THE OECD FUNDSTAT INFRASTRUCTURE: DATA, TOOLS AND APPLICATIONS TO ANALYSE R&D DIRECTIONALITY

IntelComp Project: Information day on a new era of STI policy making with AI
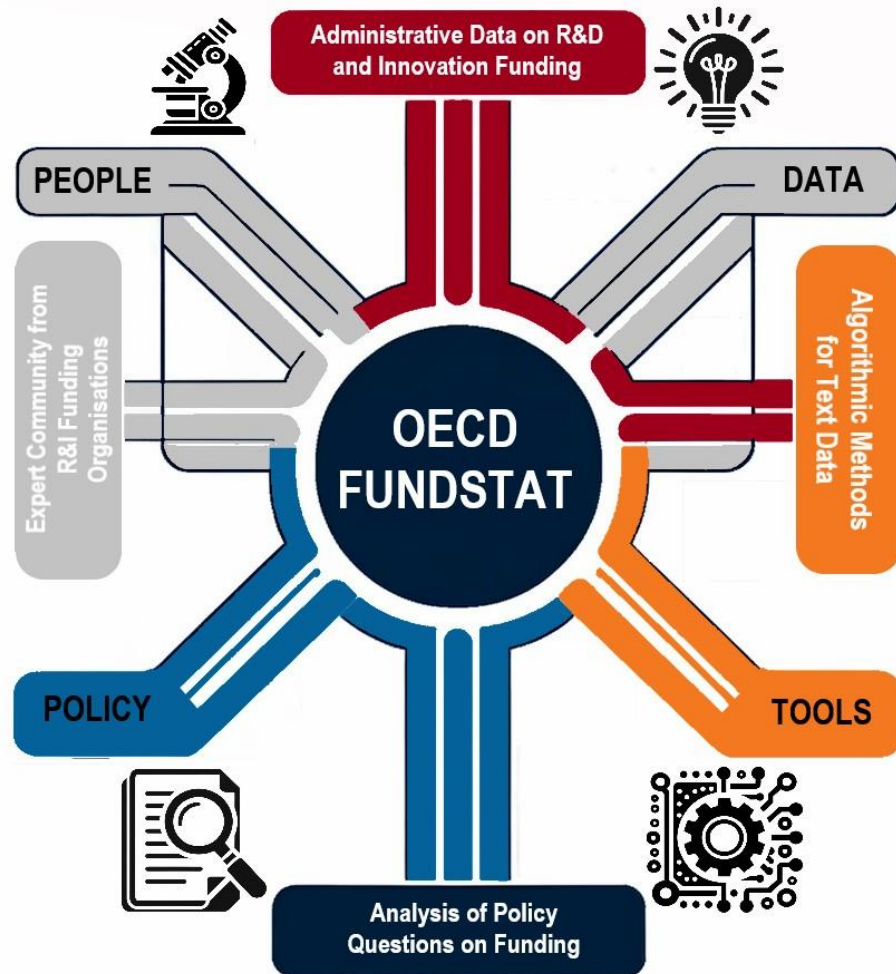Panel 2: The use of AI in STI Policy making
14th December 2023

Leonidas Aristodemou
OECD Directorate for Science, Technology and Innovation

(with Fernando Galindo-Rueda, Kuniko Matsumoto)

OECD Expert Group on the Management and Analysis of R&D and Innovation Administrative Data

- Established by the Working Party of National Experts on Science and Technology Indicators **(NESTI)** in 2021
- Focus on **administrative data on R&D and innovation funding**
- Serves as a **forum for officials** in charge of data management and analysis within R&D and innovation funding bodies to **exchange practises, challenges and co-ordinate common approaches**
- Responsible for the **OECD Fundstat** Infrastructure.

