

# 2019 Summer School in Public Auditing and Accountability

## Module 3 – Building Digital Audit Capabilities

### A day in the Lab of Data Scientists



EUROPEAN  
COURT  
OF AUDITORS

10 July 2019

Processing and visualising data for auditors  
The ECA practice

Emanuele Fossati

# The Inevitable Introduction, because many people asked about my background



Class: ECA Special Report  
Score: 92%

Class: person (70%)  
Subclass: official (50%)  
speaker (50%)  
Age: 41-45  
Gender: Male



1994

Small Internet Startup

2000

Telecommunications  
Engineering Degree

2001

Business Consultant

2007

European Court of  
Auditors – Service Mgr

2017

European Court of  
Auditors - ECALab

# Daily challenges when doing applied research

1

Many different internal customers -> very different expectations

2

Colleagues ask for cooperation at different phases of the Audit lifecycle

3

Data availability and quality

4

Compliance with existing Audit methodologies, habits and deadlines

5

Compliance with IT security processes and GDPR

6

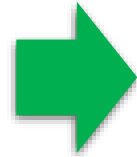
Building solutions with heterogeneous “Lego blocks”

# Daily challenges when doing applied research

2

Colleagues ask for cooperation at different phases of the Audit lifecycle

Definition of  
Audit  
Questions

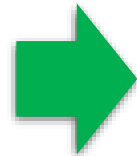


Request Data



Receive Data

Clean-up Data



Analysis



Visualisation  
and  
Discussion





# Closing the Expectations Gap



**Auditors**

- Refine Audit Questions
- Request help for data gathering and preparation
- Set-up analysis and verify results
- Derive Audit Findings



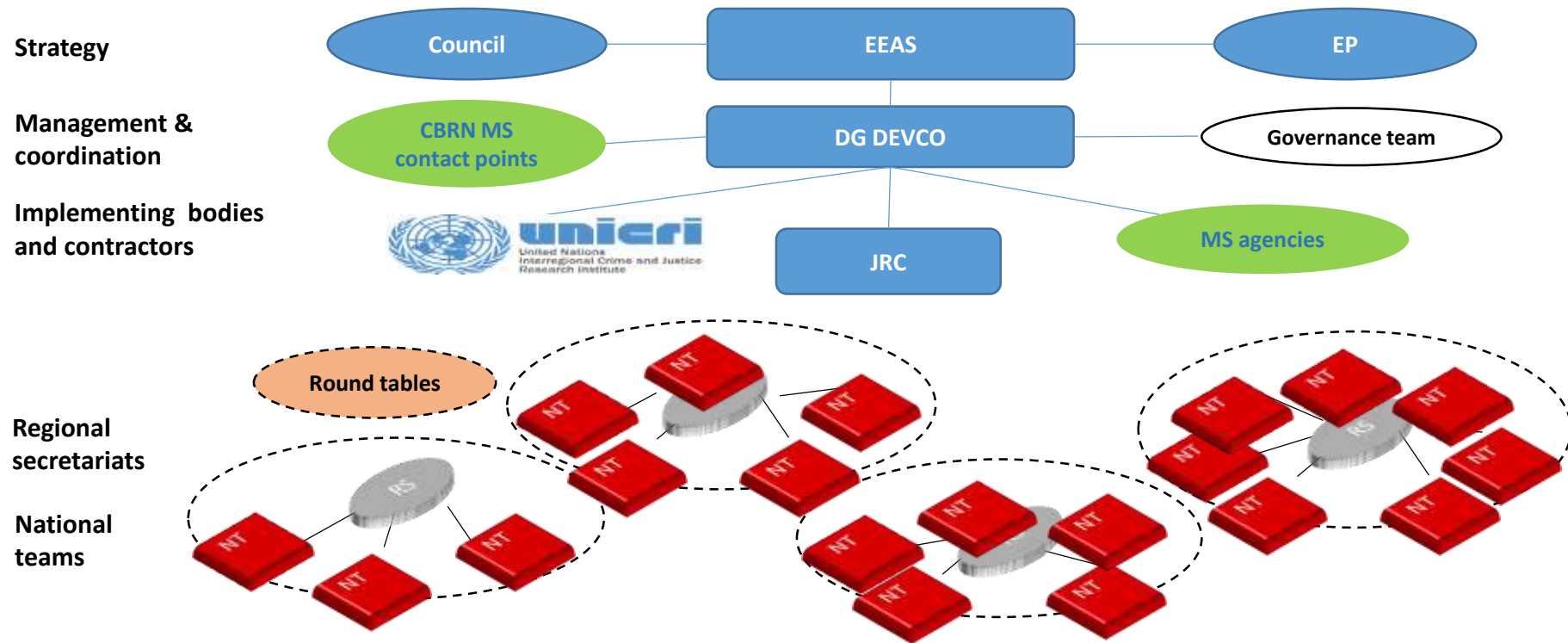
- Test several solutions until the best one is chosen
- Develop prototypes
- Prepare ad-hoc analysis



**Data Scientists**

# Case 1: Chemical, Biological, Radiological, Nuclear Centres Of Excellence (1/3)

Can the EU's Centres of Excellence initiative contribute effectively to mitigating chemical, biological, radiological and nuclear risks from outside the EU?



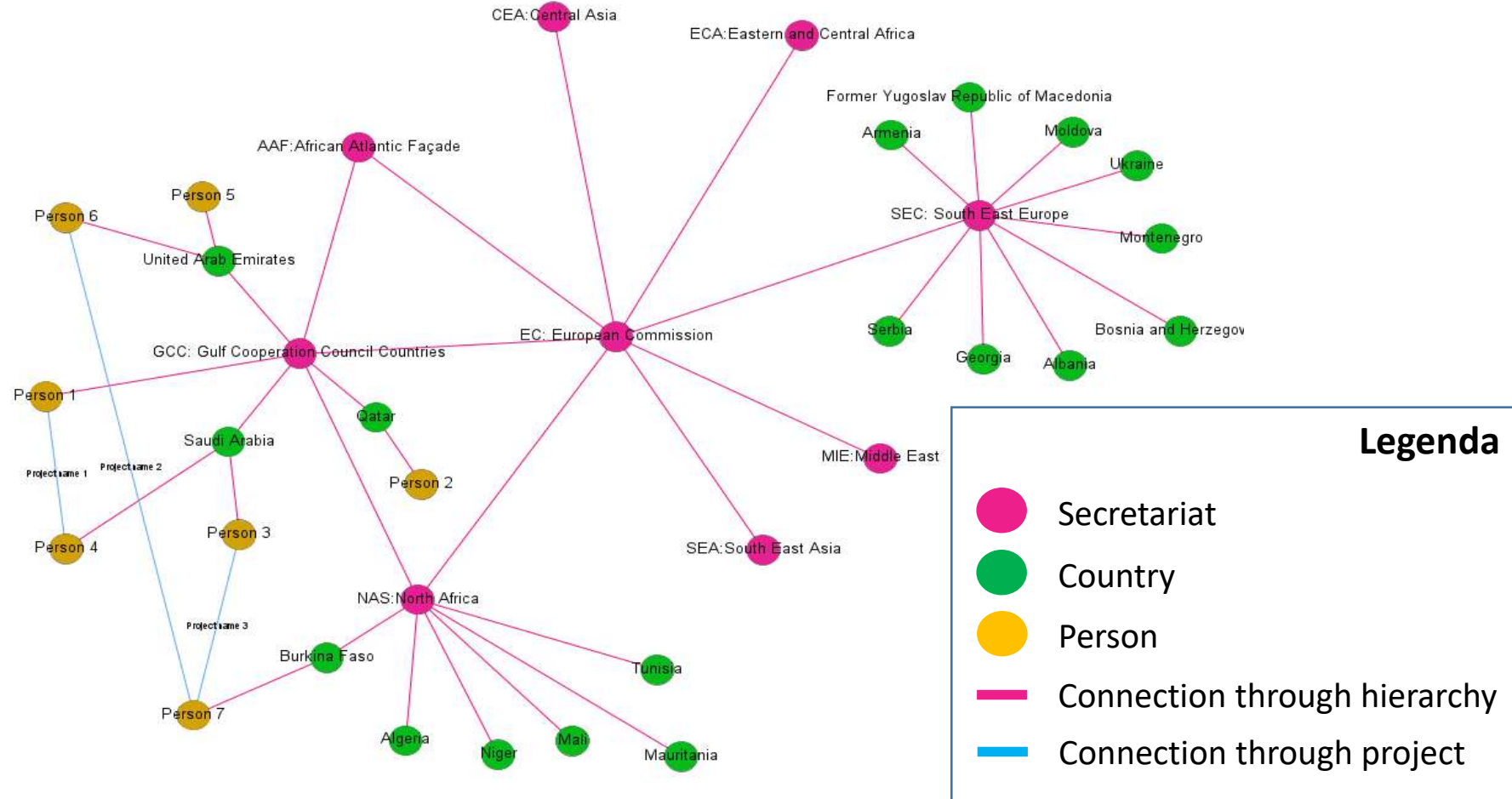
Follow-up of a previous audit task

# Case 1: Chemical, Biological, Radiological, Nuclear Centres Of Excellence (2/3)

How «Person 3» is connected to «Person 7»?

- Person 3 -> Saudi Arabia -> GCC -> NAS -> Burkina Faso -> Person 7 (through hierarchy)
- Person 3 -> Project 3 -> Person 7 (through project)

How different is the Network in 2017, compared to 2013?

