

# L'importanza dell'analisi del dato durante una pandemia





<https://twitter.com/anBenedetti>



<https://github.com/anbened>



<https://www.linkedin.com/in/abenedetti/>



Andrea Benedetti

Andrea Benedetti

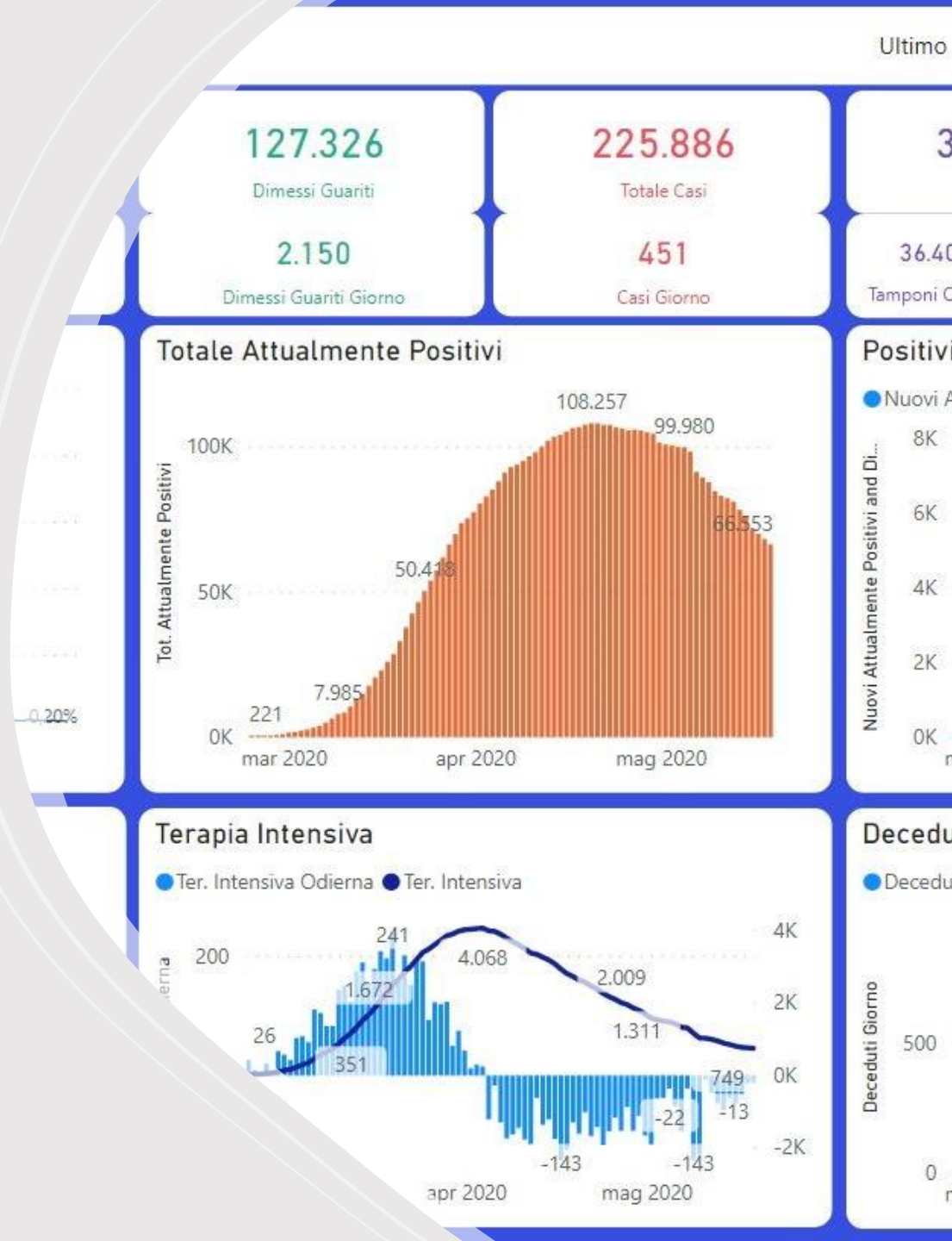
Sr. Cloud Solution Architect | Data & AI Engineer, Microsoft

