Hackathon on data science for STI policy

A STIP Data Lab and OECD-TIP event

Kick-off event
23-24 May 2022

The STIP Data Lab (and more broadly STIP Compass) has benefited from H2020 grant 101018243

European Commission

OECD

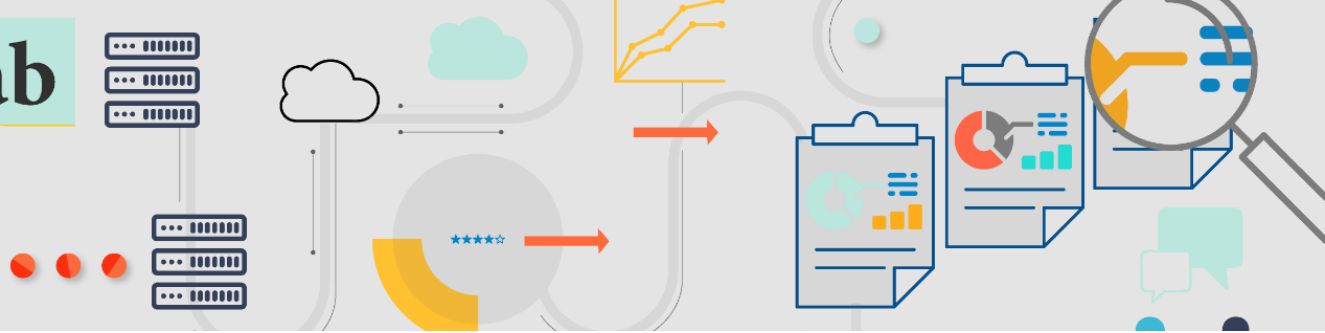
Agenda

- **Welcome** and structure of the hackathon (15')
- Presentation of the **datasets** (15')
- **Round table introduction** of the teams (one speaker per team, 1-2' per team) and team leaders
- Introduction and discussion of the **policy questions** (40’)
- **Logistical issues** and next steps (5')
- Other questions and answers (5')

Welcome and structure of the hackathon

STIP Data Lab

Analysis
Use cases
Stories



- Launched in 2020 to undertake and support **various types of analysis** of the **STIP Compass policy database**, including close to 7000 policy initiatives from 57 countries and the European Union (more on this dataset later).
- It aims to **build a community** of "lead users" of the database, involving policy makers, analysts and academics in these analyses.
- To do so, it brings forth **a new suite of tools and case studies** aiming to help leverage the data gathered by STIP Compass.
- It also **builds new partnerships**
  - With specific "volunteer" countries to produce on-demand analyses.
  - With the broader STI policy community through OECD internships, joint studies hackathons.

Working Party on Innovation and Technology Policy (TIP)

- TIP provides advice to OECD member countries to **improve STI policy-making** since 1993
- TIP is a **forum for member countries** to exchange information and best practices in the field of STI policy
- TIP is also involved in **country-specific activities** such as reviews of national STI systems
- 2021-22 project “Collaborative transitions” explores the role of STI policy in transitions with a focus on the role of co-creation.

Key questions

- What **transition goals and STI policies** are countries setting in response to the COVID-19 shock?
- What is essential to **boost collaborative research** and innovation in times of crisis and beyond?
- How can **alternative approaches, and tools** be used effectively by policy?

Semantic Analysis & the TIP

TIP@50 (2018)

Demonstration of semantic analysis

Collaborative Transitions Project (2021-22)

Application of semantic analysis & exploration of tools and approaches for STI policy

Semantic Analysis Guidelines Group

Build on the TIP's progress to promote **meaningful** and **efficient application of semantic analysis tools** for STI policy analysis.

Structure of the hackathon

- **Today's kick-off**: each team will receive and discuss a policy question from a policymaker or policy analyst to tackle during the hackathon.
- **Next two weeks**: allow at least two full days (equivalent) of work time, ideally scheduling two sprint meetings with your team leader.
- **Closure event** (7 June, 3:00pm-5:00pm CEST): Presentation and discussion of results.

In addition, we will be organising a **separate debriefing seminar** where participating teams will be able to elaborate and exchange on their technical choices and experiences working during the hackathon.

Presentations of the datasets

Datasets made available for this hackathon

Use at least one of these two:

- **TIP STI strategies database**, a text corpus including 314 STI policy strategy documents (several million words overall) from across 24 OECD countries.
- **STIP Compass policy database** (ed. 2021), made up of close to 7000 initiatives from 57 countries and addresses all main areas of STI policy.