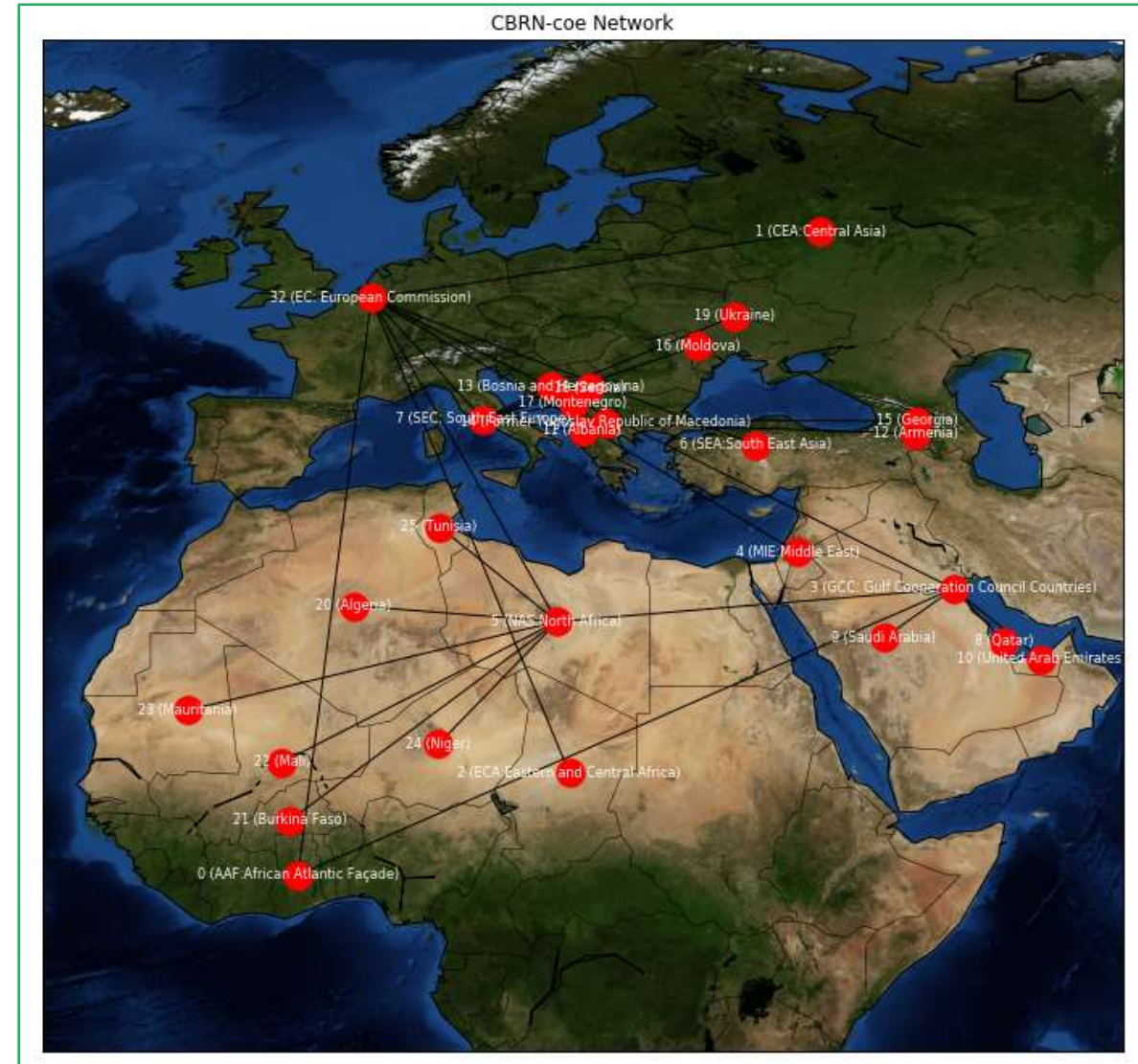


## Creation of a Graph for Network Analysis

# Case 1: Chemical, Biological, Radiological, Nuclear Centres Of Excellence (3/3)

## Information required to complete the Network Model?

- List of projects and countries
- List of people involved in each project, with start and end date
- list of meetings with participants, place and date



The Audit Task was already in an advanced phase, not possible to add info



## Case 2: Analysis of a Survey, import data and apply statistical corrections (1/2)



Import and clean-up  
Excel file containing  
the results

```
# Which columns contain open text information?
```

```
open_questions = ["Other, please specify:.1",  
                  "Other, please specify:.2",  
                  "Other, please specify:.3",  
                  "Other, please specify:.4",  
                  "14. Comments:",  
                  "Other, please specify:.5",  
                  "22. If not, please tell us why:",  
                  "24. Comment:",  
                  "26. Comment:"]
```

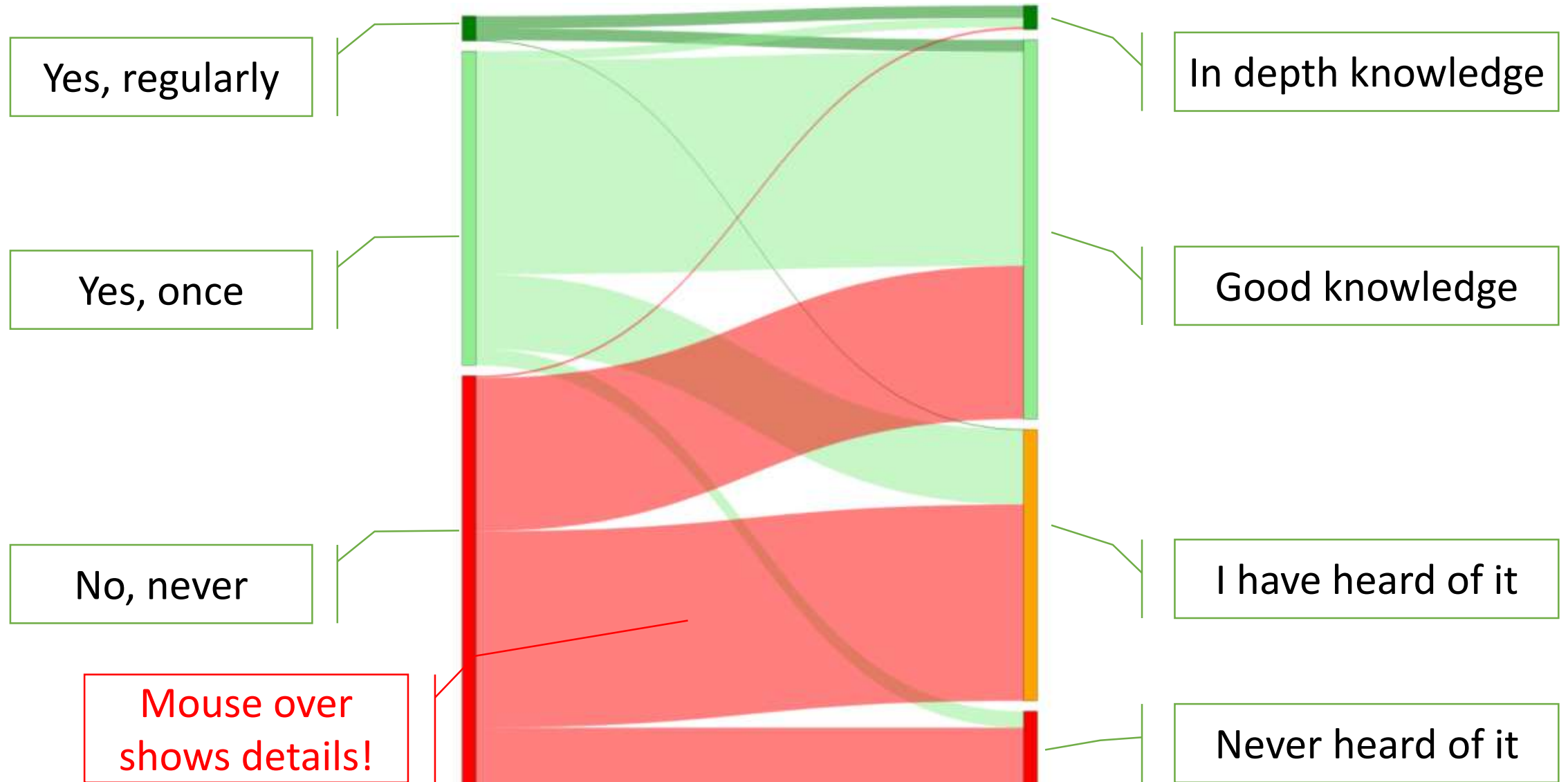
```
# Which columns contain multiple choices?
```

```
multiple_choices = ["8. If you need advice on ethical matters, to whom could you turn, according to your inst  
                    "9. If you wish to report unethical conduct, to whom are you expected to turn, according  
                    "10. If you report unethical conduct, what protective measures can you expect from your i  
                    "11. How did you learn about the ethical framework applicable within your institution? Pl  
                    ]
```