OECD Fundstat: Data and Tools Infrastructure

- Comprises of **government-funded R&D projects**
- Consists of **27 (+20) different data sources from 13 (+5) OECD member countries and the EU**: AUS, AUT, BEL, CAN, FRA, DEU, JPN, LVA, NOR, SWE, CHE, GBR, USA and EU-EC, in **7 languages**
- **Data coverage** benchmarked against Government Budget Allocations for R&D (**GBARD**) is constantly being improved by incorporating more sources

Mapping of the R&D landscape across countries to facilitate **international comparisons** and **transparency**

Description of R&D awards and funding in the Fundstat database, 2019-21

| Fundstat	FY2019	FY2020	FY2021 (partial)[1]	Total
R&D project awards ("projects")	211,968	221,076	174,054	607,098
(Of which) R&D projects with missing funding information	9,665	11,412	9,256	30,333
R&D funding [USD million]	90,044	101,599	95,936	287,579
Mean project award [USD million per project][2]	0.45	0.48	0.58	0.50

Source: OECD analysis of Fundstat database, March 2023.

OECD Fundstat aspired data model

3
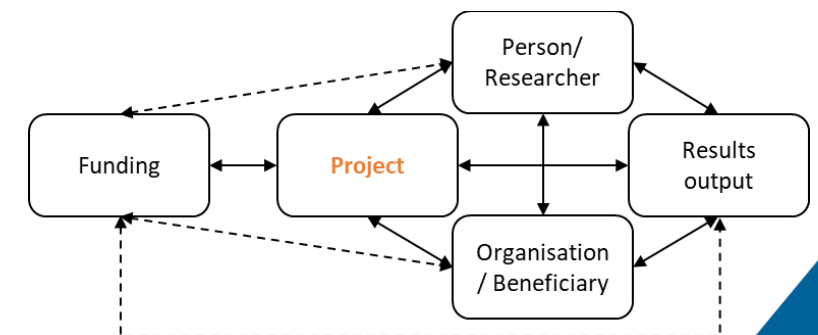
Two thematic proof of concepts: AI R&D and COVID-19 R&D
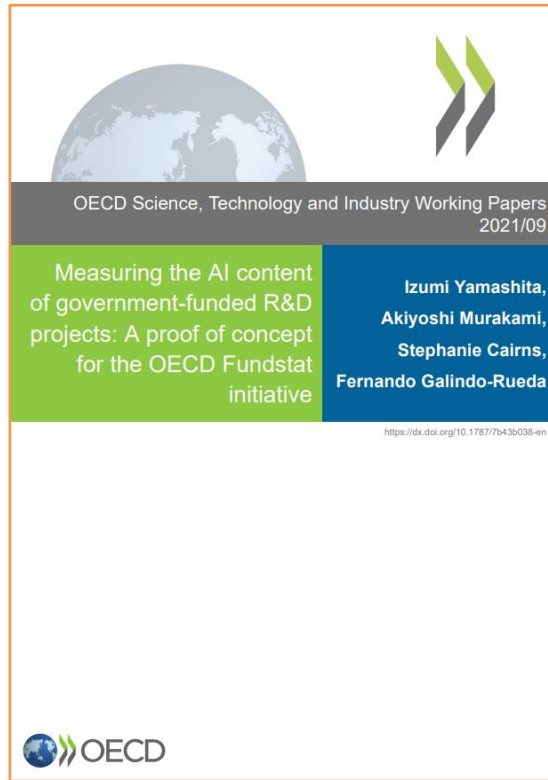
Yamashita, I., et al. (2021), "Measuring the AI content of government-funded R&D projects: A proof of concept for the OECD Fundstat initiative", OECD Science, Technology and Industry Working Papers, No. 2021/09, OECD Publishing, Paris, https://doi.org/10.1787/7b43b038-en

Aristodemou, L., et al. (2023), "Measuring governments' R&D funding response to COVID-19: An application of the OECD Fundstat infrastructure to the analysis of R&D directionality", OECD Science, Technology and Industry Working Papers, No. 2023/06, OECD Publishing, Paris, https://doi.org/10.1787/4889f5f2-en

Next phase:

- Enhance the OECD Fundstat to increase **country and database coverage**

- Focus on measuring the contribution of science and innovation to the **Green Transition**

