

Unlocking Diaspora Potentials: Towards a Sustainable Big Data Platform for Engagement and Exploitation



Professor Lukumon O. Oyedele

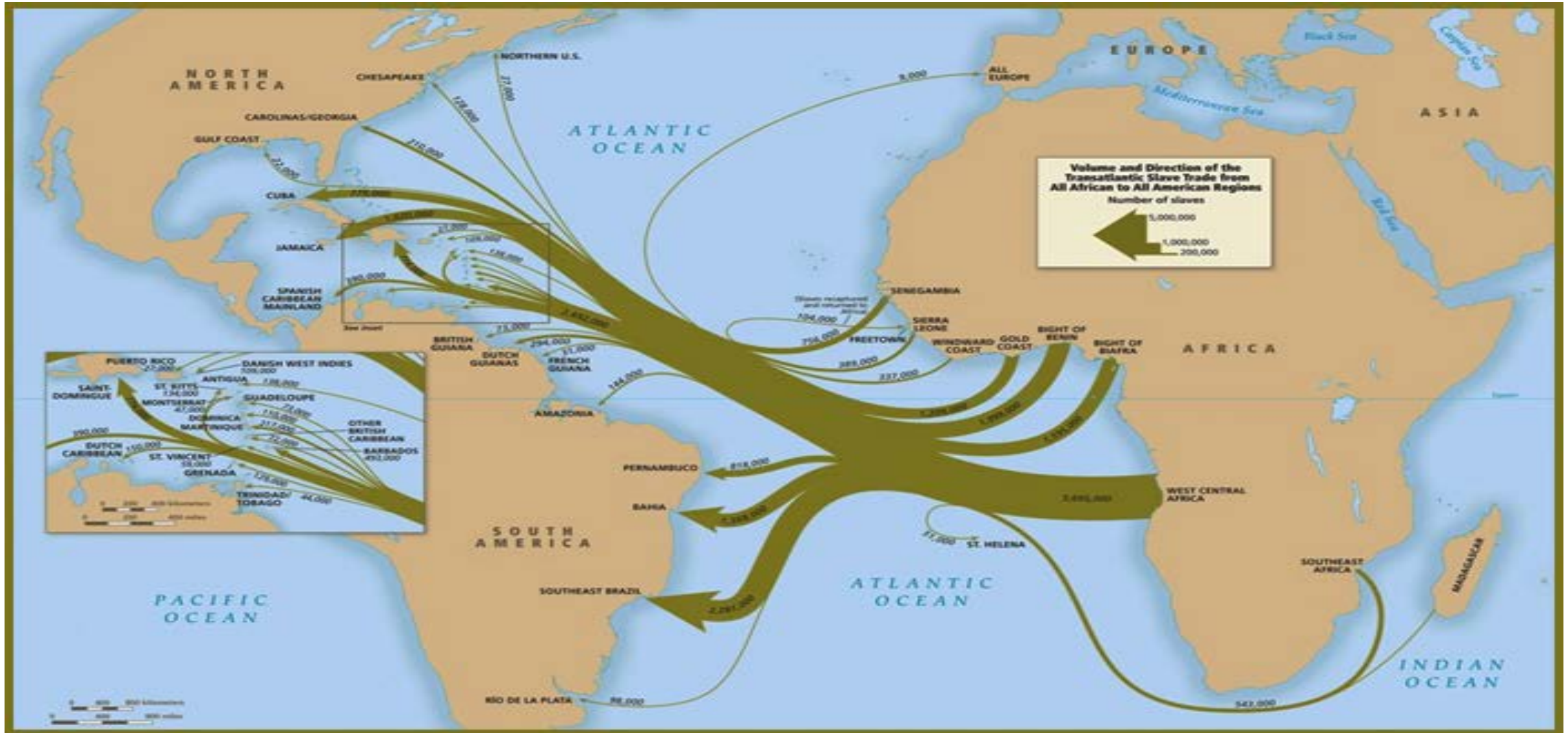
Director of Bristol Enterprise, Research and Innovation Centre (BERIC)
University of the West of England (UWE), Bristol
United kingdom

July 2016

Outline

- Critical Questions for all Diasporas
- Big Data Analytics (BDA)
- Engagement
- Exploitation

There is no doubt Diasporas has got Huge Potentials



7 Critical Questions for All Diasporas

1. Do the Government of your homeland know who you are and where you are?
2. Do your homeland government know your expertise?
3. Which sectors are your expertise needed within your homeland ?
4. If you need to return to your homeland, what support do you have to quickly settle down?
5. Must you go back home before you contribute to your homeland?
6. How can you connect other Diasporas in other countries
7. Do you need the government all the time to contribute to your homeland?