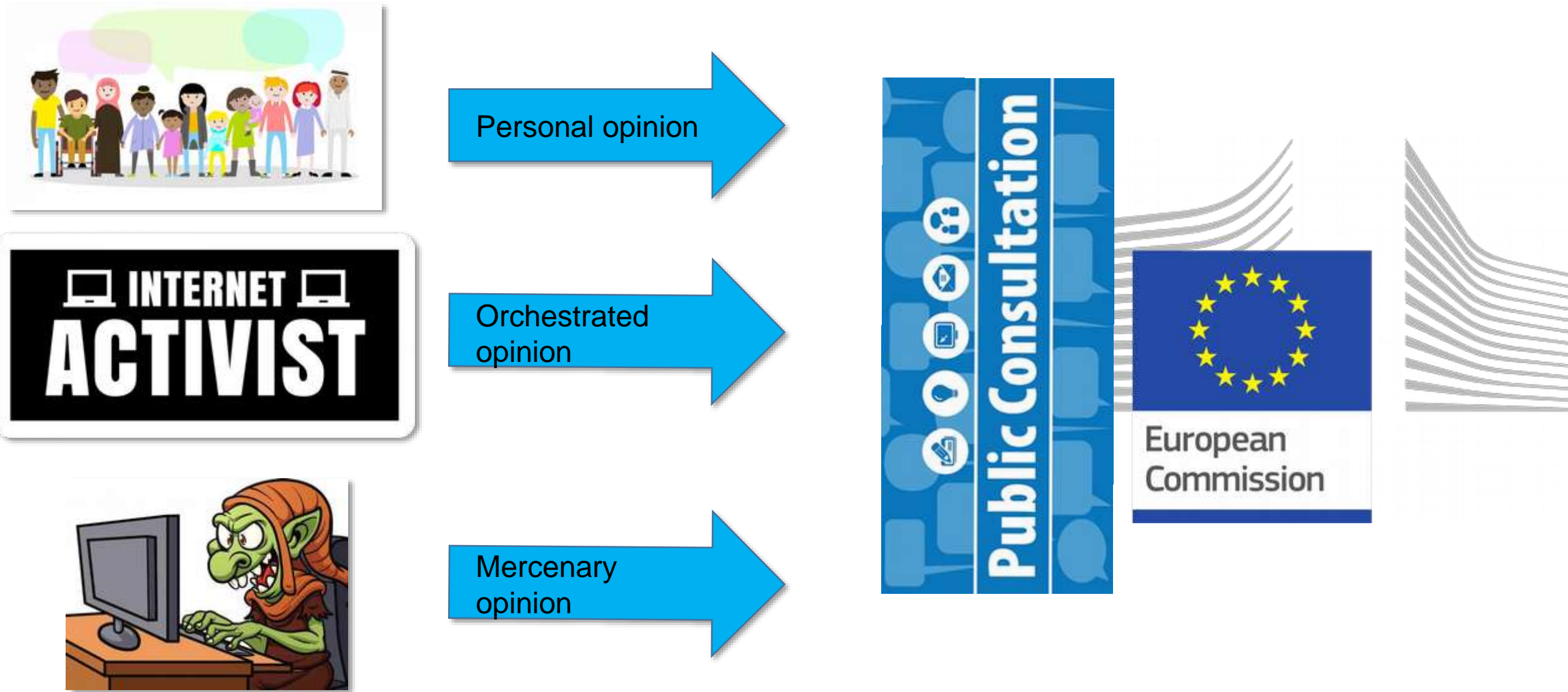
```
min_words = 2 # Set the minimum amount of words that a text has to have in order to be sent to the language c
```

The Python script acts as formal audit trail, we never modify the evidence doc!

# Case 2: Analysis of a Survey, relations between two answers (2/2)



# Case 3: Large Datasets from EC Public Consultations (1/5)



Understanding the context, knowledge transfer to the Audit Team

# Case 3: Trolls, difficulty “well-funded NLP expert” (2/5)



NATURAL LANGUAGE PROCESSING



Botnet based on hacked PC and Internet of Things (IOT) objects

Contribution at 10:00 am

{ Random delay

Similar concept contribution at 10:03 am

{ Random delay

Similar concept contribution at 10:10 am



“Legit” IP addresses, similar concepts contributions, random timing



## Case 3: Large Datasets from EC Public Consultations (3/5)

**1**

Imported around 4.500.000 individual contributions. Auditee was very happy to cooperate

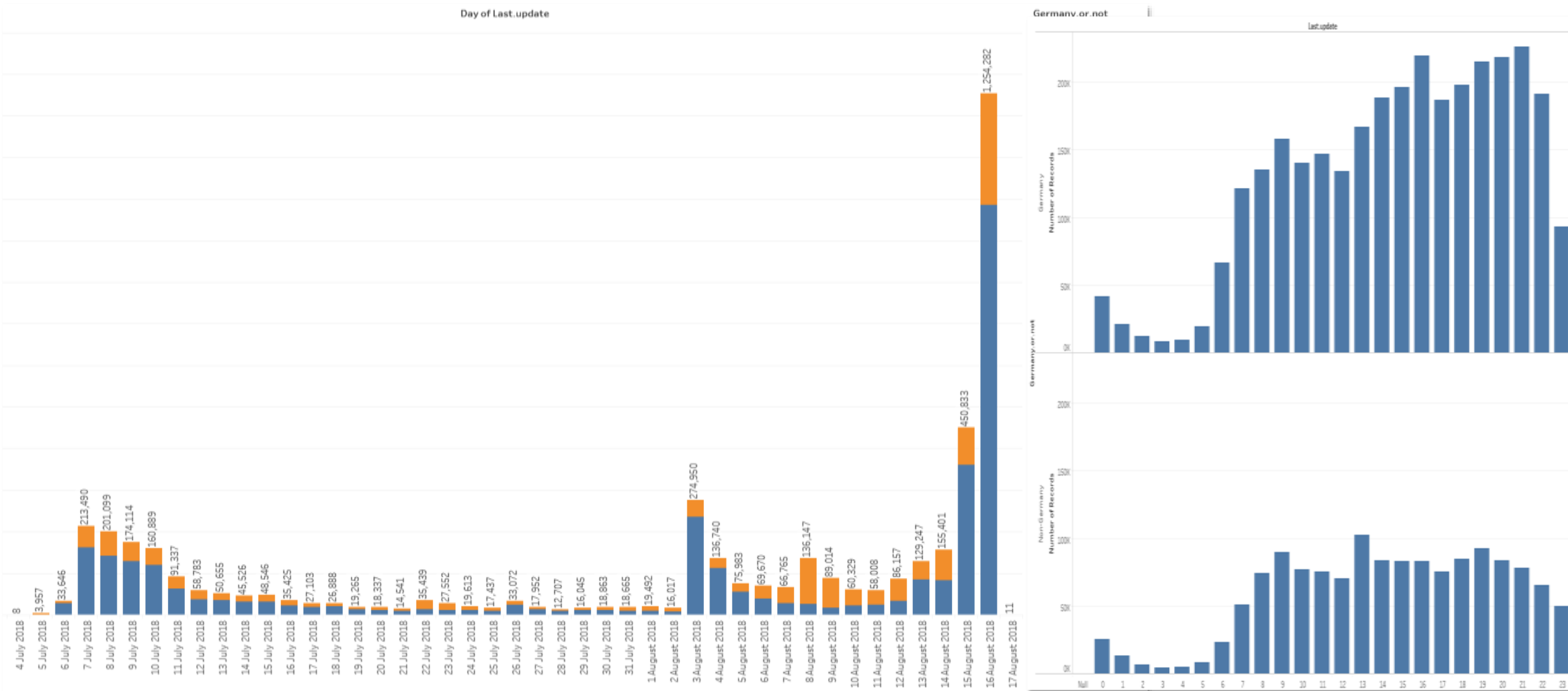
Analysis on dates and times, looking for anomalies

**2**

**3**

Language detection, machine translation and text analysis of around 800.000 open answers

# Case 3: Large Datasets from EC Public Consultations (4/5)



Date and Time analysis

## Case 3: Large Datasets from EC Public Consultations (5/5)

It's bad for my clocks, and a nuisance. The "verify my humanity" IS BLOODY BADLY IMPLEMENTED. IT GOES ON AND ON AND ON!  
DONT DO THIS!

Eingriff in die Natur mithilfe nahezu wirkungslosem Aktionismus zu Lasten auch von Kleinkindern. Weshalb haben wir den Sommerzeit-Blödsinn solange mitgemacht, obwohl er keine erkennbaren Vorteile brachte? Und: Wer hat daran verdient?

Simply abolishing the daylight savings time is not enough; I would heartily recommend a punishment for those who introduced this abysmal crap in the first place!!

Fact about summer time is that it is just a power demonstration from the so called "elite", and it actually doesn't help anybody. So let's abolish it!

summer time power demonstration  
anybody

It would be high time for European citizens to be consulted on all the issues that concern them, as our Members never have the last word and the unelected Members of the Commission have!

time citizen issue member word members  
commission

Inappropriate speech detection in the most used 5 languages

## Small break: how the ECALab looks like?

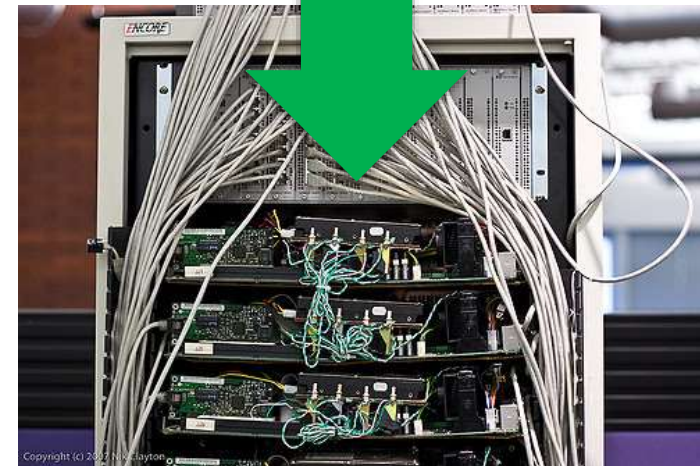
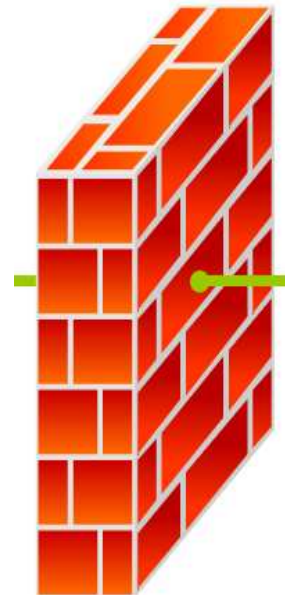
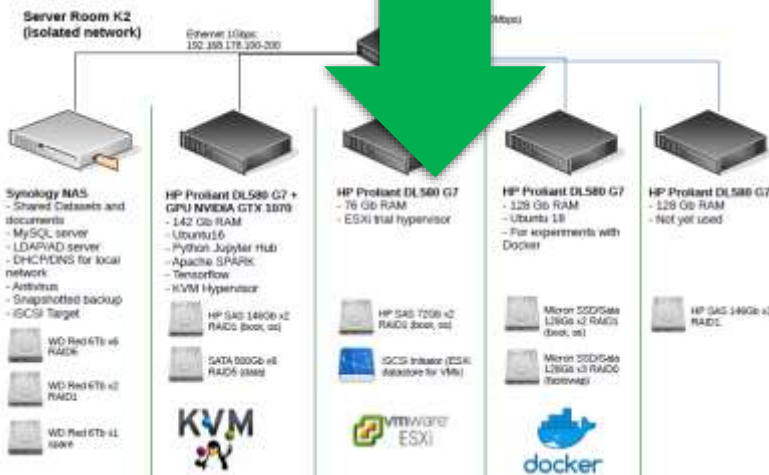