4

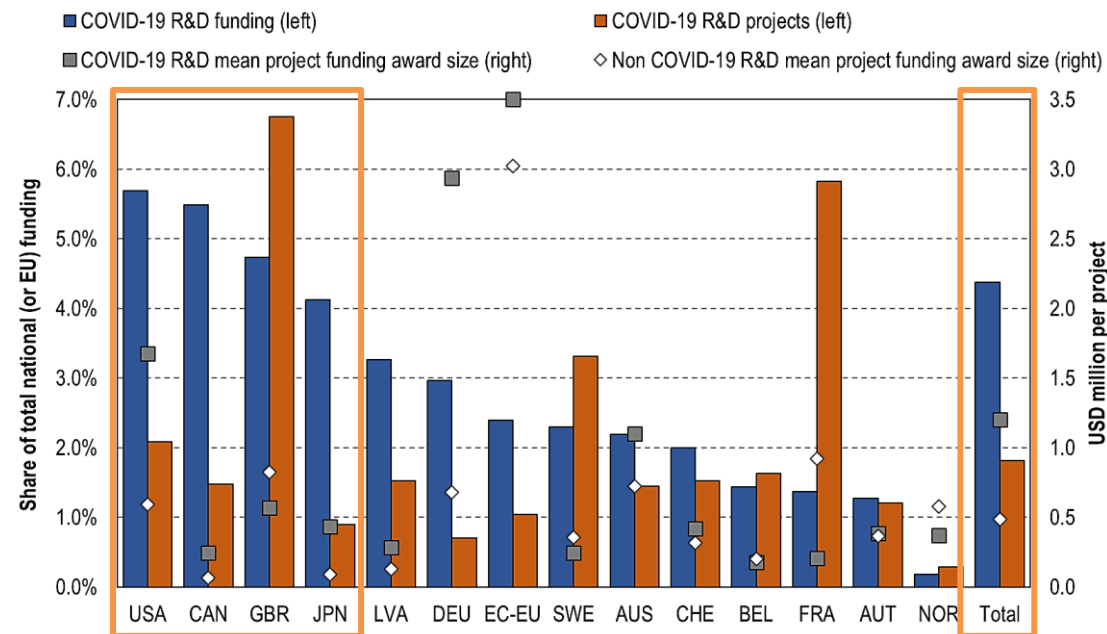# OECD Fundstat to measure how much government support for COVID-19 R&D?

- COVID-19 pandemic was a **major challenge**, with **767 million cases** and **6.9 million deaths** (May 2023)

- **COVID-19 R&D** identification:
  - Key term identification, expansion, and extraction
  - Identification of COVID-19 contextual projects (**"covidwashing"**)
  - Identification of projects with limited content

- **11,886 projects** accounting for over **USD 12.5 billion**

COVID-19 R&D represents 4% of the total funding and 2% of projects in Fundstat

Source: OECD analysis of Fundstat database, March 2023.

Retention of COVID-19 R&D projects following data cleaning, 2019-21

|	R&D projects	(Of which) projects with missing funding data	R&D funding [USD million]	Mean project award [USD million per project][1]
Total retrieved 'candidate' COVID-19 R&D projects	16,346	2,823	16,604	1.23
(-) 'Candidate' projects with contextual descriptions[2]	2,378	429	2,605	1.34
(-) 'Candidate' projects with limited content[3]	2,082	980	1,411	1.28
Total retained COVID-19 R&D projects	11,886	1,414	12,588	1.20

Source: OECD analysis of Fundstat database, March 2023.

