

The Nesta logo is the word "nesta" in a bold, white, lowercase sans-serif font. It is positioned in the upper left corner of the slide. The background is a vibrant teal with a subtle, grainy texture. Scattered across the background are several large, semi-transparent circles in various colors: a large blue circle at the top left, a large red circle at the bottom right, a large purple circle at the bottom center, a large orange semi-circle at the top right, a large black circle on the right side, a large red circle on the left side, a small yellow circle in the lower left, and a small purple circle in the lower center. A faint rainbow arc is visible on the left side of the slide.

Profiles of sustainable development research in Europe: A machine learning approach

George Richardson

[nesta.org.uk](https://www.nesta.org.uk)

 [@nesta_uk](https://twitter.com/nesta_uk)

Aim: to understand the geographic and organisational profile of SDG related research funded by the EU

Outline

1. Initial approach to modelling for detecting the Goals
2. Exploratory analysis of SDG R&I in Europe
3. Modelling revisited
4. Open questions

Advantages

- State-of-the-art text embedding methods capture semantic properties more broadly than keywords
- Pre-trained models can be applied to data with out-of-vocabulary terms

Disadvantages

- Supervised method needs a labelled dataset
- Always the possibility of missing specific or implicit concepts not included in training data

Method

1. Find labelled data on the web
2. Train a model
3. Apply to R&I datasets

Data Source - IISD Knowledge Hub

STORY HIGHLIGHTS

- The note identifies four types of cross-border challenges in scaling a global, circular plastics economy, related to regulations, standards and data, investment, and processes.
- The paper proposes three groups of trade policy actions to scale the shift towards a more circular global plastics economy: border measures, internal mechanisms, and increased transparency.
- The note observes that the COVID-19 pandemic has increased the demand for single-use plastics for personal protective equipment and contributed to an increase in disposable packaging.

- Over 10,000 SDG news articles (after cleaning)
- Paginated and structured website layout amenable to scraping
- Articles are tagged by human annotators with one or more SDGs
- Extra meta-data for further analysis

Article text

SDG labels

Issues

Partnerships

Actors

Actions

Topic tags

The World Economic Forum (WEF) has released a briefing note that analyzes opportunities to scale a more circular economy for plastics. The note draws on the expertise of environment and trade experts to identify investment and trade solutions to tackle cross-border challenges and move towards a sustainable circular plastics economy.

The note titled, 'Plastics, the Circular Economy and Global Trade,' states that the world produces over 400 million tonnes of plastics annually, much of which is "mismanaged after use... causing damage to environment and societies." Only 14-18% of plastic waste is formally recycled, and a large percentage of plastic waste intended for recycling is exported. Within this context, the note explores ways to enable legitimate, responsible trade to establish recycling economies of scale.

The note identifies four types of cross-border challenges in scaling a global, circular plastics economy, related to: regulations, such as bans or requirements on plastic waste imports; standards and data; investment; and processes. Differences among countries in each of these areas contributes to challenges in moving towards a more circular economy. Differences in standards on labeling and use, for instance, require recyclers to create different recycled plastic grades, which add to the overall cost. On processes, the note states that the Basel Convention will regulate most plastic waste trade across borders through a prior informed consent (PIC) procedure, beginning in

SDGS

- 12 RESPONSIBLE CONSUMPTION & PRODUCTION
- 14 LIFE BELOW WATER
- 17 PARTNERSHIPS FOR THE GOALS

ISSUES

Governance, Sustainable Consumption & Production, Economics & Investment, Chemicals & Wastes

GLOBAL PARTNERSHIPS

Means of Implementation, Trade

ACTORS

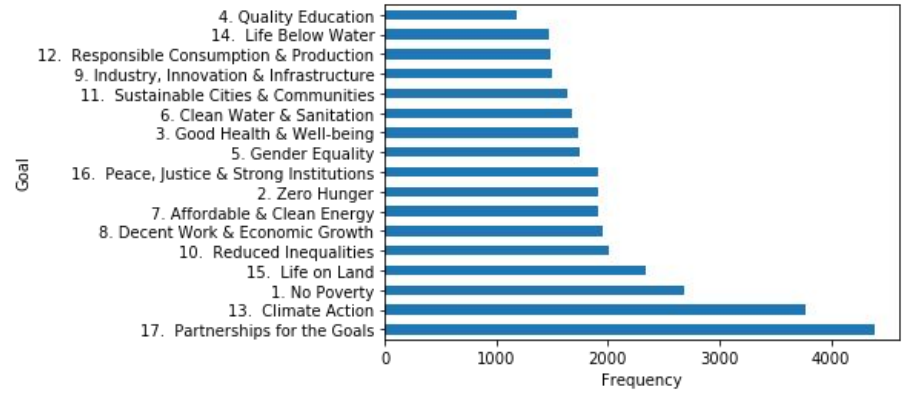
Stakeholders and Major Groups

ACTIONS

Publication

TAGS

COVID-19 Pandemic, Trade and



Model Creation

Features			Labels															
	title	text	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	A Comprehensive Guide to Solar Energy Systems ...	This chapter examines the growth of solar phot...	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
1	Sustainable Protein Sources - Chapter 1: Prote...	Food is essential for sustenance and, in parti...	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Advances in Renewable Energies and Power Techn...	Generation of electric power from renewable en...	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
3	Advances in Renewable Energies and Power Techn...	This chapter aims to present an overview of th...	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
4	Introduction to Petroleum Biotechnology - Chap...	This chapter deals extensively with the effect...	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0