### Fact sheet:

- “Startup” founded in early 2017
- 2.5 full time resources, plus Principal Manager
- Around 10 collaborators
- External consultant just started
- Intern from Pisa starting from September
- 25 ~ 30 activities



# Small break: the ECA Lab IT Infrastructure, visually



**ECA Lab Infrastructure**

**ECA Secured Environment**

# Case 4: Analysis of public impact / external scrutiny of EU agencies (1/5)



**Aggregated Content**  
(text and time analysis)

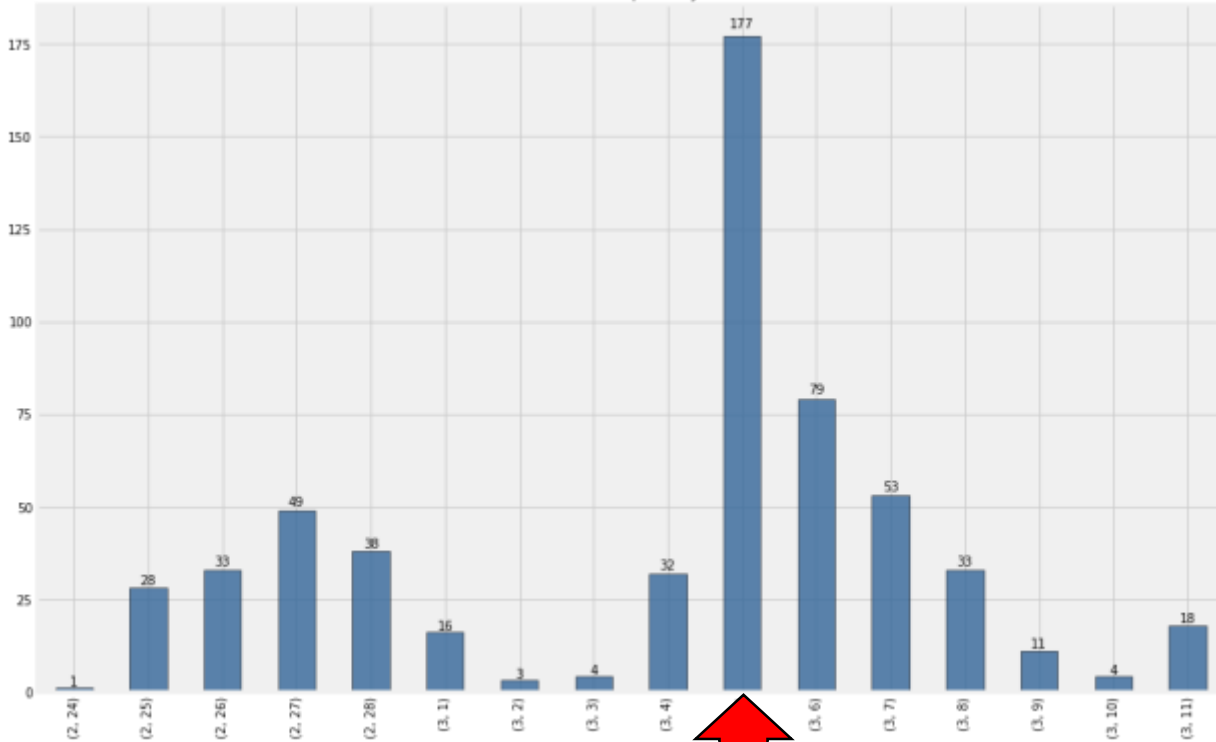


**Connections between users**  
(network analysis)

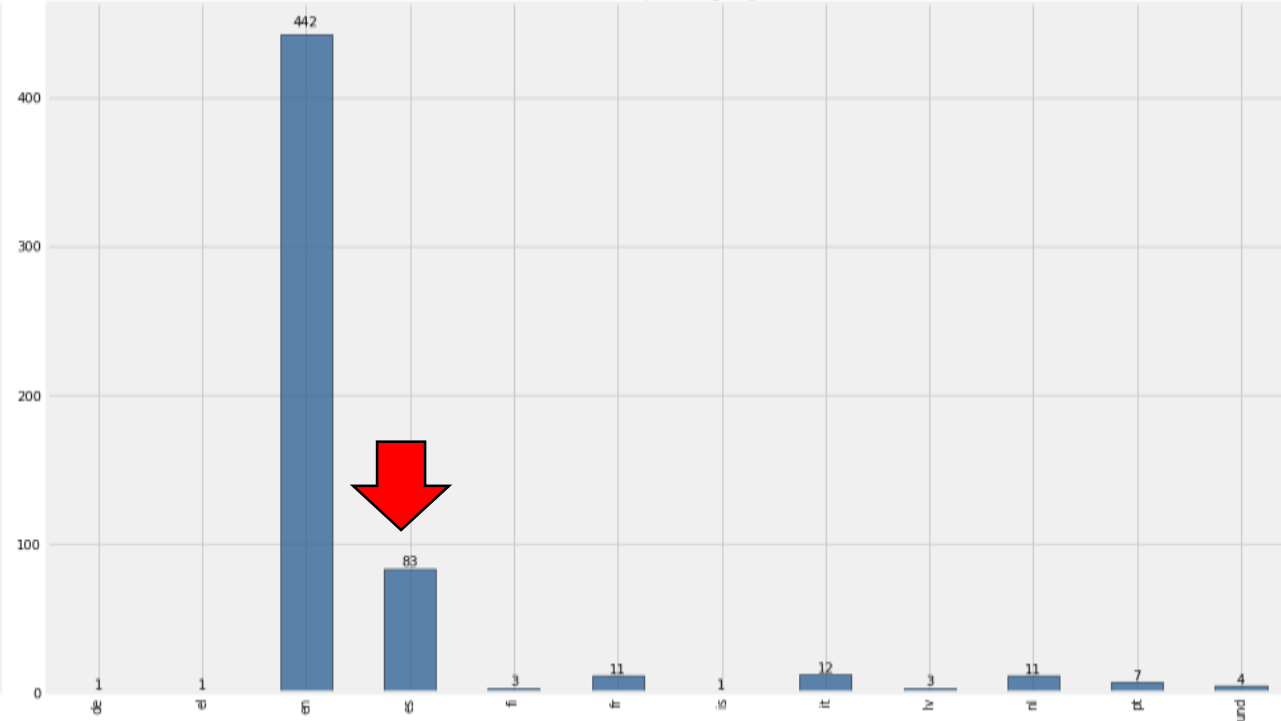
Sources: Twitter, Europe Media Monitor...

# Case 4: Analysis of public impact of EU agencies (2/5)

Tweets per day



Tweets per language



- Insight 1: what happened the 5/3? An event!
- Insight 2: an important part of the public conversation is in Spanish (EU-OSHA is based in Spain, so that makes sense)

