AGID
Agenzia per l'Italia Digitale

FormezPA

FORMAZIONE AGID – FORMEZ SULLA TRANSIZIONE DIGITALE DELLA PA

Progetto Informazione e formazione per la transizione digitale della PA nell'ambito del progetto «Italia Login – la casa del cittadino»

(A valere sul PON Governance e Capacità Istituzionale 2014-2020)

UNIONE EUROPEA
Dipartimento della Funzione Pubblica
PON
AGID Agenzia per l'Italia digitale

AGID
Agenzia per l'Italia Digitale

FormezPA

Dati aperti dalla teoria alla pratica: dati che creano valore

Dati della ricerca, l'esperienza di apertura dei dati epidemiologici

31/01/2023

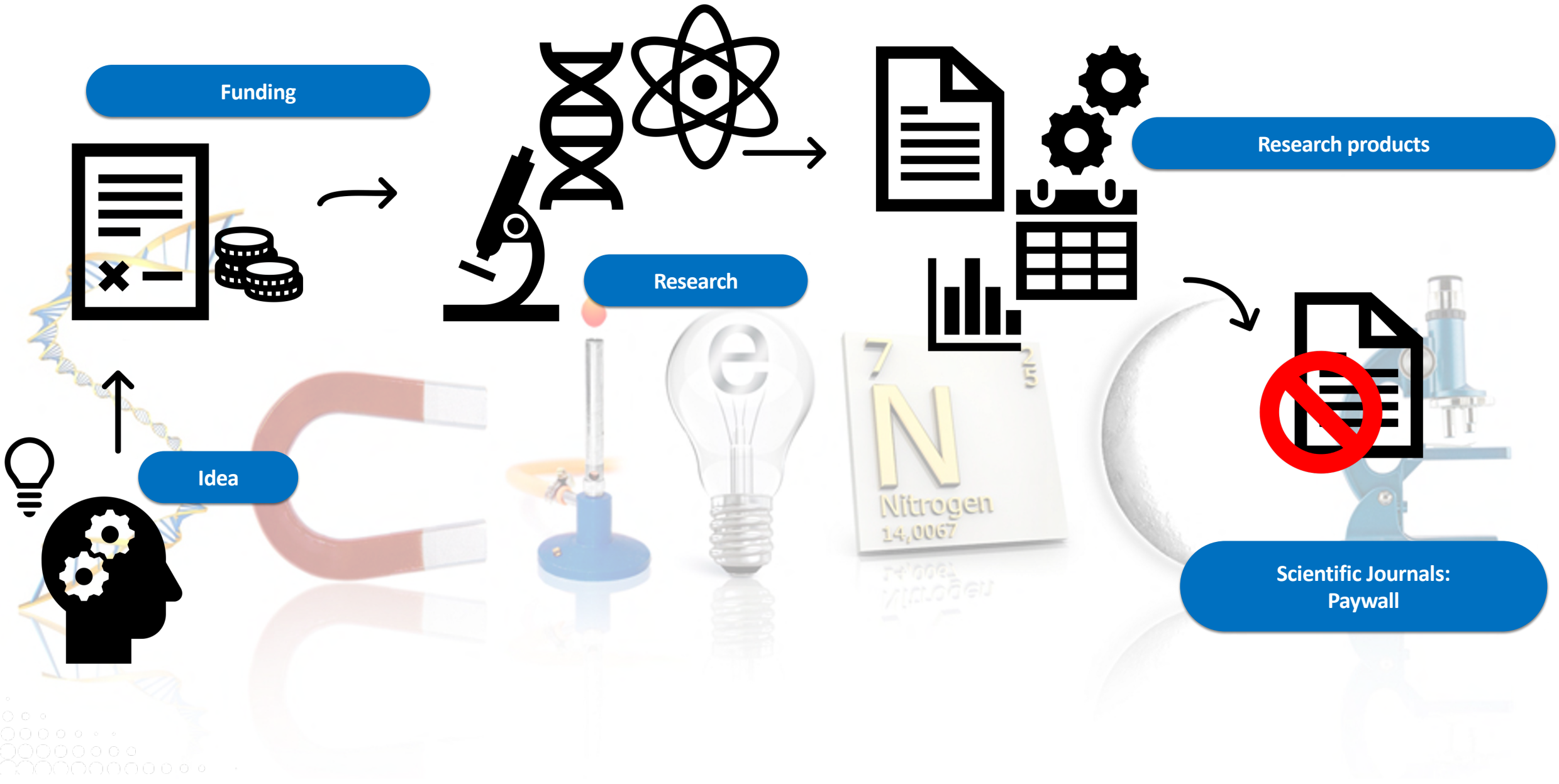
Francesco Branda

UNIVERSITÀ DELLA CALABRIA

UNIONE EUROPEA

Dipartimento della Funzione Pubblica

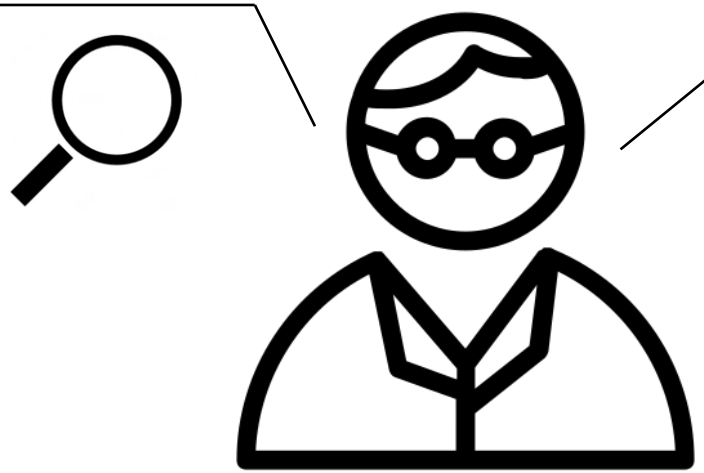
AGID Agenzia per l'Italia digitale


Funding
Research
Research products
Idea
Scientific Journals:
Paywall

FAIR principles

Findable

It must be clear where the data are located and can be cited.



Accessible

Data must be accessible for at least 10 years! It does not mean that the data is open, but it must be clear who and how can access the data.

Interoperable

Data must be easily integrated with other data, mechanically searchable, and linked to other search results.

Reusable

Data must be reusable for re-analysis or new research.

What is the cost of improper data management?

Time spent, cost of storage, licence costs, research retraction, double funding, interdisciplinarity and potential economic growth.

Following this approach, we found that the annual cost of not having FAIR research data costs the European economy at least €10.2bn every year. In addition, we also listed a number of consequences from not having FAIR which could not be reliably estimated, such as an impact on research quality, economic turnover, or machine readability of research data. By drawing a rough parallel with the European open data economy, we concluded that these unquantified elements could account for another €16bn annually on top of what we estimated. These results relied on a combination of desk research, interviews with the subject matter experts and our most conservative assumptions.

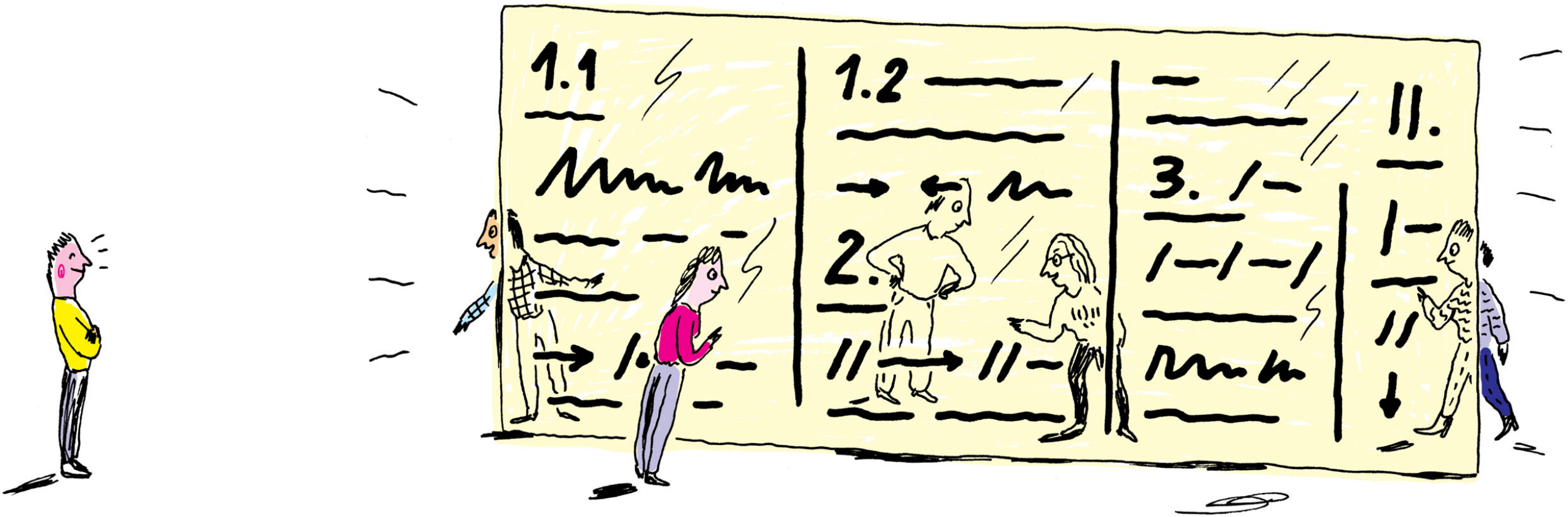
Moreover, while building on top of other available studies and being heavily reliant on existing material, we have come to realise ourselves how important is to have FAIR research data. Not only the time invested in this study could have been reduced by a significant amount, but the content could have been enhanced if more material had been accessible and reusable.

€10.2bn + €16bn every year in Europe!

Published: 2019-01-16
Corporate author(s): Directorate-General for Research and Innovation (European Commission) , PwC EU Services
Cost of not having FAIR research data

Why do we need to share research data?

