AN INTRODUCTION TO THE REITER PROJECT
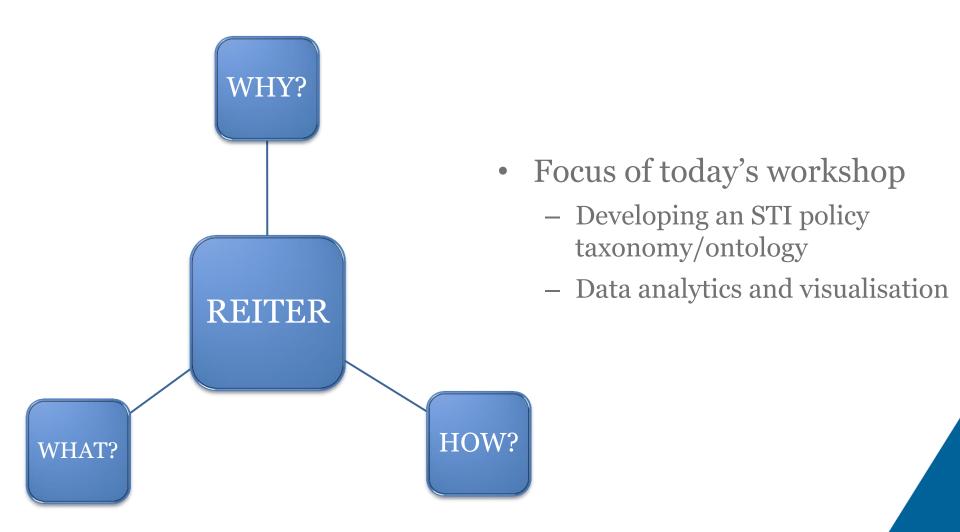
Ana Nieto, DG RTD (ana.nieto@ec.europa.eu)
Michael Keenan, OECD (michael.keenan@oecd.org)

Workshop on Semantic Technologies and Semantic Web: Structuring Data for STI Policy Analysis

Brussels, 19 June 2017

Outline

WHY?

REITER

WHAT?

HOW?

- Focus of today's workshop
  – Developing an STI policy taxonomy/ontology
  – Data analytics and visualisation

Support to countries' mutual policy learning: some of the challenges we face

- Collecting quality data on countries

- Data is in silos

- The quality of the data we're collecting needs to be improved through better survey questions and tools

- We're not exploiting the data we're collecting. In practice, the data remains closed and vastly under-utilised

- Improved search and discovery capabilities would help data analysts enormously

- The data cycle, i.e., the process of collecting, cleaning and using the data, is too slow

- All in all, we need to lighten the load on analysts and countries in collecting and exploiting the data
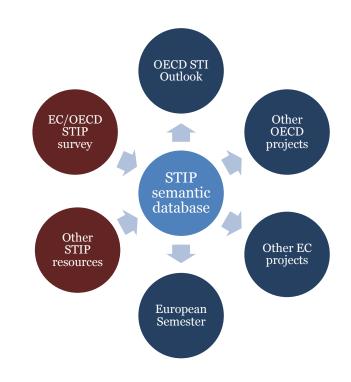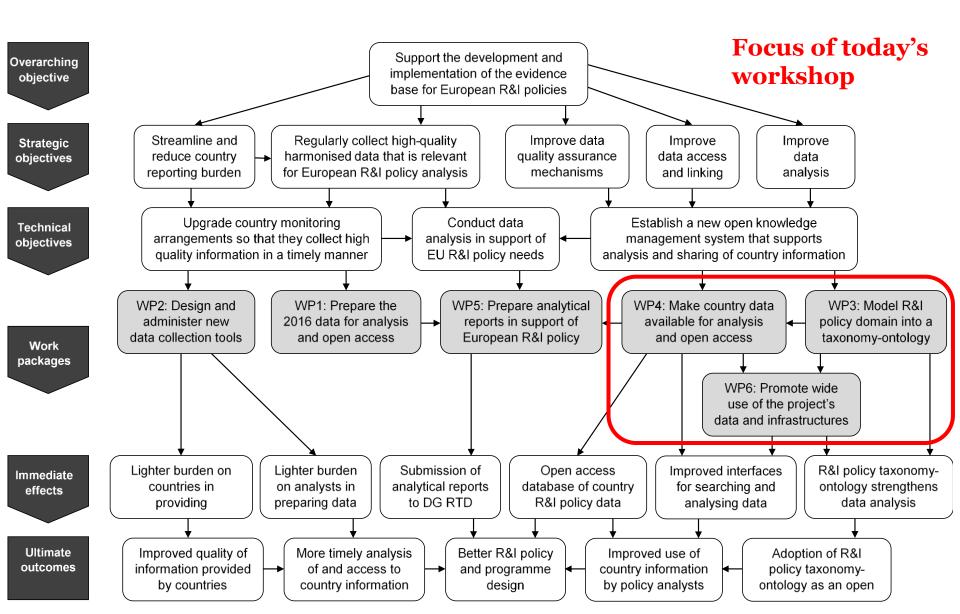
New project