Optionally (and where useful), a third data source can be used:

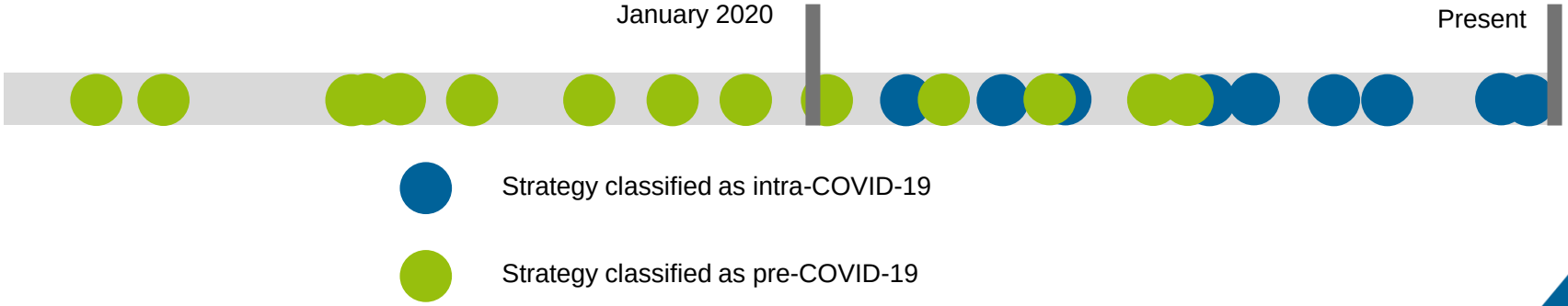
- **The STI.Scoreboard dataset**, containing over 400 unique STI indicators, drawing on the very latest, quality assured statistics from OECD and partner organisations.

TIP database of STI policy strategies

Data collection

1. Strategy documents were found through **desk research** and **classified** as “intra-COVID-19” or “pre-COVID-19” based on their date of publication

2. National delegations were consulted to **review the periodization** of documents and provide corrections as well as additional documents to **ensure representativeness**

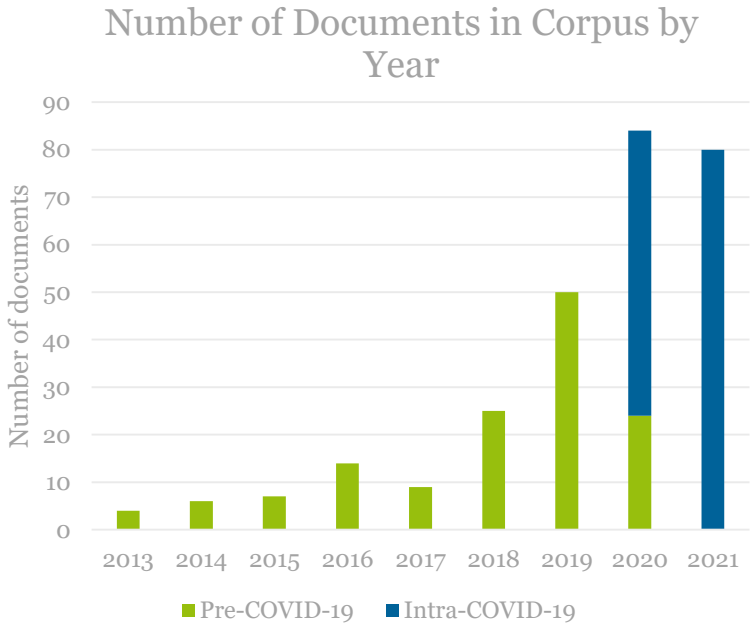


TIP database of STI policy strategies

Coverage and document types

314 documents
from **24** countries

- A National STI strategies
- B Other national strategies implicating STI
- C STI ministry and agency strategies
- D Key pieces of legislation



TIP database of STI policy strategies

Metadata variables

Country name

Publication period

Original document title

country <chr>	year <dbl>	period <chr>	doc_id <chr>	title <chr>
Austria	2020	intra-covid	AT_AC1	FTI-Strategie 2030
Austria	2021	intra-covid	AT_B1C1	Österreichischer Aufbau- und Resilienzplan 2020-2026
Austria	2020	intra-covid	AT_B1C2	Digitaler Aktionsplan Austria
Austria	2016	pre-covid	AT_B1P1	Open Innovation Strategie für Österreich
Austria	2020	intra-covid	AT_B2C1	FTI-Strategie Mobilität
Austria	2021	intra-covid	AT_B2C2	Strategie der Bundesregierung für Künstliche Intelligenz

Year of publication

Document ID
First letter after _ denotes the document type
Second letter after _ denotes publication period