## EU-OSHA good practice exchange event and award ceremony, Brussels

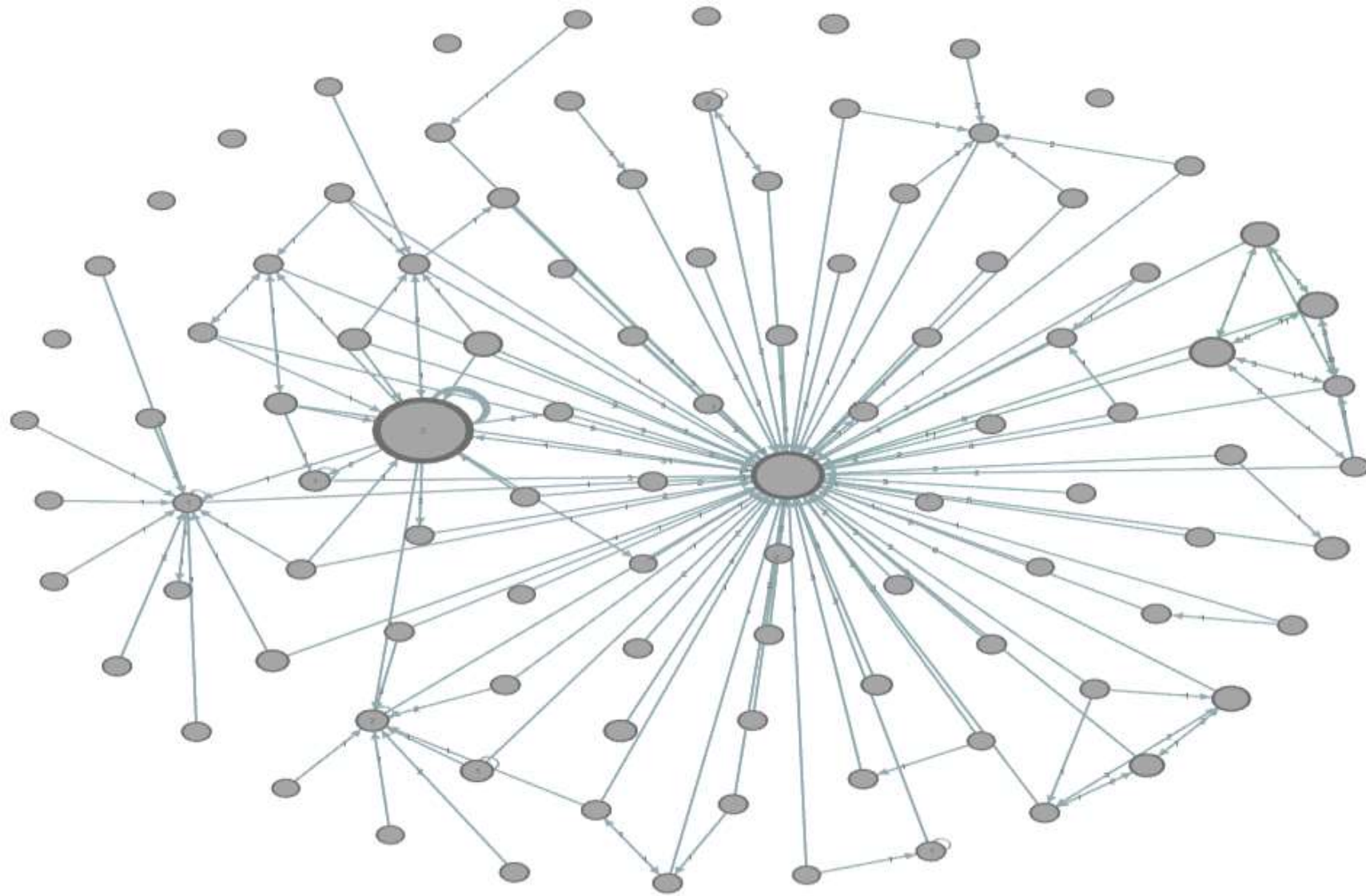
date	March 5, 2019 - March 6, 2019
city	Brussels, Belgium
location	Residence Palace, Brussels, Belgium
organisation	European Agency for Safety and Health at Work (EU-OSHA)



© Kevin Bergenhenegouwen

Start date : 05/03/2019

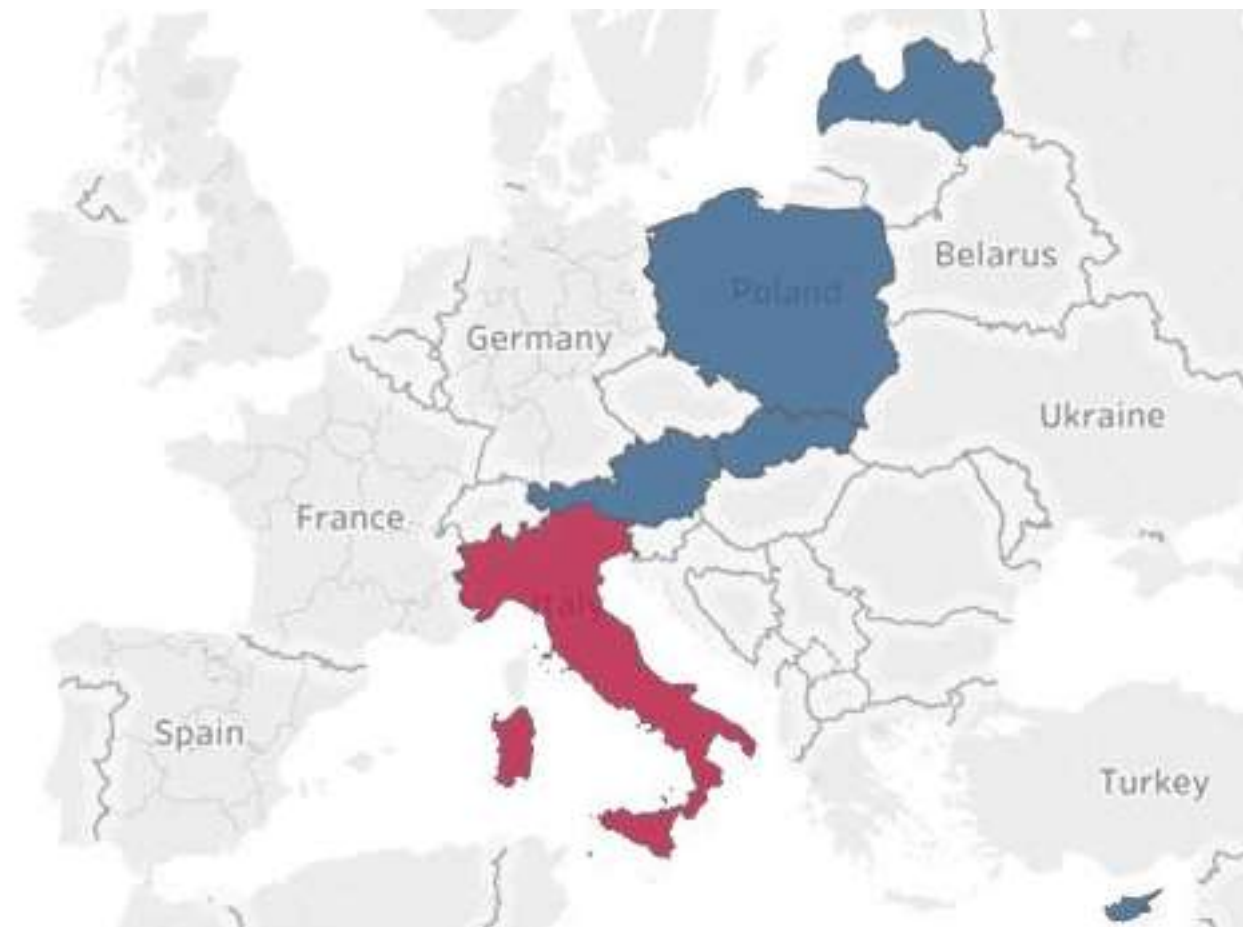
## Case 4: Analysis of public impact of EU agencies (3/5)



**Social Media Analysis: Network Analysis**



## Case 4: Analysis of public impact of EU agencies (4/5)



Source: Europe Media Monitor

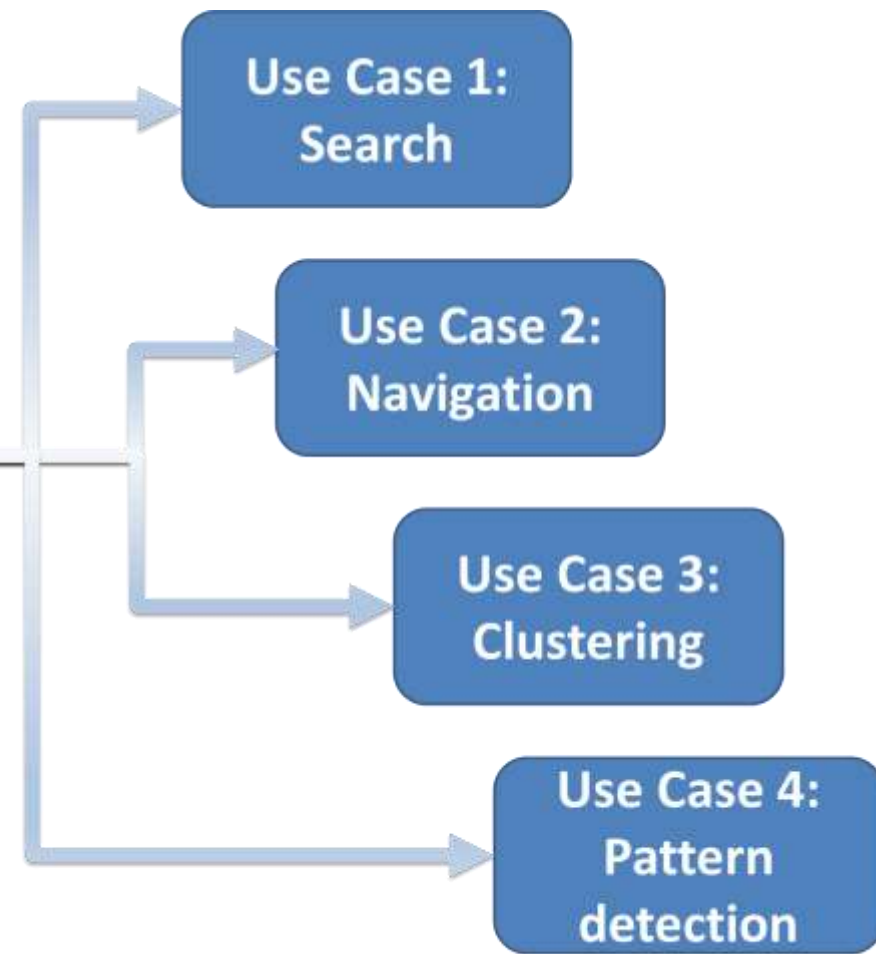
# Case 4: Analysis of public impact of EU agencies (5/5)