Support to the development and analysis of a research and innovation policy taxonomy and questionnaire (REITER)

- Joint EC-OECD project over 18 months (11/16 – 04/18)

- Main objectives:
  - Develop a new survey tool and use this to administer the next round of joint EC-OECD STIP survey(s)
  - Develop a new database / interface for semantically-linked data
  - Develop a taxonomy-ontology for linking data semantically
  - Deliver a series of analytical reports based on analysis of the database

The new infrastructure will support several projects, notably the OECD STI Outlook and European Semester

REITER's objectives, tasks and expected outcomes

Overarching objective

Strategic objectives

Technical objectives

Work packages

Immediate effects

Ultimate outcomes

Focus of today's workshop

Support the development and implementation of the evidence base for European R&I policies

Streamline and reduce country reporting burden

Regularly collect high-quality harmonised data that is relevant for European R&I policy analysis

Improve data quality assurance mechanisms

Improve data access and linking

Improve data analysis

Upgrade country monitoring arrangements so that they collect high quality information in a timely manner

Conduct data analysis in support of EU R&I policy needs

Establish a new open knowledge management system that supports analysis and sharing of country information

WP2: Design and administer new data collection tools

WP1: Prepare the 2016 data for analysis and open access

WP5: Prepare analytical reports in support of European R&I policy

WP4: Make country data available for analysis and open access

WP3: Model R&I policy domain into a taxonomy-ontology

WP6: Promote wide use of the project's data and infrastructures

Lighter burden on countries in providing

Lighter burden on analysts in preparing data

Submission of analytical reports to DG RTD

Open access database of country R&I policy data

Improved interfaces for searching and analysing data

R&I policy taxonomy-ontology strengthens data analysis

Improved quality of information provided by countries

More timely analysis of and access to country information

Better R&I policy and programme design

Improved use of country information by policy analysts

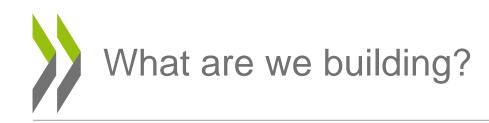Adoption of R&I policy taxonomy-ontology as an open

The project is innovative and therefore risky

- The timeframe is tight with many interconnected tasks performed in parallel

- The ability to deliver the required IT infrastructure to specification and on time

  Focus of today's workshop

  - Survey tool
  - Database with intuitive interfaces for search, analysis and visualisation
  - Open taxonomy-ontology for others to inspect and re-use
  - Seamless interconnectivity of these IT elements

- The semantic web tools actually delivering on their promise

- Countries accepting new reporting arrangements

  - New survey tool
  - New survey design
  - Adopting regular reporting routines

What are we building?

