1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
17-07-2017	Final Report	15-Mar-2015 - 31-Dec-2016

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Final Report: The Western Jihadism Data Collection	W911NF-15-1-0097
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
	611103

6. AUTHORS	5d. PROJECT NUMBER
Jytte Klausen, Ph.D.	
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
Brandeis University 415 South Street Waltham, MA 02453 -2728	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	ARO
U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
	66652-LS.6

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
The Western Jihadism Project Data Collection is a web-based database designed to study Al-Qaeda-inspired terrorist offenders in Western Europe, the Antipodes, Canada, and the United States. The project developed a secure online portal for a relational database comprised of qualitative and quantitative data files from the archives of The Western Jihadism Project (WJP). The database began as a spreadsheet in Microsoft Excel. Since then it has since evolved into a relational database in PostgreSQL and using a Django-based administrative framework for editing the data stored within. The collection also contains numerous audio-visual files including photographs

15. SUBJECT TERMS
Database, terrorist networks, al Qaeda, ISIS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	UU		Jytte Klausen
UU	UU	UU			19b. TELEPHONE NUMBER
					781-736-2762

Report Title

Final Report: The Western Jihadism Data Collection

ABSTRACT

The Western Jihadism Project Data Collection is a web-based database designed to study Al-Qaeda-inspired terrorist offenders in Western Europe, the Antipodes, Canada, and the United States. The project developed a secure online portal for a relational database comprised of qualitative and quantitative data files from the archives of The Western Jihadism Project (WJP). The database began as a spreadsheet in Microsoft Excel. Since then it has since evolved into a relational database in PostgreSQL and using a Django-based administrative framework for editing the data stored within. The collection also contains numerous audio-visual files including photographs, videos, propaganda, magazines, sermons, and news reports. The project aim was to develop the data collection into an online portal capable of supporting collaborative multi-team research and data sharing within a secure online environment.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received	Paper
07/14/2017	1 Jytte Klausen, Ph.D.. The Myth of Homegrown Terrorism, The Georgetown Security Studies Review, Special Issue: What the New Administration Needs to Know about Terrorism and Counterterrorism, ( ): . doi:
	1,050,416.00
TOTAL:	**1**

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received	Paper
TOTAL:	**2**

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Guest Lecture. Laboratory for Analytic Sciences. North Carolina State University. June 28, 2017. https://ncsu-las.org/event/guest-lecture/

[Note, the PDF version of the PowerPoint you uploaded was my presentation at this event.]

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received   Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received   Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received   Paper

TOTAL:

Number of Manuscripts:

Books

Received   Book

TOTAL:

<u>Received</u>   <u>Book Chapter</u>

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

<u>NAME</u>	<u>PERCENT_SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Names of Post Doctorates

<u>NAME</u>	<u>PERCENT_SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

<u>NAME</u>	<u>PERCENT_SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Names of Under Graduate students supported

<u>NAME</u>	<u>PERCENT_SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 6.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:...... 4.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:...... 8.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):...... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields: 2.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME   PERCENT_SUPPORTED

FTE Equivalent:
Total Number:

Sub Contractors (DD882)

Inventions (DD882)

Project description:

The database records data from public data sources in many languages. The purpose of the data collection is to support Social Network Analysis (SNA) and to develop a format for coding events and individuals that captured complex relationships in transnational and highly mobile networks.

Over 50 descriptive variables about individual jihadism-related terrorism offenders are coded. Information is drawn from public sources and is manually coded. Sources include court records, government press releases, and news media. The coding protocol supports time series analyses of the evolution and the demographic transformation of jihadist recruitment in Western democracies. Records date back to the early 1990s and are updated through early 2017. Research assistants follow a detailed codebook instructing them how to infer and code demographic data about individual perpetrators of terrorism-related criminal acts, including also information about the nature of the terrorism-related acts and previous criminal acts, charges filed, convictions, and verifiable communicative acts between terrorists and contacts with international terrorist organizations. New entries are added and existing entries are continually updated as new information about terrorism-related acts becomes available.

Data collection has been ongoing since 2006. The database has grown to include, at the present time, information about 6,000 individual jihadists of Western origin or residence from 20 countries, +700 terrorist conspiracies (foiled and successful), and close to 400 local and international terrorist organizations, and +25,000 person-to-person and person-to-organization contact links.

A more detailed description of the collection and the methodology guiding the gathering of data as well as available utilities is available on the project website: http://www.brandeis.edu/klausen-jihadism/

Technical description:

The web portal is a front-end database user interface written in Django, an open source, high-level Python Web framework. Access to the portal utilizes a unique double lock-and-key system requiring encryption and white-listing of all user access. Access is tracked and recorded.