#	Paragraph	NLTK/Vader				IBM Watson AI cloud service		
		neg	neu	pos	com	tone	sentiment	emotions
1	Notes that, with regard to one comment made in the Court's 2012 report regarding the education contribution, which was marked as "ongoing" in the Court's 2013 and 2014 reports, corrective actions were taken by the Authority and contracts were signed with 26 schools attended by children of members of staff, and negotiations with further 4 schools are ongoing;	0	0,977	0,023	0,0772	{'score': 0.752665, 'tone_id': 'confident', 'tone_name': 'Confident'}, {'score': 0.747663, 'tone_id': 'analytical', 'tone_name': 'Analytical'}}	{'score': 0, 'label': 'neutral'}	{'sadness': 0.163275, 'joy': 0.023784, 'fear': 0.076716, 'disgust': 0.235633, 'anger': 0.170385}
2	Acknowledges that the budget monitoring efforts during the financial year 2016 resulted in a budget implementation rate of 96,76 %, representing a decrease of 2,58 % compared to 2015, and that the rate of execution of payment appropriations was 88,67 %, representing a decrease of 1,03 %; acknowledges from the Authority that this continued high rate of execution is due to good budget planning and monitoring;	0	0,935	0,065	0,4939	{'score': 0.944233, 'tone_id': 'analytical', 'tone_name': 'Analytical'}}	{'score': -0.602495, 'label': 'negative'}	{'sadness': 0.182305, 'joy': 0.211506, 'fear': 0.037336, 'disgust': 0.059797, 'anger': 0.085478}
3	Notes that, due to the increase in the value of the Euro against the British Pound in 2016, the Authority requested a decreasing amending budget of EUR 1 572 000;	0	0,806	0,194	0,6124	{'score': 0.654266, 'tone_id': 'analytical', 'tone_name': 'Analytical'}}	{'score': -0.836134, 'label': 'negative'}	{'sadness': 0.198944, 'joy': 0.070583, 'fear': 0.281833, 'disgust': 0.159205, 'anger': 0.117089}

# Case 5: Large document set for preparation of the Annual Work Program (1/5)



- Stack of Documents
- Known (SR, Regulations,...)
  - Unknown (ED)



Goals: what I want to achieve

Document Navigation (improved) prototype



# Case 5: Large document set for preparation of the Annual Work Program (2/5)

Document listText FeaturesClusters ( 4 7 12 15 )Topics ( 5 10 15 20 25 )Clear all filters

Dataset: Evidence Documents for 18CH3020, Briefing Paper on the EU Cybersecurity policy  
Emanuele Fossati / ECALab / Oct 2018

Documents (367)	Original	Paragraphs	Eurovoc Categories
<input type="radio"/> 171002-end-6-2017-dgd-2-letter-pr-annex-en[1][1].pdf	<input type="radio"/>	<input type="radio"/> 1	<ul style="list-style-type: none"><li><input type="radio"/> 0.15 - Council of the European Union</li><li><input type="radio"/> 0.13 - armed forces</li><li><input type="radio"/> 0.13 - EU Member State</li><li><input type="radio"/> 0.11 - information transfer</li><li><input type="radio"/> 0.11 - vocational training</li><li><input type="radio"/> 0.10 - self-employed person</li></ul>
<input type="radio"/> 18_7925 Council conclusions on on malicious cyber activities[1].pdf	<input type="radio"/>	<input type="radio"/> 1	<ul style="list-style-type: none"><li><input type="radio"/> 0.14 - information technology</li><li><input type="radio"/> 0.11 - data protection</li><li><input type="radio"/> 0.11 - cooperation policy</li><li><input type="radio"/> 0.10 - information society</li><li><input type="radio"/> 0.10 - international cooperation</li><li><input type="radio"/> 0.09 - electronic mail</li></ul>

Council conclusions on malicious cyber activities - approval  
Delegations will find in the Annex Council conclusions on malicious cyber activities adopted by the Foreign Affairs Council at its meeting held on 16 April 2018.

ANNEX Council Conclusions on malicious cyber activities The Council of the European Union adopted the following conclusions: The EU stresses the importance of a global, open, free, stable and secure cyberspace where human rights and fundamental freedoms and the rule of law fully apply for the social well-being, economic growth, prosperity and integrity of our free and democratic societies.

The EU recalls its Conclusions on the Framework for a Joint EU Diplomatic Response to Malicious Cyber Activities [1] which contributes to conflict prevention, cooperation and stability in cyberspace by setting out the measures within the EU's Common Foreign and Security Policy, including restrictive measures, that can be used to prevent and respond to malicious cyber activities.

The EU expresses its serious concern about the increased ability and willingness of third states and non-state actors to pursue their objectives by undertaking malicious cyber activities and will continue to bolster its capabilities to address cyber threats.

The EU underlines that States must not use proxies to commit internationally wrongful acts using ICTs, and should seek to ensure that their territory is not used by non-state actors to commit such acts as expressed in the 2015 report of the United Nations Groups of Governmental Experts in the Field of Information and Telecommunications in the Context of International Security (UNGGE).

The EU stresses that compliance with voluntary non-binding norms of responsible state behaviour in cyberspace contribute to an open, secure, stable, accessible and peaceful ICT environment.

The EU expresses its willingness to continue working on the further development and implementation of the voluntary non-binding norms, rules and principles for the responsible State behaviour in cyberspace as articulated in the 2010, 2013 and 2015 reports of the respective UNGGE, within the UN and other appropriate international fora.

Summary

Full document

Eurovoc categories



# Case 5: Large document set for preparation of the Annual Work Program (3/5)

Pages selected (10)

(1 of 1)

Page level

07\_The\_EUs\_Multiannual\_Financial\_Framework\_feei\_2018\_q3\_screen.pdf - page 1

Search similar pages



Keywords

Top 3 Keywords:

- European Union s budget ✘
- cohesion spending ✘
- budgets of federal countries ✘

Top 3 Categories:

- /society/work/unemployment ✘
- /law, govt and politics ✘
- /finance/investing/funds ✘

Top 3 Concepts:

- European Union ✘
- European Commission ✘
- Common Agricultural Policy ✘

Top 3 Entities:

- European Union (Organization) ✘
- EU (Location) ✘
- European Commission (Organization) ✘

Entities

☛ The European Union s budget which is fundamentally different from the budgets of federal countries and amounts to only about 1% of the EU s gross national income continues to be heavy on agricultural and cohesion spending. The literature shows that the EU s common agricultural policy (accounting for 38% of EU spending from the current budget) provides good income support, especially for richer farmers, but is less effective for greening and biodiversity and is unevenly distributed. The EU s ...

07\_The\_EUs\_Multiannual\_Financial\_Framework\_feei\_2018\_q3\_screen.pdf - page 2

Search similar pages



Top 3 Keywords:

- EU budget ✘
- EU budget revenues ✘
- stabilization policy ✘

Top 3 Categories:

- /finance/investing/funds ✘
- /society/work/unemployment ✘
- /finance/investing/beginning investing ✘

Top 3 Concepts:

- European Union ✘
- European Commission ✘
- Common Agricultural Policy ✘

Top 3 Entities:

- European Union (Organization) ✘
- EU (Location) ✘
- EUR (Location) ✘

Document Navigation prototype, page view

# Case 5: Large document set for preparation of the Annual Work Program (4/5)

07\_The\_EUs\_Multiannual\_Financial\_Framework\_feei\_2018\_q3\_screen.pdf - page 1

Search similar pages



#### Top 3 Keywords:

- European Union s budget ✕
- cohesion spending ✕
- budgets of federal countries ✕

#### Top 3 Categories:

- /society/work/unemployment ✕
- /law, govt and politics ✕
- /finance/investing/funds ✕

#### Top 3 Concepts:

- European Union ✕
- European Commission ✕
- Common Agricultural Policy ✕

#### Top 3 Entities:

- European Union (Organization) ✕
- EU (Location) ✕
- European Commission (Organization) ✕

• The European Union s budget which is fundamentally different from the budgets of federal countries and amounts to only about 1% of the EU s gross national income continues to be heavy on agricultural and cohesion spending. The literature shows that the EU s common agricultural policy (accounting for 38% of EU spending from the current budget) provides good income support, especially for richer farmers, but is less effective for greening and biodiversity and is unevenly distributed. The EU s...

## The EU's Multiannual Financial Framework and some implications for CESEE countries

*The European Union's budget – which is fundamentally different from the budgets of federal countries and amounts to only about 1% of the EU's gross national income – continues to be heavy on agricultural and cohesion spending. The literature shows that the EU's common agricultural policy (accounting for 38% of EU spending from the current budget) provides good income support, especially for richer farmers, but is less effective for greening and biodiversity and is unevenly distributed. The EU's cohesion policy (accounting for 34% of current EU spending) contributes to convergence, but it is unclear how strong and long-lasting the effects are. Spending on new priorities such as border control could require additional funds of at least EUR 100 billion in the 2021–2027 period, but there will be a EUR 94 billion Brexit-related hole in the EU budget for 2021–2027 if the EU loses the United Kingdom's share of contributions and the EU's work program as a share of gross national income remains unchanged. The European Commission's May 2, 2018, proposal for the 2021–2027 budget makes several welcome steps in reforming the EU budget, e.g. by reorganizing spending commitments toward priorities which have gained more importance recently, while reducing the share of spending on agriculture and cohesion policies. But many details remain quite fuzzy.*

Zsolt Darvas,  
Guntram B. Wolff<sup>1</sup>

Document Navigation prototype, page view

# Case 5: Large document set for preparation of the Annual Work Program (5/5)

## Search Text Features:



## Top global keywords:

countries (508) ✗  
Moldova (328) ✗  
Source (299) ✗  
European Commission (272) ✗  
country (263) ✗  
Germany (262) ✗  
example (256) ✗  
Figure (227) ✗  
addition (224) ✗  
member states (215) ✗

## Top global categories:

/society/work/unemployment (962) ✗  
/law, govt and politics (946) ✗  
/finance/investing (504) ✗  
/law, govt and politics/politics (442) ✗  
/finance/financial news (413) ✗  
/education/homework and study tips (354) ✗  
/finance/investing/beginning investing (339) ✗  
/society/unrest and war (313) ✗  
/law, govt and politics/government contracting and procurement (277) ✗  
/finance/investing/funds (233) ✗