Big Data



Definition

McKinsey Global Institute (2011)

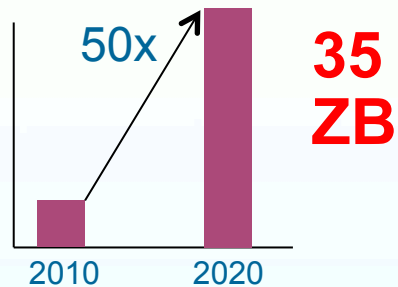
Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze.

Various Aspects

- Large dataset (Megabyte, Gigabyte, Terabyte, Petabyte, Exabyte)
- Unstructured data (networked data but fuzzy relationships)
- Data-driven research, business & decisions
- High skills (IT, statistics, etc.)

Big data characteristics – The 5Vs

Cost efficiently processing the growing **Volume**



Responding to the increasing **Velocity**



30 Billion
RFID sensors
and counting

Collectively analyzing the broadening **Variety**



80% of the
world's data is
unstructured

Identifying hidden data
Of **Value**



Almost **every manager** is
concerned about the
money they spent



Establishing the
Veracity of big
data sources

1 in 3 business leaders don't trust the
information they use to make decisions



Big data analytics



Big data analytics

- **Big data in vacuum is useless**
- **Big data analytics** is the process of examining large **data** sets (**Big data**) to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful business information.
- Real game changer