TIP database of STI policy strategies

Text variables

Original text of the STI policy strategy document

text_original

<chr>

FTI-Strategie 2030 Strategie der Bundesregierung für Forschung, Technolo
Österreichischer Aufbau- und Resilienzplan 2020-2026 Wien, 30. April 202
Digitaler Aktionsplan Austria Z IE L E ,\tL E I T L INIE N\t&\tP R INZ IP IE N In
Open Innovation Strategy for Austria Open Innovation Strategy for Austria (
FTI-Strategie Mobilität Innovationen in und aus Österreich für ein klimaneu
Strategie der Bundesregierung für Künstliche Intelligenz Artificial Intelligen

Google translated text of the document

text_translated

<chr>

"x" "1" "FTI - Strategy 2030 Federal government strategy for research, technolo
"x" "1" "Austrian development and resilience plan 2020-2026 Vienna, 30. April
"x" "1" "More digital Action plan Austria Z IE L E, L E I T L INIE N & P R INZ IP IE
"x" "1" "Open Innovation Strategy for Austria Open Innovation Strategy for Aust
"x" "1" "RTI - Mobility Strategy Innovations in and from Austria for a climate-ne
"x" "1" "german Austria intra-covid 2021 B2 Strategy of the federal governmen

Cleaned text of the document
(punctuation, numbers, symbols removed; lemmatization; no stopwords; n-grams)

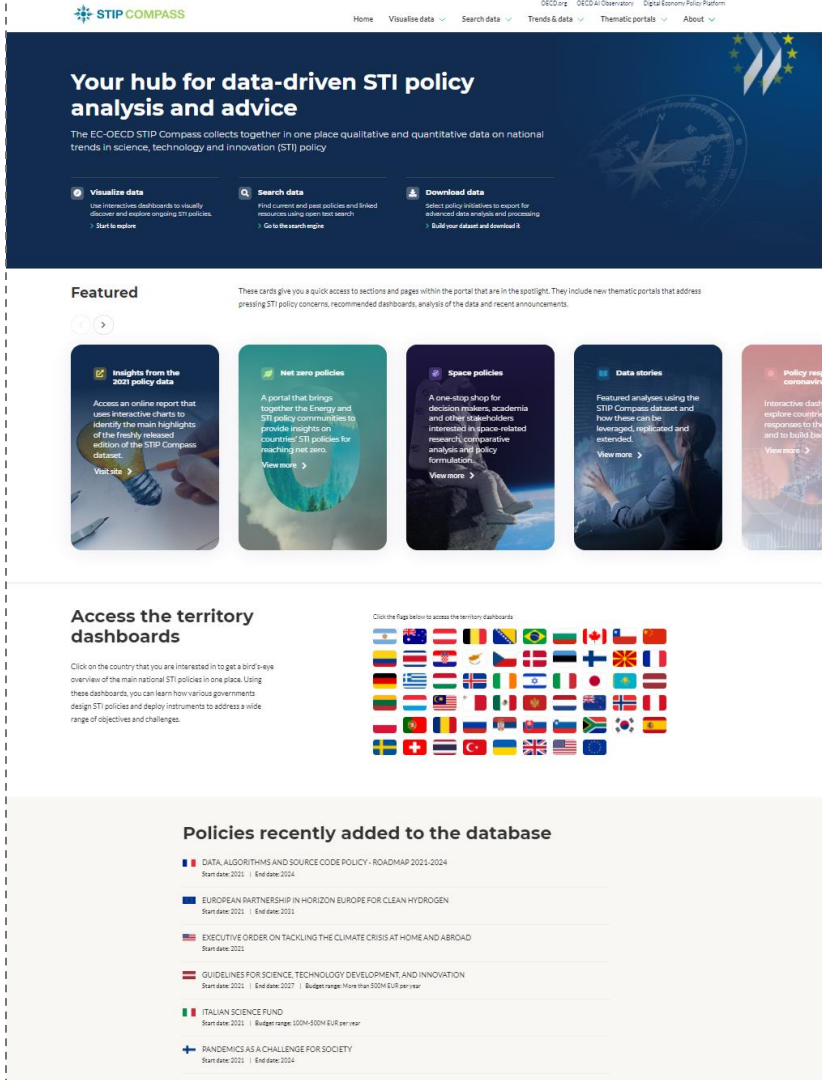
text_clean

<chr>

government_strategy research technology innovation government_strategy r
development resilience_plan content objective coherence plan measure deve
digital plan content future crisis-proof innovative manner topic digital actio
great strength country innovative capability people company research_instit
climate-neutral mobility_system imprint owner publisher editor project_mar
intra-covid federal_government artificial_intelligence imprint owner publish

