

› Big Data

Session 10: Challenges & Opportunities

Frank Hopfgartner
Institute for Web Science and Technologies

- Big Data Analytics has emerged to become a driving force with significant impact on our society
- Research efforts show us what is possible already and what could be achieved in the near future
- The field develops quickly and it's important to stay abreast of new developments, rising challenges, and emerging opportunities

Intended Learning Outcomes

At the end of this lecture, you will be able to:

- Identify current research challenges and opportunities in the field of big data analytics
- Outline research efforts related to climate change
- Appreciate the special circumstances surrounding research advances in the context of COVID-19

Outline

- Keeping track of research advances
- Case Study: Climate change
- Case Study: COVID-19

What is research?

- Research is a systematic inquiry that investigates hypotheses, suggests new interpretations of data or texts, and poses new questions for future research to explore.
- Research consists of:
 - Asking a question that nobody has asked before;
 - Doing the necessary work to find the answer; and
 - Communicating the findings to an audience.

Why research?

- Research contributes to knowledge
- Research helps growth in business
- Academic research facilitates learning
- Academic research leads to personal growth of students

Research leads to innovation

“The researcher studies
thermodynamics; the innovator
develops the toaster”



Academic dissemination

- Research Papers
- Talks:
 - Conference
 - Invited
 - Industry
- Self-knowledge
- Commercial Outcomes

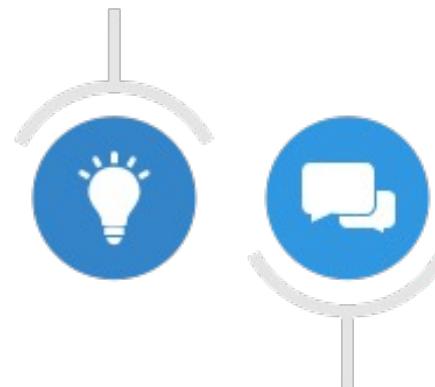
The screenshot shows a research article from the journal "Computers in Biology and Medicine". The article title is "Pre-hospital prediction of adverse outcomes in patients with suspected COVID-19: Development, application and comparison of machine learning and deep learning methods". The authors listed are M. Hasan ^{a,*}, P.A. Bath ^{a,b}, C. Marincowitz ^a, L. Sutton ^a, R. Pilbery ^c, F. Hopfgartner ^d, S. Mazumdar ^b, R. Campbell ^b, T. Stone ^a, B. Thomas ^a, F. Bell ^a, J. Turner ^a, K. Biggs ^a, J. Petrie ^a, and S. Goodacre ^a. The article is published in Computers in Biology and Medicine, volume 151 (2022) 106024. It includes sections for Article Info, Background, Method, Results, and Conclusions. The background section discusses the COVID-19 pandemic and the need for rapid decision-making. The method section details the use of machine learning models like SVM, EGB, ANN, and Stacking ensemble. The results section shows improvements in sensitivity over baseline results using these models. The conclusions section notes potential reductions in adverse outcomes and the need for further work.

A short paper
describing a
software

demonstrator that
should be brought
to the conference.

Test it first!

Demo

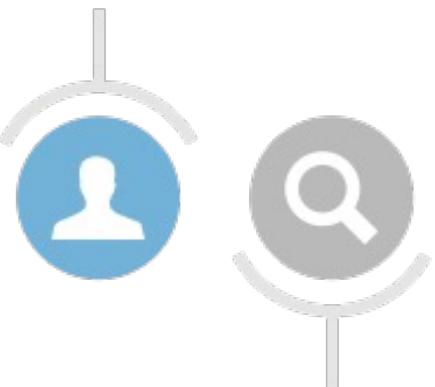


Poster

A short paper
describing an A0 or
A1 poster.

A short (half-length)
paper with short
oral/poster delivery.

Short Paper

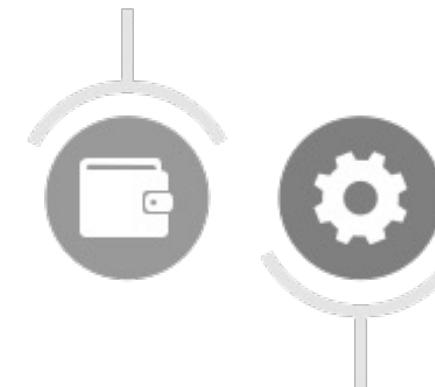


Paper

A regular length
conference paper

A non-peer-
reviewed self-
published body of
work.

Working P.



Abstract

For some
conferences (mostly
in different domains)
, an abstract is
sufficient to submit
for consideration for
oral presentation.

Top venues

Top 20 publications matching *big data*

Publication	<u>h5-index</u>	<u>h5-median</u>
1. Big Data & Society	<u>45</u>	69
2. IEEE International Conference on Big Data	<u>41</u>	52
3. IEEE Transactions on Big Data	<u>38</u>	58
4. Journal of Big Data	<u>34</u>	84
5. EPJ Data Science	<u>30</u>	44
6. Big Data Research	<u>29</u>	50
7. Big Data	<u>25</u>	39
8. IEEE International Conference on Data Science and Advanced Analytics (DSAA)	<u>23</u>	41
9. IEEE International Congress on Big Data	<u>21</u>	34
10. International Conference on Big Data and Smart Computing (BIGCOMP)	<u>19</u>	31
11. International Conference on Big Data Computing Service and Applications (BigDataService)	<u>18</u>	28
12. International Journal of Data Science and Analytics	<u>18</u>	28
13. Statistical Analysis and Data Mining: The ASA Data Science Journal	<u>17</u>	24
14. IEEE International Conference on Big Data and Cloud Computing (BdCloud)	<u>16</u>	24
15. International Conference Cloud System and Big Data Engineering	<u>15</u>	30
16. IEEE International Conference on Multimedia Big Data	<u>15</u>	27
17. IEEE International Conference on Data Science and Data Intensive Systems	<u>15</u>	25

https://scholar.google.de/citations?hl=en&view_op=search_venues&vq=big+data

Example: IEEE Big Data Conference

“IEEE Big Data brings together leading researchers and developers from academia, research and the industry from all over the world to facilitate innovation, knowledge transfer and technical progress in addressing the 5 V's (Velocity, Volume, Variety, Value and Veracity) of Big Data.”

Topics of Interest of IEEE Big Data

2013

- Big Data Science and Foundations
- Big Data Infrastructure
- Big Data Management
- Big Data Search and Mining
- Big Data Applications
- **Big Data Security & Privacy**

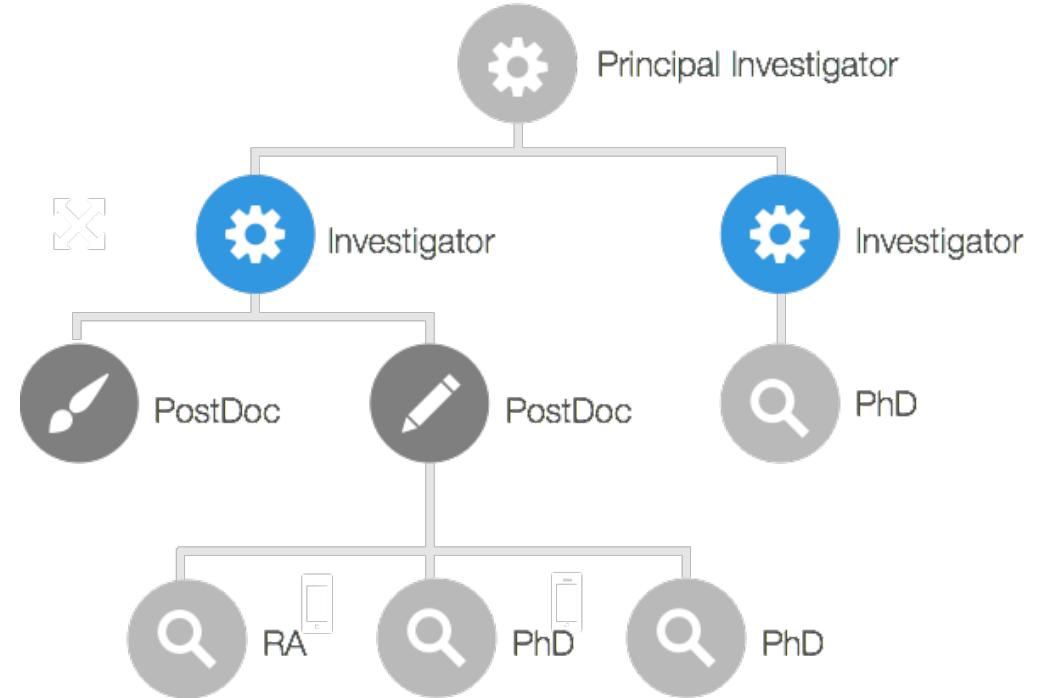
2023

- Big Data Science and Foundations
- Big Data Infrastructure
- Big Data Management
- Big Data Search and Mining
- Big Data Applications
- **Big Data Learning and Analytics**
- **Data Ecosystems**