## Top global concepts:

European Union (1379) ✗  
United Kingdom (415) ✗  
Europe (358) ✗  
Economics (284) ✗  
European Commission (276) ✗  
International trade (249) ✗  
Treaty of Lisbon (219) ✗  
European Parliament (208) ✗  
Euro (190) ✗  
Slovakia (188) ✗

## Top global entities:

EU (Organization) (1088) ✗  
Germany (Location) (434) ✗  
Europe (Location) (402) ✗  
Moldova (Location) (389) ✗  
European Commission (Organization) (295) ✗  
Ukraine (Location) (272) ✗  
France (Location) (239) ✗  
UK (Location) (213) ✗  
Georgia (Location) (206) ✗  
European Union (Organization) (202) ✗

## Least common global keywords:

match ✗  
location worker ✗  
skilled work MyHammer ✗  
routine work ✗  
Testbirds ✗  
de Online client-determined  
specialist work CrowdFlower

## Least common categories:

/education/distance learning ✗  
/education/language learning ✗  
/art and entertainment/shows and events/exhibition ✗  
/art and entertainment/music/music reference/music composition and theory ✗

## Least common concepts:

Computing platform ✗  
Humboldt University of Berlin ✗  
Die Zukunft ✗  
North Rhine-Westphalia ✗  
Bochum ✗  
French Third Republic ✗  
Rhine-Ruhr ✗

## Least common entities:

Ruhr University (Organization) ✗  
Workers Protection National (Organization) ✗  
Bochum (Location) ✗  
Waltermann (Person) ✗  
Welche (Company) ✗  
Arbeitswelt (Location) ✗

# Case 6: Search projects related to "Industry 4.0" (1/4)

```
import pandas as pd
from sklearn.metrics.pairwise import cosine_similarity
import spacy
```

```
pd.set_option('display.max_colwidth', -1)
pd.set_option('display.max_columns', None)
```

```
# Import file
filename = '2014DE16M20P001_CH2109757DE01-19AA-Industry_info_request-TR.XLSX'
projects = pd.read_excel("Excel_files/"+filename, skiprows=1).fillna('')
projects.shape
```

```
(806, 16)
```

```
projects.dtypes
```

Bezeichnung des Projekts	object
Prioritätsachse	object
Maßnahme	object
Thematisches Ziel	object
Interventions-\ncode	int64
Kurze Beschreibung des Projekts	object
Endbegünstigter	object
Projektstandort	object
Projektbeginn (Datum)	int64
(Geplantes) Projektende (Datum)	int64
Ursprünglich vereinbarte Gesamtkosten	float64
Ursprünglich vereinbarter\n EU-Beitrag	float64
Von der Verwaltungs-behörde bescheinigte Ausgaben (Stand: Ende Februar 2019)	float64
Geplante Projektziele und Indikatoren	object
Erzielte Projektoutputs und -ergebnisse	object
Antrags-Nr.	int64
dtype: object	



Import excel file with project descriptions (in German) as Pandas DataFrame



## Case 6: Machines love numbers, but we have texts (2/4)

Word

Word2Vector

100  
numbers

“Similar” words have similar vector representation!

Any text

Word2Vector

100  
numbers

“Similar” texts have similar vector representation!

Word2Vector -> Sentence2Vector -> Document2Vector

## Case 6: Search projects related to “Industry 4.0” (3/4)

```
# Definition of Industrie 4.0 from Wikipedia DE
```

```
definition_de = ""
```

```
Industrie 4.0 ist die Bezeichnung für ein Zukunftsprojekt zur umfassenden Digitalisierung der industriellen Produktion, um sie  
Inhaltsverzeichnis
```

- 1 Bezeichnung
- 2 Organisationsgestaltungsprinzipien
- 3 Geschichte und Akteure
- 4 Bedeutung
- 5 Herausforderungen
- 6 Entwicklungen im Ausland
- 7 Rezeption der Plattform Industrie 4.0
- 8 Filmische Dokumentationen
- 9 Literatur
- 10 Dokumente
- 11 Weblinks
- 12 Einzelnachweise

```
Bezeichnung
```

```
Mit der Bezeichnung Industrie 4.0 soll das Ziel zum Ausdruck gebracht werden, eine vierte industrielle Revolution einzuleiten:
```

```
Die erste industrielle Revolution bestand in der Mechanisierung mittels Wasser- und Dampfkraft; darauf folgte  
die zweite industrielle Revolution, geprägt durch Massenfertigung mit Hilfe von Fließbändern und elektrischer Energie, sowie  
die dritte industrielle Revolution oder digitale Revolution mit Einsatz von Elektronik und IT (v. a. die Speicherprogramme)
```

Look for the definition of “Industrie 4.0” from German Wikipedia

# Case 6: Search projects related to "Industry 4.0" (4/4)

```
shown_columns = ['Bezeichnung des Projekts', 'Kurze Beschreibung des Projekts', 'Lemmas', 'score', 'rank']
projects[shown_columns].sort_values(by='rank')
```

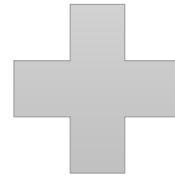
	Bezeichnung des Projekts	Kurze Beschreibung des Projekts	Lemmas	score	rank
325	PersonA-PP – Persönliche Assistenz für Patienten in der Pflege	Entwicklung eines Assistenten für Pflegepatienten auf der Basis eines kollaborativen Industrieroboters mit innovativer Mensch-Maschine-Schnittstelle	PersonA-PP Assistenz Patient Pflege Entwicklung Assistent Pflegepatienten Basis Industrieroboters Mensch-Maschine-Schnittstelle	0.986279	1.0
74	Entwicklung und Zulassung eines Um- und Nachrüstbausatzes für Jetantrieb	Entwicklung von dem Segelflugzeug LS-4 für die ergänzende Musterzulassung mit einem Um- und Nachrüstbausatz eines luffahrtzugelassenen Jetantriebs	Entwicklung Zulassung Nachrüstbausatzes Jetantrieb Entwicklung Segelflugzeug LS-4 Musterzulassung Nachrüstbausatz Jetantriebs	0.984917	2.0
124	Entwicklung eines Produktionsplanungs- und Steuerungssystems	Folgende Ziele werden verfolgt: 1. Termintreue mit Toleranz von 1 AT, Reduzierung Durchlaufzeiten um 10%, 3. Erhöhung der Auslastung um 15%.	Entwicklung Steuerungssystems zielen Termintreue Toleranz AT Reduzierung Durchlaufzeiten Erhöhung Auslastung	0.983456	3.0
583	Erweiterung einer bestehenden Betriebsstätte	Neubau einer Lager- sowie einer Produktionshalle und Anschaffung von Maschinen, Anlagen, Einrichtungen und Software	Erweiterung Betriebsstätte Neubau Produktionshalle Anschaffung Maschine anliegen Einrichtung Software	0.982145	4.0
91	Entwicklung eines Messsystems zum Fernmonitoring von Schraubverbindungen	Entwicklung eines technischen Produkts, der eine Drehung der Mutter messen kann und die Messdaten in ein Fernüberwachungssystem übermitteln kann	Entwicklung Messsystems Fernmonitoring Schraubverbindungen Entwicklung Produkt Drehung Mutter Messdaten Fernüberwachungssystem	0.981989	5.0
298	ReKuTe - Partizipative Wissenschaft für Region, Kultur, Technik TP V	Aktionspläne, Dialogformate und beispielhafte Pilotprojekte für ein nachhaltiges Netzwerk zur Förderung Partizipativer Wissenschaft in der Region	ReKuTe Wissenschaft Region Kultur Technik TP Aktionspläne Dialogformate Pilotprojekte Netzwerk Förderung Wissenschaft Region	0.981571	6.0
238	Ad-Hoc Vernetzung zur Gefahrenerkennung und -abwehr	Entwicklung von Hardware, Kommunikations-Protokollen und Software zum Aufbau teil-autonomer, intelligenter Ad-Hoc Netzwerke zur Gefahrenabwehr.	Vernetzung Gefahrenerkennung -abwehr Entwicklung Hardware Kommunikations-Protokollen Software Aufbau Ad-Hoc Netzwerke Gefahrenabwehr	0.981199	7.0
359	ReKuTe - Partizipative Wissenschaft für Region, Kultur, Technik TP JHS	Aktionspläne, Dialogformate und beispielhafte Pilotprojekte für ein nachhaltiges Netzwerk zur Förderung Partizipativer Wissenschaft in der Region.	ReKuTe Wissenschaft Region Kultur Technik TP JHS Aktionspläne Dialogformate Pilotprojekte Netzwerk Förderung Wissenschaft Region	0.981131	8.0