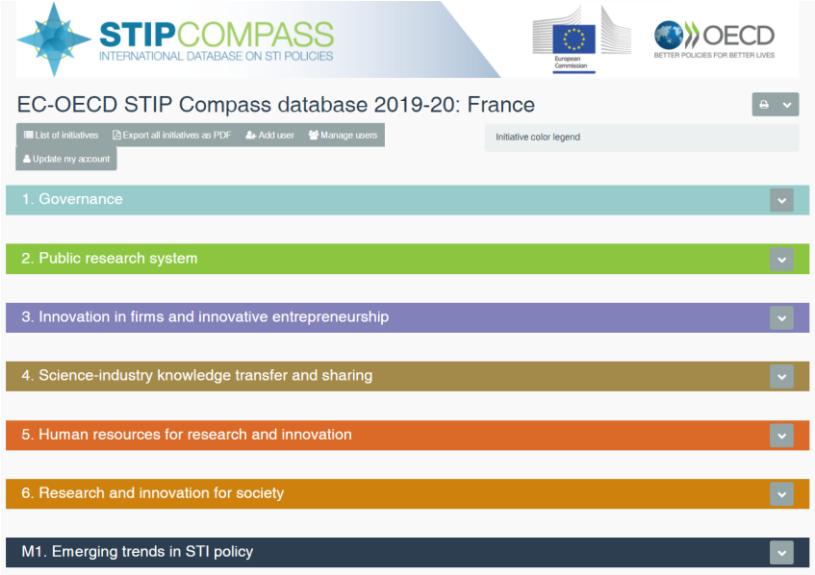
STIP COMPASS

- An **EC-OECD joint initiative** to collect together in one place qualitative and quantitative data on national trends in **STI policy**.
- It uses a **semantic database** for storing, accessing and linking various kinds of data (policies, statistics, publication metadata, twitter feeds…).
- It is **publicly available** online at stip.oecd.org.



The EC-OECD STIP Survey

- The **main data source** for STIP Compass.
- Data collection is firmly **structured** by **taxonomies** to characterise **policy initiatives**, making responses more **comparable** and **facilitating analysis**.
- Conducted **every two years**, it aims to collect data on **national STI policy initiatives**.
  - The survey is distributed across ministries and funding agencies by national contact points, designated mainly by delegates of our Committee for Scientific and Technological Policy (CSTP).
  - The survey is run for 3 months. Thereafter, it is re-opened as a monitoring tool that countries can continue to use to keep the data up-to-date.



“Policy initiative” as the unit of reporting

Defined as a public action that:

i. aims to achieve one or several public **policy goals**

ii. is expected to **modify or frame the behaviours** of actors and stakeholders

iii. is **implemented** with a minimum time horizon or on a continuous basis (i.e. not as a one-off “event”).

Characterised by taxonomies

- Basic fields for **textual data** (name, short description, objectives, background)
- A **policy instrument** taxonomy built based on Lit. review and OECD workshops
- A taxonomy for **target groups** and list of **budget ranges**
- An **STI thematic taxonomy** that structures the questions of the STIP Survey

Policy instruments taxonomy v2021

Governance

- Strategies, agendas and plans
- Creation or reform of governance structure or public body
- Policy intelligence (e.g. evaluations, benchmarks and forecasts)
- Formal consultation of stakeholders or experts
- Horizontal STI coordination bodies
- Regulatory oversight and ethical advice bodies
- Standards and certification for technology development and adoption
- Public awareness campaigns and other outreach activities

Direct financial support

- Institutional funding for public research
- Project grants for public research
- Grants for business R&D and innovation
- Centres of excellence grants
- Procurement programmes for R&D and innovation
- Fellowships and postgraduate loans and scholarships
- Loans and credits for innovation in firms
- Equity financing
- Innovation vouchers

Indirect financial support

- Tax or social contributions relief for R&D and innovation
- Tax relief for individuals supporting R&D and innovation
- Debt guarantees and risk sharing schemes

Collaborative infrastructures (soft and physical)

- Networking and collaborative platforms
- Dedicated support to research infrastructures
- Information services and access to datasets