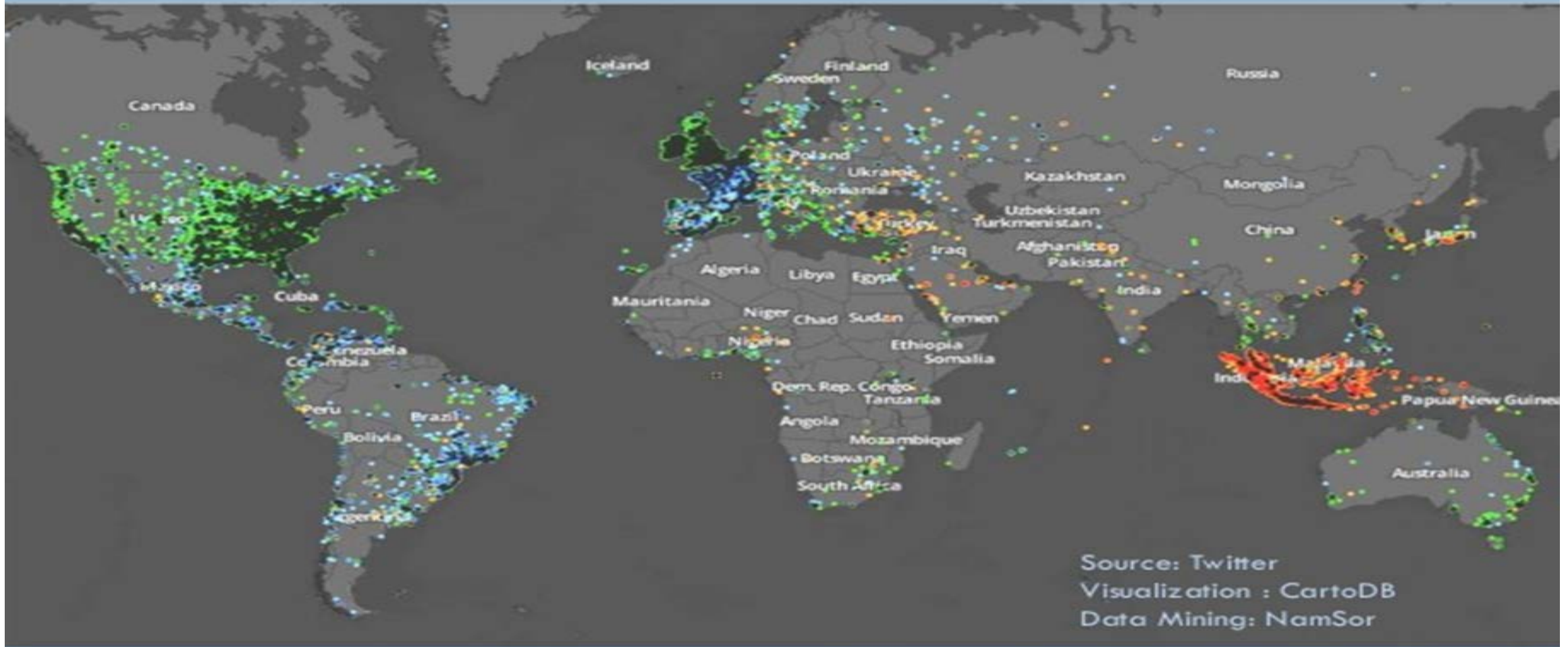


Mutual Opportunities

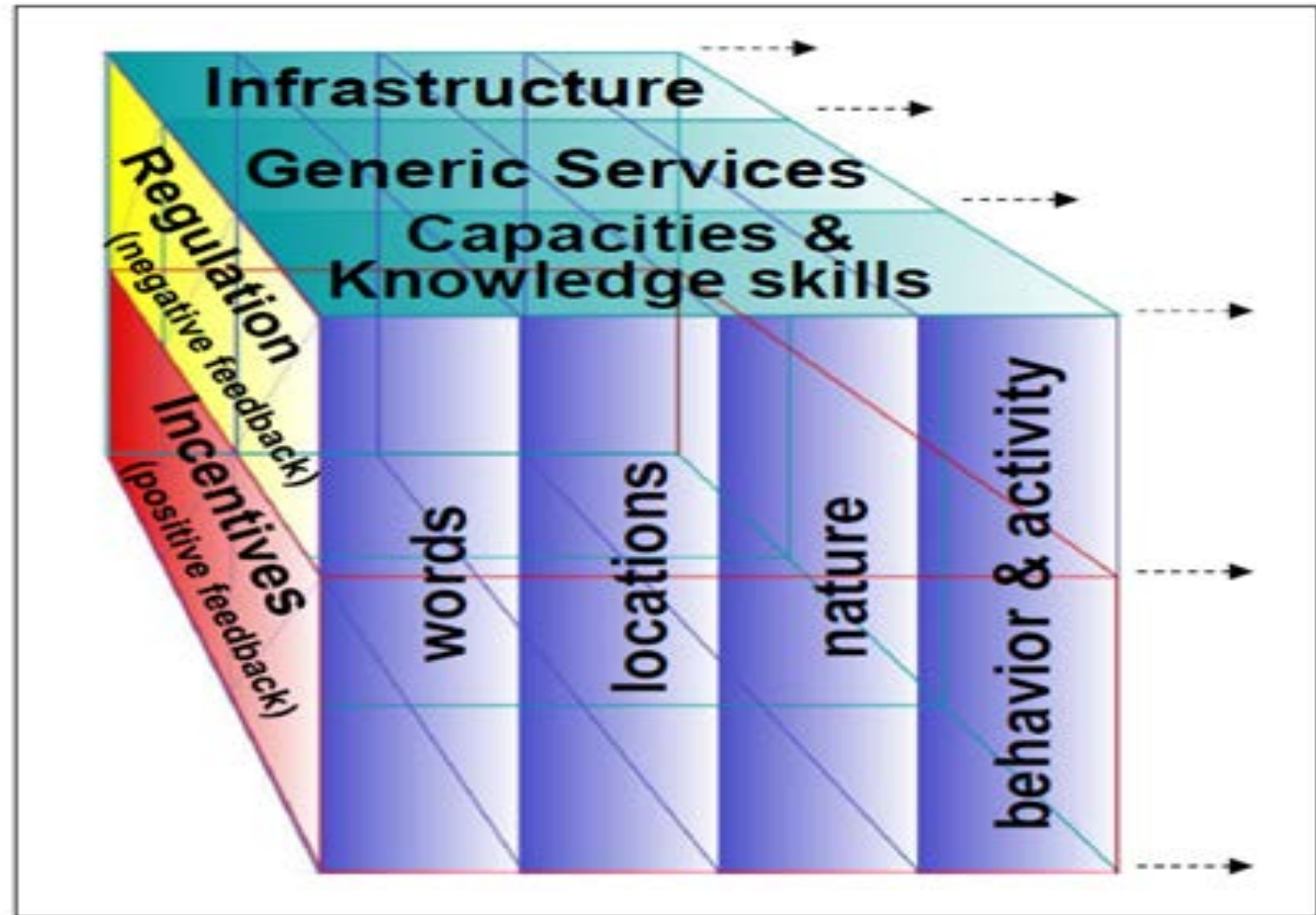


Mapping Diasporas

Mining 3M twitter names to map Diasporas
Who are they, where are they and what are they doing?



The three-dimensional “ICT-for development-cube” framework applied to Big Data



- Names reflect cultural *Identity*
- NamSor** data mining software recognizes the **linguistic or cultural origin of names** in any alphabet / language, with fine grain and high accuracy.

- Personal names are meaningful : we use sociolinguistics to extract their semantics and deliver actionable intelligence.