«Spiegare le battute è come  
sezionare una rana: a  
nessuno piace, e nel processo  
la rana muore»

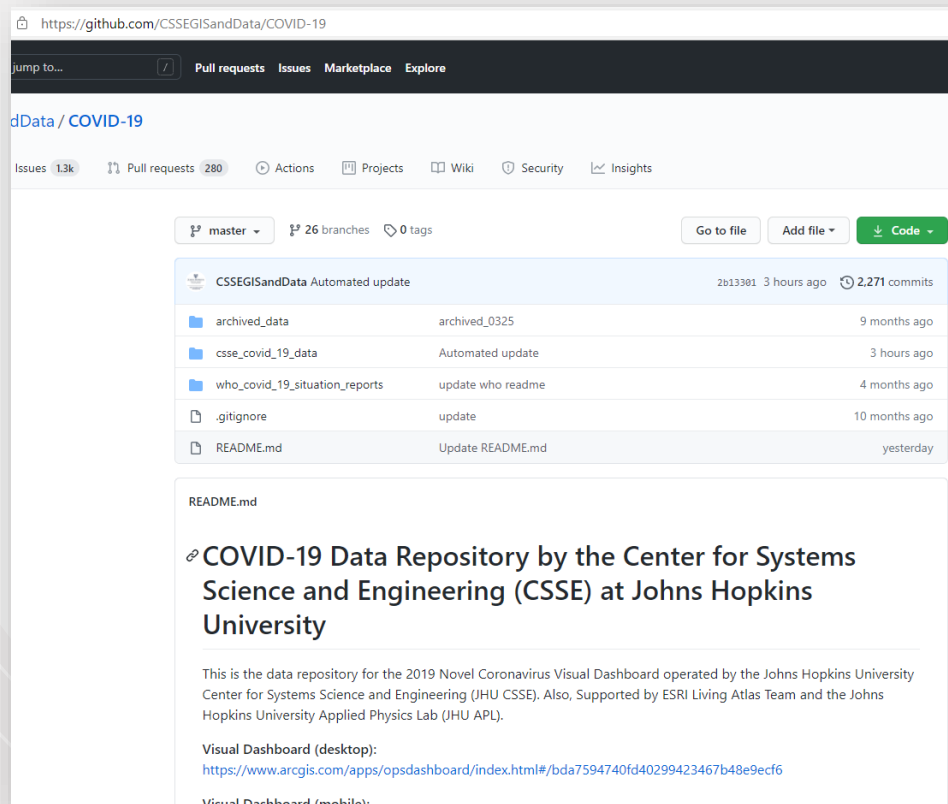
Sabato 7 marzo

<https://aka.ms/covidreport>

- Per capire qualcosa, per studiare il fenomeno, per condividere in maniera chiara - senza giri di parole o interpretazioni - cosa avevamo intorno
- Un esercizio di comprensione che aveva dato vita a una serie di analisi e di report che sono stati resi pubblici immediatamente



# Le fonti dato



The screenshot shows the GitHub repository page for CSSEGISandData/COVID-19. The repository has 1.3k issues, 280 pull requests, and 2,271 commits. The main branch is master, with 26 branches and 0 tags. The repository contains a README.md file and a directory structure including archived\_data, csse\_covid\_19\_data, who\_covid\_19\_situation\_reports, .gitignore, and README.md. The README.md file describes the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, mentioning the 2019 Novel Coronavirus Visual Dashboard and the ESRI Living Atlas Team.

https://github.com/CSSEGISandData/COVID-19

Jump to... Pull requests Issues Marketplace Explore

CSSEGISandData / COVID-19

Issues 1.3k Pull requests 280 Actions Projects Wiki Security Insights

master 26 branches 0 tags Go to file Add file Code

CSSEGISandData Automated update 2b13301 3 hours ago 2,271 commits

archived_data	archived_0325	9 months ago
csse_covid_19_data	Automated update	3 hours ago
who_covid_19_situation_reports	update who readme	4 months ago
.gitignore	update	10 months ago
README.md	Update README.md	yesterday