Data is stored in PostgreSQL, an object-relational database management system. The project is hosted on two Red Hat Enterprise Linux servers, one for development and one for production, and deployed using an uWSGI application container. The Git version control system is used for tracking changes and coordinating work between developers.

The SQL-format facilitates the integration of quantitative and qualitative data, and the web-based interface allows for easy integration with a wide variety of analysis platforms and third-party extensions.

The primary objects in the database schema are People, Plots, and Organizations. Primary keys to records in supplementary tables, such as Country, Ethnicity, and City, provide further information to describe primary object records. Primary keys also link People to records of unique supplementary information such as arrests, foreign fighter attempts, mugshot images, and source documents. Links used in the Generic Queries tool, such as Person-to-Plot links and Person-to-Organization links, are saved by entering records containing the primary keys of the desired objects. Triggers prevent users from entering data, which would compromise the integrity of the database.
The web portal consists of a graphical user interface that presents a human-readable, object-oriented method for viewing joined records. Included in the object model are tools for administration to manage user access and monitor security. The framework allows research assistants to add and update records in the database without knowledge of the Postgres structured query language.

Generic Queries, also called "Primary Queries", is a generalized query tool that allows users to access and filter data with the same functionality as a spreadsheet application like Microsoft Excel or macOS Numbers. A pivot table provides the interface for querying the data. The Django API is used to retrieve and filter the data. Users are able to select from six major topics of data analysis and select what columns will be included. Users can filter rows based on keywords and rudimentary Boolean logic. The table can be exported to a .csv file for data analysis with applications such as Gephi, R, or *ORA, or passed to the web portal's browser-based data visualization tools.

Data recorded with matching record IDs may be exported and imported between users allowing for the pooling of data between independently developed databases.

The web portal includes a series of data visualization tools written in D3.js, a JavaScript library for building data-driven graphics using HTML, SVG, and CSS in modern web browsers. The tools are integrated with the Django database queries allowing users control over the data filtered into the visualization. Charts and network graphs are included with plans to expand to suit user needs.

Intended Users

The online database portal is now capable of supporting online collaboration and data sharing with research teams based outside Brandeis. Two such collaborations are in progress, one based at the Laboratory for Analytic Sciences at North Carolina State University and the other at Program on Global Terrorism at the Elcano Royal Institute in Madrid. Partial one-way data sharing arrangements have been established with research teams at Colorado State University (Benjamin Hung, PhD candidate, and Professor Anura Jayasumana, PhD) and researchers at the Massachusetts Institute of Technology (Christopher Marks, PhD candidate, and Professor Tauhid Zaman. For more information about the collaboration, see https://topdata. news/how-mit-sloan-researchers-are-using-machine-learning-to-take-down-isis-on-twitter/.) Dr. Klausen's team provided data to the researchers, who did not wish or need to access to the online portal to obtain the needed data.

Researchers who wish to use the data are required to sign a non-disclosure agreement and to accept responsibility for obtaining approved IRB protocols and for complying with national and local rules regarding research with human subjects. Graduate students wishing to access the data are required to obtain the consent of their supervisor to assume the legal and practical responsibility for meeting the requirements. Users are additionally required to cover the hourly costs on the part of Dr. Klausen and her team of providing technical assistance related to their use of the data and the data portal.

Technology Transfer

Jytte Klausen

Dr. Klausen founded the Western Jihadism Project in 2006 with the aim of providing an integrated approach to demographic research and network analysis of Western violent extremists associated with Al Qaeda and the study of extremist Islamist ideology.

She leads a team at Brandeis University who are studying Islamist terrorist networks, funded initially by the UK Home Office, the National Institute of Justice at the U.S. Department of Justice, the MINERVA Initiative at the U.S. Department of Defense, and the National Science Foundation.

METHODOLOGY: Historical Method & Quantitative Analysis

- *Cliometrics* (def.): the statistical analysis of data mined from archival sources.

- *Clio*: The muse of history. One of nine daughters of Zeus and Mnemosyne. Each daughter became a sponsor of a domain of knowledge.

- Every data point in the database is coded individually.

- ~5,700 individual records; 23,000 person-to-person and person-to-organization links.

What is recorded?

Total of 36 variables

- Demographic data (age, conversion, education, gender, ethnicity etc.).

- Arrests dates and outcomes, content of charges, length of sentences, appeals, pre-radicalization criminality, and immigration status.

- Legal status: Nationality is defined simply as legal status. Residency is defined as an individual's last known country of long-term domicile.

- Relationship data intended to facilitate the analysis of communication structures (co-habitation, training camp attendances, email communications etc.).

Case: Karim Mejjati

A French citizen born in Casablanca, who at times lived in France, Spain, Germany, and possibly Britain; married an American woman in Morocco and a Belgian woman who lived in Britain; fought in Bosnia, trained four times with Al-Qaeda at Khalden, studied under Abu Qatada, sentenced in absentia in Morocco for involvement in the 2003 Casablanca bombings, linked to the 2004 Madrid bombings, and was eventually killed in Saudi Arabia where he was living under a fake Qatari identity.