Visibility and impact



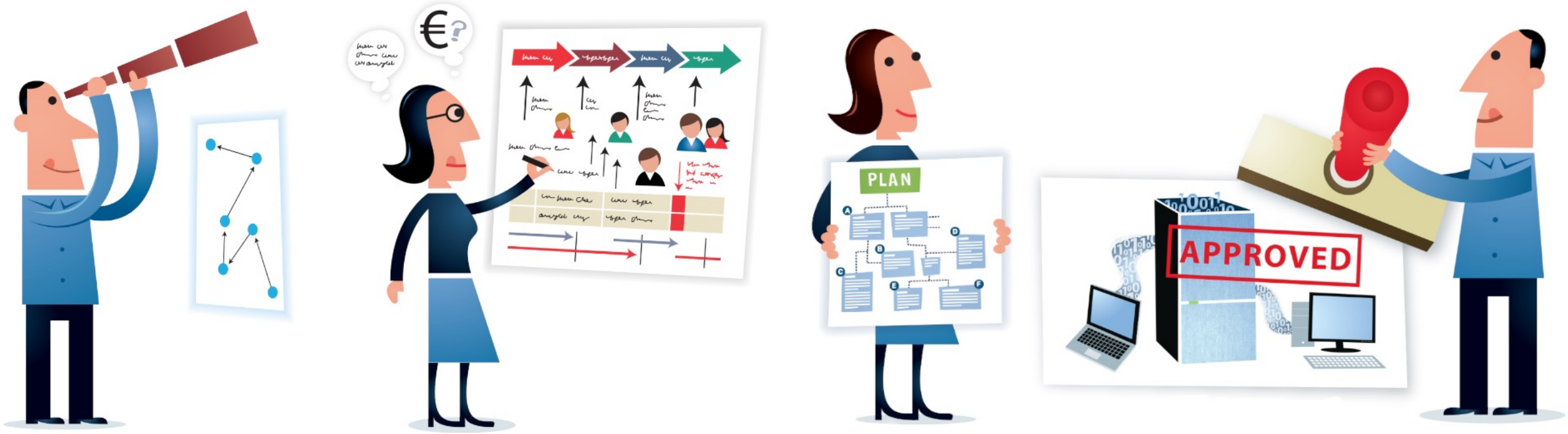
Reproducibility

Share

Preserve

Reliability

Data is the proof of your papers: how can others trust your research without accessing the data?



Collaboration

«It is imperative to leverage scientific innovations and support principles of openness and inclusiveness in processes that generate solutions to the severe health menace that is likely to bring significant hardships to humanity.»

UNESCO - Open access to facilitate research and information on COVID-19

Letter from the Elsevier Journals Leadership Team

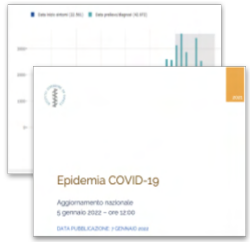
"Thank you for publishing with Elsevier in 2022. Publishing academic research is no small feat, so we would like to celebrate your achievement. Your contribution is further advancing global research and the feedback you provide helps us improve the products and services we can offer your communities.

We recognize you have a choice of journals to publish your work in, so we are honoured that you put your trust in us. We look forward to working together in the coming year.

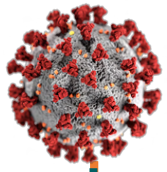
Thank you again for the hard work you devote to your field."

Laura Hassink
Managing Director Elsevier Journals

Personal initiatives for share research data



Epidemiological surveillance



All SARS-CoV-2 sequences

VoC/VoI

- Integrated surveillance (case-based data sent to TESSy)
- Monthly prevalence surveys (cross-sectoral)
- National vaccination repository
- Genomic surveillance and analysis platform
- School monitoring

Ministry of Education

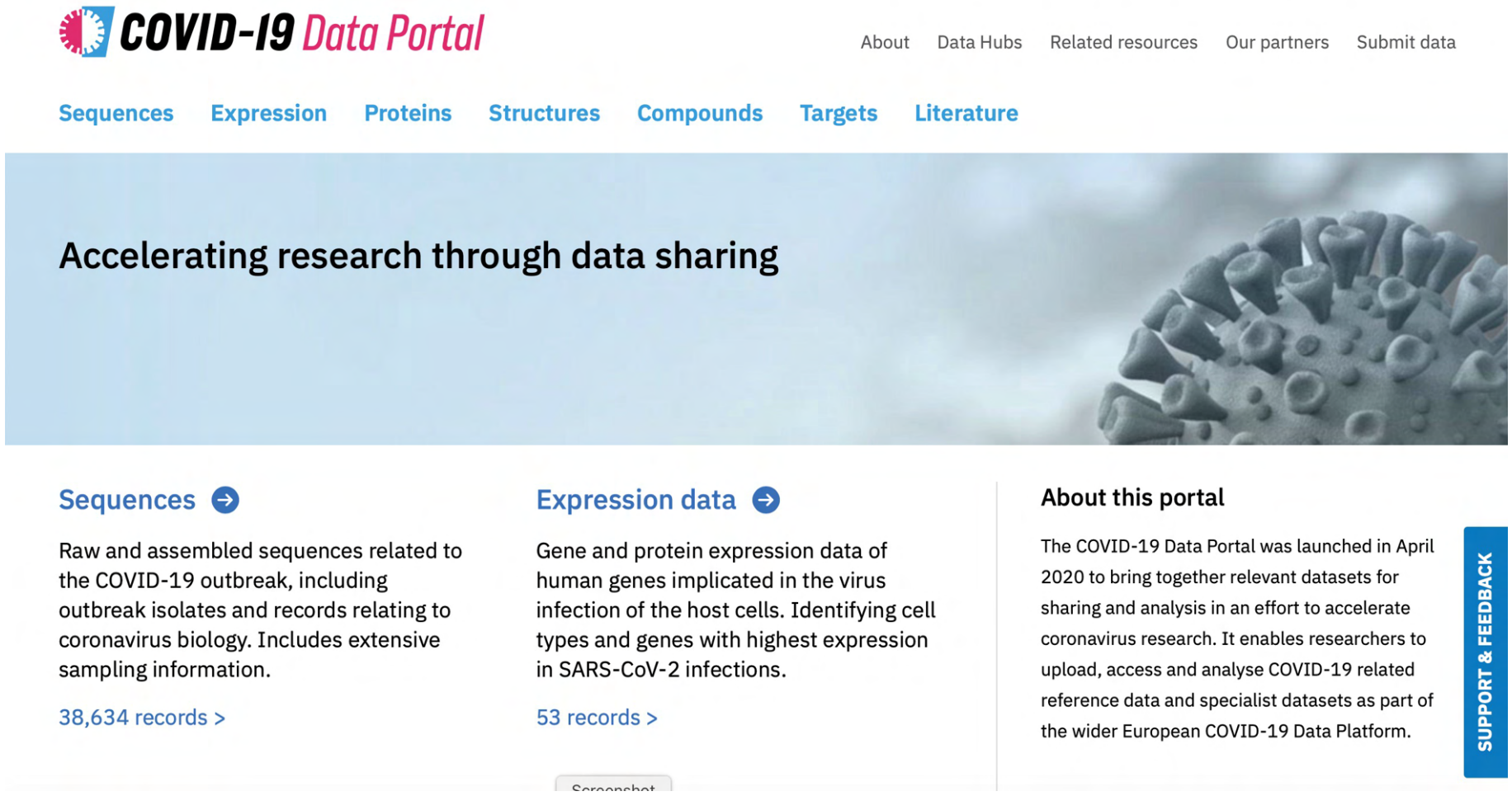
National Institute of Health (ISS)

Regional weekly monitoring

European COVID-19 Data Platform

https://www.covid19dataportal.org/

- Launched in April 2020
- Bring together relevant datasets for sharing and analysis in an effort to accelerate COVID-19 research
- Enables researchers to upload, access and analyse COVID-19 related data



https://www.covid19dataportal.it/it/data_types/forms/

App | Deep-Learning-Pr... | Predictive Policing... | Mapping Crime: U... | echen102/us-pres... | owid/covid-19-dat... | MatthewDaws/SE... | 5 Things You Nee... | Elenco di lettura

COVID-19 Data Portal ITALY

Chi siamo | Portale Europeo | Supporto & Feedback | Cerca | en | it

Genomica & Trascrittomica | Dati sulle Proteine | Dati di Imaging | Dati Sanitari | Ricerca | Eventi

Ricerca Italiana

Home / Tipologia di Dati / Ricerca Italiana

Progetti di Ricerca

Visualizzazione 1-3 di 3 elementi.

Ricerca: Argomento :

Descrizione	Leader del progetto	Contatto del Repository	Istituzione	Argomento
From the infection report to the vaccines: all the data on the Covid emergency in Calabria on a single platform	Francesco Branda	Francesco Branda francesco.branda@unical.it	Università della Calabria Rende	Health data
Descrizione completa	The COVIDA project is dedicated to the collection and visualization of data related to the COVID-19 emergency in Calabria and makes available to the scientific community all the necessary information, such as the total number of infections recorded on the territory, with a series of detailed indications (hospitalized, cured, deceased, number of swabs performed) to monitor and classify the epidemic risk, and the number of subjects vaccinated with the first dose, those vaccinated with a full cycle, the progress of vaccinations by category and age group, to evaluate the progress of the vaccination campaign. The platform can be reached at the link https://covida.tk/			
Full-genome sequencing of SARS-CoV-2 strains from the province of Parma	Enrico Silini	Riccardo Percudani riccardo.percudani@unipr.it	Università di Parma Parma	Genomics/Trascrittomics
Synthetic Antibodies neutralize SARS-CoV-2 infection of mammalian cells	Giuseppe Novelli	Giuseppe Novelli novelli@med.uniroma2.it	Università degli Studi di Roma Tor Vergata Roma	Health data Other

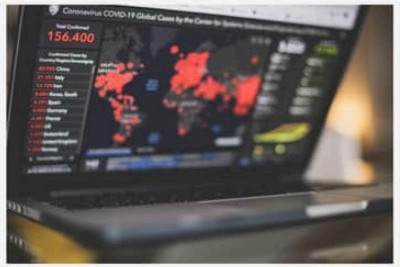
Visualizza 5 elementi per pagina

« 1 »

Portale COVID-19

COVIDA

Seguici su:



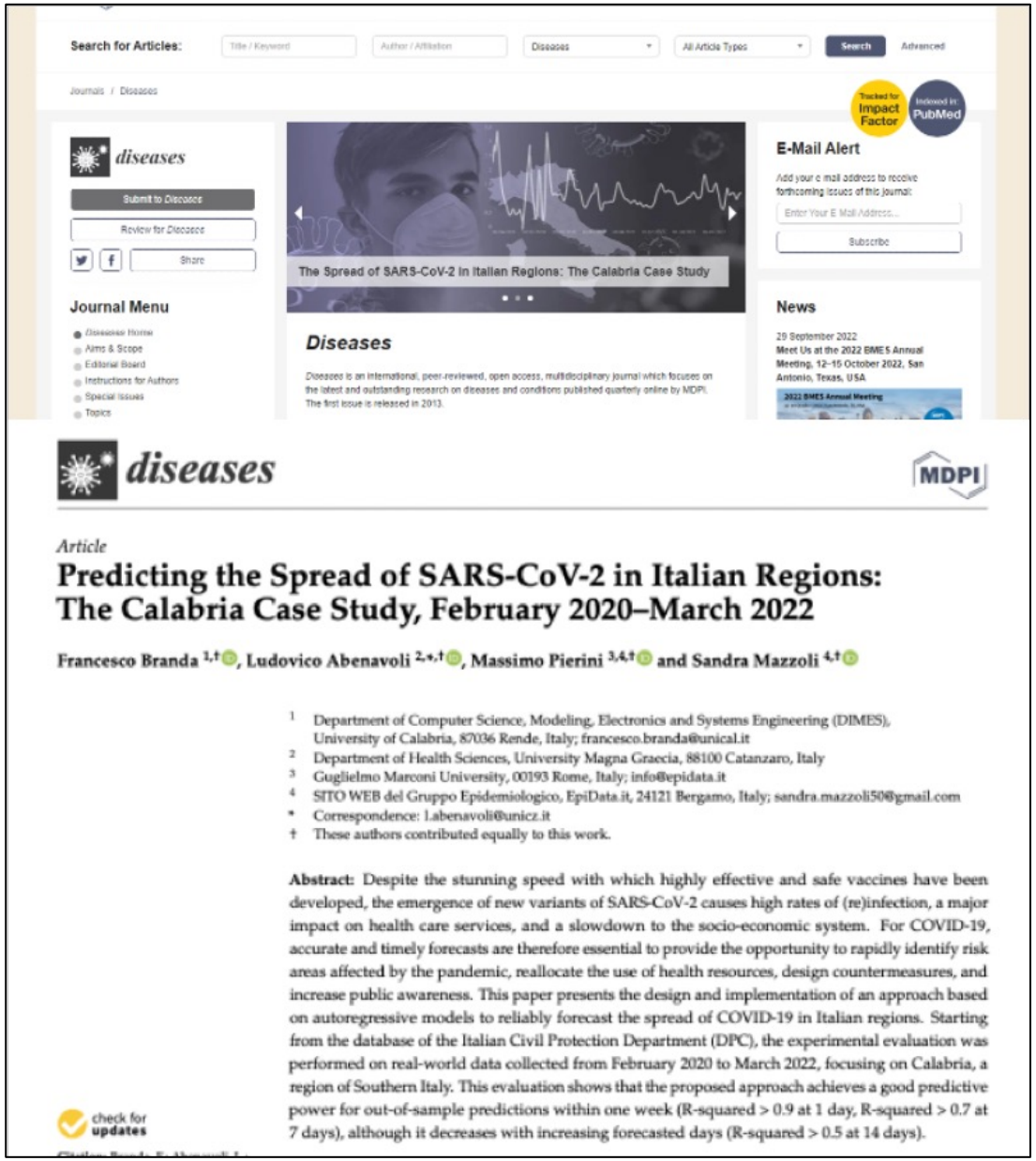
Andamento vaccinazione | Andamento epidemiologico | Bollettino giornaliero

COVID-19 Regione Calabria

Ultimo aggiornamento: 21-07-2022

+3.020*	+4	+1.503	72.904
Nuovi casi	Deceduti	Dimessi / Guariti	Casi attivi
315 (-8)	18 (-1)		
Ricoverati	Terapia intensiva		

*Numero tamponi: 12.208 (+3.870 rispetto a ieri)



Search for Articles: Title / Keyword | Author / Affiliation | Diseases | All Article Types | Search | Advanced

Journals / Diseases

Tracked for Impact Factor | Indexed in PubMed

diseases

Submit to Diseases | Review for Diseases | Share

Journal Menu

- Diseases Home
- Aims & Scope
- Editorial Board
- Instructions for Authors
- Special Issues
- Topics

The Spread of SARS-CoV-2 in Italian Regions: The Calabria Case Study

Diseases

Diseases is an international, peer-reviewed, open access, multidisciplinary journal which focuses on the latest and outstanding research on diseases and conditions published quarterly online by MDPI. The first issue is released in 2013.

E-Mail Alert

Add your e-mail address to receive forthcoming issues of this journal:

Enter Your E-Mail Address...

Subscribe

News

29 September 2022

Meet Us at the 2022 BMES Annual Meeting, 12–15 October 2022, San Antonio, Texas, USA

2022 BMES Annual Meeting

diseases

MDPI

Article

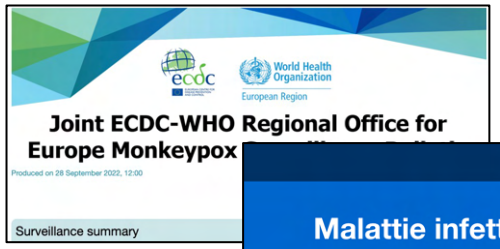
Predicting the Spread of SARS-CoV-2 in Italian Regions: The Calabria Case Study, February 2020–March 2022

Francesco Branda [1,†], Ludovico Abenavoli [2,*,†], Massimo Pierini [3,4,†] and Sandra Mazzoli [4,†]

1 Department of Computer Science, Modeling, Electronics and Systems Engineering (DIMES), University of Calabria, 87036 Rende, Italy; francesco.branda@unical.it
2 Department of Health Sciences, University Magna Graecia, 88100 Catanzaro, Italy
3 Guglielmo Marconi University, 00193 Rome, Italy; info@epidata.it
4 SITO WEB del Gruppo Epidemiologico, EpiData.it, 24121 Bergamo, Italy; sandra.mazzoli50@gmail.com
* Correspondence: l.abenavoli@unicz.it
† These authors contributed equally to this work.

Abstract: Despite the stunning speed with which highly effective and safe vaccines have been developed, the emergence of new variants of SARS-CoV-2 causes high rates of (re)infection, a major impact on health care services, and a slowdown to the socio-economic system. For COVID-19, accurate and timely forecasts are therefore essential to provide the opportunity to rapidly identify risk areas affected by the pandemic, reallocate the use of health resources, design countermeasures, and increase public awareness. This paper presents the design and implementation of an approach based on autoregressive models to reliably forecast the spread of COVID-19 in Italian regions. Starting from the database of the Italian Civil Protection Department (DPC), the experimental evaluation was performed on real-world data collected from February 2020 to March 2022, focusing on Calabria, a region of Southern Italy. This evaluation shows that the proposed approach achieves a good predictive power for out-of-sample predictions within one week (R-squared > 0.9 at 1 day, R-squared > 0.7 at 7 days), although it decreases with increasing forecasted days (R-squared > 0.5 at 14 days).

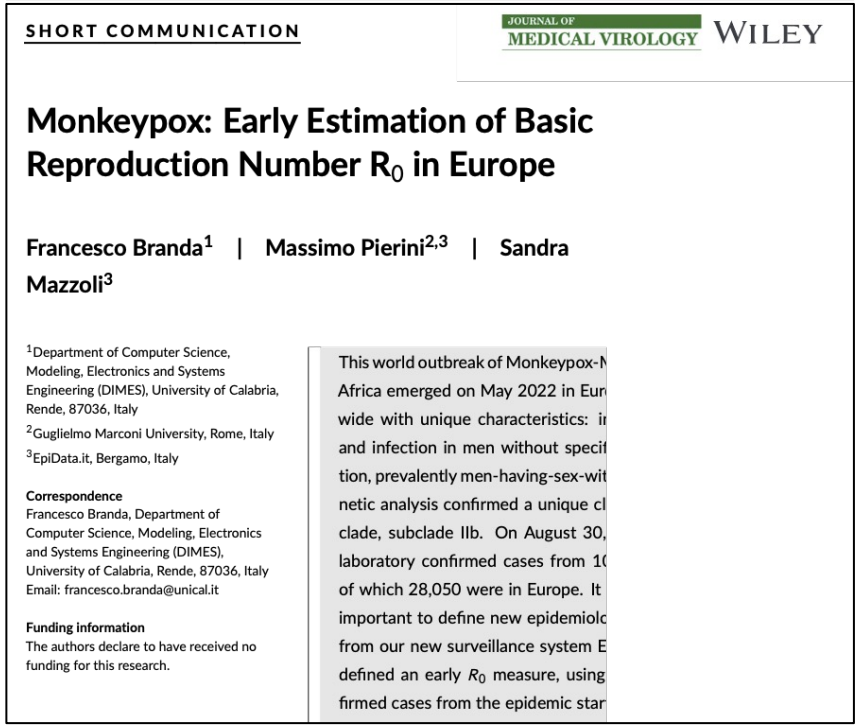
check for updates

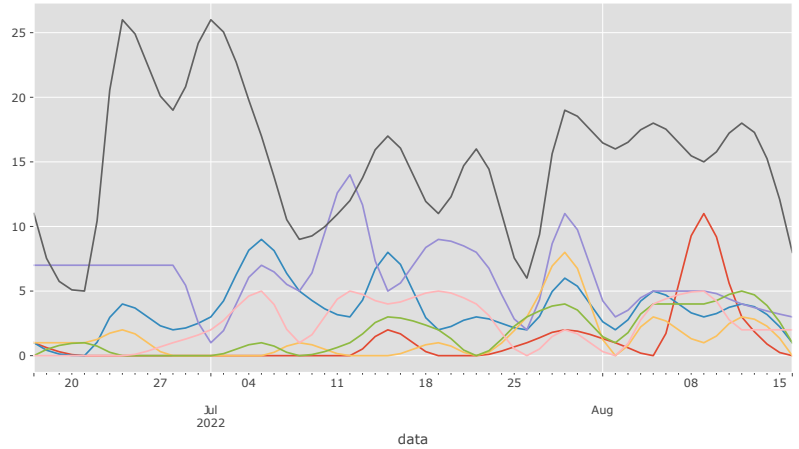

Joint ECDC-WHO Regional Office for Europe Monkeypox
Surveillance summary

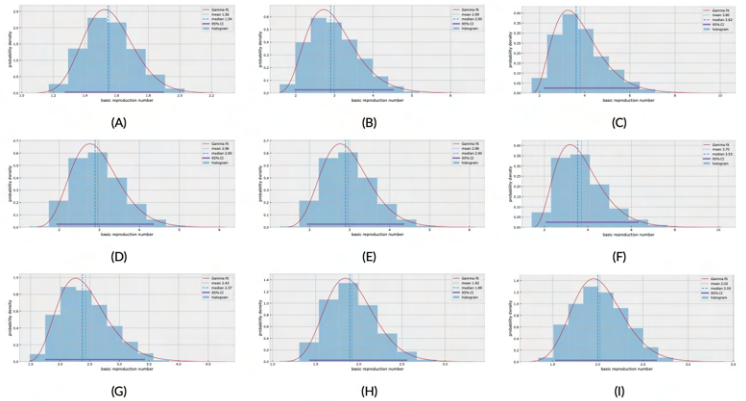

Ministero della Salute
Malattie infettive
Home / Argomenti - Malattie infettive A-Z / Vaiolo delle scimmie - Mpox
Vaiolo delle scimmie - Mpox