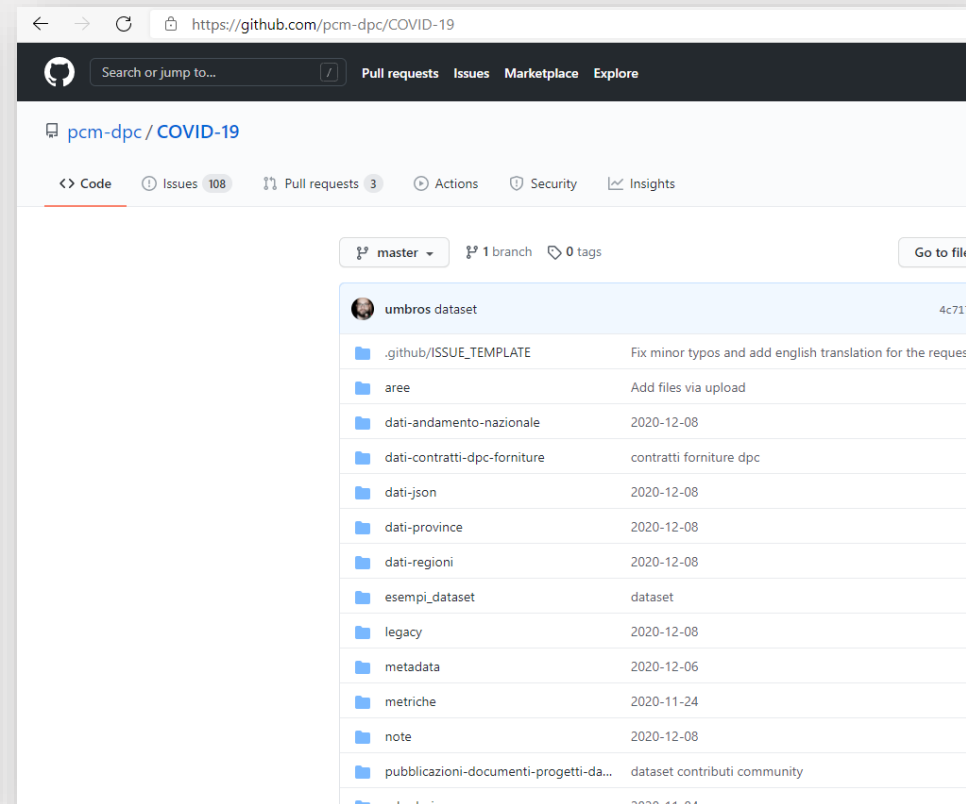
README.md

COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University

This is the data repository for the 2019 Novel Coronavirus Visual Dashboard operated by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). Also, Supported by ESRI Living Atlas Team and the Johns Hopkins University Applied Physics Lab (JHU APL).

Visual Dashboard (desktop):  
<https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>

Visual Dashboard (mobile):



The screenshot shows the GitHub repository page for pcm-dpc/COVID-19. The repository has 108 issues, 3 pull requests, and 0 tags. The main branch is master, with 1 branch and 0 tags. The repository contains a directory structure including .github/ISSUE\_TEMPLATE, aree, dati-andamento-nazionale, dati-contratti-dpc-forniture, dati-json, dati-province, dati-regioni, esempi\_dataset, legacy, metadata, metriche, note, pubblicazioni-documenti-progetti-da..., and scheda-isc. The README.md file describes the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, mentioning the 2019 Novel Coronavirus Visual Dashboard and the ESRI Living Atlas Team.