DATA COLLECTION: How do we get our information?

- Data comes from our own research based upon public records and compiled by means of archival research.

- All the information recorded in our database derives from publicly available documents:

  1. Primary sources: court records, indictments, judgments, social media sites, and private and government reports.

  2. Secondary sources: press articles and books written by experts and scholars. Media coverage and primary sources derive, for the most part, one way or another from law enforcement activities, arrests, judicial proceedings in national courts or international intergovernmental organizations.

- Routinely benchmarked against national and international arrest statistics, and other research projects.

The Western Jihadism Project: Queries

	Country	Ethnicity	Plot Links	Foreign Insurgency	Person Links	Organization Links

Showing 1-20 of 1336 (0.508 seconds)

	Legacy ID	Name	Start Date	Country Name
1	IT345	Anas El Abboubi	2013	Syria - Iraq
2	US965	Zakia Nasrin	2011	Syria - Iraq
3	FR647	David Drugeon	2014	Syria - Iraq
4	NL157		2014	Syria - Iraq
5	AUS140	Jasmina Milovanov	2015	Syria - Iraq
6	AU1	Mohammed Shawqi Mahmoud	2012	Syria - Iraq
7	AU21	Gokhan Cambaz	2013	Syria - Iraq
8	AU22	Hasan B.	2013	Syria - Iraq
9	AU23	Tamima Yasira	2014	Syria - Iraq
10	AU24	Samra Kesinovic	2014	Syria - Iraq
11	AU25	Sabina Selimovic	2014	Syria - Iraq
12	AU26	Ramzan	2013	Syria - Iraq
13	AU27	Firas Abdullah Al-Slimani	2014	Syria - Iraq
14	AU29	Albaba Hani	2014	Syria - Iraq
15	AU30	Yasir Firat	2014	Syria - Iraq