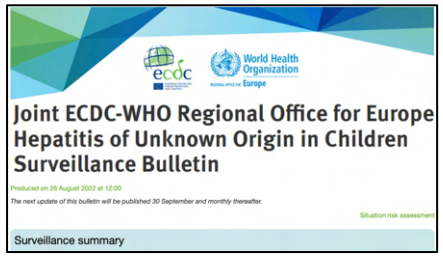

Joint ECDC-WHO Regional Office for Europe Mpox Surveillance Bulletin
Mpox
Esempio d'uso dei dati
Contributi
Licenza


Journal of Clinical Virology Plus 2 (2022) 100114
Contents lists available at ScienceDirect
Journal of Clinical Virology Plus
journal homepage: www.elsevier.com/locate/jcvp
Monkeypox: EpiMPX Surveillance System and Open Data with a Special Focus on European and Italian Epidemic
Francesco Branda, Massimo Pierini, Sandra Mazzoli


SHORT COMMUNICATION
JOURNAL OF MEDICAL VIROLOGY
WILEY
Monkeypox: Early Estimation of Basic Reproduction Number R0 in Europe
Francesco Branda | Massimo Pierini | Sandra Mazzoli
Correspondence
Funding information

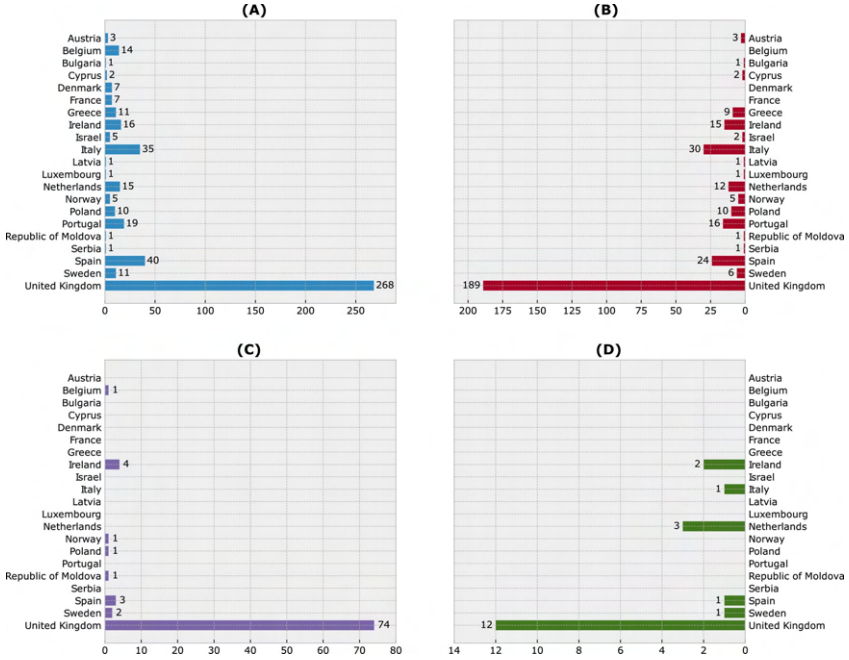

data
Jul
2022
Aug

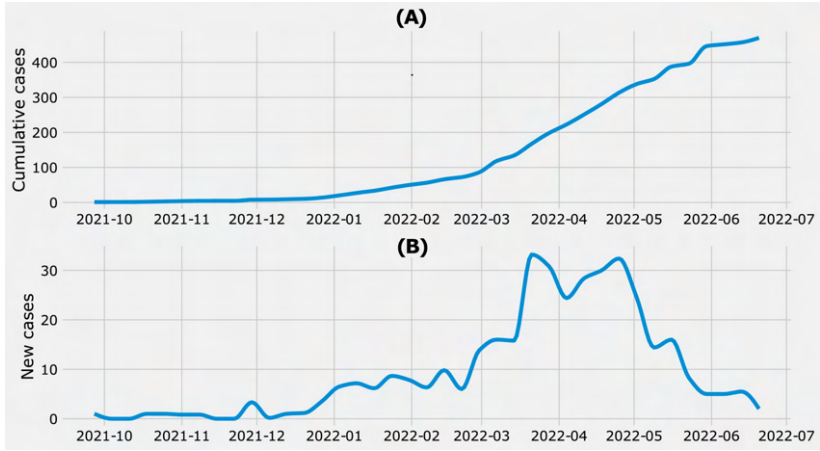

(A)
(B)
(C)
(D)
(E)
(F)
(G)
(H)
(I)


World Health Organization
Europe
Joint ECDC-WHO Regional Office for Europe Hepatitis of Unknown Origin in Children Surveillance Bulletin
Produced on 26 August 2022 at 12:00
The next update of this bulletin will be published 30 September and monthly thereafter.
Situation risk assessment
Surveillance summary

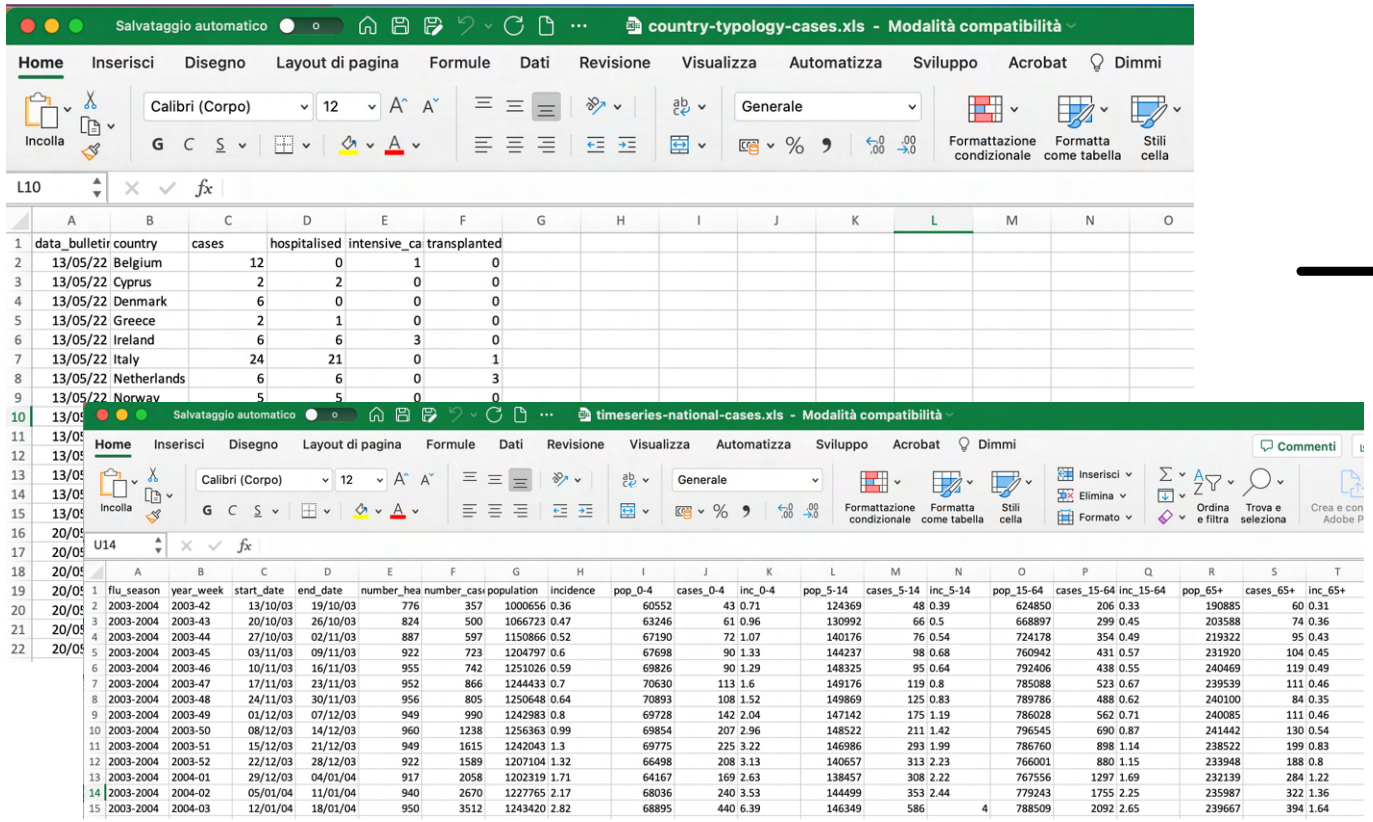

unicalportale Add files via upload
88d9137 last week
History
ECDC-WHO
Fit Hepatitis.ipynb
LICENSE.md
README.md
README_EN.md
hepatitis_model.drawio.png
README.md
Joint ECDC-WHO Regional Office for Europe Hepatitis of Unknown Origin in Children Surveillance Bulletin
Italiano - English
Epatite di origine sconosciuta nei bambini
Questo repository contiene i dati estratti dai bollettini ECDC-WHO/Europe.
Esempio d'uso dei dati
Download diretto (CSV): https://raw.githubusercontent.com/fbranda/dati-epidemiologici/main/epatite-di-origine-sconosciuta-nei-bambini/ECDC-WHO-Regional-Office-for-Europe/case-count.csv
Python (richiede pandas):
import pandas as pd
df = pd.read_csv("https://raw.githubusercontent.com/fbranda/dati-epidemiologici/main/epatite-di-origine-sconosciuta-nei-bambini/ECDC-WHO-Regional-Office-for-Europe/case-count.csv")
R (richiede httr):
library(httr)
df <- read.csv(text=content(GET("https://raw.githubusercontent.com/fbranda/dati-epidemiologici/main/epatite-di-origine-sconosciuta-nei-bambini/ECDC-WHO-Regional-Office-for-Europe/case-count.csv")))
Contributi
1. Branda F, Pierini M, Mazzoli S. Hepatitis of unknown origin in children: Why and how to create an open access database. Journal of Clinical Virology Plus. 2022 Aug 1;2(3):100102.
Licenza
CC-BY-4.0 - Visualizza licenza