https://github.com/pcm-dpc/COVID-19

Search or jump to... Pull requests Issues Marketplace Explore

pcm-dpc / COVID-19

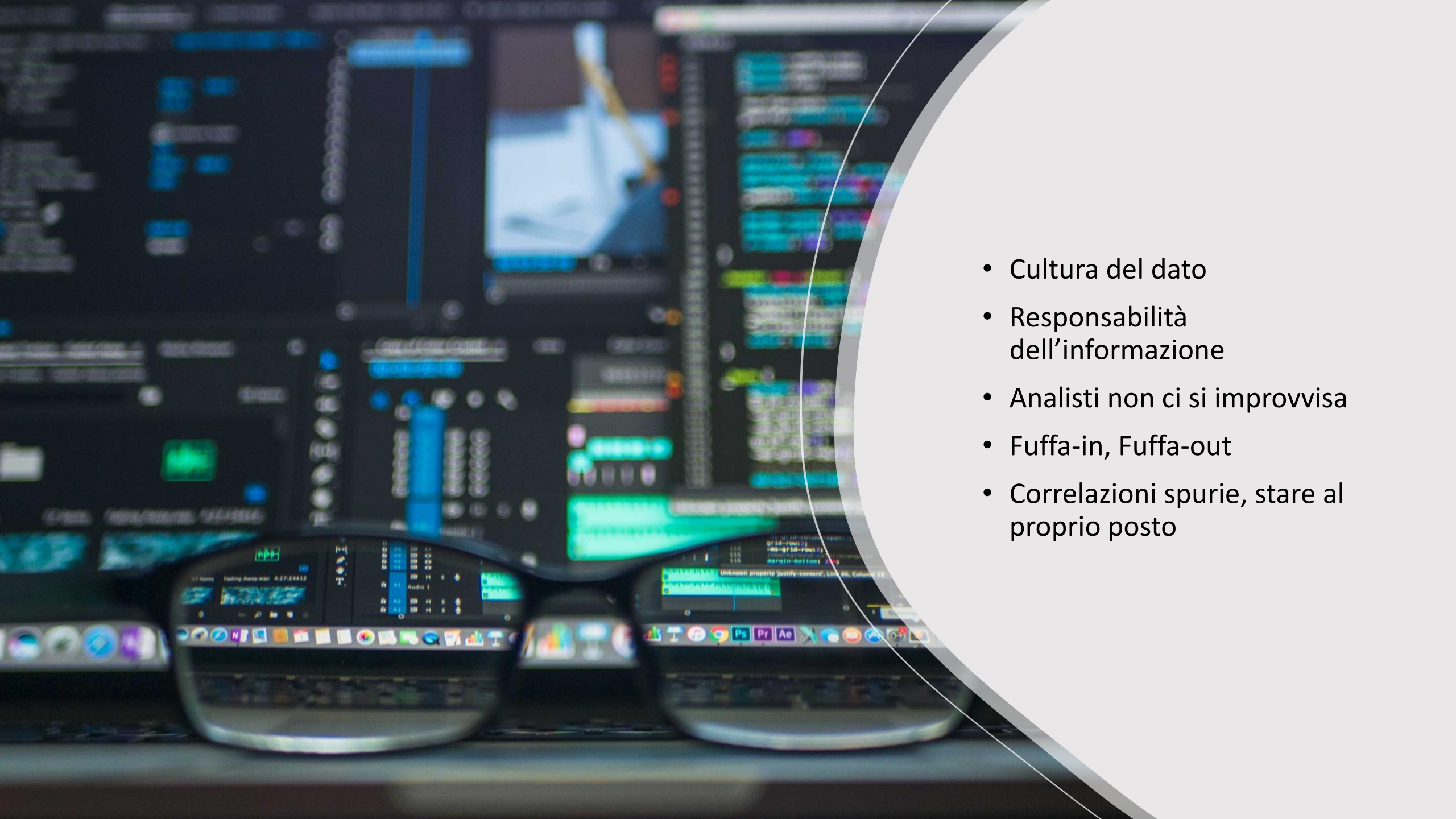
<> Code Issues 108 Pull requests 3 Actions Security Insights

master 1 branch 0 tags Go to file

umbros dataset 4c717

.github/ISSUE_TEMPLATE	Fix minor typos and add english translation for the request
aree	Add files via upload
dati-andamento-nazionale	2020-12-08
dati-contratti-dpc-forniture	contratti forniture dpc
dati-json	2020-12-08
dati-province	2020-12-08
dati-regioni	2020-12-08
esempi_dataset	dataset
legacy	2020-12-08
metadata	2020-12-06
metriche	2020-11-24
note	2020-12-08
pubblicazioni-documenti-progetti-da...	dataset contributi community
scheda-isc	2020-11-04





- Cultura del dato
- Responsabilità dell'informazione
- Analisti non ci si improvvisa
- Fuffa-in, Fuffa-out
- Correlazioni spurie, stare al proprio posto



I dati  
sono **determinanti** e lo  
sono in ogni ambito.