Sentence2Vector (Spacy, DE model) and rank

# Case 7: Visualisation of investment coverage (1/3)



Excel file from the Auditee



Excel file extracted by Audit Team to complete missing information

Verify data correctness

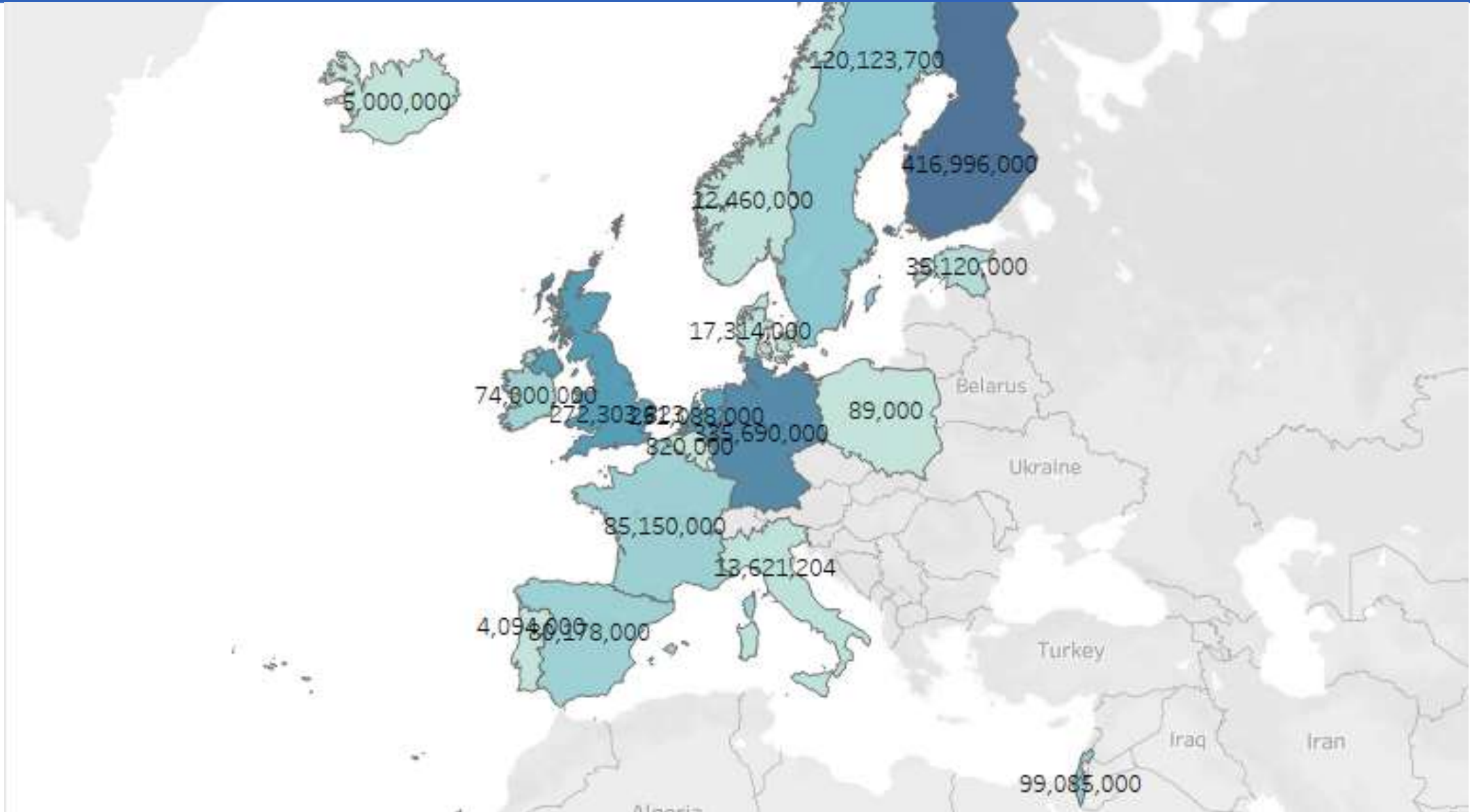
The screenshot shows the dealroom.co website interface. The top navigation bar includes the logo, a search bar with a lock icon and the text "Login to lookup a person or entity", and two tabs: "Enza biotech funding" (selected) and "Enza biotech investors". A left sidebar contains navigation links: Dashboard, Startups & scaleups, Corporates, Investors, Funding rounds, Exits, and People. The main content area displays a table of funding data for Enza biotech.

DATE	INVESTORS	AMOUNT	ROUND
Dec 2014	Horizon 2020 SME instrument	€50k	GRANT
Mar 2015	N/A	N/A	
Jul 2017	Croda International Plc	N/A	ACQUISITION
<b>Total Funding*</b>		<b>€50k</b>	

Web scraping, Python clean-up, total amounts split

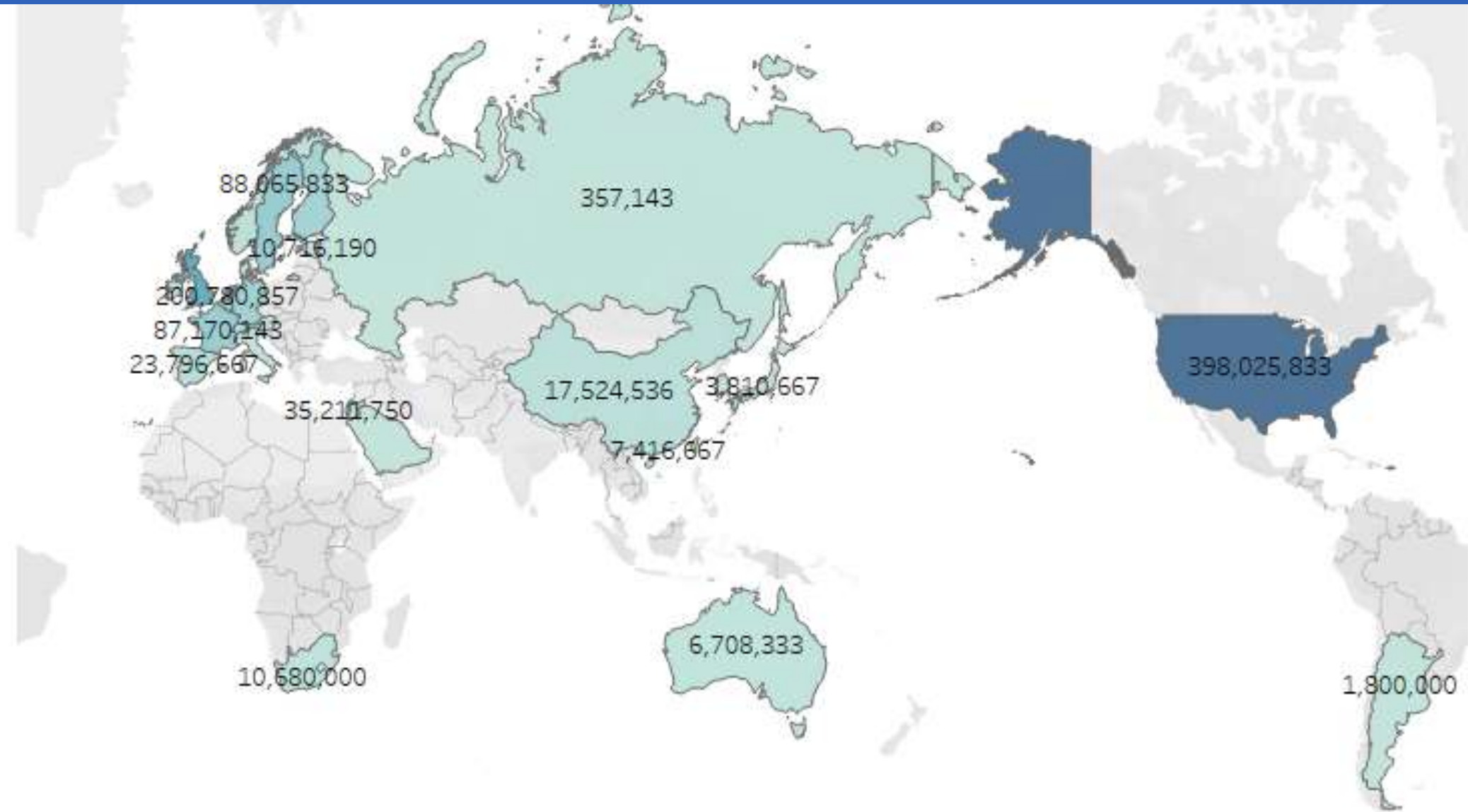


## Case 7: Visualisation of investment coverage (2/3)



Amount received by Legal Entities with Tableau Desktop

## Case 7: Visualisation of investment coverage (3/3)



Investments by Investors HQ with Tableau Desktop

# Case 8: Use Policy Scan Documents to model Policy Areas (1/2)

Place source documents in a folder



Train classifier on vectorialised documents

Inspection of results

Tag groups

Policy Areas model

Fine-tuning of clean-up

Doc2Vec Document Classifier

# Case 8: Use Policy Scan Documents to model Policy Areas (2/2)

News sources

EC Press Releases

Reference Documents

Policy Areas Model

Did I miss anything?

Alerts

## EC Press Releases Analyser - 30 days

Title	Description	Guessed ECA Policy Area
<b>Strengthening the competitiveness of the European retail sector</b> Thu, 19 Apr 2018 12:00:00 +0200	European Commission - Press release Brussels, 19 April 2018 Today the Commission is publishing a set of best practices to support Member States' efforts to create a more open, integrated and competitive retail sector. The retail sector is one of the biggest sectors in the EU economy, with almost one...	4.8 Single market
<b>Trade: European Commission proposes signature and conclusion of Japan and Singapore agreements</b> Wed, 18 Apr 2018 10:04:00 +0200	European Commission - Press release Strasbourg, 18 April 2018 The Commission today presented the outcome of negotiations for the Economic Partnership Agreement with Japan and the trade and investment agreements with Singapore to the Council. This is the first step towards the signature and conclusion of these agreements.	3.3 Development and cooperation policy
<b>Integration of migrants: Commission and OECD publish check list to support local, regional and national authorities</b> Wed, 18 Apr 2018 12:04:00 +0200	European Commission - Press release Brussels, 18 April 2018 Today the European Commission and OECD are publishing a report that identifies the main challenges to the integration of migrants and sets out concrete policy recommendations in response.	2.4 Employment, poverty and social inclusion

Document2Vector Document Classifier



In conclusion, my message to you is...

**IT... COULD... WORK!**

# Appendix I: The ECA Lab toolbox



## Cloud Service Providers



## Artificial Intelligence languages and libraries



## Data Analysis and Visualisation