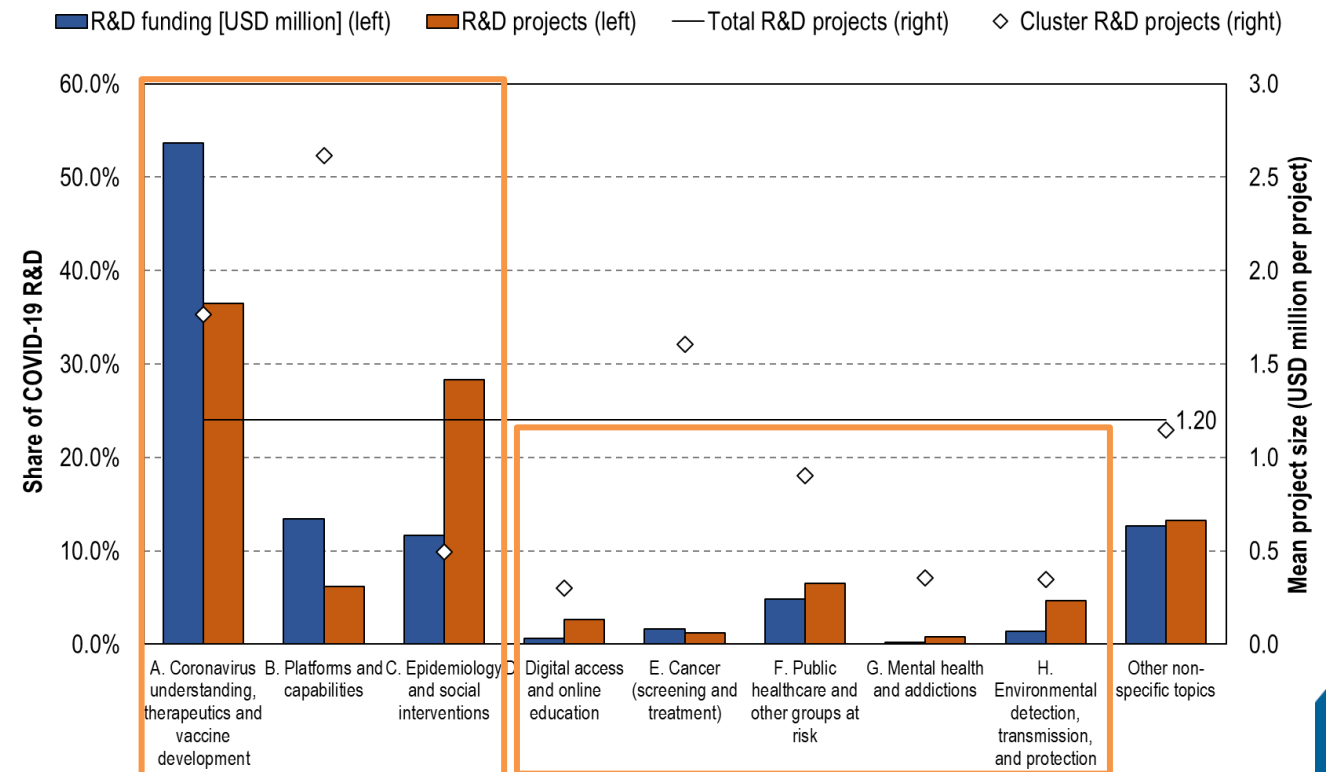# COVID-19 R&D projects and their dominant high-level clusters (machine classification)