- A new semantic database for storing and analysing country STI policy information, with
  - enhanced data analysis capabilities and interoperability with other data sources
  - improved search and analysis interfaces that make country information more accessible to policy makers and analysts alike
    - these will utilise visualisation tools, dashboards and standard queries that will help unlock the data's analysis potential
    - they are expected to increase the data's utilisation in future policy analysis
- Implementing the FAIR guiding principles, the database will be freely open for anyone to use or link to
- OPEN ACCESS/OPEN DATA to have the greatest utility

The backbone of the system is an STI policy taxonomy/ontology

- The creation of an STI policy taxonomy/ontology is driven by the need to discover more in the qualitative data collected via the EC-OECD STIP survey

- Country information will be "semantically enriched" using the STI policy taxonomy/ontology

- It will allow analysts to gain more insights more quickly and avoid the time-consuming task of reading through large amounts of qualitative data

- Designing a taxonomy/ontology is an exercise based on an expert level of comprehension of the STI policy field . . .

- . . . but considerable further effort is required to move from the conceptual world of the STI policy analyst to the machine-readable world of the semantic web – requires guidance by taxonomists

- It will allow us to comply more easily with the FAIR guiding principles

- Open question: could we move towards an STI policy taxonomy/ontology "standard"? If so, what could this look like?
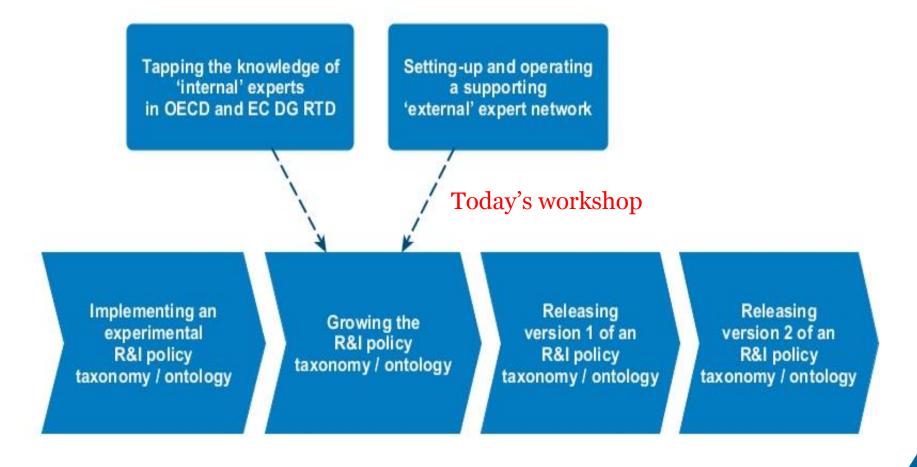
What we'd like to gain from today's workshop

- Raise awareness among the practitioner / research community of what we're doing in REITER and our ambition

- Invite critical feedback on REITER's activities from this community, particularly on its STI policy taxonomy/ontology work

- Explore ways to leverage related activities and progress of this community to move forward more quickly in REITER and the area more generally

- Contribute to structuring a community of practitioners, researchers and policy makers through an EC-OECD "community of practice" that we are launching today
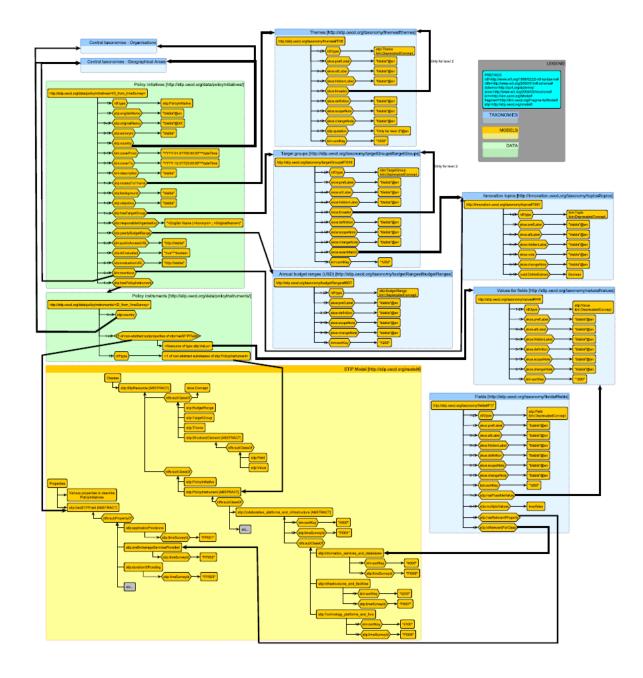
STI POLICY TAXONOMY/ONTOLOGY