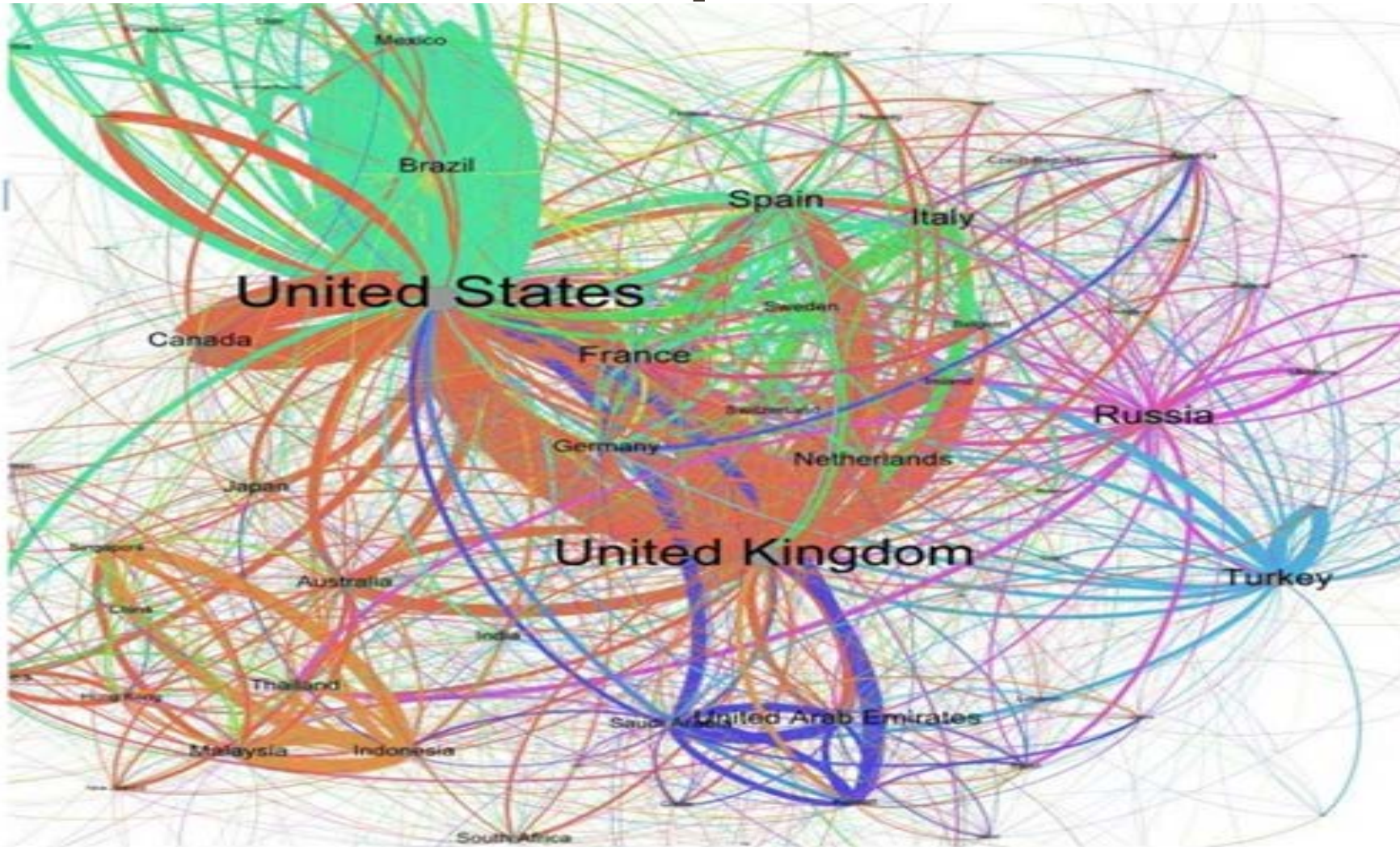
Diaspora Location Data

Flow view –
who travels
where?

Source	Target	Type	Id	Onoma	Weight
United Kingdom	France	Directed	16	Great Britain	37
Spain	France	Directed	55	Spain	14
United States	France	Directed	75	Great Britain	12
Turkey	France	Directed	79	Turkey	11
Brazil	France	Directed	87	Portugal	10
United Kingdom	France	Directed	112	Ireland	9
Italy	France	Directed	152	Italy	7
Switzerland	France	Directed	226	France	5
Belgium	France	Directed	247	France	5
United Kingdom	France	Directed	258	France	5
Mexico	France	Directed	287	Spain	4
Ireland	France	Directed	317	Great Britain	4
United Kingdom	France	Directed	333	Italy	4
United States	France	Directed	375	France	4

Source: Twitter
Visualization : Gephi
Data Mining: NamSor

Diasporas Location and Data



**Flow view –
who travels
where?**

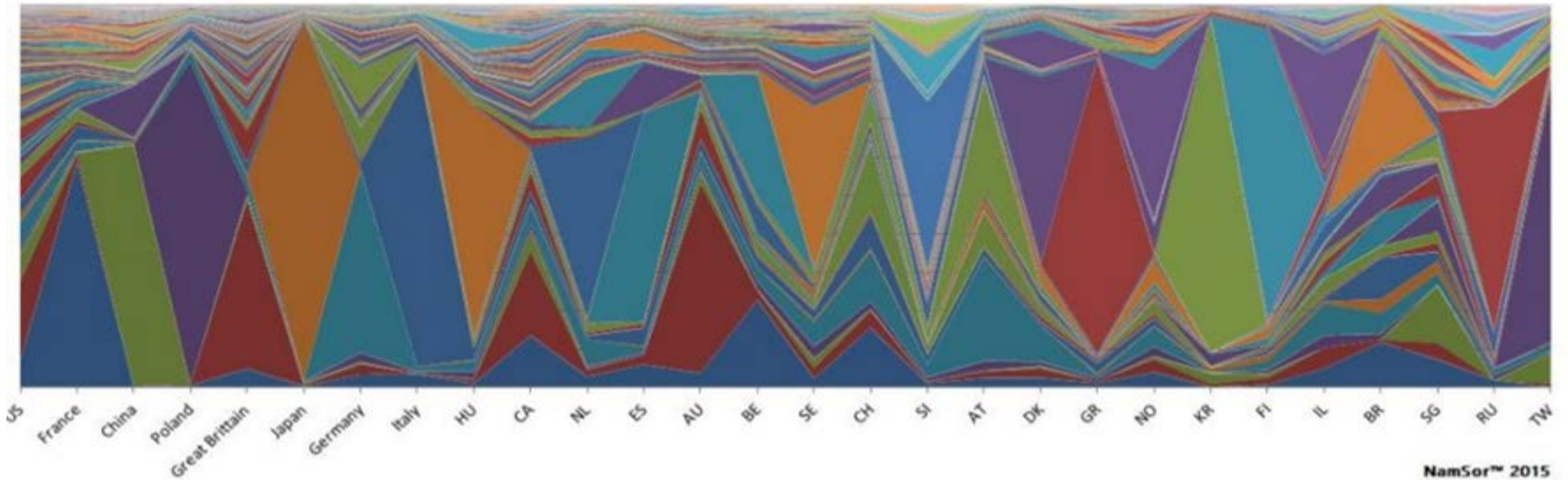
Source: Twitter
Visualization : Gephi
Data Mining: NamSor