- Data ecosystem concepts, theory, structure, and process
- Ecosystem services and management
- Methods for data exchange, monetization, and pricing
- Trust, resilience, privacy, and security issues
- Privacy preserving Big Data collection/analytics
- Trust management in Big Data systems
- Ecosystem assessment, valuation, and sustainability
- Experimental studies of fairness, diversity, accountability, and transparency

Research is costly

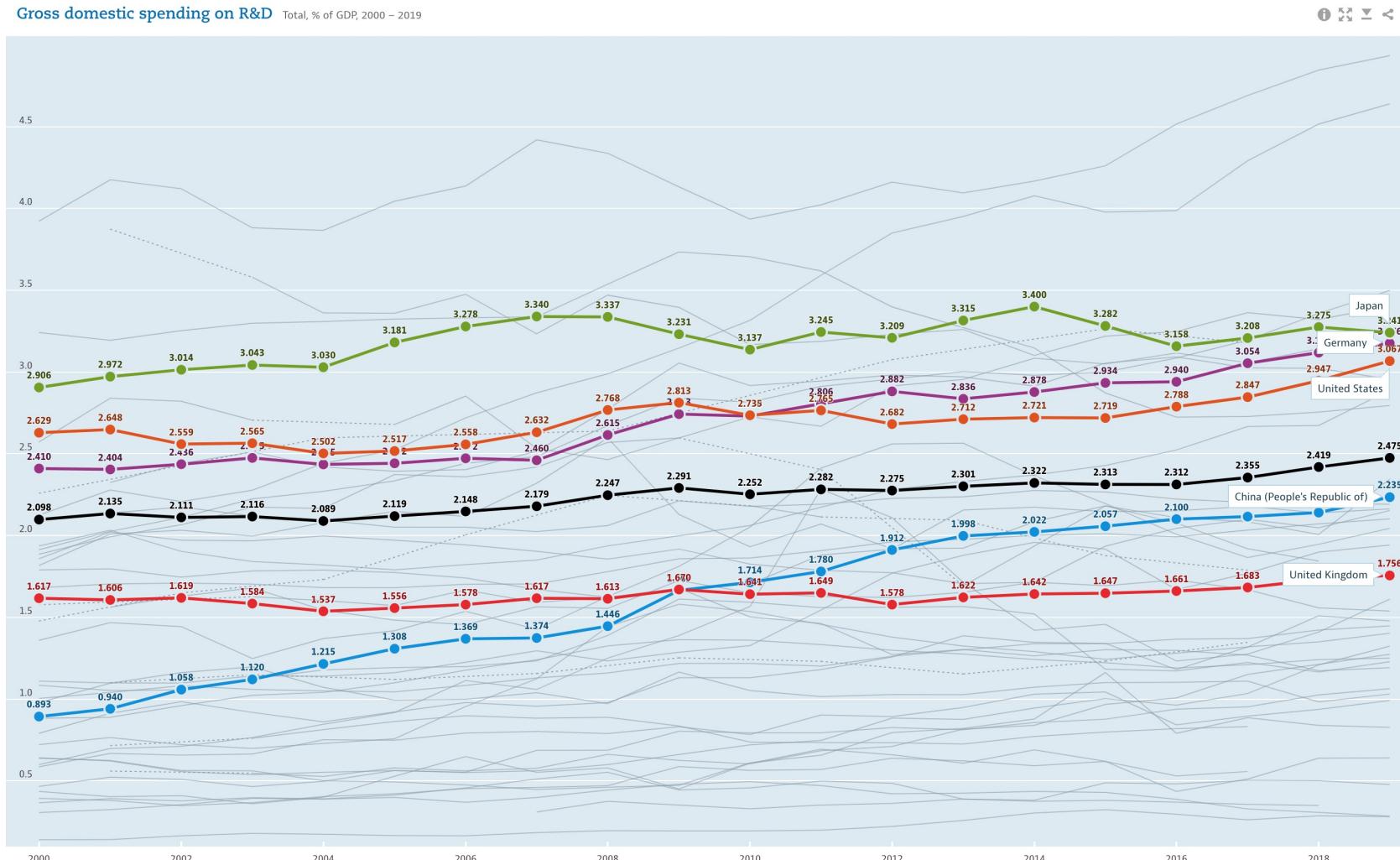
- Salaries
- Equipment
- Infrastructure
- Access to data / resources
- Conference registration fees
- Travel, Subsistence
- ...



Money rules!

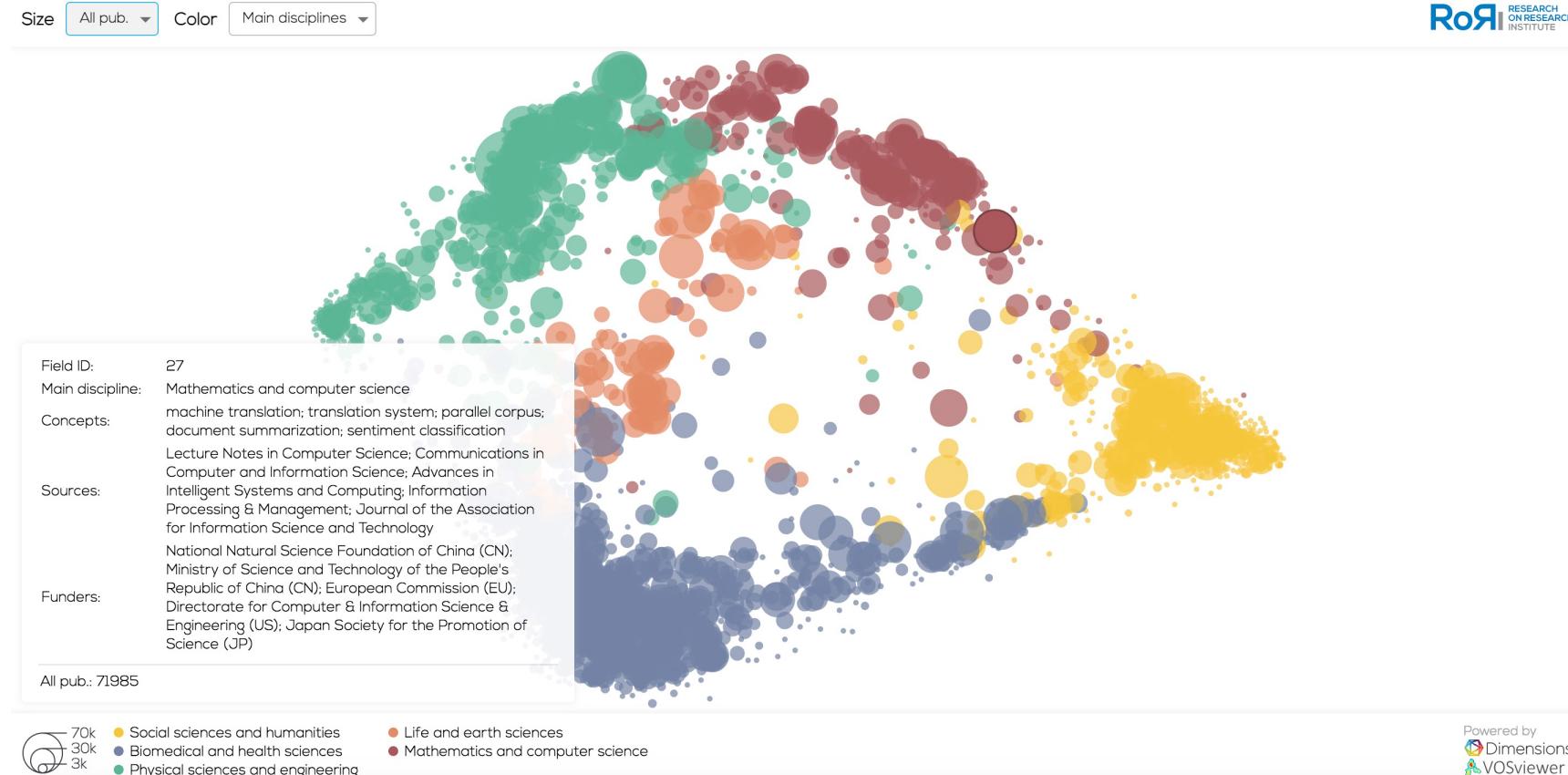
"The battle for funding has led to a change in the types of experiment that are being proposed by scientists. Grant panels are most commonly interested in proposals that are focused on solving problems, such as novel targets to prevent pathogen infection, or are guaranteed to be successful in the short term."

Who funds research?