“La nostra capacità di fare grandi cose con i dati farà davvero la differenza in ogni aspetto della nostra vita”

Jennifer Pahlka,  
fondatrice e CEO di  
Code for America



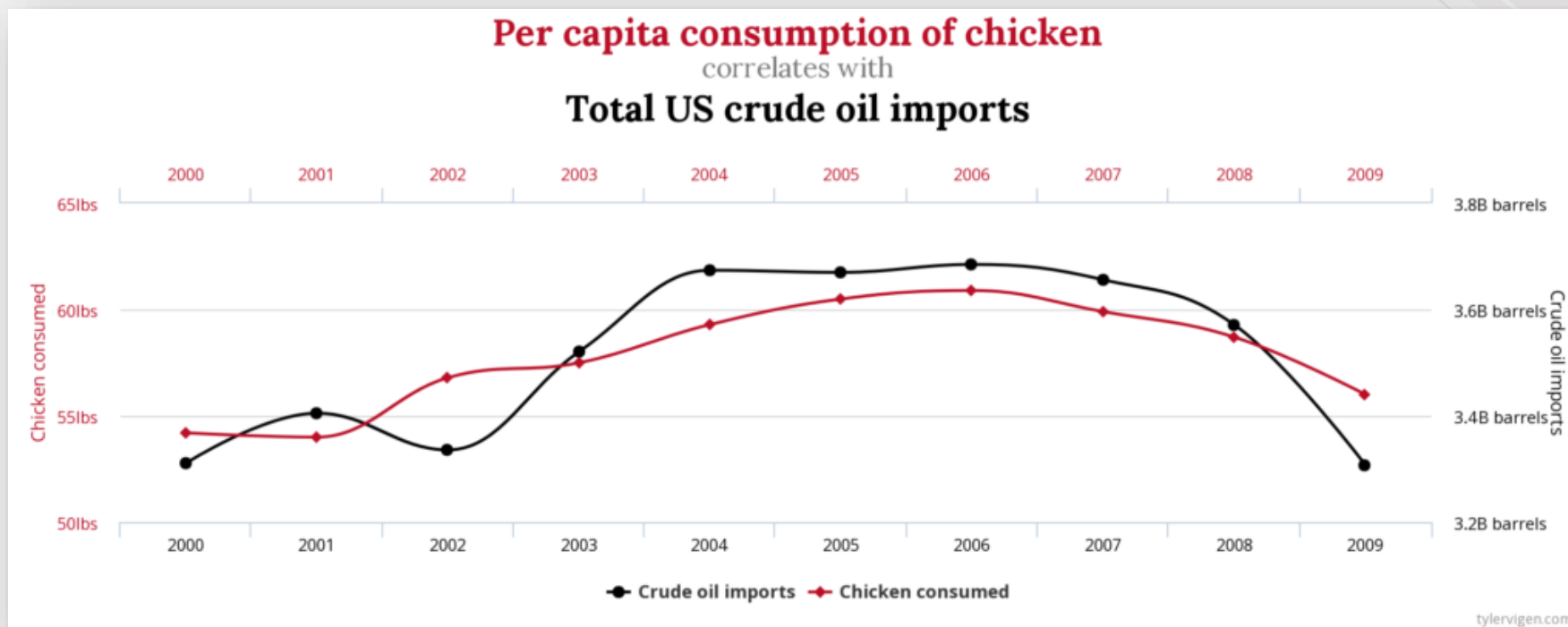
# Data quality

Se entra spazzatura nei sistemi, potrai avere la spazzatura più veloce e più sicura del mondo. Ma resta spazzatura.

Data	Citta	Casi Confermati
01/05/2020	Brescia	12
01/05/2020	Bergamo	21
02/05/2020	Brescia	14
02/05/2020	Bergamo	21
02/05/2020	Brescia	14
03/05/2020	Brescia	16
03/05/2020	Bergamo	22

# Fuffa-in = Fuffa-out

Se l'input è inaffidabile lo è anche il risultato della sua  
elaborazione (output)



# Correlazioni spurie

Tra il 2000 e il 2009, la quantità di formaggio consumata pro-capite negli Stati Uniti ha la stessa tendenza del numero di persone morte asfissiate nelle loro stesse lenzuola.