Guidance, regulation and incentives

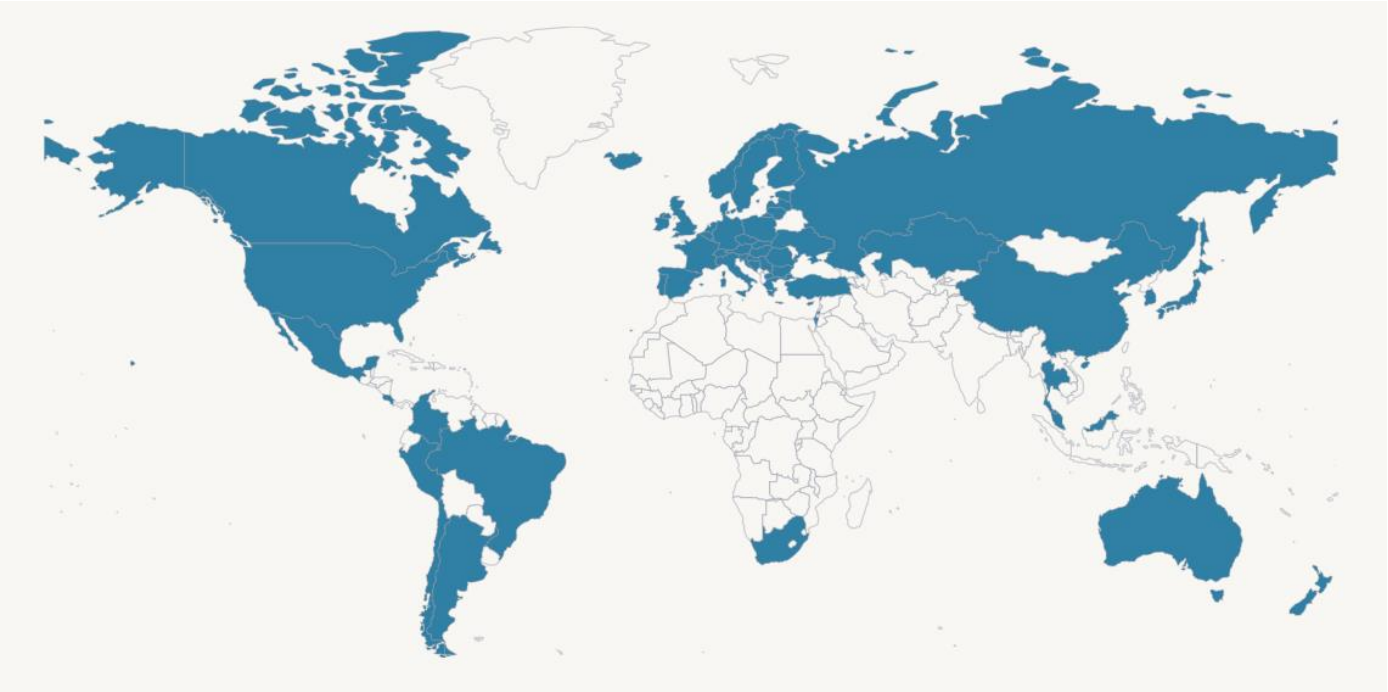
- Technology extension and business advisory services
- Science and technology regulation
- Labour mobility regulation and incentives
- Intellectual property regulation and incentives
- Science and innovation challenges, prizes and awards

Science and technology regulation instrument

Facet	Facet choices
Objective	Market regulation (e.g. antitrust law)
	Enable technology/innovation (e.g. interoperability standards)
	Risk mitigation (e.g. consumer and social protection)
	Regulate the delivery of public services (e.g. requirements in procurement, education)
	Promote research integrity
	Protect public values
Challenge(s) addressed	Risks to human safety
	Environmental sustainability
	Privacy protection
	Social disruption (e.g. job insecurity)
	Unethical practices (e.g. discrimination)
	Security (e.g. dual-use technologies)
	Limited competition (e.g. monopolies, oligopolies)
	Other

	Formal law or regulation
Type(s) of regulation	International agreement
	Self-regulation (e.g. codes of conduct, scientific advice, standards)
	Regulatory experiments (e.g. sandboxes)
	Other
Regulatory approach	Technology-based regulation (e.g. moratoria, standards of use)
	Performance or output-based regulation (e.g. safety thresholds)
Level of governance	Local
	Regional
	National
	International
Approach to monitor compliance	The regulator develops and maintains technologies for data collection, transmission and/or analytics
	Regulated parties are incentivised to adopt monitoring technology that is not managed by the regulator
	Regulated parties are simply required to share compliance data (no regulator support)

+6600 policy initiatives and +9100 policy instruments collected from 57 countries + EU in the 2021 survey (published in Nov)



Taken together, the countries covered in the STIP survey/STIP Compass database account for an estimated 97% of global R&D

2021 STIP survey data

STIP Compass database indicators

	2021 edition	Δ2021-19
Number of users in the online questionnaire tool	1600†	+33%
Number of organisations managing policy initiatives	2100	+11%
Number of policy initiatives	6700	+18%
Updated initiatives (prefilled from prior edition)	3700	
New initiatives (created in this edition	2100	
Number of policy instruments	9100	+18%
Country coverage	57 + EU	-8%
Median number of policy initiatives by country	96	+13%
Share of policy initiatives including budget data	77%	+6 p.p.