Emergence of distinct topics from the Fundstat data

Source: OECD analysis of Fundstat database, March 2023.
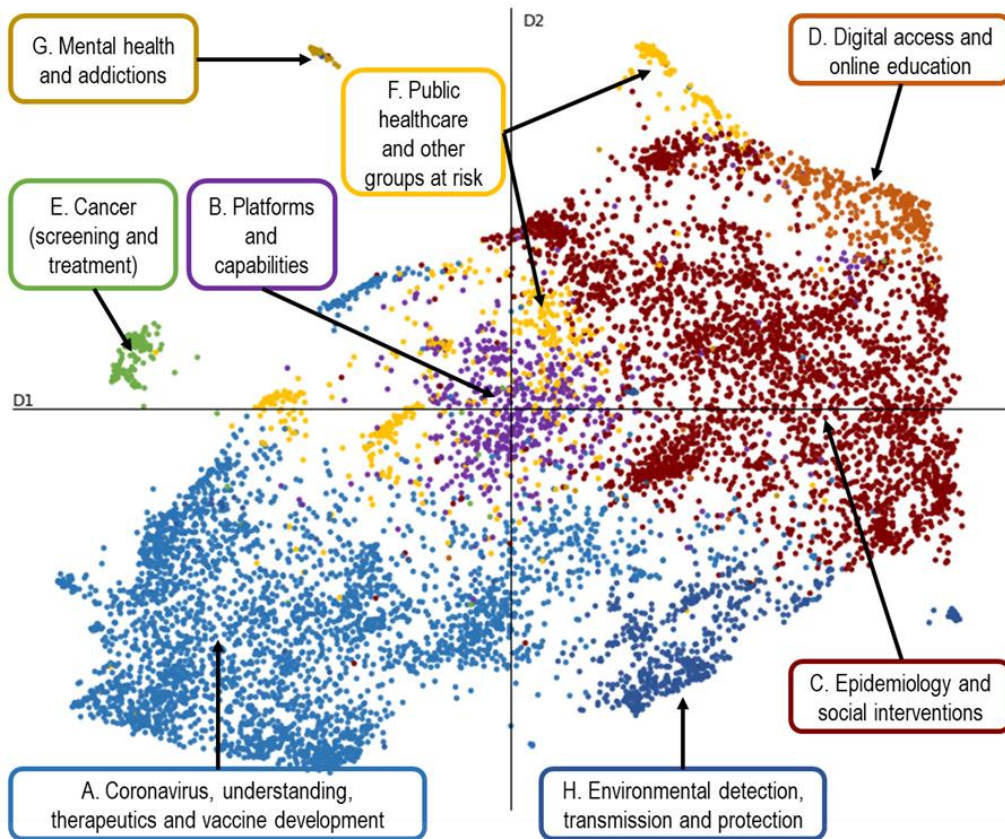
R&D funding and project allocation to high-level clusters (machine classification)

*Source: OECD analysis of Fundstat database, March 2023.*

Triangulation with other data sources and combining machine and human classifications

Funding allocation according to WHO categories from UKCDR & GloPID-R COVID-19 tracker
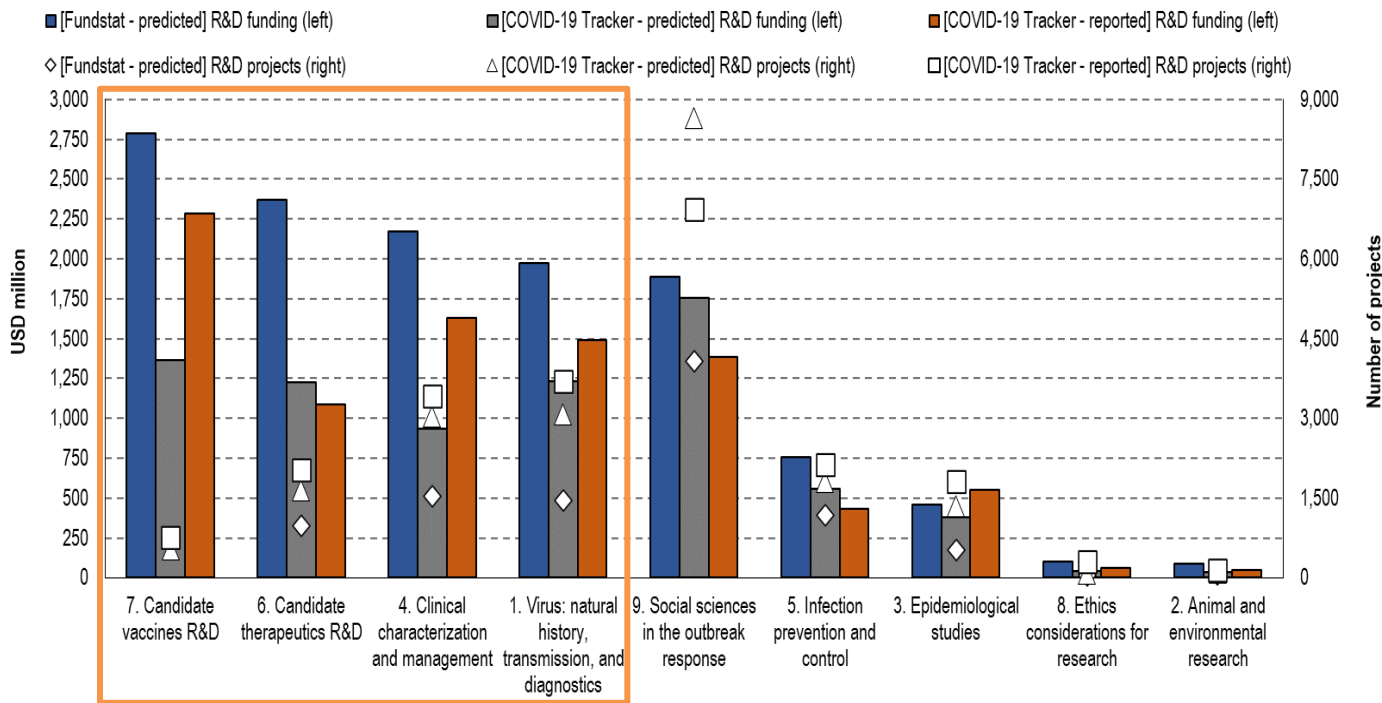
Source: OECD analysis of Fundstat database, March 2023.