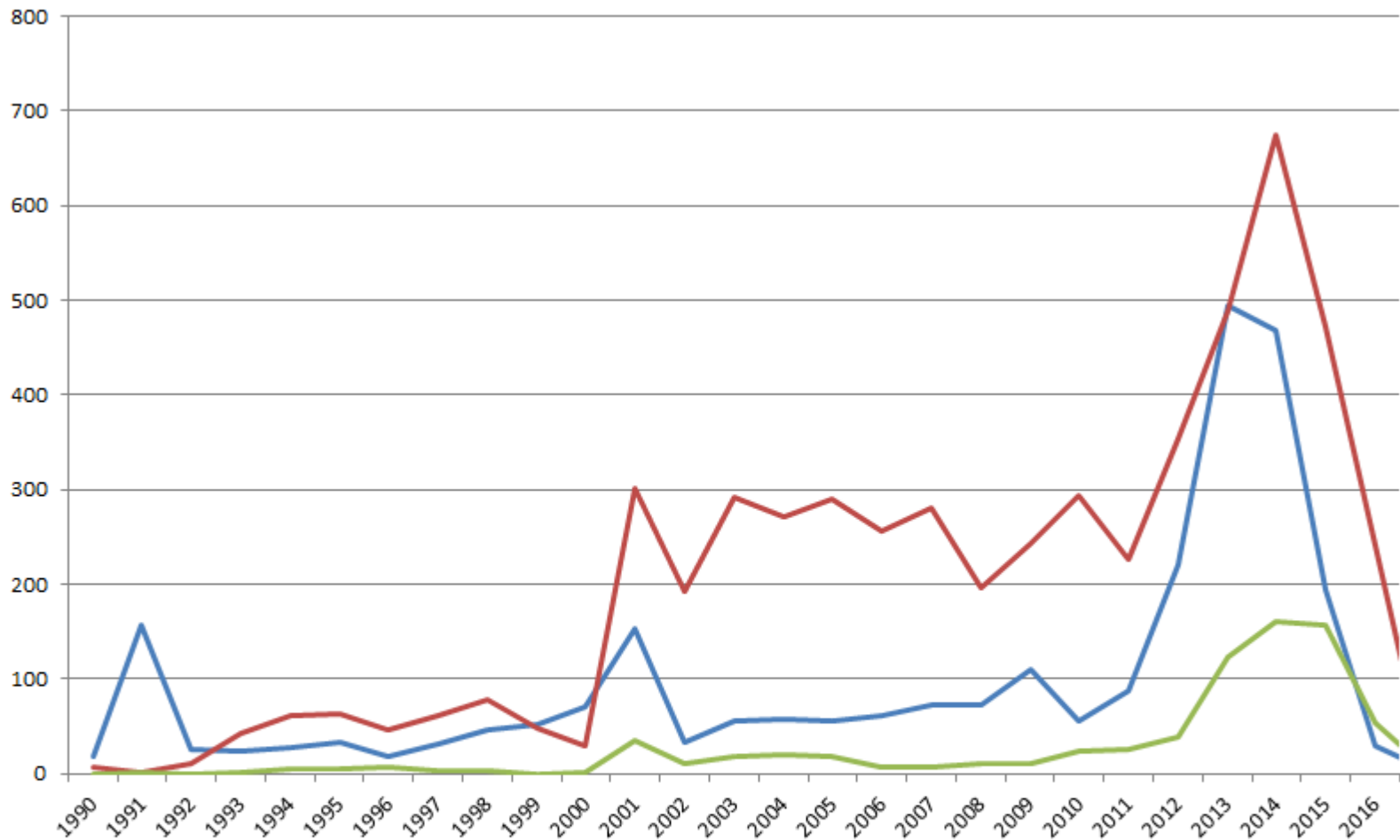# The Western Jihadism Project Database: Graphing Tool
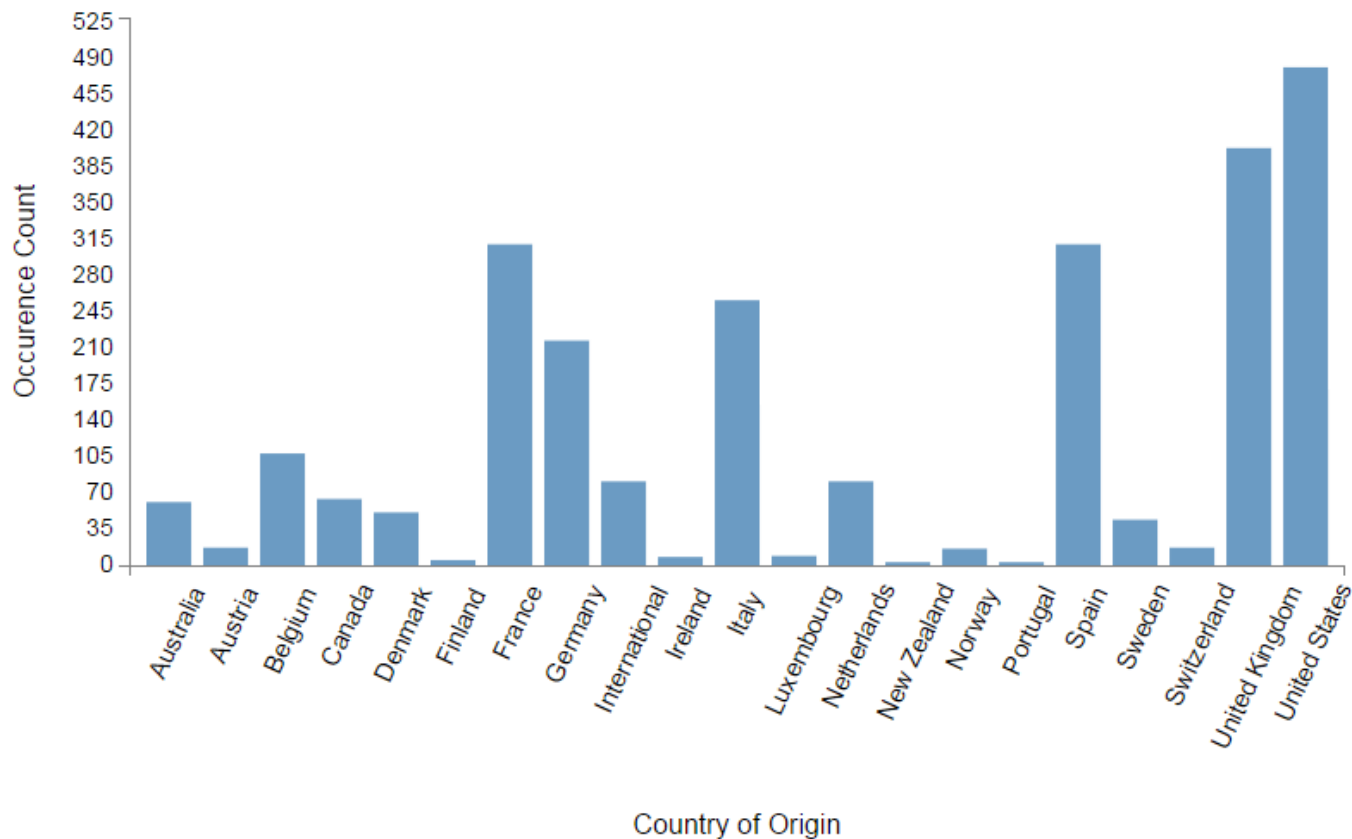
Number of Individual Terrorists by Action Type (N=5,544)

Note: Red = First law enforcement contact, which refers to the date of first known arrest or other public law enforcement engagement with a jihadist extremist.
Blue=Instances of going abroad to fight in a jihadist insurgency.
Green=Deceased.