<https://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm>

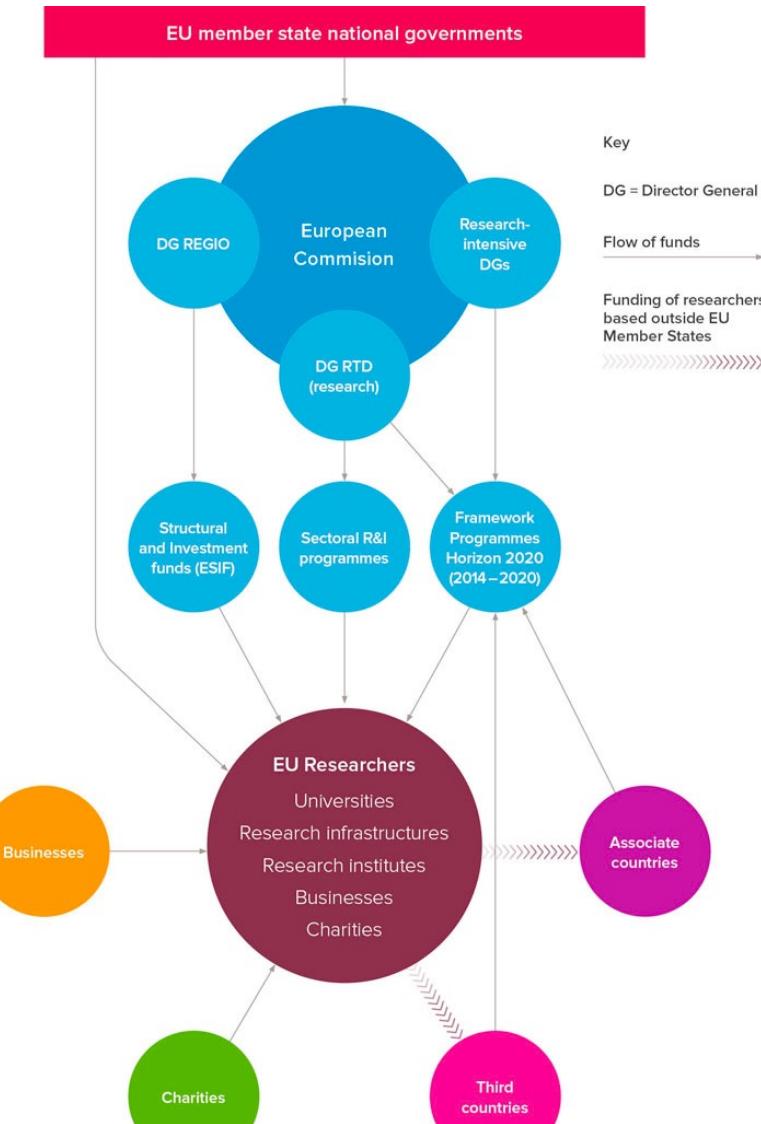
Funding landscape



European context

Approx. 15% of publicly funded research conducted in EU member states comes from, or is coordinated, by the EU

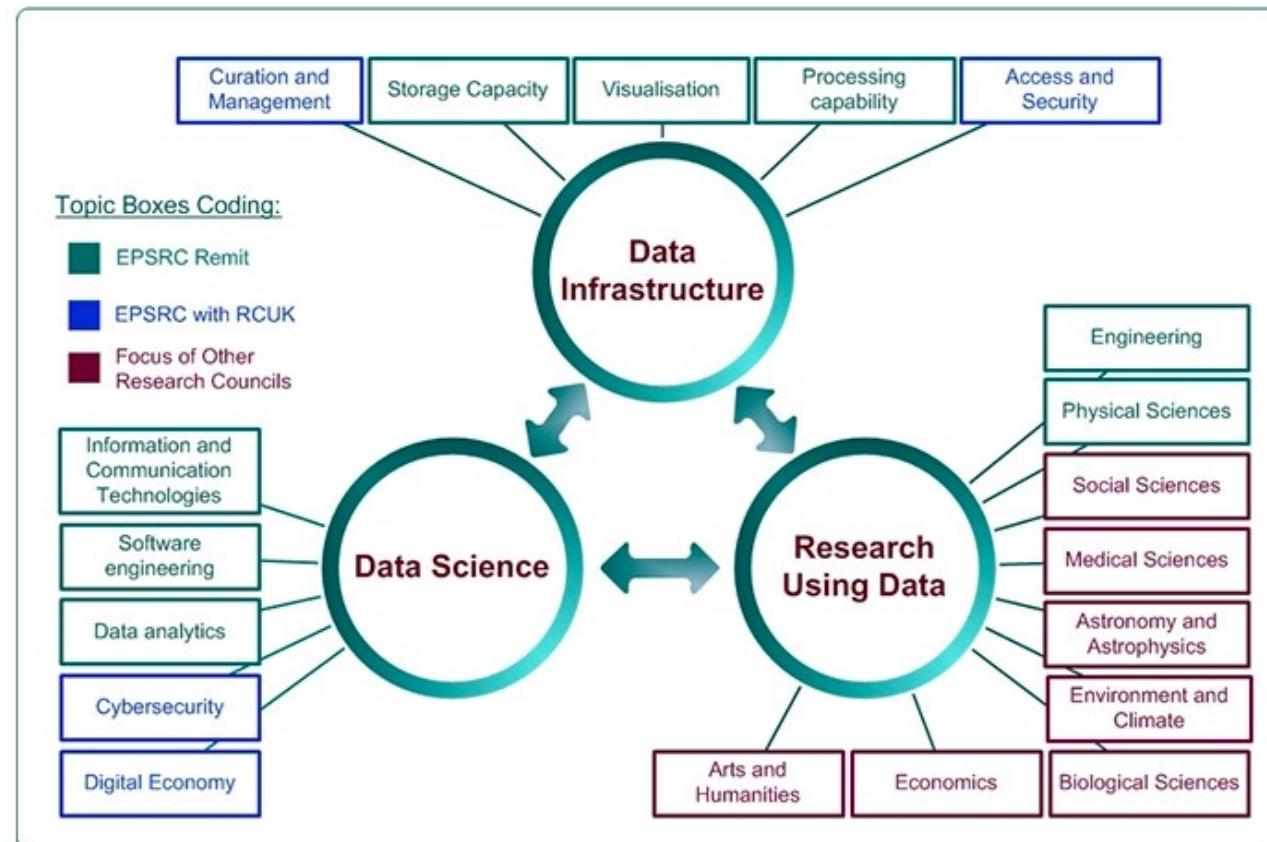
<https://royalsociety.org/topics-policy/projects/uk-research-and-european-union/role-of-EU-in-funding-UK-research/who-funds-research-in-europe/>