Steps and timeline

Tapping the knowledge of 'internal' experts in OECD and EC DG RTD

Setting-up and operating a supporting 'external' expert network

Today's workshop

Implementing an experimental R&I policy taxonomy / ontology

Growing the R&I policy taxonomy / ontology

Releasing version 1 of an R&I policy taxonomy / ontology

Releasing version 2 of an R&I policy taxonomy / ontology

Nov 2017   Apr 2018

Developing the STI policy taxonomy-ontology

- Scope of the STIP taxonomy/ontology?
    - Determined largely by the queries we expect to have . . .
    - . . . and by the scope of the survey data we collect (and any other data sources we're able to link to)
- Building the STIP taxonomy/ontology
    - OECD analysts and taxonomists will lead in modelling the STI policy world
    - Limited potential to re-use existing schemas?
    - Inputs from the wider community?
- Two generations of taxonomy/ontology during the project
- The taxonomy-ontology will be available for inspection and download / linking through APIs

First starting point: the project's "data model", used to create the EC-OECD STIP survey and database search facets

Governance

- Horizontal policy coordination
- Strategic policy intelligence
- Responsible research and innovation
- Evaluation and impact assessment
- International STI policy governance
- International STI policy strategy and framework

Public research system

- Competitive research funding
- Non-competitive research funding
- Third-party funding
- Structural change of the public research system
- Large research infrastructures and equipment
- Open science and open access
- Exploratory and high risk research
- Interdisciplinary research
- Internationalisation in public research
- Research integrity and reproducibility
- Gender dimension in research

Innovation in firms and innovative entrepreneurship

- Financial support to business R&D and innovation
- Non-financial support to business R&D and innovation
- Support to innovative SMEs
- Targeted support to young innovative enterprises
- Foreign direct investment
- Access to innovative finance
- Entrepreneurship capabilities and culture
- Demand for innovation and market creation

Human resources for innovation

- STEM skills
- Doctoral and Postdoctoral
- Research careers
- Gender balance
- International mobility

Public-private knowledge transfers and linkages

- Collaborative research
- Intersectoral mobility
- Intellectual property rights in public research
- Cluster policies
- Commercialisation of public research results

Challenge and technology-oriented policies

- Innovation for health and healthcare
- Innovation for sustainable development
- Innovation for developing countries
- Science, technology and innovation culture
- Multi-stakeholder engagement

Taxonomy of themes

Taxonomy of target groups of policy intervention (v.0)

Researchers, students and teachers	Researchers
	Undergraduate and master students
	Secondary education students
	PhD students
	Teachers
Social groups especially emphasised	Women
	Disadvantaged and excluded groups
	Civil society
Capital and labour	Private investors
	Entrepreneurs
	Labour force
Research and education institutions	Higher education institutes
	Public research institutes
	Private research and development lab
Governmental entities	National government
	Subnational government
Firms by age	Firms of any age
	New firms
	Young firms (less than 5 years old)
	Established firms (5 years old or more)
Firms by size	Firms of any size
	Micro-enterprises
	SMEs
	Large firms
	Multinational enterprises
Intermediaries	Incubators, science parks or technoparks
	Technology transfer offices
	Industry associations
	Academic societies / academies

Taxonomy of policy instruments (v.0)