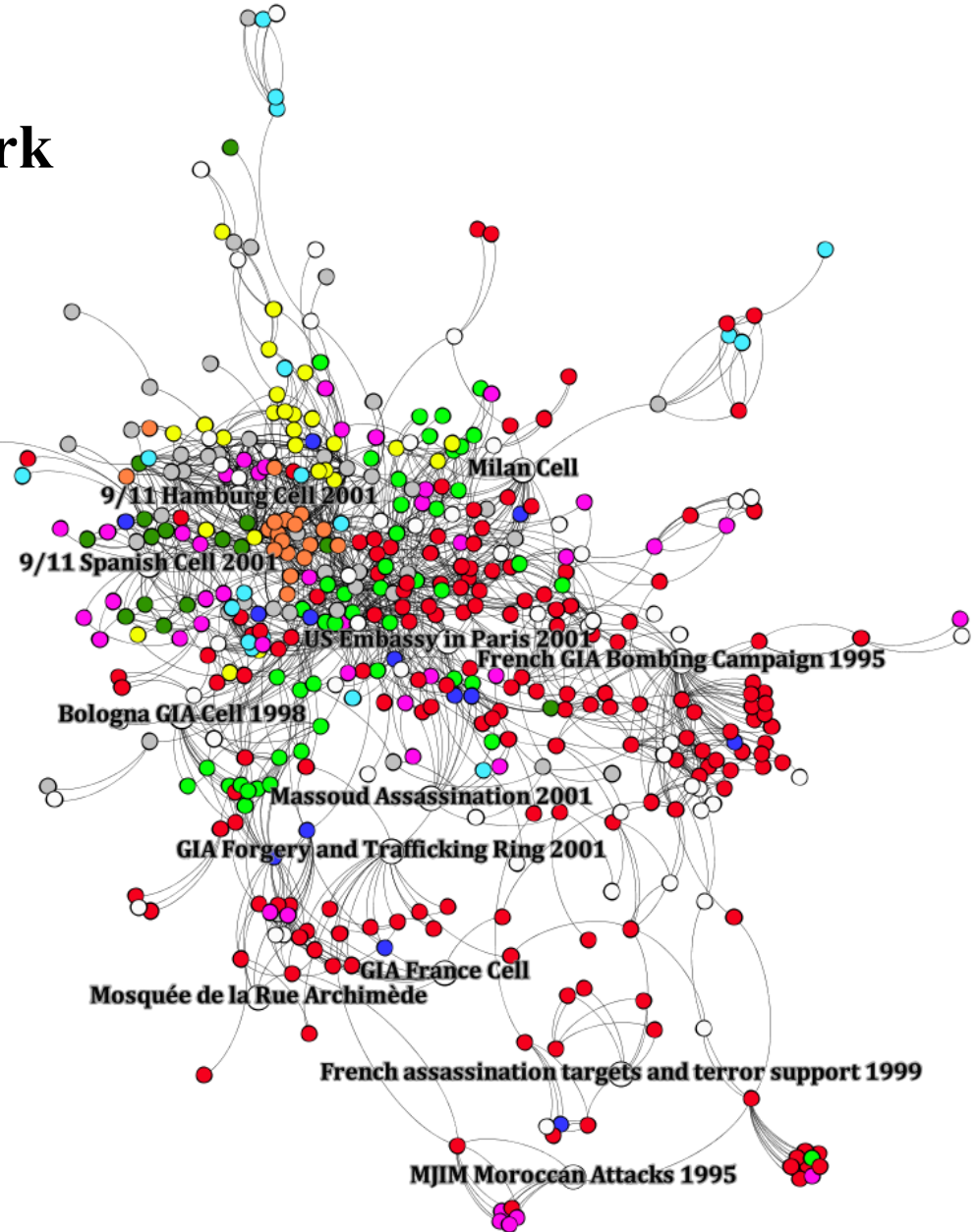# First Terrorism Related Law Enforcement & average age
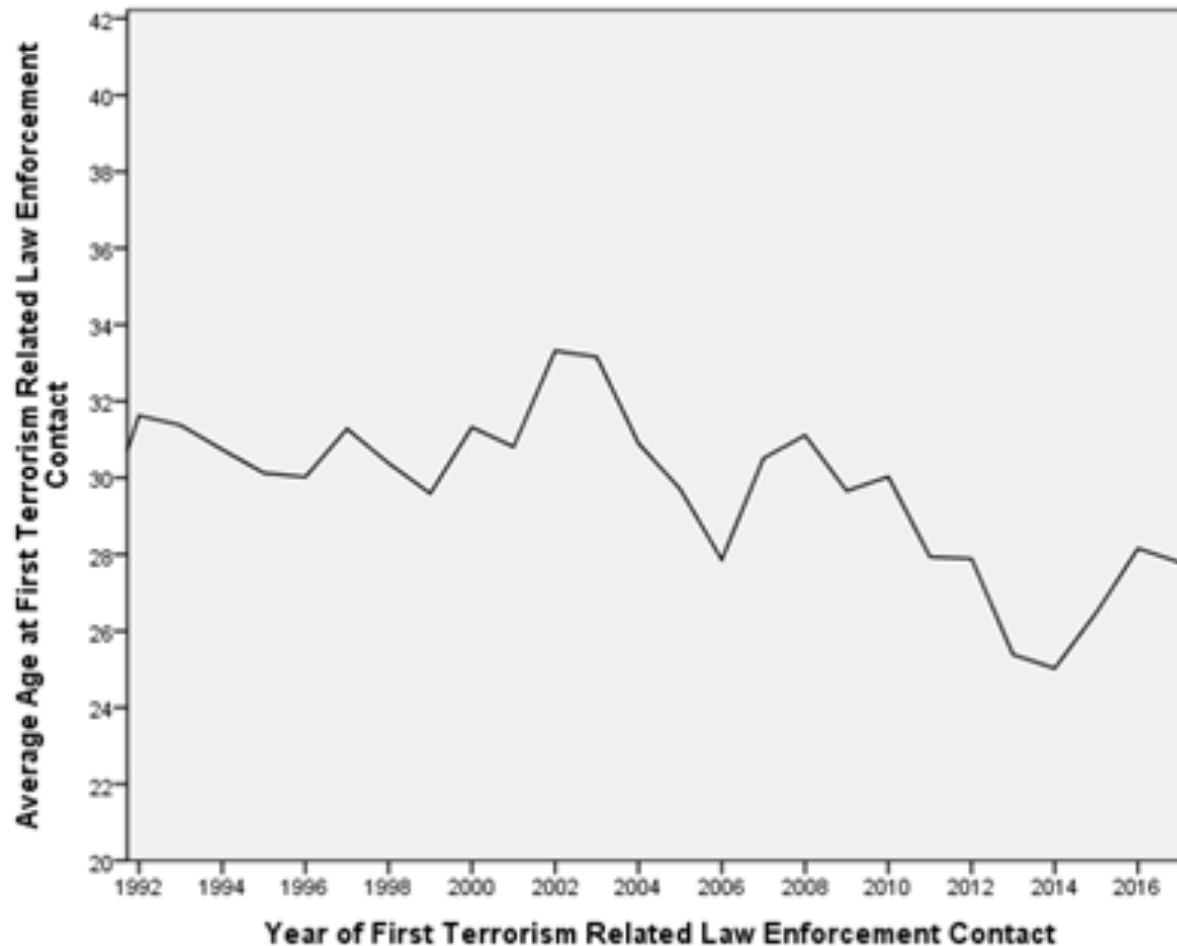
- N= 4,938 observations
- Estimations: 870 estimated using average age of arrest
- Dropped values: 674 INTs due to identification issues or because of YEAR FLEC