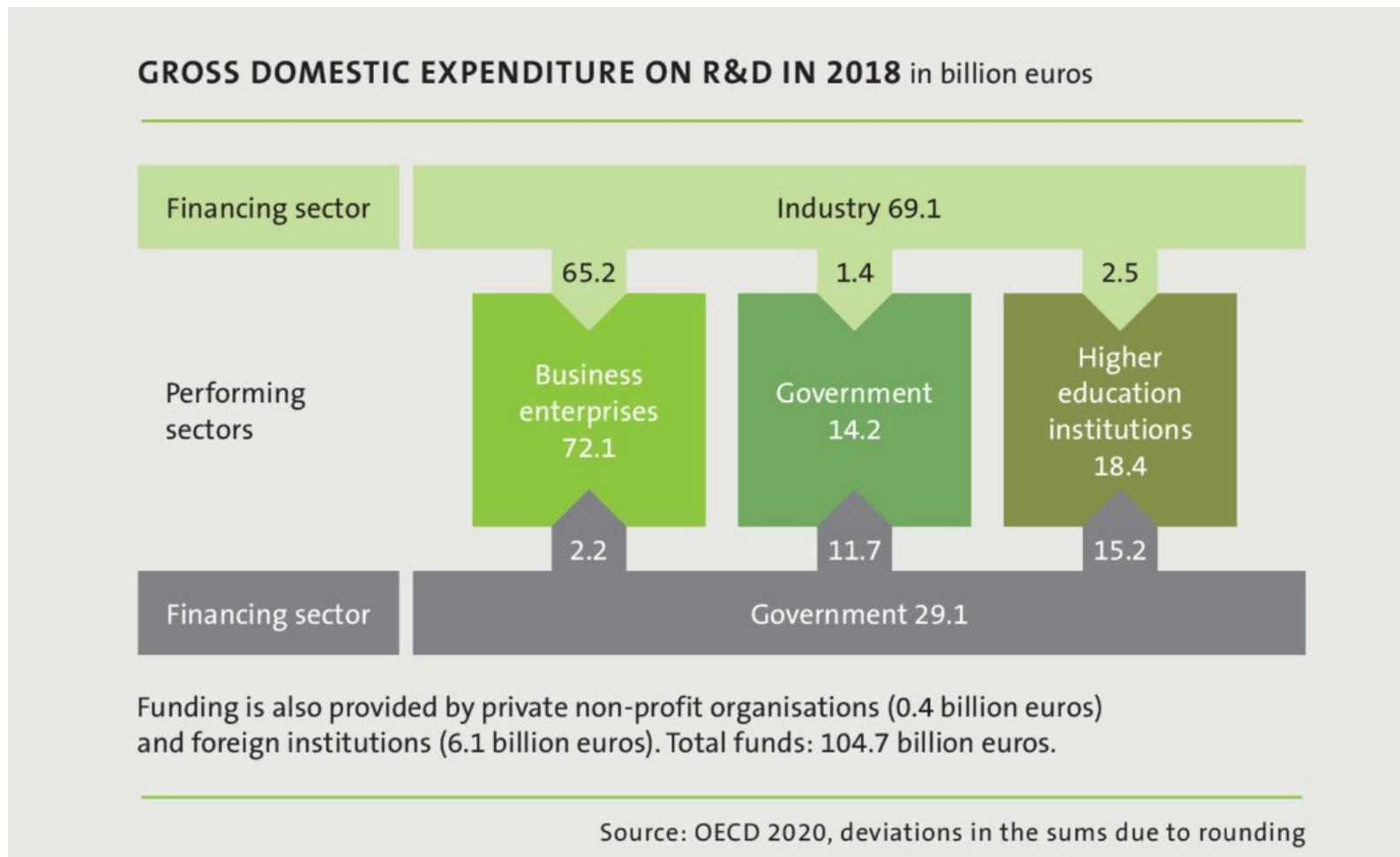
Example: Horizon Europe

The screenshot shows a newsroom agenda page from the European Commission's website. At the top left is the European Union flag icon. To its right is a search bar with a grey "Search" button. Below the search bar, the breadcrumb navigation reads: European Commission > Strategy > Shaping Europe's digital future > Newsroom Agenda > Newsroom Agenda >. The main title "Shaping Europe's digital future" is followed by a large, bold heading "Funding opportunities about Big data". On the left side, there are three dropdown filters: "Type" (set to "Funding opportunity"), "Topic" (set to "-Big data"), and "Date" (set to "26-03-2021"). Below these filters is a blue "Filter" button. On the right side, under the heading "About Big data", there is a vertical list of four categories: "Policies", "Blog posts", "News", and "Events".

Example: EPSRC



Who conducts research?



White papers

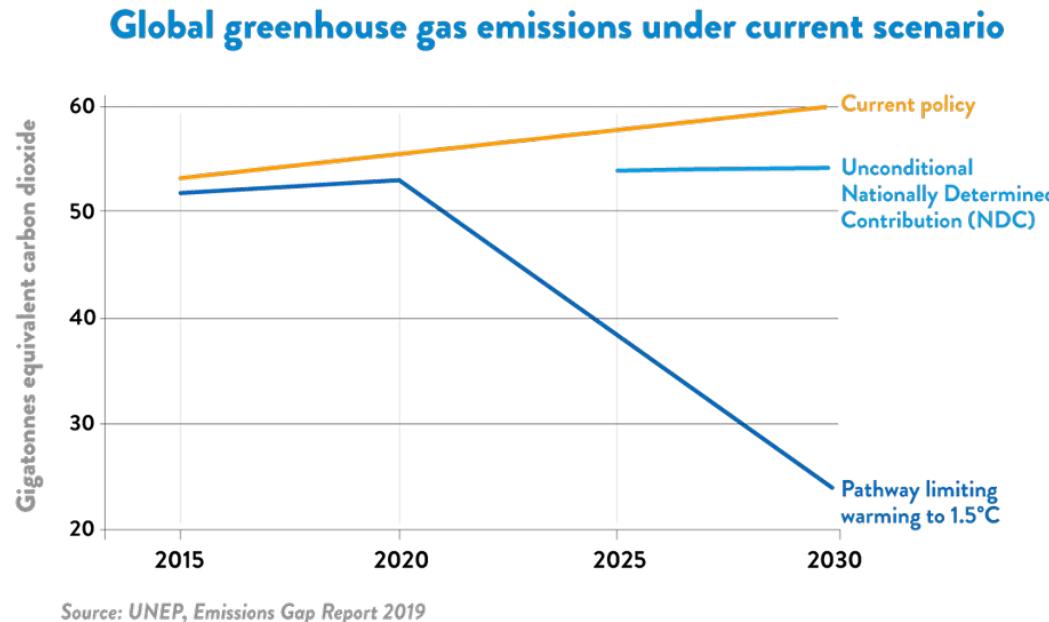
- Strategy papers, e.g., written by industry representatives, governments, funding agencies, etc.
- For example:
 - Big Data Value Association
 - Data Science Foundation
 - Chinese Academy of Information and Communications Technology



Outline

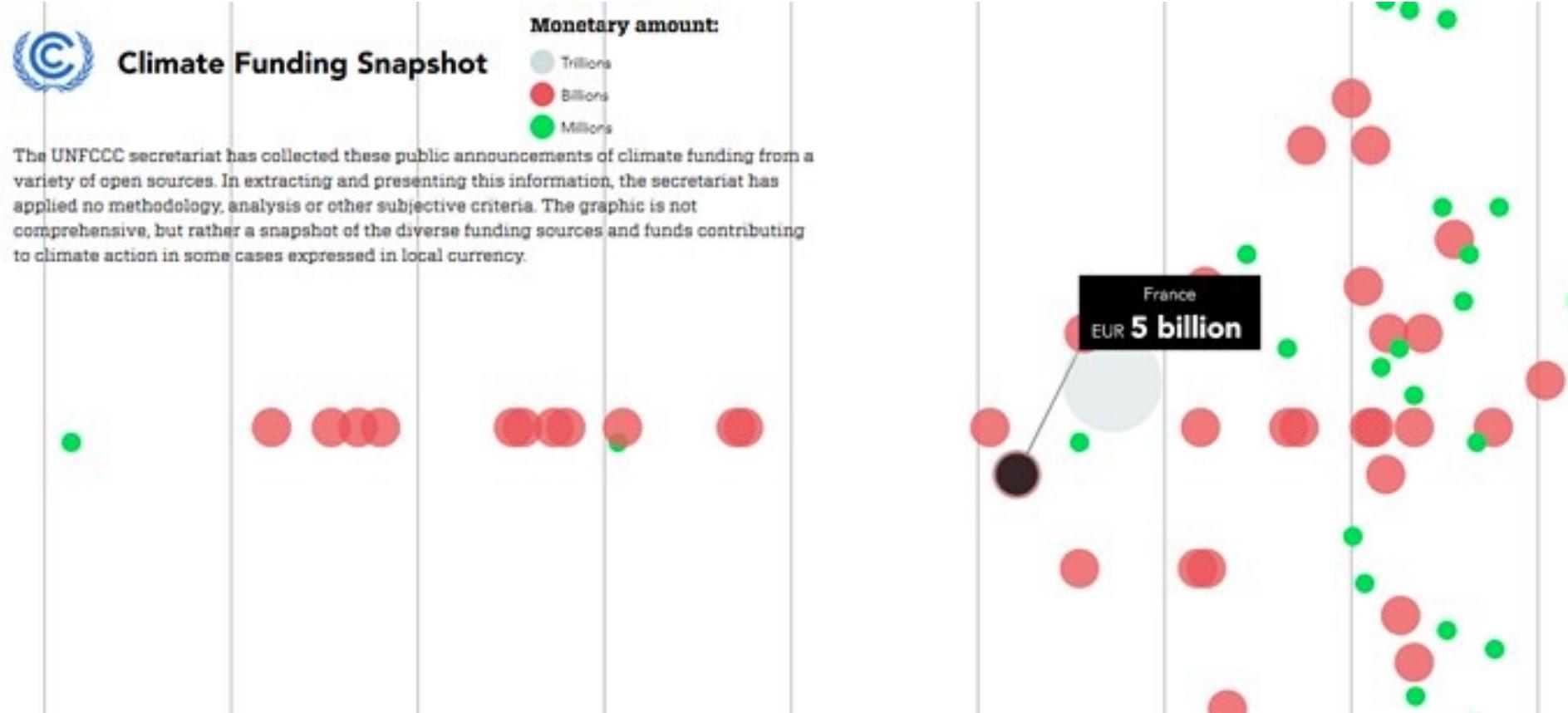
- Keeping track of research advances
- **Case Study: Climate change**
- Case Study: COVID-19

Climate action



“We have a decade to significantly curb carbon emissions and avoid catastrophe. Because of years of delayed action, we face an even more pressing mandate. We need to halve global emissions by 2030 but the emissions gap between what is needed and our current commitments is significant. Starting this year, we need to cut emissions by 7.6% every year for the next 10 years to limit warming to 1.5 degrees.”

Research funding opportunities



Big data analytics for the rescue?

“It is time for the development community and policymakers around the world to recognise and seize this historical opportunity to address twenty-first century challenges, including the effects of global volatility, climate change, and demographic shifts, with twenty-first century tools.”



How to cite this paper: UN Global Pulse (May 2012) Big Data for Development: Challenges and Opportunities.