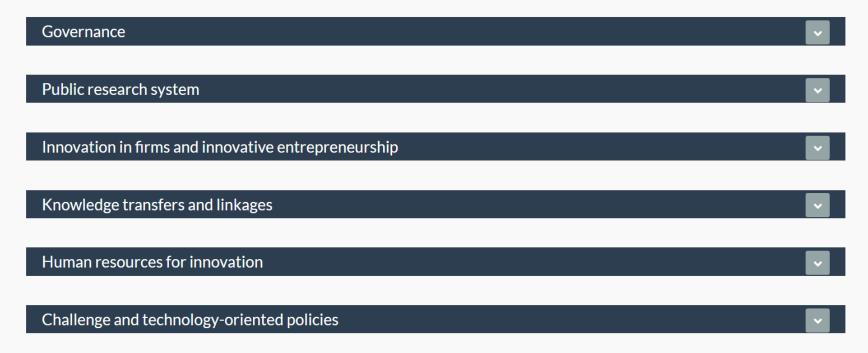|	
Direct financial support	Competitive grants for R&D and innovation projects
	Postgraduate loans, scholarships and fellowships
	Loans and credits for innovative firms
	Equity financing
	Block funding for public research
	Centres for excellence grants
	Procurement programmes for R&D and innovation
	Innovation vouchers
	Innovation prizes and awards
Indirect financial support	Tax relief
	Debt guarantees and risk sharing schemes
Non-financial support	Business innovation and technology advice
	Labour mobility incentives and regulation
Collaborative platforms and infrastructure	Collaborative platforms and fora
	Support to research infrastructures
	Information services and databases
Governance	National strategies, agendas and plans
	Public consultation of stakeholders
	Horizontal STI coordination bodies
	Performance contracts
	Intellectual property frameworks
	Standards and certification for innovation
	Public outreach activities

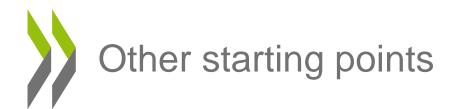# Prefilled survey is generated by the data model structuring the database

Other starting points

The Innovation Policy Platform

euroCRIS — Current Research Information Systems

casrai

VIVO — connect · share · discover

ORCiD

SAPIENZA — Università di Roma

RISIS — Research infrastructure for research and innovation policy studies

KNOW/\/\4K

- IPP tagging vocabulary of more than 1200 terms

- Existing taxonomies and ontologies

Expert advisory group

- Drawn from community of analysts and practitioners active in STI policy data

- Role: make a link to related activities and resources and provide feedback on the project's emerging findings

- First invitations sent today

- The group will also advise on the OECD Project "Digital Science and Innovation Policy and Governance (DSIP)"
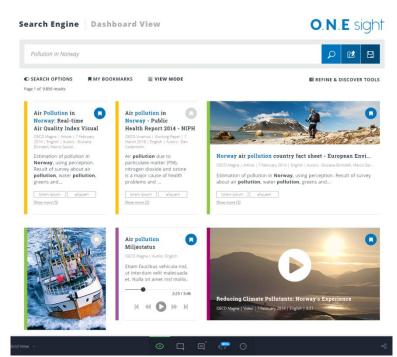
DATABASE AND USER INTERFACE
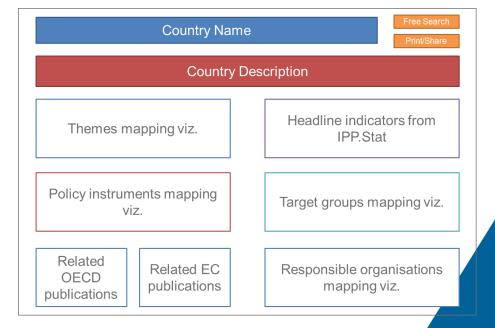
As an analyst, what I'd really like to have

- A search and discovery tool that gives me the information I need from the piles of country responses to the EC-OECD STIP survey

- Ways for me to gain a top-level overview of the data and the ability to drill down as I need

- A tool that produces attractive charts that I can play with and download for my reports

- Combinations of qualitative and quantitative data that provide context on the countries and themes I'm interested in

- Links to existing publications – internal and external – that relate to the themes and countries I'm interested in

We're using a MarkLogic database with bespoke UI

What data and how to find it?