Pre-2002: European Network

Algerian	168	42.75%
Egyptian	26	6.62%
Lebanese	6	1.53%
Libyan	8	2.04%
Moroccan	43	10.94%
Pakistani	18	4.58%
Palestinian	7	1.78%
Saudi	19	4.83%
Syrian	15	3.82%
Tunisian	46	11.70%
White-European	14	3.56%

Red: Algerian, Purple – Moroccan, Turquoise – Pakistani, Yellow – Egyptians, Light green – Tunisian, Blue – White European, Grey – Arrest events and Plots. All unconnected nodes and cells comprised of less than XX were removed from the depicted graph.

Labels in figure: Milan Cell; 9/11 Hamburg Cell 2001; 9/11 Spanish Cell 2001; US Embassy in Paris 2001; French GIA Bombing Campaign 1995; Bologna GIA Cell 1998; Massoud Assassination 2001; GIA Forgery and Trafficking Ring 2001; GIA France Cell; Mosquée de la Rue Archimède; French assassination targets and terror support 1999; MJIM Moroccan Attacks 1995

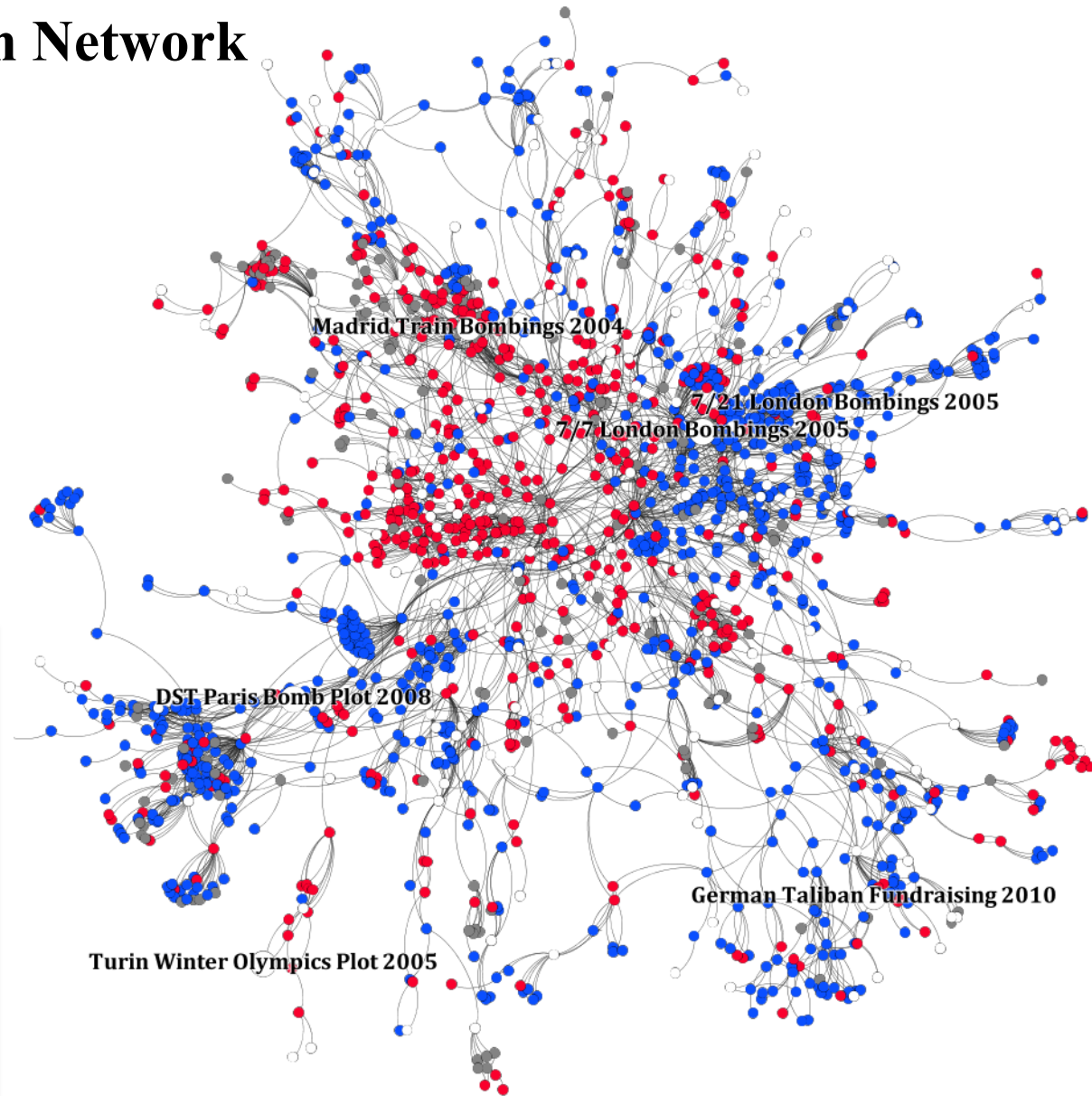# European Network Before 9/11 – Data Selection

Inclusion criteria:

European citizen and/or resident

First law enforcement contact or death in 2001 or before

Min. one link to another node

Created in Gephi
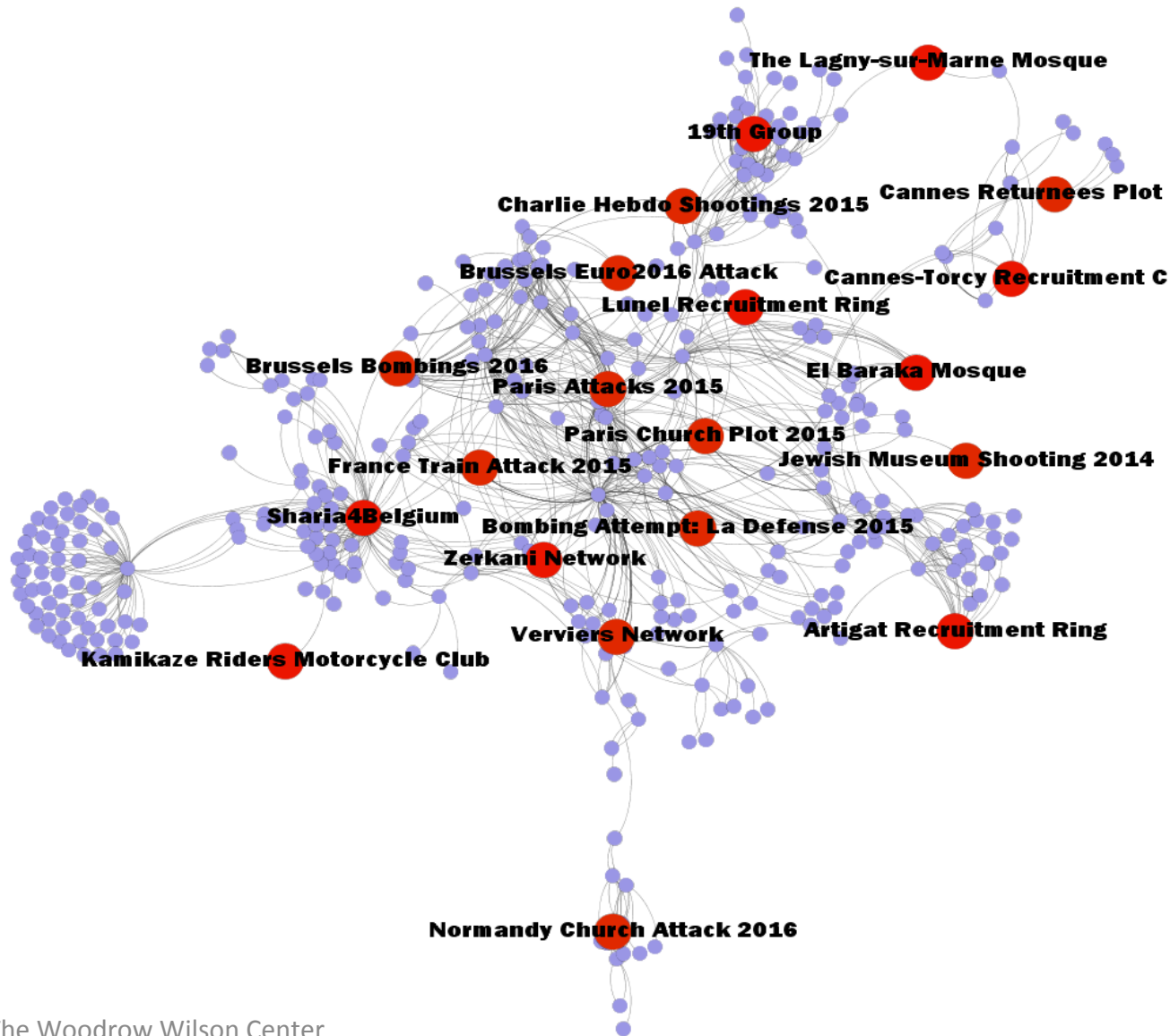
2002-2011: European Network

Number of people in Giant Component of Network (pictured): 1,720

Number of people not connected to giant component and

removed: 283

Native Born

Status	# of People	Percentage
IMMIGRANT	543	34.87%
NATIVE	847	54.40%
UNKNOWN	167	10.73%
Total People	**1557**	**100.00%**

Madrid Train Bombings 2004

7/21 London Bombings 2005

7/7 London Bombings 2005

DST Paris Bomb Plot 2008

German Taliban Fundraising 2010

Turin Winter Olympics Plot 2005

Sociogram of Paris-Brussels Attack Network (2014-2016)

The Lagny-sur-Marne Mosque

19th Group

Cannes Returnees Plot

Charlie Hebdo Shootings 2015

Brussels Euro2016 Attack

Cannes-Torcy Recruitment C

Lunel Recruitment Ring

Brussels Bombings 2016

El Baraka Mosque

Paris Attacks 2015

Paris Church Plot 2015

France Train Attack 2015

Jewish Museum Shooting 2014

Sharia4Belgium

Bombing Attempt: La Defense 2015

Zerkani Network

Verviers Network

Artigat Recruitment Ring

Kamikaze Riders Motorcycle Club

Normandy Church Attack 2016

NETWORK STATISTICS:

352 Total number of people in the network
85 Individuals Directly Linked to Attack Plots

86 Deceased
26 of whom are in the "Inner Circle"
319 People Arrested

749 Interpersonal Links
11 Plots—6 Successful; 5 Foiled
9 Organizations

12 Different Ethnicities Among European Residents and Citizens
Moroccan—23%, Algerian—14%, White-European (converts)—8%,
Tunisian—6%, Sub-Saharan—3%, Unknown—27%

41 Non-Western Individuals
Countries of Origin: Saudi Arabia—10, Morocco—7, Algeria—4, Tunisia—4,
All other non-Western countries—1 or 2 Individuals

Heat Map of "Hometowns" of Individuals Linked to the Paris-Brussels Attack Network