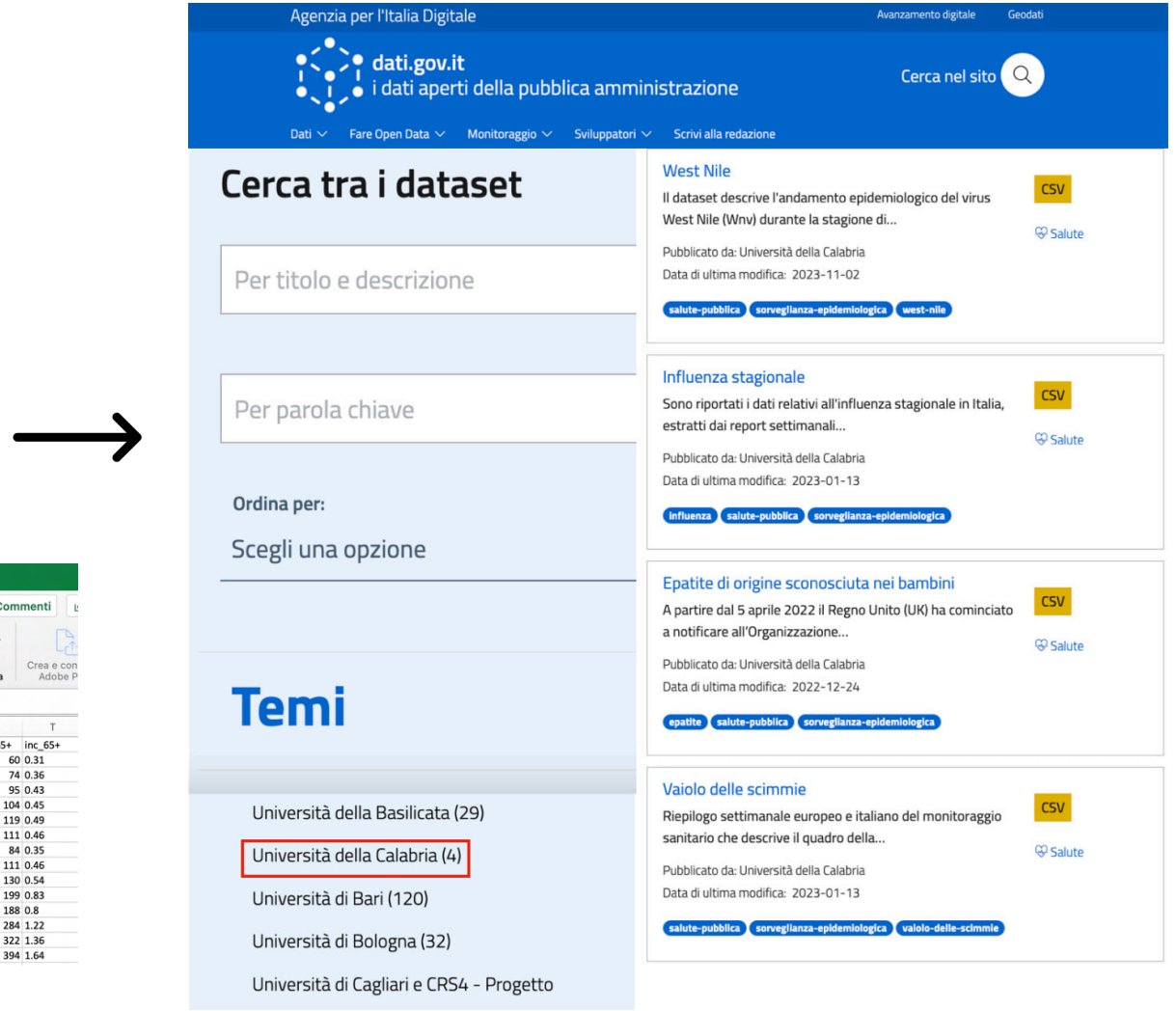

ELSEVIER
Journal of Clinical Virology Plus
Volume 2, Issue 3, August 2022, 100102
CLINICAL VIROLOGY PLUS
Hepatitis of unknown origin in children: Why and how to create an open access database
Francesco Branda [a], Massimo Pierini [b, c], Sandra Mazzoli [b]


(A)
Austria 3
Belgium 14
Bulgaria 1
Cyprus 2
Denmark 7
France 7
Greece 11
Ireland 16
Israel 5
Italy 35
Latvia 1
Luxembourg 1
Netherlands 15
Norway 5
Poland 10
Portugal 19
Republic of Moldova 1
Serbia 1
Spain 40
Sweden 11
United Kingdom 268
0 50 100 150 200 250
(B)
Austria 3
Belgium 1
Bulgaria 2
Cyprus
Denmark
France
Greece 9
Ireland 15
Israel 2
Italy 30
Latvia 1
Luxembourg 1
Netherlands 12
Norway 5
Poland 10
Portugal 16
Republic of Moldova 1
Serbia 1
Spain 24
Sweden 6
United Kingdom 189
200 175 150 125 100 75 50 25 0
(C)
Austria
Belgium 1
Bulgaria
Cyprus
Denmark
France
Greece
Ireland 4
Israel
Italy
Latvia
Luxembourg
Netherlands
Norway 1
Poland 1
Portugal
Republic of Moldova 1
Serbia
Spain 3
Sweden 2
United Kingdom 74
0 10 20 30 40 50 60 70 80
(D)
Austria
Belgium
Bulgaria
Cyprus
Denmark
France
Greece
Ireland 2
Israel
Italy 1
Latvia
Luxembourg
Netherlands 3
Norway
Poland
Portugal
Republic of Moldova
Serbia
Spain 1
Sweden 1
United Kingdom 12
14 12 10 8 6 4 2 0


(A)
Cumulative cases
0 100 200 300 400
2021-10 2021-11 2021-12 2022-01 2022-02 2022-03 2022-04 2022-05 2022-06 2022-07
(B)
New cases
0 10 20 30
2021-10 2021-11 2021-12 2022-01 2022-02 2022-03 2022-04 2022-05 2022-06 2022-07