Tra il 1999 e il 2009, la spesa degli Stati Uniti nella tecnologia spaziale ha la stessa tendenza dei suicidi avvenuti per strangolamento e impiccagione.

Tra il 2008 e il 2010, il numero di morti per aggressione da parte di un cane ha la stessa tendenza del fatturato delle vendite online del Black Friday

## Ci sono note positive, in ogni caso

- Il tema della disponibilità dei dati nel nostro Bel Paese, almeno nel passato, è sempre stato sentito poco per un motivo molto semplice e, direi, culturale (dal lat. cultura, der. di colĕre «coltivare»): un tema di proprietà privata.
- Finché si continuerà a ragionare e a trattare i dati come un capitale personale, riservato, non condivisibile, sperando di ottenerne vantaggi solo con un utilizzo esclusivo, andremo poco lontano.

# Il dato è una moneta.


Pregiata (tra l'altro) e, come per il denaro, il suo valore  
accresce solo se viene messo in circolazione.



«Okay, Houston, we've had a problem here» ,  
John Swigert

(«Okay, Houston, abbiamo avuto un problema  
qui»)





Un solo, chiaro,  
obiettivo: poter prendere  
decisioni migliori.

(obiettivo che è poi quello fondamentale della raccolta,  
dell'analisi e della distribuzione dei dati)

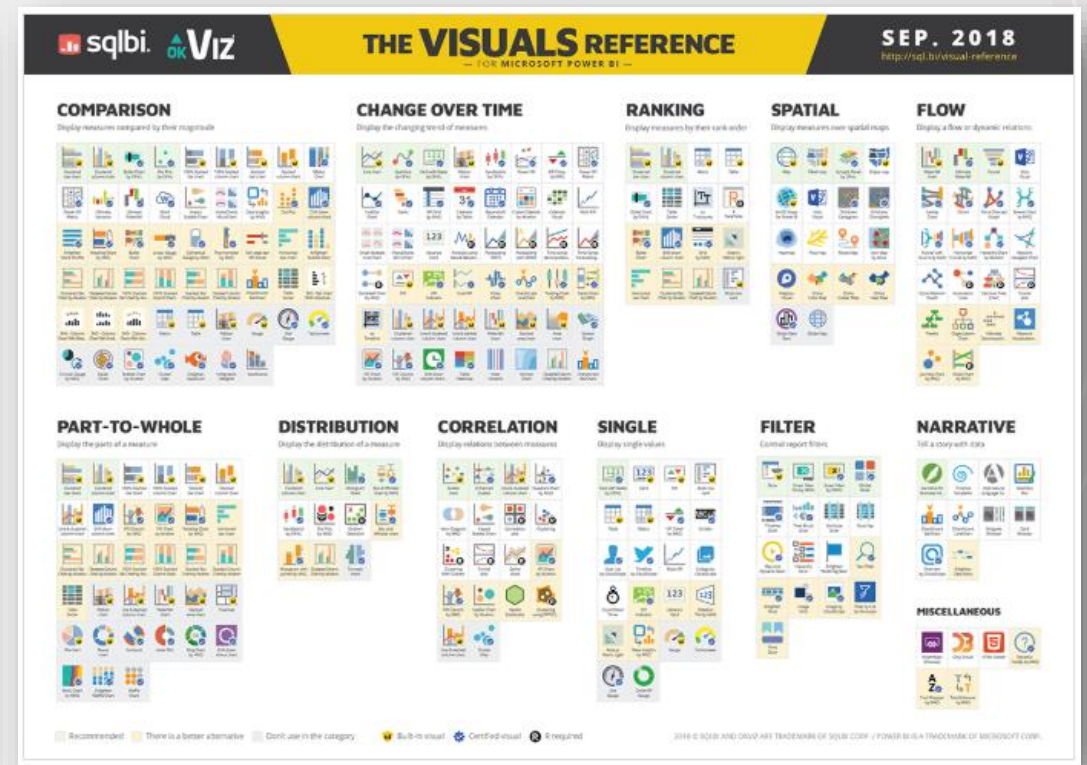
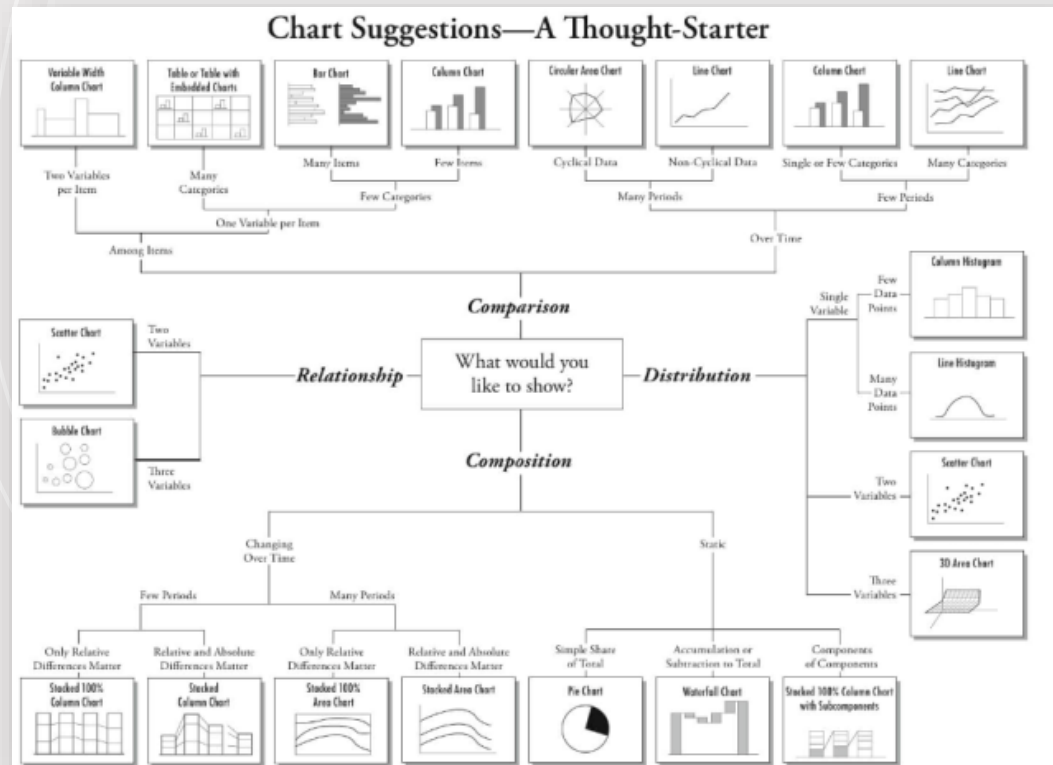
«Non si può gestire ciò  
che non si misura".