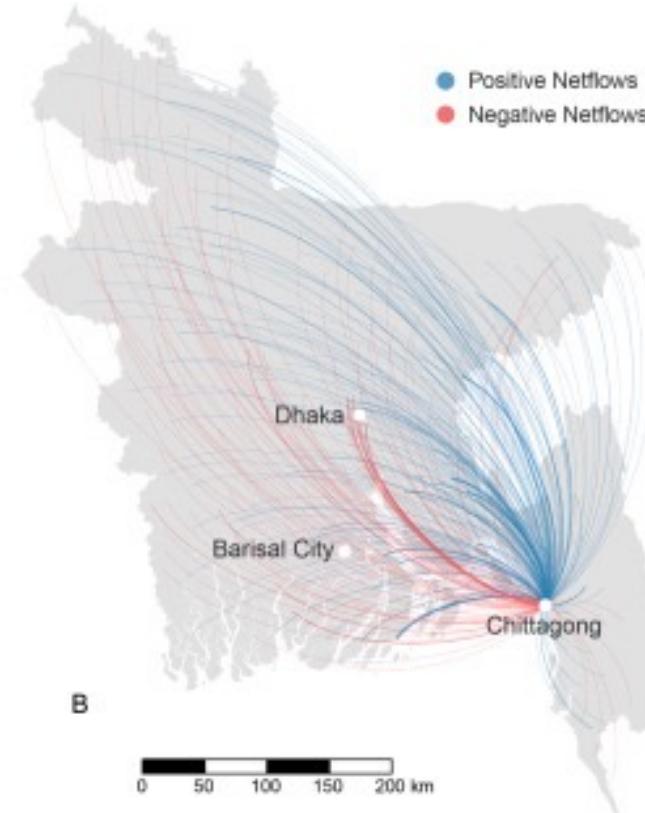
Application opportunities

- Vulnerability assessment
 - identify who and what are vulnerable to climate change
- Early warning
 - hazard warning systems
- Monitoring and evaluation
 - environmental change and people's behaviour
- Combining small and big data

J. D. Ford et al. Opinion: Big data has big potential for applications to climate change adaptation. *Proceedings of the National Academy of Sciences of the USA*, 113(39):10729-10732, 2016.

Example: Mobility patterns via phone records

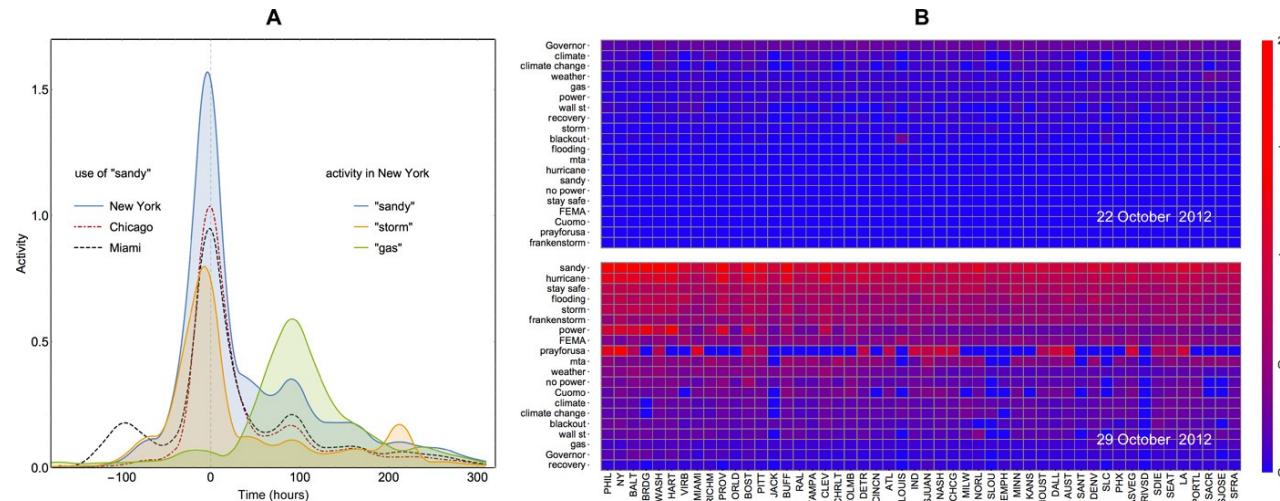
- Lack of data to monitor mobility patterns in most vulnerable regions
- Analysis of anonymized use of mobile phones following Cyclone Mahasen in Bangladesh



X. Lu et al. Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh, *Global Environmental Change*, 38:1-7, 2016

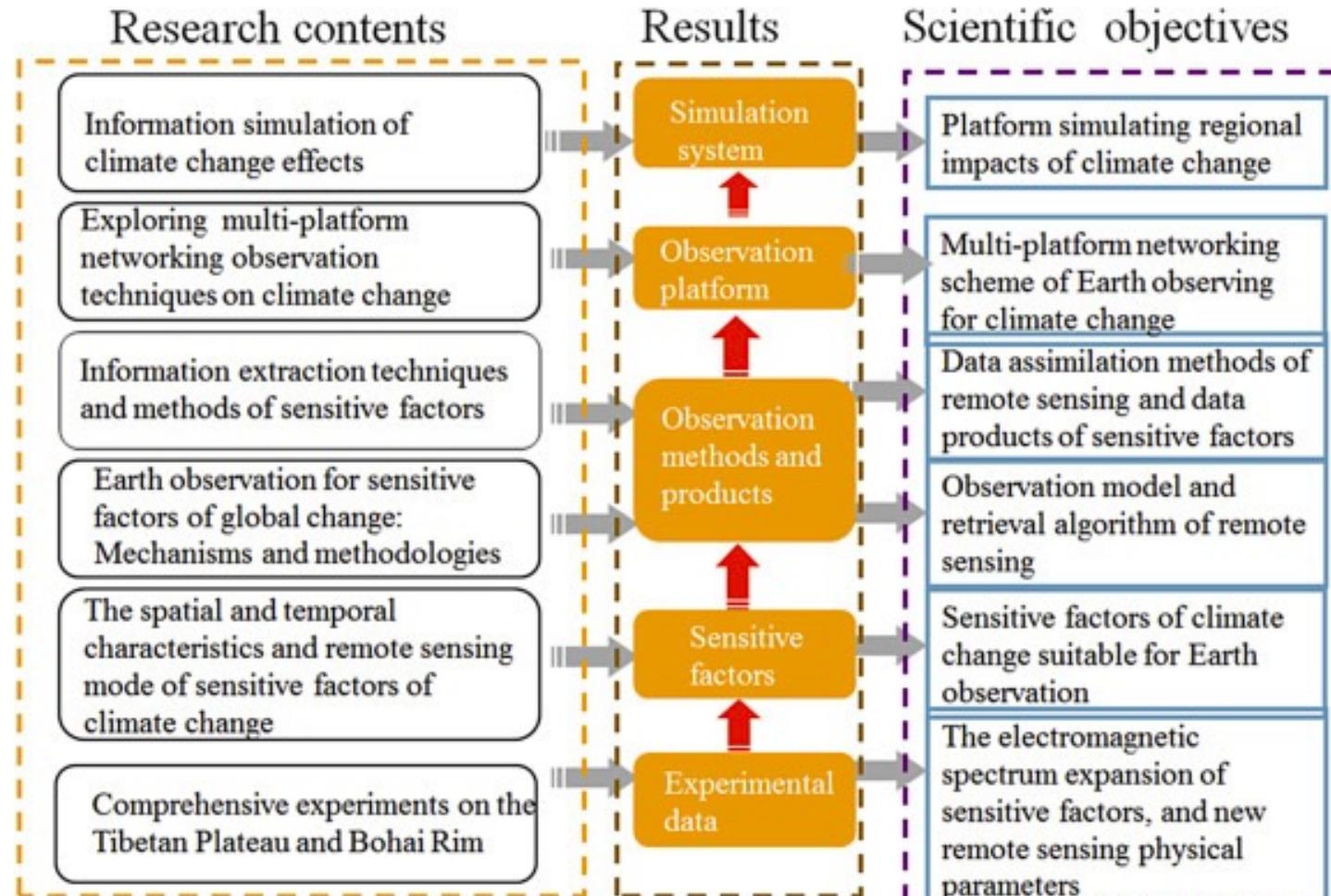
Example: Assess disaster through the lens of Twitter

- Analysis of Twitter activity before, during, and after hurricane Sandy
- Strong relationship between Sandy's path and Twitter activity