Open data

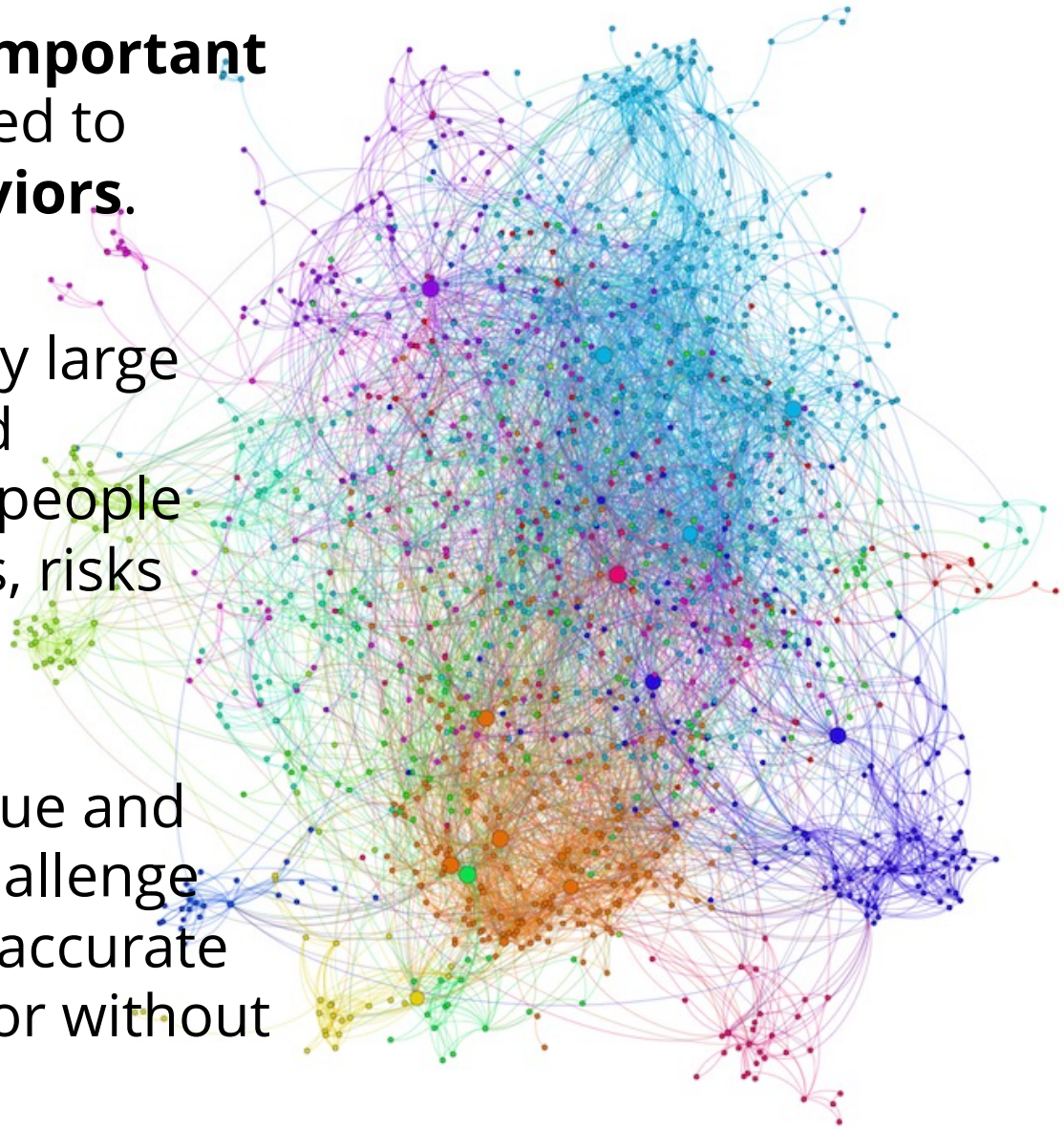

Salvataggio automatico
country-typology-cases.xls - Modalità compatibilità
Home Inserisci Disegno Layout di pagina Formule Dati Revisione Visualizza Automatizza Sviluppo Acrobat Dimmi
Calibri (Corpo) 12 Generale
Incolla
Formattazione condizionale
Formatta come tabella
Stili cella
L10
data_bulletir country cases hospitalised intensive_ca transplanted
13/05/22 Belgium 12 0 1 0
13/05/22 Cyprus 2 2 0 0
13/05/22 Denmark 6 0 0 0
13/05/22 Greece 2 1 0 0
13/05/22 Ireland 6 6 3 0
13/05/22 Italy 24 21 0 1
13/05/22 Netherlands 6 6 0 3
13/05/22 Norway 5 5 0 0
Salvataggio automatico
timeseries-national-cases.xls - Modalità compatibilità
Home Inserisci Disegno Layout di pagina Formule Dati Revisione Visualizza Automatizza Sviluppo Acrobat Dimmi
Commenti
Calibri (Corpo) 12 Generale
Incolla
Formattazione condizionale
Formatta come tabella
Stili cella
Inserisci
Elimina
Formato
Ordina e filtra
Trova e seleziona
U14
flu_season year_week start_date end_date number_hea number_cas population incidence pop_0-4 cases_0-4 inc_0-4 pop_5-14 cases_5-14 inc_5-14 pop_15-64 cases_15-64 inc_15-64 pop_65+ cases_65+ inc_65+
2003-2004 2003-42 13/10/03 19/10/03 776 357 1000656 0.36 60552 43 0.71 124369 48 0.39 624850 206 0.33 190885 60 0.31
2003-2004 2003-43 20/10/03 26/10/03 824 500 1066723 0.47 63246 61 0.96 130992 66 0.5 668897 299 0.45 203588 74 0.36
2003-2004 2003-44 27/10/03 02/11/03 887 597 1150866 0.52 67190 72 1.07 140176 76 0.54 724178 354 0.49 219322 95 0.43
2003-2004 2003-45 03/11/03 09/11/03 922 723 1204797 0.6 67698 90 1.33 144237 98 0.68 760942 431 0.57 231920 104 0.45
2003-2004 2003-46 10/11/03 16/11/03 955 742 1251026 0.59 69826 90 1.29 148325 95 0.64 792406 438 0.55 240469 119 0.49
2003-2004 2003-47 17/11/03 23/11/03 952 866 1244433 0.7 70630 113 1.6 149176 119 0.8 785088 523 0.67 239539 111 0.46
2003-2004 2003-48 24/11/03 30/11/03 956 805 1250648 0.64 70893 108 1.52 149869 125 0.83 789786 488 0.62 240100 84 0.35
2003-2004 2003-49 01/12/03 07/12/03 949 990 1242983 0.8 69728 142 2.04 147142 175 1.19 786028 562 0.71 240085 111 0.46
2003-2004 2003-50 08/12/03 14/12/03 960 1238 1256363 0.99 69854 207 2.96 148522 211 1.42 796545 690 0.87 241442 130 0.54
2003-2004 2003-51 15/12/03 21/12/03 949 1615 1242043 1.3 69775 225 3.22 146986 293 1.99 786760 898 1.14 238522 199 0.83
2003-2004 2003-52 22/12/03 28/12/03 922 1589 1207104 1.32 66498 208 3.13 140657 313 2.23 766001 880 1.15 233948 188 0.8
2003-2004 2004-01 29/12/03 04/01/04 917 2058 1202319 1.71 64167 169 2.63 138457 308 2.22 767556 1297 1.69 232139 284 1.22
2003-2004 2004-02 05/01/04 11/01/04 940 2670 1227765 2.17 68036 240 3.53 144499 353 2.44 779243 1755 2.25 235987 322 1.36
2003-2004 2004-03 12/01/04 18/01/04 950 3512 1243420 2.82 68895 440 6.39 146349 586 4 788509 2092 2.65 239667 394 1.64


Agenzia per l'Italia Digitale
Avanzamento digitale
Geodati
dati.gov.it
i dati aperti della pubblica amministrazione
Cerca nel sito
Dati
Fare Open Data
Monitoraggio
Sviluppatori
Scrivi alla redazione
Cerca tra i dataset
Per titolo e descrizione
Per parola chiave
Ordina per:
Scegli una opzione
Temi
Università della Basilicata (29)
Università della Calabria (4)
Università di Bari (120)
Università di Bologna (32)
Università di Cagliari e CRS4 - Progetto
West Nile
Il dataset descrive l'andamento epidemiologico del virus West Nile (Wnv) durante la stagione di...
CSV
Salute
Pubblicato da: Università della Calabria
Data di ultima modifica: 2023-11-02
salute-pubblica sorveglianza-epidemiologica west-nile
Influenza stagionale
Sono riportati i dati relativi all'influenza stagionale in Italia, estratti dai report settimanali...
CSV
Salute
Pubblicato da: Università della Calabria
Data di ultima modifica: 2023-01-13
influenza salute-pubblica sorveglianza-epidemiologica
Epatite di origine sconosciuta nei bambini
A partire dal 5 aprile 2022 il Regno Unito (UK) ha cominciato a notificare all'Organizzazione...
CSV
Salute
Pubblicato da: Università della Calabria
Data di ultima modifica: 2022-12-24
epatite salute-pubblica sorveglianza-epidemiologica
Vaiolo delle scimmie
Riepilogo settimanale europeo e italiano del monitoraggio sanitario che descrive il quadro della...
CSV
Salute
Pubblicato da: Università della Calabria
Data di ultima modifica: 2023-01-13
salute-pubblica sorveglianza-epidemiologica vaiolo-delle-scimmie

- Social media platforms have become an **important source** of information that can be exploited to understand **human dynamics** and **behaviors**.

- In the context of natural disasters, the very large use of social media platforms has enabled eyewitnesses and other disaster-affected people to share information about their damages, risks and emergencies in real time.

- The use of social media posts to help rescue and intervention activities remains an open challenge as users often publish posts containing inaccurate information, slang or abbreviated words, or without using geolocalization.



Belcastro *et al. J Big Data* (2021) 8:79
https://doi.org/10.1186/s40537-021-00467-1

Journal of Big Data

RESEARCH Open Access

Using social media for sub-event detection during disasters

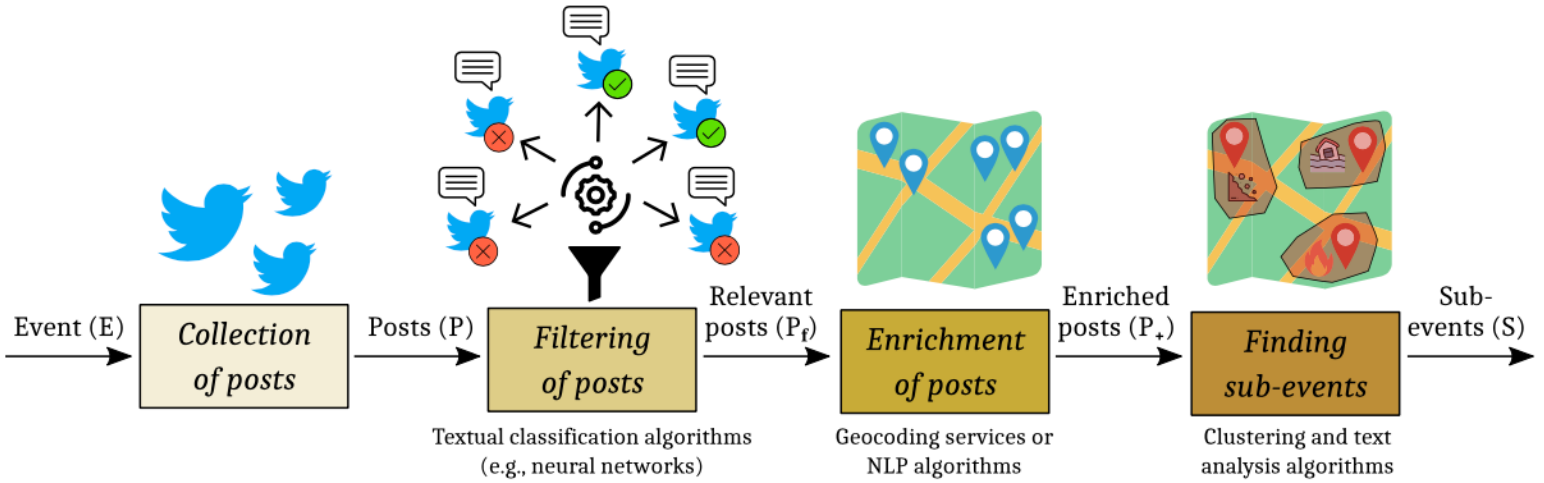
Loris Belcastro[1], Fabrizio Marozzo[1*], Domenico Talia[1], Paolo Trunfio[1], Francesco Branda[1], Themis Palpanas[2,3] and Muhammad Imran[4]

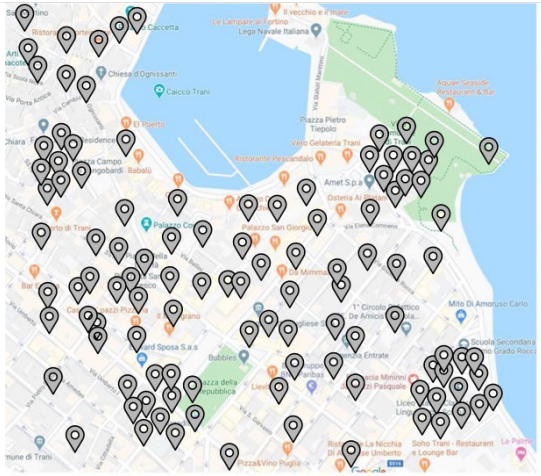
*Correspondence:
fmarozzo@dimes.unical.it
[1] University of Calabria, Rende, Italy
Full list of author information is available at the end of the article

Abstract

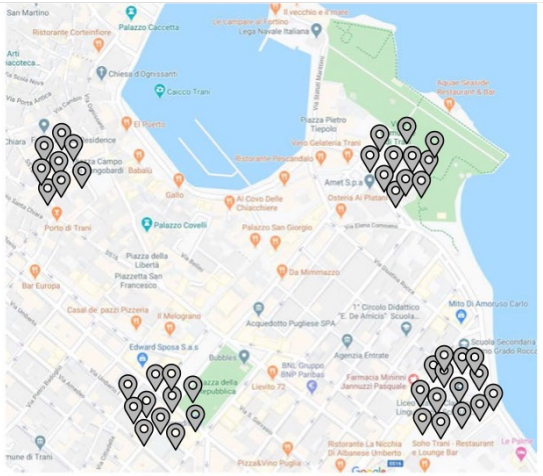
Social media platforms have become fundamental tools for sharing information during natural disasters or catastrophic events. This paper presents SEDOM-DD (Sub-Events Detection on sOcial Media During Disasters), a new method that analyzes user posts to discover sub-events that occurred after a disaster (e.g., collapsed buildings, broken gas pipes, floods). SEDOM-DD has been evaluated with datasets of different sizes that contain real posts from social media related to different natural disasters (e.g., earthquakes, floods and hurricanes). Starting from such data, we generated synthetic datasets with different features, such as different percentages of relevant posts and/or geotagged posts. Experiments performed on both real and synthetic datasets showed that SEDOM-DD is able to identify sub-events with high accuracy. For example, with a percentage of relevant posts of 80% and geotagged posts of 15%, our method detects the sub-events and their areas with an accuracy of 85%, revealing the high accuracy and effectiveness of the proposed approach.