- The primary data source for the database will be country responses to the EC-OECD STIP survey, but these can be semantically linked to other types of data, including
  - texts from existing OECD and EC publications
  - statistics
  - information in other databases, e.g. SIPER
- Information will be findable through search and discovery tools
- The user interfaces will also make extensive use of dashboards:
  - Provide a top-level visual overview of data on (i) countries, (ii) themes, (iii) policy instruments and (iv) target groups
  - Interactive, providing details on demand, and enabling faceted search and in-situ analysis
  - Built around "standard queries"

Data model variables that could be exploited in dashboards

Broader Categories	Primary Dashboards	Policy Initiative and Instrument Facets
Country Grouping	Country	Responsible Organisation
Broad Theme	Theme	Evaluated (Y/N)
		Budget Range
Broad Type of Policy Instrument	Policy Instrument	Policy Instrument Facets
Broad Target Group	Target Group	Time

Country dashboard example
Panels could be arranged differently

Country Name	Free Search
	Print/Share

Country Description

Themes mapping viz.	Headline indicators from IPP.Stat
Policy instruments mapping viz.	Target groups mapping viz.

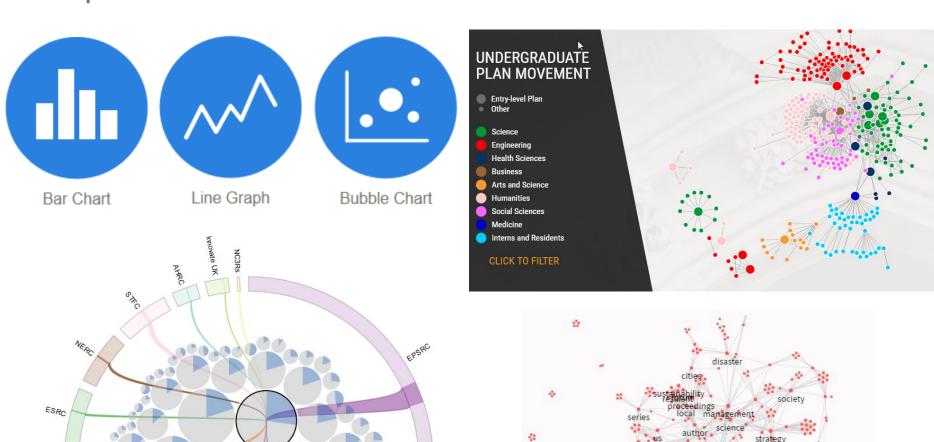| Related OECD publications	Related EC publications	Responsible organisations mapping viz.

Examples of "standard queries" that could be embedded in various dashboards

Country queries	• How many policy initiatives does a given country use to address a given theme? • How does a given country's policy instrument mix compare to other countries? • What is the distribution of budgets to different policy target groups? • What is the relative influence of different ministries and agencies in STI policy?
Theme queries	• On what themes do countries have the most policy initiatives? • On what themes do countries assign the highest budgets? • Which policy instruments are most commonly used to address a given theme? • Which target groups are most commonly targeted in a given theme?
Policy instrument queries	• Which policy instruments are most prevalent in national STI policy? • Which policy instruments have the highest budgets? • How does policy instrument design vary by country? • Are some policy instruments more commonly found with other instruments?
Target group queries	• To which target groups do countries assign the highest budgets? • Which policy instruments are most commonly used to address a given target group? • To what extent are different target groups addressed by the same initiatives? • How has policy attention to a given target group shifted over time?

Examples of sorts of interactive visualisations for the dashboards

Bar Chart

Line Graph

Bubble Chart

UNDERGRADUATE PLAN MOVEMENT

- Entry-level Plan
- Other
- Science
- Engineering
- Health Sciences
- Business
- Arts and Science
- Humanities
- Social Sciences
- Medicine
- Interns and Residents

CLICK TO FILTER

FURTHER INFORMATION

https://innovationpolicyplatform.org/stip-monitoring-and-analysis-ec-oecd-project