Y. Kryvasheyeu et al. Rapid assessment of disaster damage using social media activity. *Science Advances*, 2(3), 2016

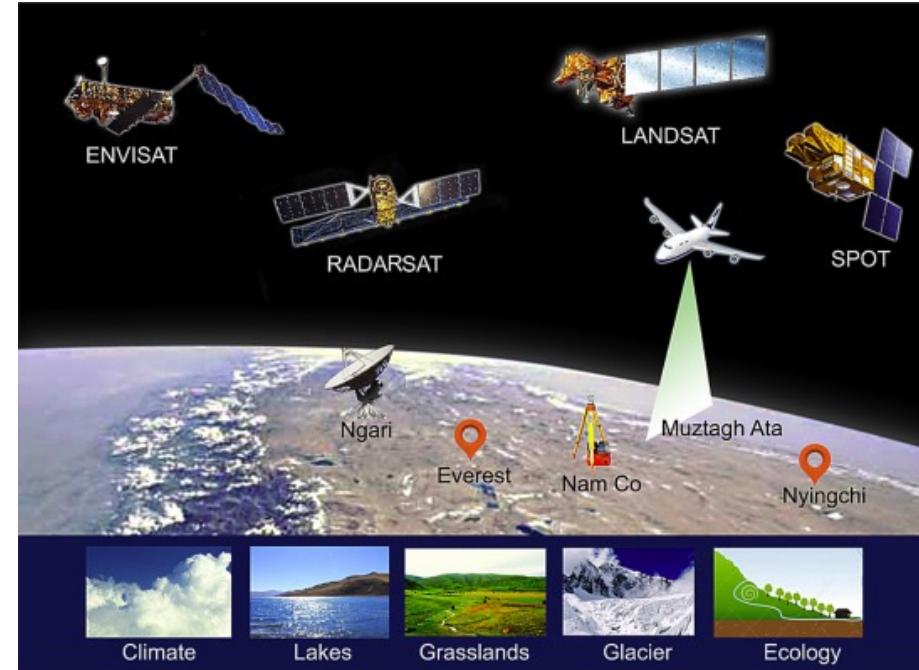
Example: Earth Observation for Sensitive Variables of Global Change Programme



H. D. Guo, L. Zhang, L. W. Zhu. Earth observation big data for climate change research. *Advances in Climate Change Research*, 6(2):108-117, 2015

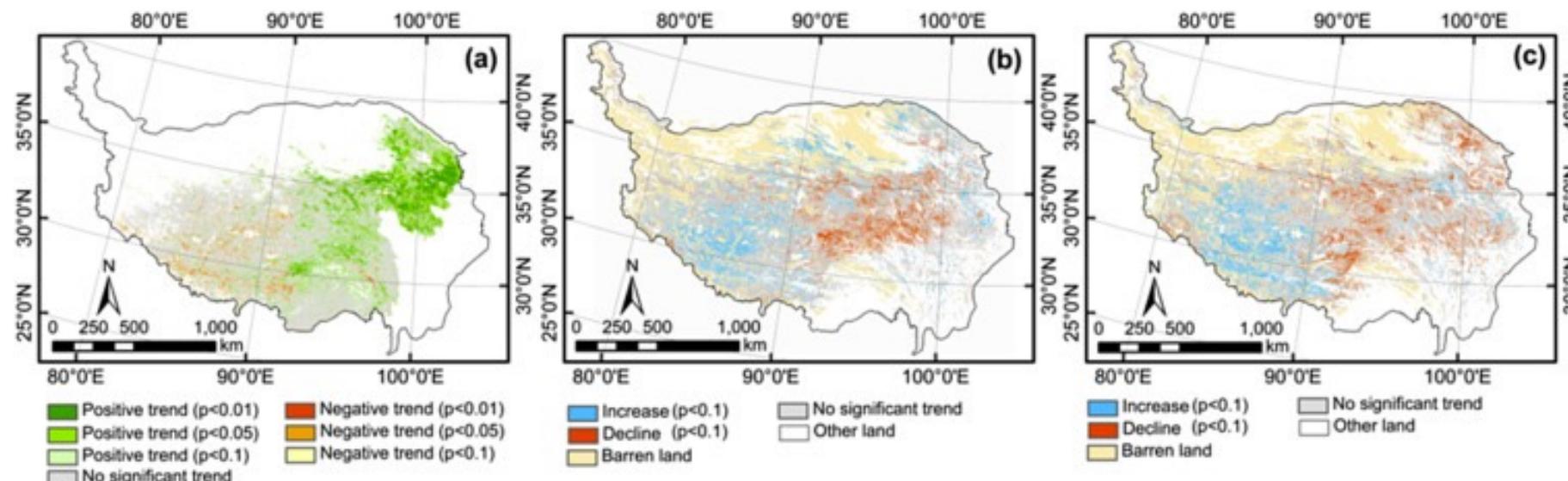
Experiments on Qinghai-Tibetan plateau

- Synchronous satellite-aerial ground observation
- Data has typical characteristics of big data



Earth observation

- Glaciers
- Lakes
- Vegetation
- Radiation
- Urbanisation



H. D. Guo, L. Zhang, L. W. Zhu. Earth observation big data for climate change research. *Advances in Climate Change Research*, 6(2):108-117, 2015

UN Climate Change Conference 2021

Sustainability Governing Principles

1. Actively manage potential impacts on the environment and local community and identify opportunities to deliver environmental and social value
2. Provide an accessible and inclusive setting for all
3. Encourage healthy living
4. Ensure a safe and secure atmosphere
5. Encourage more sustainable behaviour
6. Promote the use of responsible sources and responsible use of resources throughout the supply chain
7. Leave a positive legacy





EN English Search

Home > Horizon Europe > Missions in Horizon Europe > Adaptation to climate change including societal transformation

Mission area: Adaptation to climate change including societal transformation

What this mission area is, how missions will be chosen, mission boards, meetings, news, events.

PAGE CONTENTS

What are missions and mission areas?

What are missions and mission areas?

EU missions are commitments to solve major societal challenges like fighting cancer, adapting to climate change, protecting our oceans, living in greener cities and ensuring soil health and food.

As integral part of the Horizon Europe framework programme beginning in 2021, they are a new way to deliver on these commitments.

Each mission will operate as a portfolio of actions – such as research projects, policy measures or even legislative initiatives - to achieve a measurable goal that could not be achieved through individual actions. The missions will contribute to the goals of the European Green Deal, Europe's Beating Cancer Plan as well as the Sustainable Development Goals.

What this mission area deals with

How specific missions are identified

More information

Example projects



SCIENCE WITH AND FOR SOCIETY **TeRRIfica**

TeRRIfica is engaging citizens and other stakeholders in six pilot regions across Europe to co-develop climate change adaptation and mitigation strategies.



SOCIETAL CHALLENGE 2 (Food security, sustainable agriculture and forestry, marine, maritime and inland water research, and the bioeconomy) **FOODSHIFT2030**

The project will launch an ambitious citizen-driven transition of the European food system towards a low carbon circular future, including a shift to less meat and more plant based diets.

More information on the projects funded by Horizon 2020 is available at cordis.europa.eu



SOCIETAL CHALLENGE 5 (Climate action, environment, resource efficiency and raw materials) **CLAIR-CITY**

CLAIR-CITY engaged citizens across Europe to better understand their environmental behaviours. The project looked at how they travelled or heated their homes and at their future aspirations. Were citizens interested in adopting greener lifestyles? And if so, how would this affect carbon emissions and air pollution? Importantly, would their involvement in decision-making processes accelerate the transition to a zero carbon future?



EUROPEAN RESEARCH COUNCIL CITIZEN SCIENCE **LICCI**

ERC-funded LICCI aims to bring indigenous and local knowledge to climate change research. Through cutting-edge science, it strives to deepen our understandings of perceived climate change impacts, and endeavours to bring indigenous and local knowledge into policy-making processes and influence international climate change negotiations.

Contributing to the challenge

- Data revolution threatens sustainable development because of its environmental footprint
- Big data requires high consumption of non-renewable energy, waste production and CO2 emissions