2021 STIP survey data

STIP Compass database indicators

	2021 edition	Δ2021-19
	1600†	+33%
ves	2100	+11%
	6700	+18%
ion)	3700	
tion	2100	
Number of policy instruments	9100	+18%
Country coverage	57 + EU	-8%
Median number of policy initiatives by country	96	+13%
Share of policy initiatives including budget data	77%	+6 p.p.

25 OECD countries reported more than 100 policies

6 OECD countries reported less than 70 policies

Dataset limitations

To be accounted for when analysing the data

i. The data consists of **self-reported descriptions** of national STI policies given by ministries, funding agencies and other public bodies that manage these initiatives.

ii. Policy initiatives **vary in scope and scale**, which can make aggregating them in groups problematic.

- What **meaning** should be given to the fact that one country uses **twice as many** policy initiatives than another to address the same policy theme?

iii. Countries also vary in the scope and scale of their **reporting**.

- The dataset is **useful to learn about existing types of STI policies** in a wide variety of topics, but it **is not representative of the actual population** of policies in countries.

Weighting policy initiatives by budget

One way to scale policies is using the budget range as a weight.

Budget range (in Million Euros)	Budget weight coefficient
Don't know / Not applicable	0 (no weight)
Less than 1 M	0.005
1 M – 5 M	0.01
5 M – 20 M	0.05
20 M – 50 M	0.2
50 M – 100 M	0.5
100 M – 500 M	1
More than 500 M	5

$$Nbw_i = \sum_{j=1}^{N_i} (1 + \text{budget weight coefficient for initiative j})$$

The STI.Scoreboard platform

- Provides policy makers, analysts and the public at large with a resource to **retrieve STI statistical indicators** across OECD countries and several other economies.
- It contains over **400 unique indicators** on a wide range of topics such as research and development, science, business innovation, patents, education and the economy.
- It draws on the very **latest, quality assured** statistics from OECD and partner international organisations.
- Accessible through a dedicated API using SDMX queries.

Accessing the hackathon's datasets

Quick access instructions are given in the main page of the hackathon's GitHub repository (link below). It provides the necessary guidance, codebooks, Python/R code snippets and examples.

We will be sharing this link in the chat.

https://github.com/STIPlab/hackathon

Round table introductions
of teams and team leaders

Kick-off day 2

Team	Team leader	TL organisation	TL email
Fraunhofer Institute	Joseba Sanmartín	Spanish Foundation for Science and Technology (FECYT)	Joseba.Sanmartin@fecyt.es
SPRU, University of Sussex Business School	Daniel Ferreira	Portuguese national funding agency for science, research and technology (FCT)	Daniel.Ferreira@fct.pt
University of Tokyo	Philippe Larrue	OECD	Philippe.LARRUE@oecd.org

Kick-off day 1

Team	Team leader	TL email
Aalborg University	Tiago Santos Pereira	tsp@ces.uc.pt
Georgia Institute of Technology	Caroline Paunov	Caroline.PAUNOV@oecd.org
Korea Advanced Institute of Science and Technology	Alan Paic	Alan.paic@oecd.org
University of Turin	Laura Kreiling	Laura.KREILING@oecd.org

Introduction and discussion of the policy questions

In what ways do **STI strategies differ between countries** around the theme of sustainability (climate change) and, in particular, around the energy innovation system?

This question may be placed at the formulation stage of the policy cycle, when policy makers need to explore different courses of action and investigate alternative paths.