Così come non si può guidare bendati, come dice  
anche il principio dell'efficienza energetica: "non si  
può gestire ciò che non si misura".

# Osservazione

I dati sono e restano il modo più intelligente di osservare un fenomeno, limitandone l'interpretazione individuale ed emotiva.

"With enough data, the numbers speak for themselves"





# Dati per prendere decisioni

I dati, utilizzati in modo opportuno, diventano potenti abilitatori in grado di consentire decisioni basate su informazioni certe, tralasciando il più possibile quelle che sono scelte dettate da sensazioni e soggettività.

Good Data + Good Information + Good  
Analysis

=

Good Judgment ==> Good Decisions

Data has a better idea.

# Link

- [Coronavirus. Due pensieri, dopo 72 giorni.](#)
- [Trasparenza sui dati. Facciamo ordine e qualche ragionamento insieme?](#)
- [Data Culture e Coronavirus. Una proposta per un Data-Strategy manifesto - Startupitalia](#)



дякую  
ขอบคุณครับ  
Dank u Wel  
благодаря  
நன்றி  
Mulțumesc  
Tack  
Köszönöm  
Gracias  
Terima Kasih  
감사합니다  
متشكراً  
Teşekkürler  
ευχαριστώ  
Grazie  
多謝晒  
شكراً  
Děkuji  
Dziękuję  
Danke  
Kiitos  
شكريه  
Cám ơn  
Obrigado  
Hvala  
спасибо  
Salamat Po  
Tack  
Tak

# Thank You!



<https://twitter.com/anBenedetti>



<https://github.com/anbened>



<https://www.linkedin.com/in/abenedetti/>