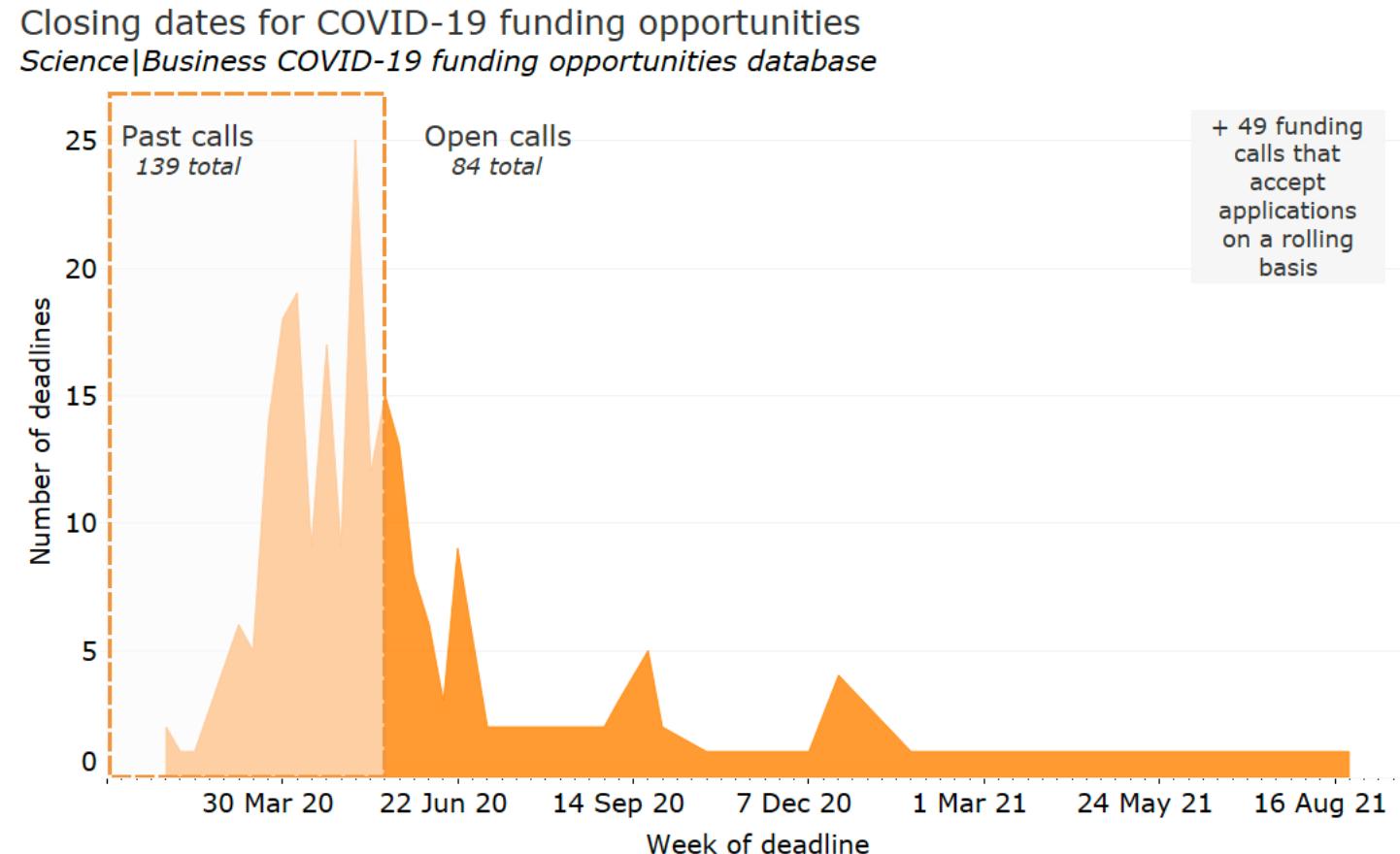
F. Lucivero. Big Data, Big Waste? A Reflection on the Environmental Sustainability of Big Data Initiatives. *Science and Engineering Ethics*, 26:1009-1030, Springer Verlag, 2020.

Outline

- Keeping track of research advances
- Case Study: Climate change
- **Case Study: COVID-19**

- “**It is not only a public health crisis, but also a humanitarian crisis in the making**” (International Red Cross)
- “**The Coronavirus crisis [...] is attacking societies at their core.**” (United Nations)
- “**The COVID-19 pandemic represents an unprecedented disruption to the global economy**” (World Trade Organisation)

Funding for COVID-19 research



Example: UKRI funding call

- UKRI called for research that:
 - has a clear impact pathway with the potential to deliver a significant contribution to the understanding of, and response to, the COVID-19 pandemic and its impacts
 - supports the manufacture or wide-scale adoption of an intervention with significant potential for COVID-19
 - gathers time-critical data and resources related to COVID-19 quickly and secures them for future research use
 - builds capability in areas that will be necessary to deliver a significant contribution to the understanding of, and response to, the COVID-19 pandemic and its impacts in the future.

Funding COVID-19 UKRI research grants



COVID-19 is a data science issue

Sections	Subsection
(§II) Datasets and Resources This section provides information about numerous datasets related to COVID-19. It also gives information about ongoing data science competitions, and online resources.	(§II-A) COVID-19 Case Data (§II-B) COVID-19 Textual Data (§II-C) COVID-19 Biomedical Data (§II-F) Other Supportive Datasets (§II-E) COVID-19 Competition Datasets
(§III) Applications of Data Science for COVID-19 This section highlights different use cases related to the application of data-driven methodologies for addressing COVID-19. It also discusses some examples of these use cases.	(§III-A) Risk Assessment and Patient Prioritisation (§III-B) Screening and Diagnosis (§III-C) Simulation and Modelling (§III-D) Contact Tracing (§III-E) Understanding Social Interventions (§III-F) Logistical Planning and Economic Interventions (§III-G) Automated Primary Care (§III-H) Supporting Drug Discovery and Treatment
(§IV) Survey of Ongoing Research This section surveys ongoing work across several types of data. It also provides brief summaries of outcomes and methodologies.	(§IV-A) Image Data Analysis (§IV-B) Textual Data Analysis (§IV-C) Voice Sound Data Analysis (§IV-D) Embedded Data Analysis (§IV-E) Pharmaceutical Research
(§V) Bibliometric Analysis of COVID-19 Research This section presents a bibliometric analysis of COVID-19 research.	(§V-A) Bibliometric Data Collection (§V-B) Peer-reviewed vs. Non-peer-reviewed publications (§V-C) Research Topics (§V-D) COVID-19 vs. Earlier Epidemics
(§VI) Cross-Cutting Challenges This section highlights challenges that researchers may face when performing data-driven research related to COVID-19.	(§VI-A) Data Limitations (§VI-B) Correctness of Results vs. Urgency (§VI-C) Security, Privacy, and Ethics (§VI-D) The Need For Multidisciplinary Collaboration (§VI-E) New Data Modalities (§VI-F) Solutions for the Developing World
(§VII) Conclusions	

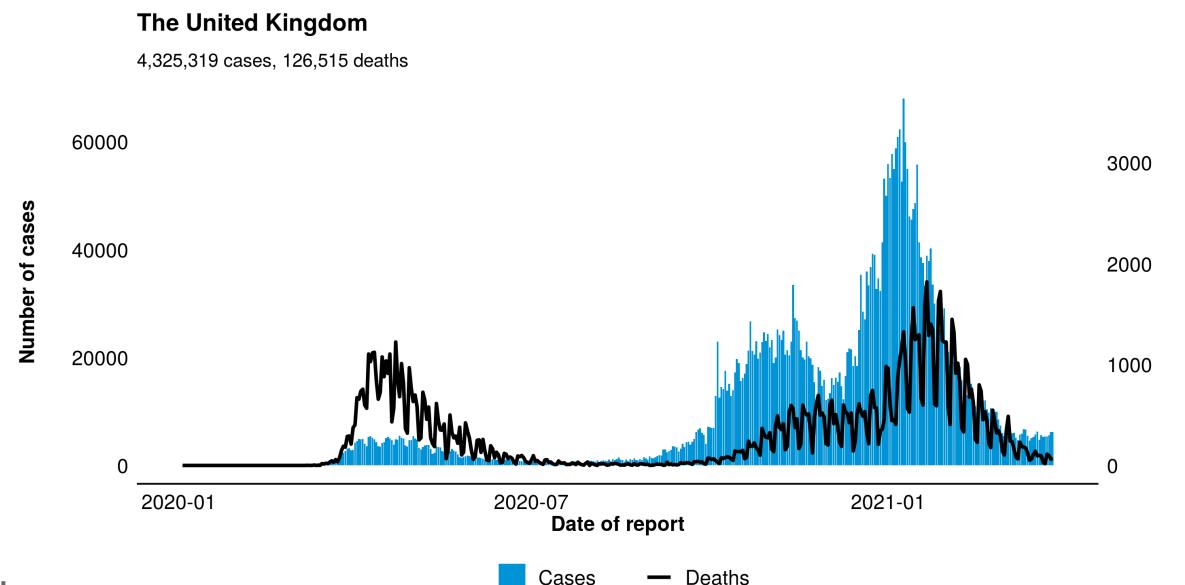
COVID-19 is a big data problem

“To make any real progress in this situation, you need to bring together people who understand the computation and AI, people who understand the biological and biomedical implications, and people who understand population models. It's a very interdisciplinary problem, and to make any headway, we need the data and we need the team.”