Fraunhofer Institute: Possible dataset slices

- TIP Strategies
  - Example keywords: sustainability; climate_change; net_zero; carbon_emissions; renewable_energy; transition; policy_strategy; …
- STIP Compass:
  - Green energy transitions (TH92, 249 policy initiatives); Strategies, agendas and plans (PI024, 1341 policies) which can be decomposed by themes, e.g.: STI Plan or strategy (TH13, 344 strategies) ;Public research strategies (TH18, 257 strategies); Business innovation policy strategies (TH30, 227 strategies); Transfer and linkages strategies (TH41, 102 strategies); STI human resource strategies (TH50, 148 strategies); Research and innovation for society strategy (TH58, 192 strategies)
- STI.Scoreboard:
  - Air emissions per GDP; Emissions per capita; Energy productivity/intensity; Climate change-related taxes; R&D budget for Planet and Infrastructure (SDG); Renewable energy sources RD&D budget, …

To what extent is it possible to characterise typologies of policy proposals on the theme of **scientific employment and research careers?**

Possibly considering,

- The **policy objectives** expressed in the databases:
  - training of researchers; researchers' labour market; scientific employment; job mobility; research careers; evaluation of researchers, etc.
- The diversity of **socio-economic contexts** involved:
  - geopolitical variables; economic structure; socio-demographic profile; data on education and qualification, institutional context, etc.
- The diversity and specificity of **each STI system**:
  - HR in Science and Technology activities; public and private investment in R&D; innovative profile of the country/region; digitalization indicators; Technological Balance of Payments, etc.

SPRU: Possible dataset slices

- TIP Strategies
  - Example keywords: research_career; higher_education; science; employment; university_employees; …
- STIP Compass:
  - STI human resource strategies (TH50, 240 policies); Research careers (TH53, 285 policies); Inter-sectoral mobility (TH44, 113 policies); Doctoral and post-doctoral research (TH52, 296 policies); International mobility of human resources (TH55, 313 policies); Gender balance and inclusiveness (TH54, 238 policies); Postdocs and other early-career researchers (TG11, 1556 policies); PhD students (TG12, 1207 policies); Established researchers (TG9, 1926 policies)
- STI.Scoreboard:
  - Share of adult population with a doctorate; Total researchers (in FTE) per thousand total employment; Various indicators on women participation in STI activities; (e.g. Percentage of women among researchers); International mobility of scientific authors (outflows, inflows, stayers); among others.

What information is provided in strategies about **policy implementation**, such as specific goals, timelines, budgetary commitments or policy actions and/or their governance/monitoring?

University of Tokyo: Possible dataset slices

- TIP Strategies
  - Example keywords: timeline, milestone, action_plan, budget_allocation, progress_evaluation, monitoring, …
- STIP Compass:
  - Strategies, agendas and plans (PI024, 1341 policies) which can be decomposed by themes, e.g.: STI Plan or strategy (TH13, 344 strategies) ;Public research strategies (TH18, 257 strategies); Business innovation policy strategies (TH30, 227 strategies); Transfer and linkages strategies (TH41, 102 strategies); STI human resource strategies (TH50, 148 strategies); Research and innovation for society strategy (TH58, 192 strategies).

Aalborg University – Tiago Santos Pereira REVISED

To what extent can we identify distinct instruments and country policy goals that reflect a novel **co-creation** approach vs. a more traditional knowledge-transfer approach?

- You could use keywords to delineate a 'novel co-creation' approach from a 'traditional knowledge transfer' approach, e.g.
  - **Co-creation**: 'joint', longer time period (e.g. institutional), involvement of different actors (e.g. including civil society)
  - **Knowledge transfer**: based on contracts, short-term, university-industry based
- STIP Compass has a thematic portal dedicated to this topic, which offers further insights on these definitions: https://stip.oecd.org/knowledge-transfer/
- Analysis could be extended to the Covid Watch policy data to explore whether the Covid-19 response initiatives are particularly characteristic of a co-creation approach?

Aalborg University: Possible dataset slices

- TIP Strategies
  - Example keywords: knowledge_transfer, co-creation, commercialisation, technology_transfer, clusters, intellectual_property
- STIP Compass:
  - Collaborative research and innovation (TH42, 537 policies); Commercialisation of public research (TH, 314 policies); Cluster policies (TH47, 278 policies); Intellectual property rights in public research (TH46, 163 policies); Intersectoral mobility (TH44, 113 policies); Transfer and linkage strategies (TH41, 183 policies)
- STI.Scoreboard:
  - HERD/GOVERD financed by firms; innovative firms cooperating with HEIs and PRIs; …

Georgia Tech – Caroline Paunov

To what extent are countries' **green transition goals**, as set out in their strategies, reflected in their STI policies?

Georgia Tech: Possible dataset slices

- TIP Strategies
  - Example keywords: energy_transition; renewable_energy; hydrogen; battery_production; mobility_transition; …
- STIP Compass:
  - Green energy transitions (TH92, 249 policy initiatives); Stimulus for STI systems (TH96, 212 policies); Building more resilient societies and economies (TH98, 73 policies); Research and innovation for society strategy (TH58, 282 policies)
- STI.Scoreboard:
  - Government support by SDG: Planet and Infrastructure; Health and Medical Sciences; Industry and Knowledge; Security

KAIST – Alan Paic

Can we characterise typologies of policies in support of making **research data from publicly funded research openly accessible** and reusable to the largest extent possible?