Mapping Talents in Cancer research

Who's in Cancer Research - an onomastic mille-feuille

abscysse: country of affiliation; ordinate: likely country of origin

Source: Thomson WoS

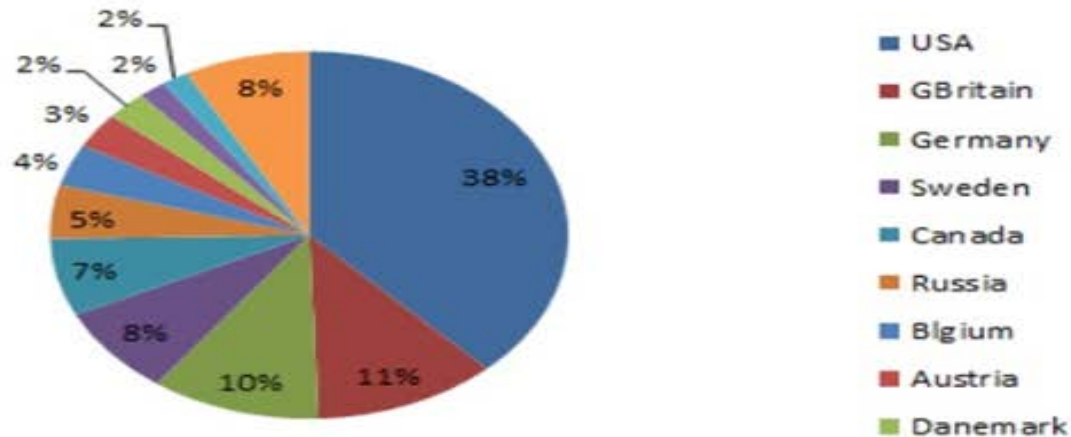


Source: WoS; Data Mining: INSERM with NamSor

Cancer Research in Poland and Slovenia Examining the 'brain drain'

1

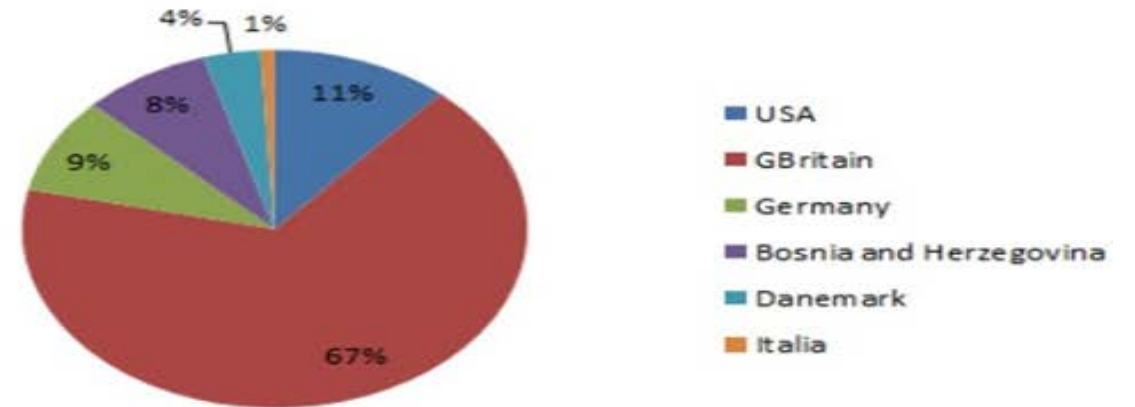
The Polish "brain drain"