Combining machine and human classifications to understand funding allocations
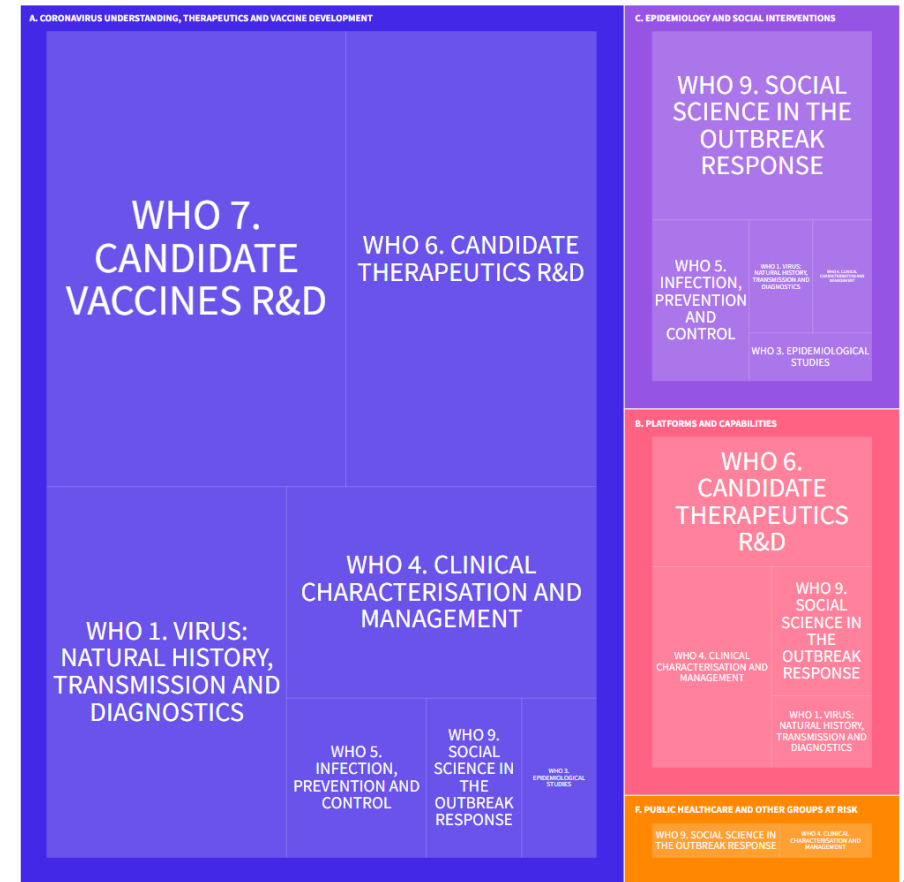
Notes: Excludes R&D funding less than USD 100 million, and other non-specific topics. Created with flourish.studio.

Thank you

OECD Secretariat
mariad@oecd.org

Leonidas Aristodemou
leonidas.aristodemou@oecd.org