Keywords: Social media, Events detection, Natural disasters, Catastrophic events, Crisis computing, Disaster management, Mass emergencies, Earthquake

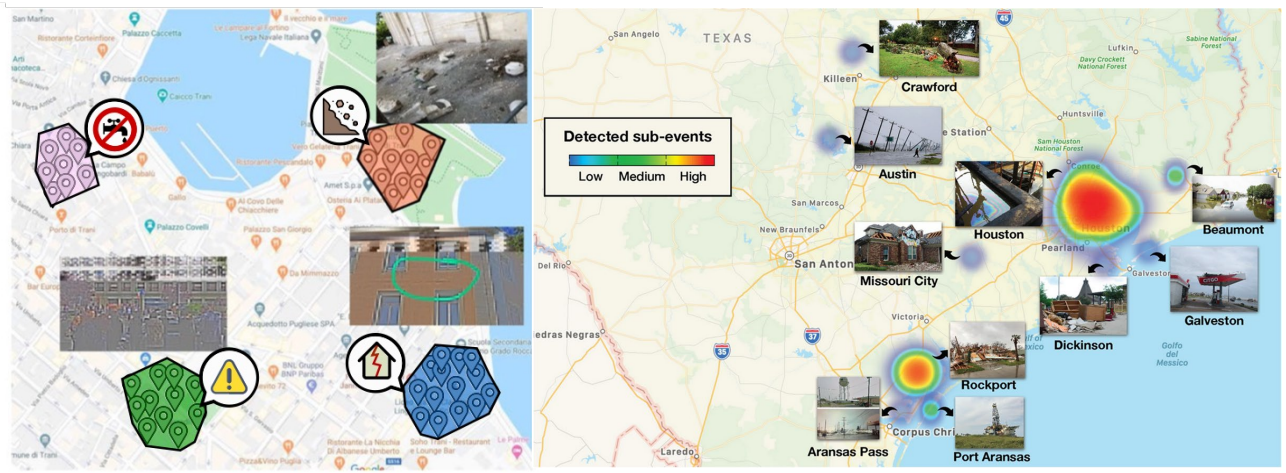




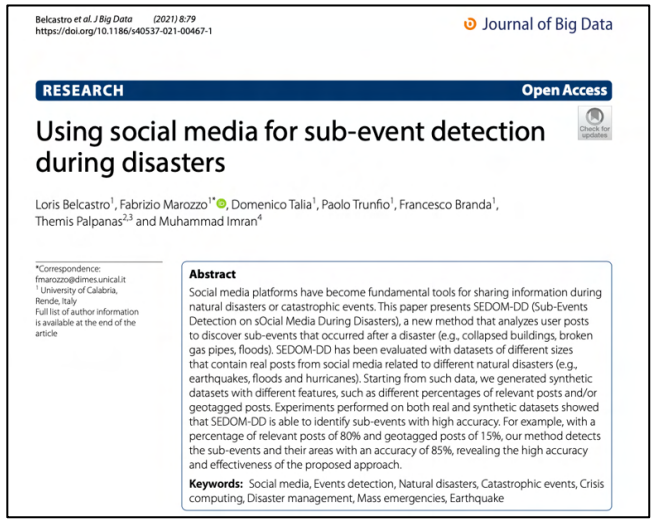
(a) Relevant posts.



(b) Filtering of posts related to sub-events.



(c) Sub-events detection.



Belcastro *et al. J Big Data* (2021) 8:79
https://doi.org/10.1186/s40537-021-00467-1

Journal of Big Data

RESEARCH Open Access

Using social media for sub-event detection during disasters

Loris Belcastro[1], Fabrizio Marozzo[1*], Domenico Talia[1], Paolo Trunfio[1], Francesco Branda[1], Themis Palpanas[2,3] and Muhammad Imran[4]

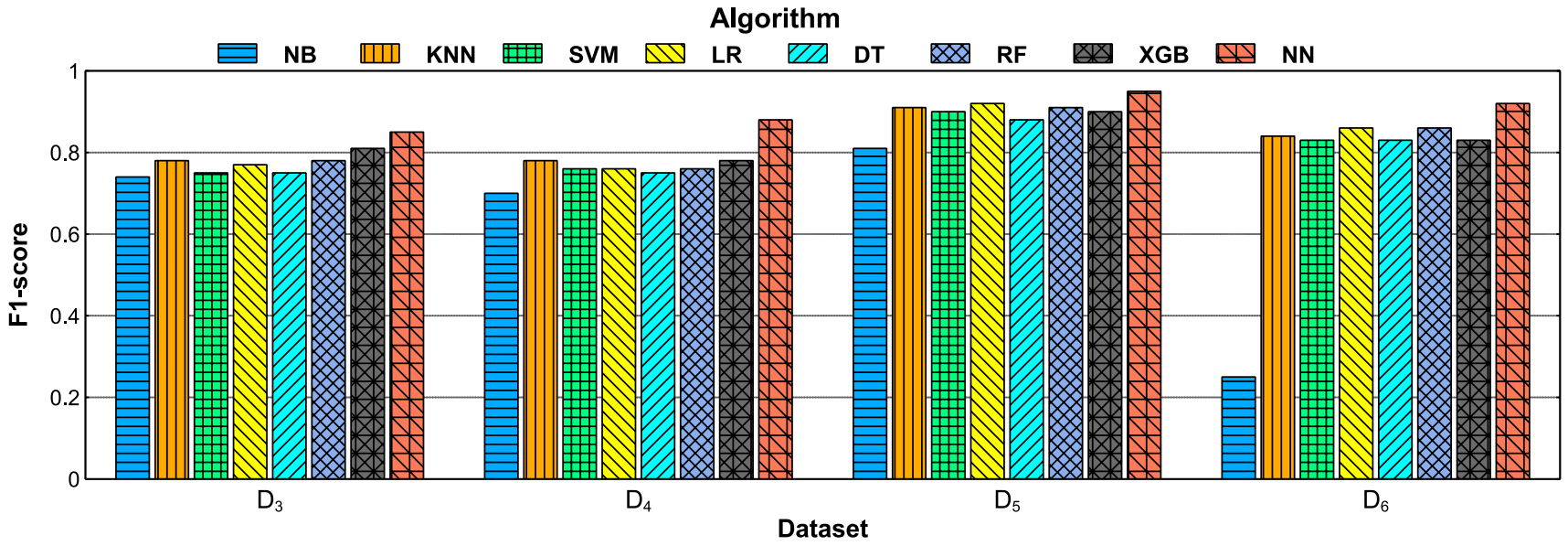
*Correspondence: fmarozzo@dimes.unical.it
[1] University of Calabria, Rende, Italy
Full list of author information is available at the end of the article

Abstract

Social media platforms have become fundamental tools for sharing information during natural disasters or catastrophic events. This paper presents SEDOM-DD (Sub-Events Detection on sOcial Media During Disasters), a new method that analyzes user posts to discover sub-events that occurred after a disaster (e.g., collapsed buildings, broken gas pipes, floods). SEDOM-DD has been evaluated with datasets of different sizes that contain real posts from social media related to different natural disasters (e.g., earthquakes, floods and hurricanes). Starting from such data, we generated synthetic datasets with different features, such as different percentages of relevant posts and/or geotagged posts. Experiments performed on both real and synthetic datasets showed that SEDOM-DD is able to identify sub-events with high accuracy. For example, with a percentage of relevant posts of 80% and geotagged posts of 15%, our method detects the sub-events and their areas with an accuracy of 85%, revealing the high accuracy and effectiveness of the proposed approach.

Keywords: Social media, Events detection, Natural disasters, Catastrophic events, Crisis computing, Disaster management, Mass emergencies, Earthquake

Algorithms	Acc	Prec	Rec	F1
Naïve Bayes	0.753	0.735	0.753	0.739
KNN	0.807	0.803	0.807	0.781
SVM	0.776	0.765	0.776	0.751
Logistic Regr.	0.790	0.773	0.790	0.766
Decision Tree	0.744	0.755	0.744	0.753
Random For.	0.795	0.794	0.790	0.783
XGBoost	0.815	0.812	0.815	0.809
Neural Net.	**0.830**	**0.826**	**0.864**	**0.845**



Social Network Analysis and Mining (2022) 12:83
https://doi.org/10.1007/s13278-022-00913-9