In the Polish Corpus, we look at co-authors with Polish names, affiliated abroad. Top countries:

1. **USA**
2. **Great-Britain**
3. **Germany**

The Slovenien "brain drain"



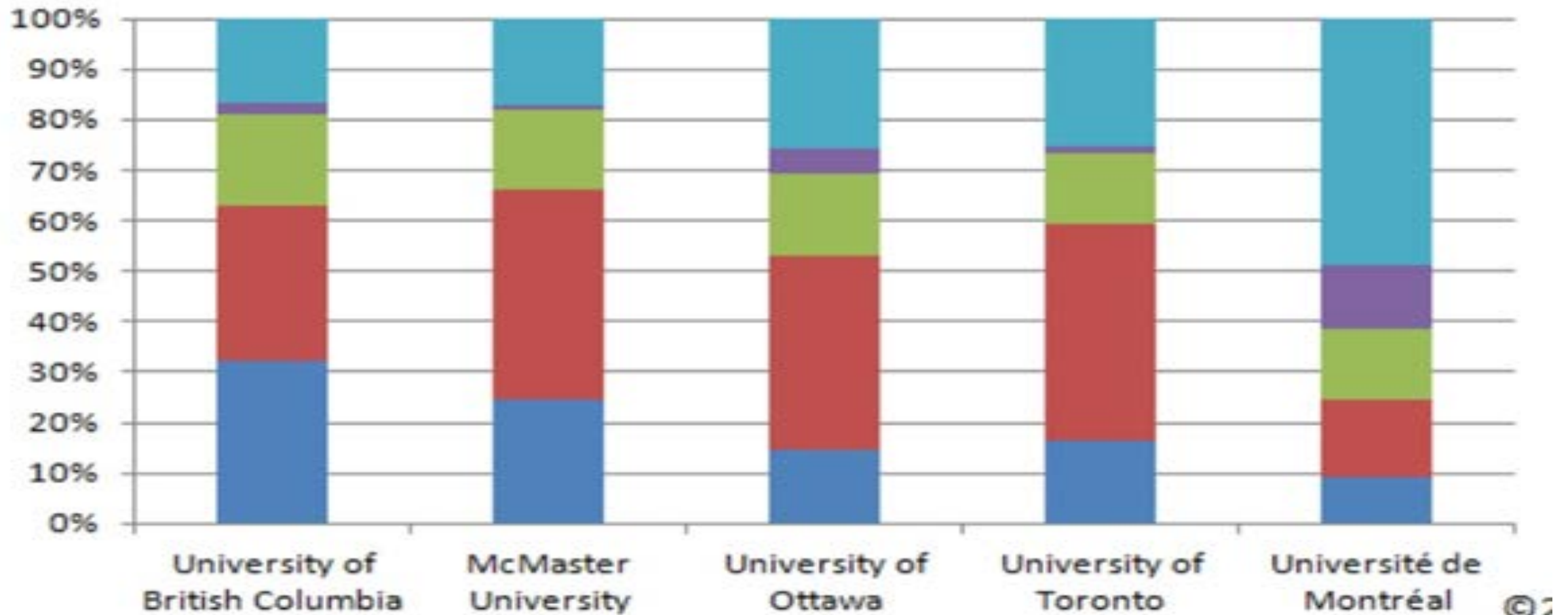
In the Slovenian Corpus, we look at co-authors with Slovenian names, affiliated abroad. Top countries:

1. **Great-Britain**
2. **USA**
3. **Germany**

Moroccan Academics in Canadian Universities

Diaspora Scholars in Canadian Universities

Source : scholarUniverse, NamSor.com

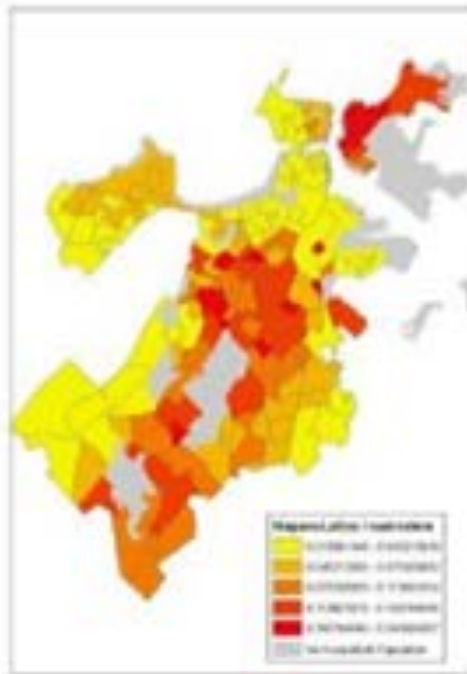


Canadian Science Policy Conference - CSPC2015

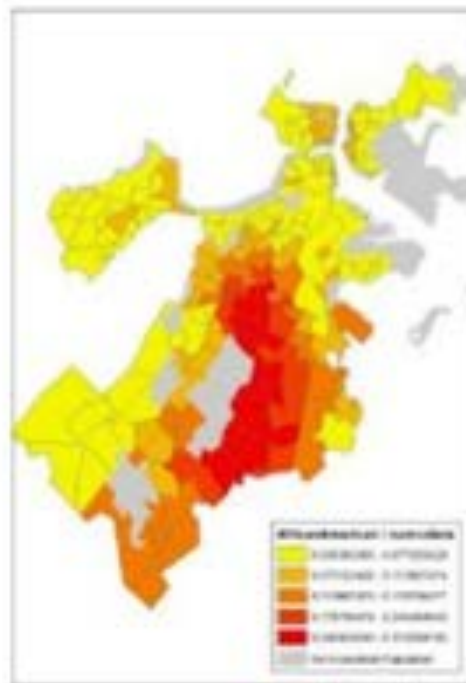
Boston Demographics



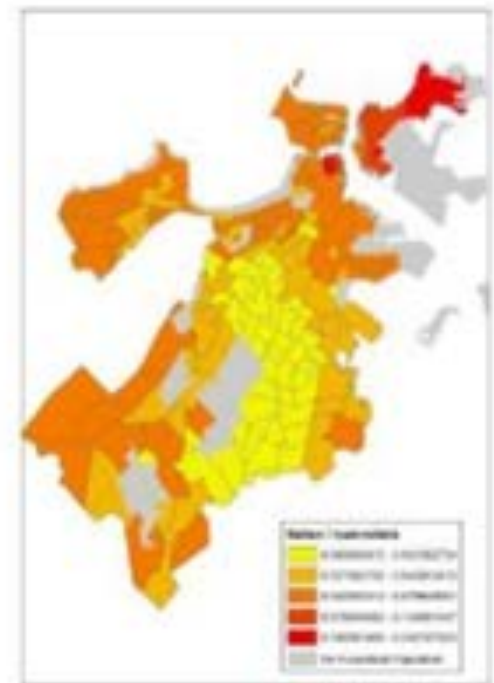
Hispanic/Latino Share, namsor



Black/African-American Share, nams



Italian Share, namsor



Analysing Patent Data – Who has filled what Patent?



Get our newsletter

Search

The Demographics of Innovation in the United States

Adams Nager, David M. Hart, Stephen Ezell, and Robert D. Atkinson February 24, 2016

A groundbreaking ITIF survey shows why the country needs to broaden and deepen its pool of potential innovators with better STEM immigration and education policies.

[View Report](#)

[View Executive Summary](#)

[Event](#)



Groundbreaking @ITIFdc survey shows why US needs to broaden and deepen pool of potential innovators



.@ITIFdc releases groundbreaking survey on who innovates in the United States and where and how it occurs

Big data business that analyses how infectious diseases spread secures Series of funding



Conclusion

Big Data Analytics + Diasporas

- Real Game Changer for Developing our homeland
 - Networking Diaspora Potentials and Expertise?
 - Being very Proactive – What can we do better for homeland?



Thank You