In particular, policies that:

- Foster and support open access by default to research data and other research-relevant digital objects from public funding (read more);
- Develop infrastructure and services to facilitate the accessibility of research data and other research-relevant digital objects from public funding within and across scientific domains and disciplines.

KAIST: Possible dataset slices

- TIP Strategies
  - Example keywords: open_data; data_access; open_access; research_data; research_infrastructure; transparency; …
- STIP Compass:
  - Open science and enhanced access to publications and research data (TH10, 328 policies); Information services and access to datasets (PI023, 398 policies); Open science measures in response to Covid-19 (THc41-44, 108 policies - available separately here)

University of Turin – Laura Kreiling

Can we characterise typologies of strategies and/or policies that aim to **foster the responsible development of emerging technologies** for societal benefit? Do such policies focus on specific technologies and, if so, which ones?

- responsible development of technologies
  - …aims to better align both the innovation process and its outcomes with the values, needs and expectations of society
  - …means taking into account effects and potential impacts on the environment and society
- what are emerging technologies?
  - referred to as: *"key", "emerging", "converging" or "critical" technologies*
  - largely unrealized or ongoing: development and/or application
  - characteristics: radical novelty, potential to change the status quo, uncertainty, ambiguous
  - examples: nanotechnology, robotics, synthetic biology, AI, quantum computing

University of Turin: Possible dataset slices

- TIP Strategies
  - Example keywords: technological_change; impact_assessment; societal_change; data_regulation; artificial_intelligence; social_impact; responsible_development …
- STIP Compass:
  - Ethics of emerging technologies (TH89, 239 policies); Science and technology regulation (PI032, 156 policies); Regulatory oversight and ethical advice bodies (PI033, 142 policies); Research and innovation for society strategy (TH58, 282 policies)

Logistical issues and next steps

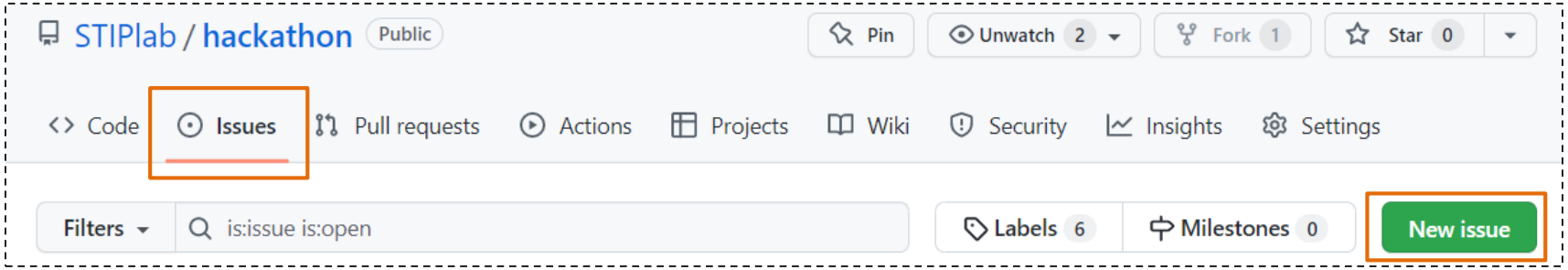
Logistical issues and next steps

- Teams may now **contact their team leaders via email** to further clarify the policy question, if needed.
- During the next couple of weeks, you should **arrange two 30-45' sprint meetings with your team leader** to discuss progress and receive feedback.
- For the closure event on 7 June, you should **prepare a 10' presentation** that will be followed by a 5' Q&A. You will soon receive a calendar invitation.
- The OECD hackathon team will be able to **answer any questions via GitHub**. Today's slides can also be found there.

https://github.com/STIPlab/hackathon

How to ask your questions via GitHub to the OECD hackathon team

- You will need a GitHub account.
- Go to the repository's "Issues" tab and create a new issue:



- You'll need to **tag the issue using the following pre-defined labels** so that the right persons in the organising team are notified: Logistics, TIP strategies dataset, STIP Compass dataset, STI.Scoreboard dataset, Other.
- If you don't use any of these labels, **we won't be notified** of your question.

Other Q&A

Hackathon on data science for STI policy

A STIP Data Lab and OECD-TIP event

Kick-off event
23-24 May 2022

The STIP Data Lab (and more broadly STIP Compass) has benefited from H2020 grant 101018243