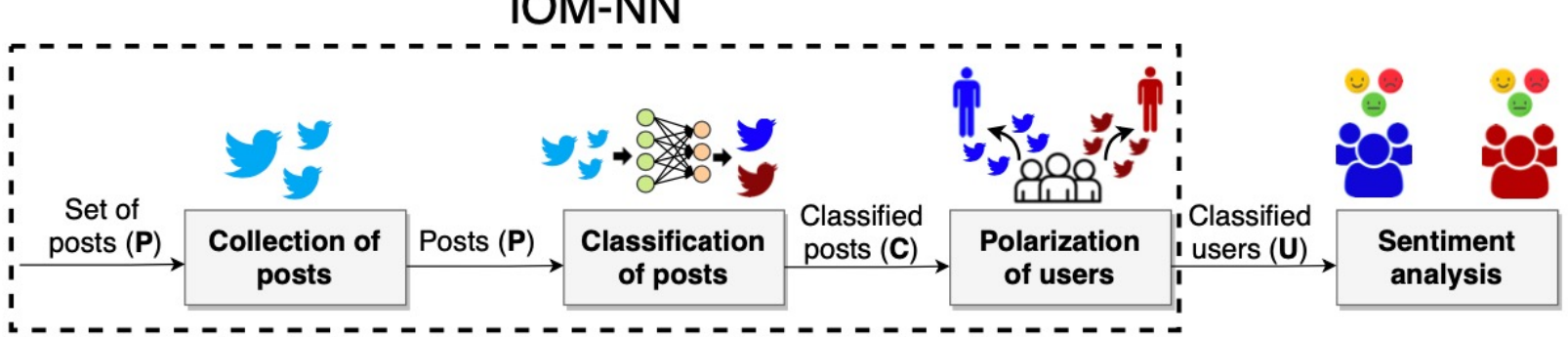
ORIGINAL ARTICLE

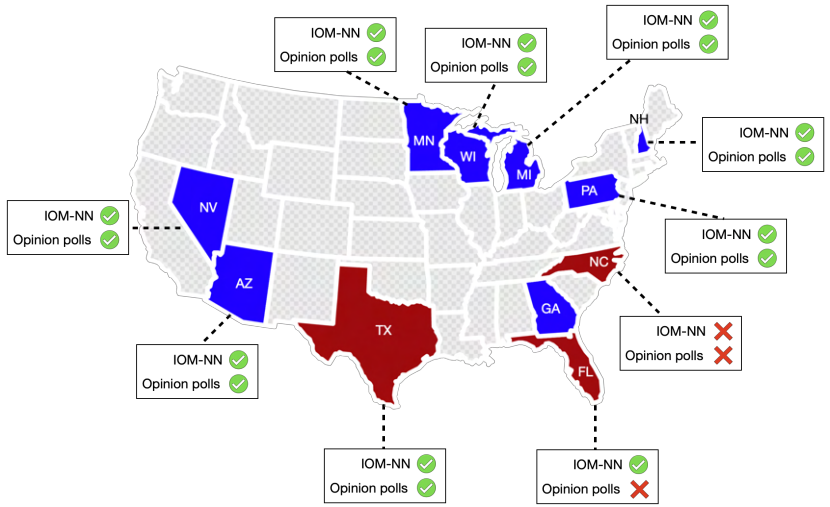
Analyzing voter behavior on social media during the 2020 US presidential election campaign

Loris Belcastro[1] · Francesco Branda[1] · Riccardo Cantini[1] · Fabrizio Marozzo[1] · Domenico Talia[1] · Paolo Trunfio[1]

Received: 20 January 2022 / Revised: 11 May 2022 / Accepted: 23 June 2022

Abstract
Every day millions of people use social media platforms by generating a very large amount of opinion-rich data, which can be exploited to extract valuable information about human dynamics and behaviors. In this context, the present manuscript provides a precise view of the 2020 US presidential election by jointly applying topic discovery, opinion mining, and emotion analysis techniques on social media data. In particular, we exploited a clustering-based technique for extracting the main discussion topics and monitoring their weekly impact on social media conversation. Afterward, we leveraged a neural-based opinion mining technique for determining the political orientation of social media users by analyzing the posts they published. In this way, we were able to determine in the weeks preceding the Election Day which candidate or party public opinion is most in favor of. We also investigated the temporal dynamics of the online discussions, by studying how users' publishing behavior is related to their political alignment. Finally, we combined sentiment analysis and text mining techniques to discover the relationship between the user polarity and sentiment expressed referring to the different candidates, thus modeling political support of social media users from an emotional viewpoint.





State	Real percentages		Opinion polls		IOM-NN	
	B	*T*	*B*	*T*	*B*	*T*
Arizona	49.4	49.1	**48.0**	45.8	**50.2**	48.3
Florida	47.9	51.2	48.7	46.0	48.0	**51.1**
Georgia	49.5	49.2	**47.6**	47.4	**52.7**	46.0
Michigan	50.6	47.8	**49.9**	44.4	**55.4**	43.0
Minnesota	52.4	45.3	**51.6**	41.8	**55.1**	42.6
Nevada	50.1	47.7	**49.4**	44.4	**49.8**	48.0
New Hampshire	52.7	45.4	**53.4**	42.4	**50.9**	47.3
North Carolina	48.6	49.9	47.8	47.5	56.6	41.9
Pennsylvania	50.0	48.8	**49.4**	45.7	**55.7**	43.1
Texas	46.5	52.1	47.5	**48.8**	46.1	**52.5**
Wisconsin	49.4	48.8	**52.0**	42.8	**56.3**	41.9
Correctly classified	-		**9/11**		**10/11**	
Tweets	-		-		670,451	
Users	-		≈ 11,000		57,116	
Avg. Acc	-		0.82		0.91	

Social Network Analysis and Mining (2022) 12:83
https://doi.org/10.1007/s13278-022-00913-9

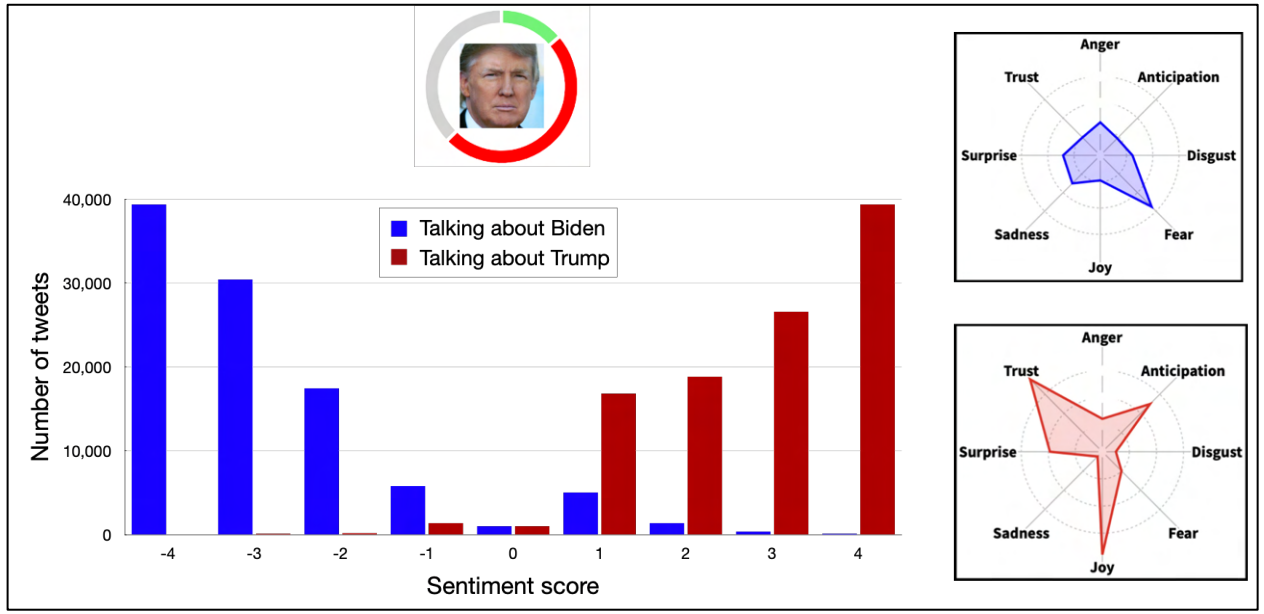
ORIGINAL ARTICLE

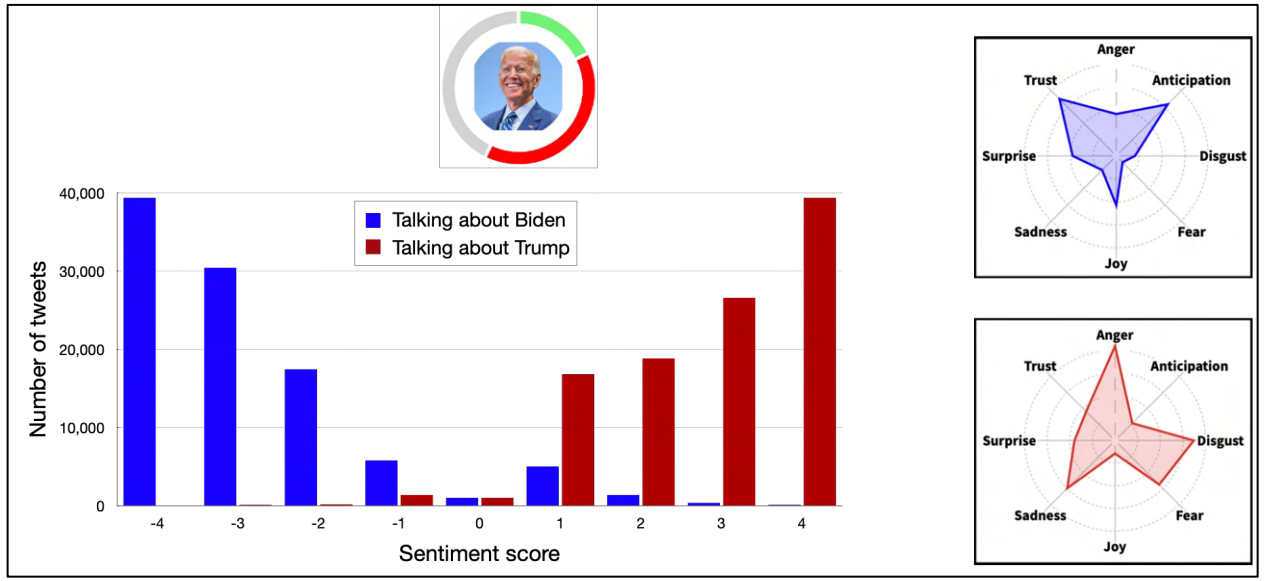
Analyzing voter behavior on social media during the 2020 US presidential election campaign

Loris Belcastro[1] · Francesco Branda[1] · Riccardo Cantini[1] · Fabrizio Marozzo[1] · Domenico Talia[1] · Paolo Trunfio[1]

Received: 20 January 2022 / Revised: 11 May 2022 / Accepted: 23 June 2022

Abstract
Every day millions of people use social media platforms by generating a very large amount of opinion-rich data, which can be exploited to extract valuable information about human dynamics and behaviors. In this context, the present manuscript provides a precise view of the 2020 US presidential election by jointly applying topic discovery, opinion mining, and emotion analysis techniques on social media data. In particular, we exploited a clustering-based technique for extracting the main discussion topics and monitoring their weekly impact on social media conversation. Afterward, we leveraged a neural-based opinion mining technique for determining the political orientation of social media users by analyzing the posts they published. In this way, we were able to determine in the weeks preceding the Election Day which candidate or party public opinion is most in favor of. We also investigated the temporal dynamics of the online discussions, by studying how users' publishing behavior is related to their political alignment. Finally, we combined sentiment analysis and text mining techniques to discover the relationship between the user polarity and sentiment expressed referring to the different candidates, thus modeling political support of social media users from an emotional viewpoint.





Publishing research without data is simply advertising, not science

Graham Steel

www.agid.gov.it

Francesco Branda (Università della Calabria)
Email: francesco.branda@unical.it
Sito Web: https://francescobranda.netlify.app