accurately recognizes that 'NamSor Applied Onomastics' ([@NomTri](#)) is probably a trade mark or a company name, whereas 'Elian Carsenat' ([@ElianCarsenat](#)) is probably a personal name – and most likely a

French name. Fancy names are also recognized and filtered out.

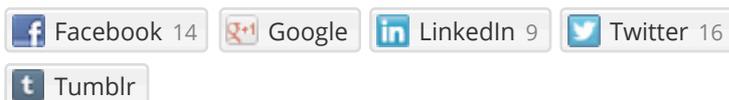
We see wide applications of such maps. When Captain James Cook explored the seas in the 18th century, having accurate maps could mean life or death for a ship and its crew. Working out latitude had been known for centuries, but measuring longitude was still tricky and inaccurate. In today's digital world, I see latitude as 'recognizing the semantics' in a message expressed in a particular language and longitude as 'recognizing the culture' of the target audience. We're full of curiosity on how and to whom this map can be useful, possibly Twitter itself. We're going from Paris to Dublin in two weeks to find out : we hope to meet people at Twitter European Headquarters. Twitter just issued its IPO but is also not clear how to make its money. We'll also meet Irish urban planners, people working in the tourism industry, investment analysts and Diaspora experts.

Read our next posts to discover more Twitter GEOnomastics maps showing Irish, French, German, Spanish, Russian, Turkish, Swedish, Italian, Dutch e-Diasporas (or cultural influence).

NB. The maps are currently interactive, so you can zoom in and out of a particular territory, however this may be shut down in a month or two.

---

**Share this:**



---

**Like this:**



Be the first to like this.

[Onomastics for Business...](#)  
In "General"

[Human societies are frac...](#)  
In "EthnoViz"

[Monitoring western inve...](#)  
In "EthnoViz"

December 3, 2013

 [Leave a reply](#)

---

[« Previous](#)

## Leave a Reply

You must be [logged in](#) to post a comment.

[View Full Site](#)

**Now Available!** [Download WordPress for Android](#)

[Blog at WordPress.com.](#)