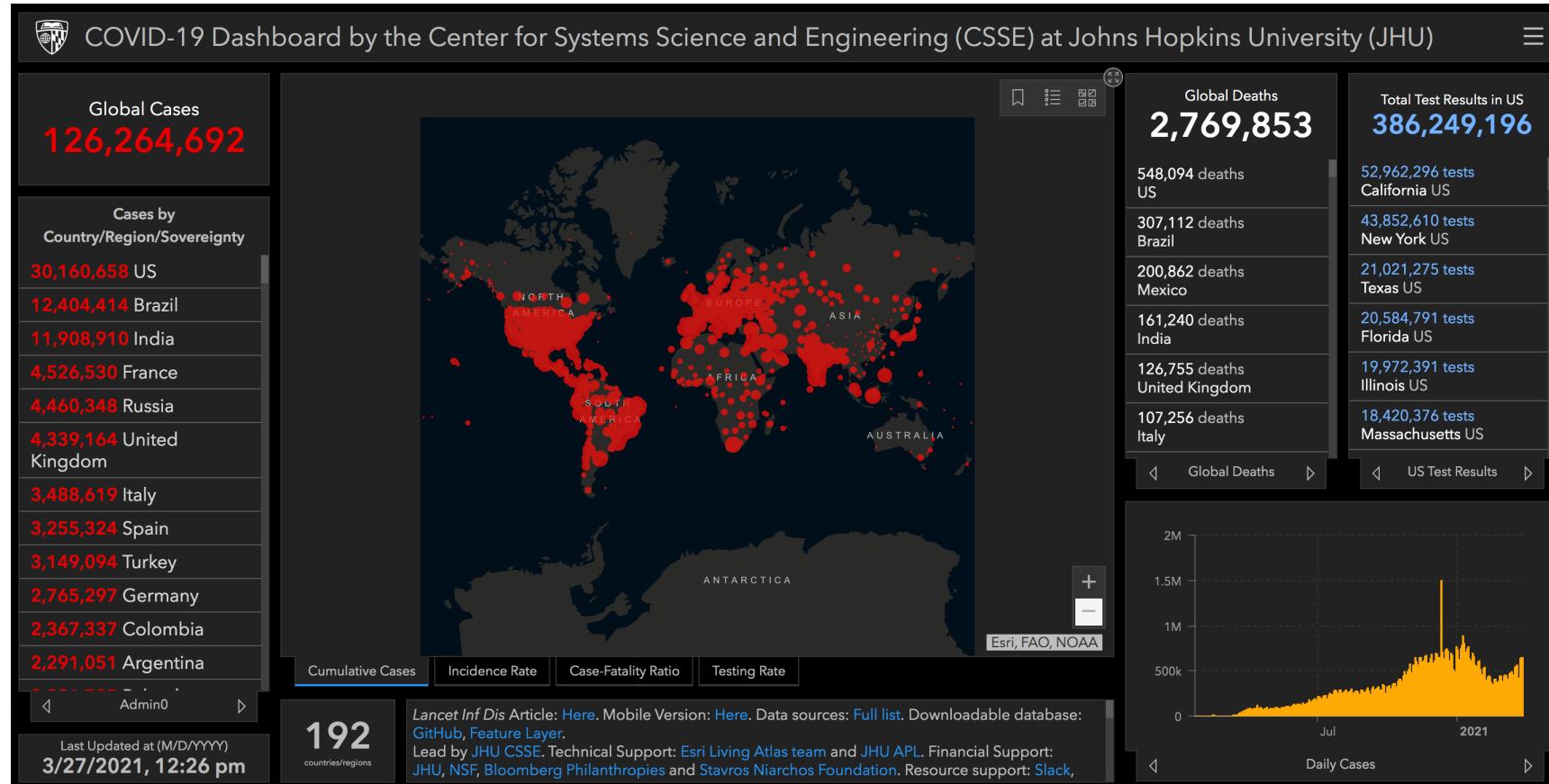
Prof. James Hendler
Rensselaer Institute for
Data Exploration and
Applications

- Countless datasets made available, e.g.,
 - World Health Organisation
 - National Governments / Health Service providers
 - Companies (e.g., Apple, Google)
 - ...



Source: WHO COVID Intel Database

Example: Real-time dashboard



Open COVID pledge

The Pledge Licenses ▾ IP ▾ Partners ▾ FAQs News About Contact Us

Make the pledge to share your intellectual property in the fight against COVID-19.

Make the Pledge Support the Pledge Featured IP

Amazon
Intel
IBM
Microsoft
SAP
...

Big Data infrastructure access

- **Amazon Web Services** committed \$20 million for customers working on diagnostics solutions
- **Google Cloud** established COVID-19 Public Dataset Program to provide access to data and their BigQuery ML tool
- Many other examples...

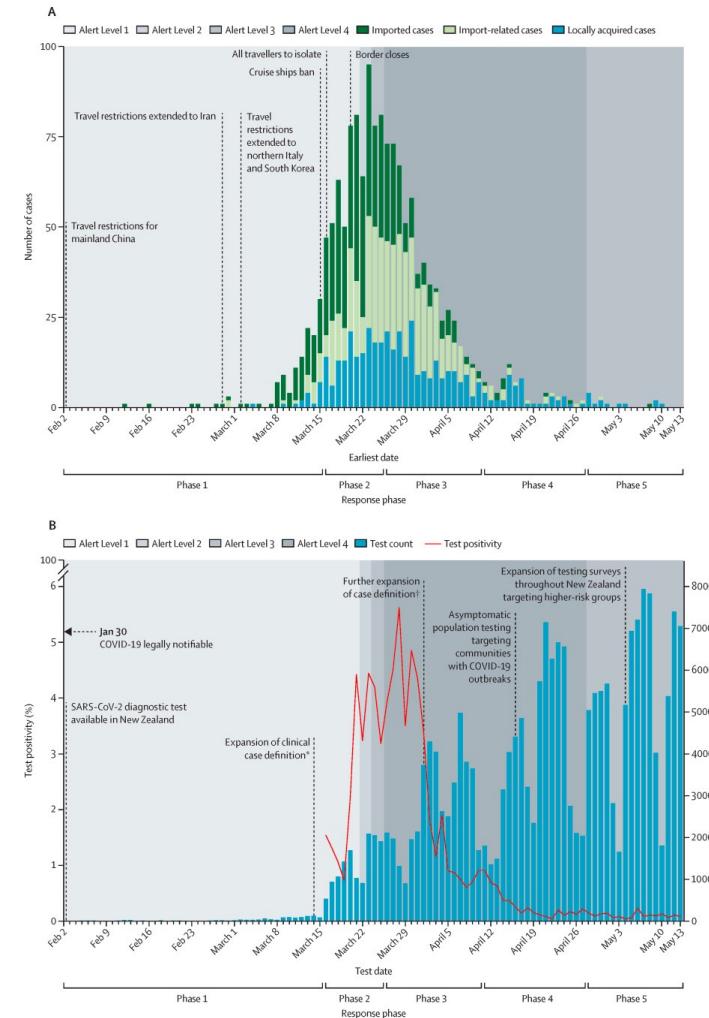
<https://www.aboutamazon.com/news/aws/aws-launches-global-initiative-to-accelerate-covid-19-diagnostics-research-and-testing>
<https://cloud.google.com/blog/products/data-analytics/free-public-datasets-for-covid19>

- Descriptive and Diagnostic Analytics
 - Analyse and summarise past data
- Predictive Analytics
 - Predictions about the future
- Prescriptive Analytics
 - Determine actions and evaluate their impact

J. Sheng, J. Amankwah-Amoah, Z. Khan, X. Wang. COVID-19 Pandemic in the New Era of Big Data Analytics: Methodological Innovations and Future Research Directions. *British Journal of Management*, 2020,
<https://doi.org/10.1111/1467-8551.12441>

Example: Impact in New Zealand

- Analysis of all lab-confirmed and probable cases of COVID-19 in the country
- Mapped with other data



S. Jeffries et al. COVID-19 in New Zealand and the impact of the national response: A descriptive epidemiological study. *The Lancet*, 5(11):e611-e623, 2020

Example: Predicting COVID-19 spread in Switzerland

- Simulation of the entire population of Switzerland, including daily activities, transmission pathways, government interventions

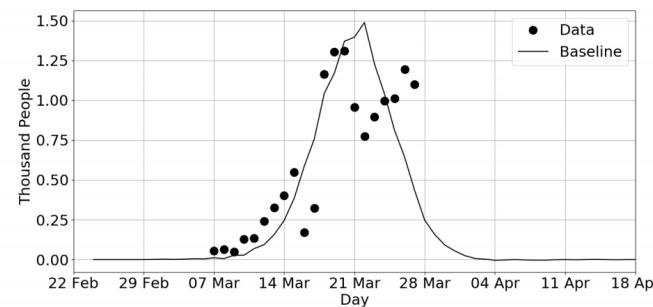


FIG 4: Comparison of predicted new infections to data for 65+ demographic for the period 22 Feb to 18 April 2020.

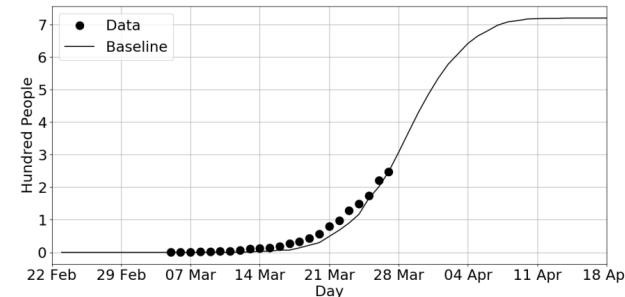


FIG 5: Comparison of predicted COVID-19 deaths to data for the period 22 February to 18 April 2020.

M. Marini et al. COVID-19 Epidemic in Switzerland: Growth Prediction and Containment Strategy Using Artificial Intelligence and Big Data, 2021, preprint on medRxiv