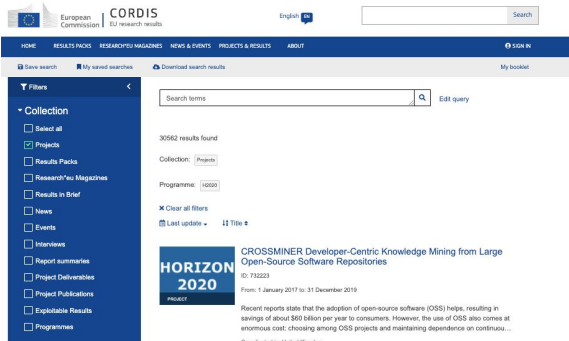
Modelling pipeline

1. Filtering data - removing short texts and articles with >3 goals
2. Text pre-processing - lemmatisation and 800 custom stop words
3. Text vectorisation - TensorFlow Universal Sentence Encoder
4. Modelling - imbalanced resampling + random forest



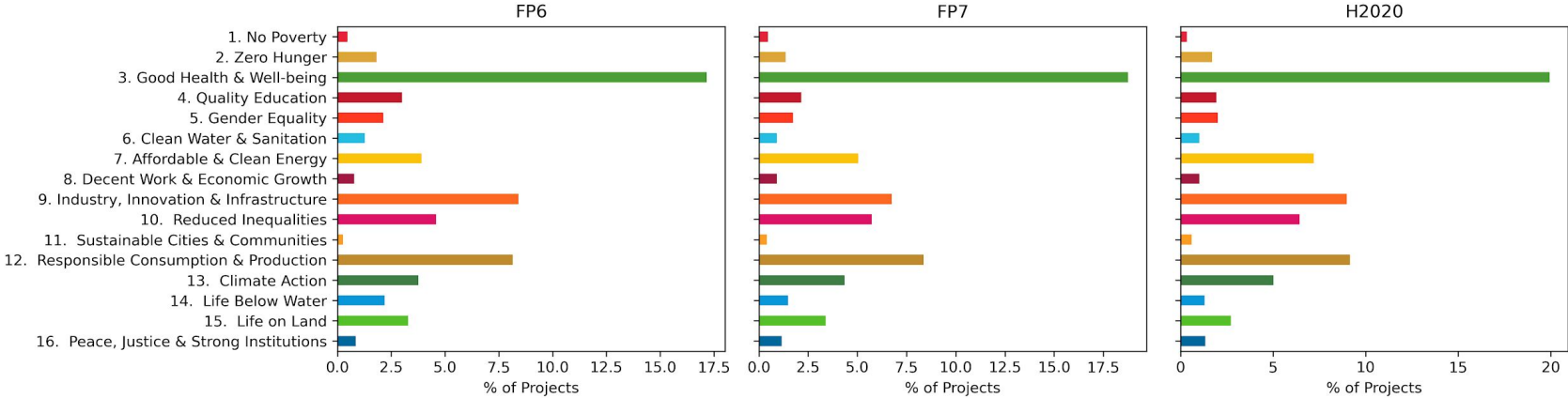
Multi-class multi-label SDG text classification model

Deployment on R&I Abstracts

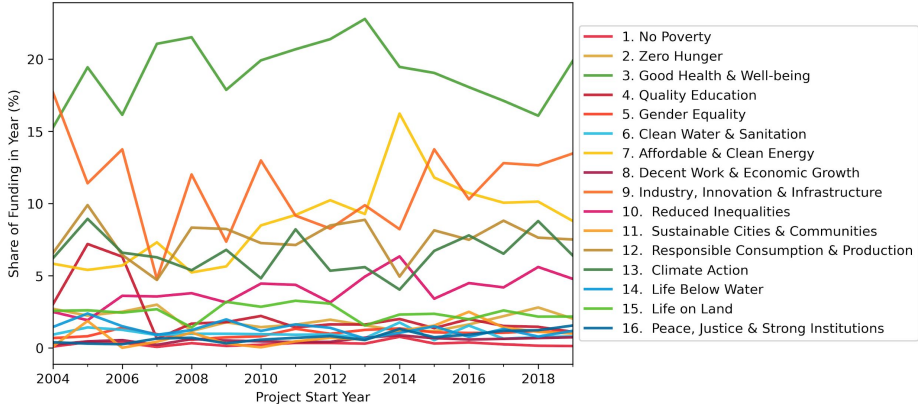
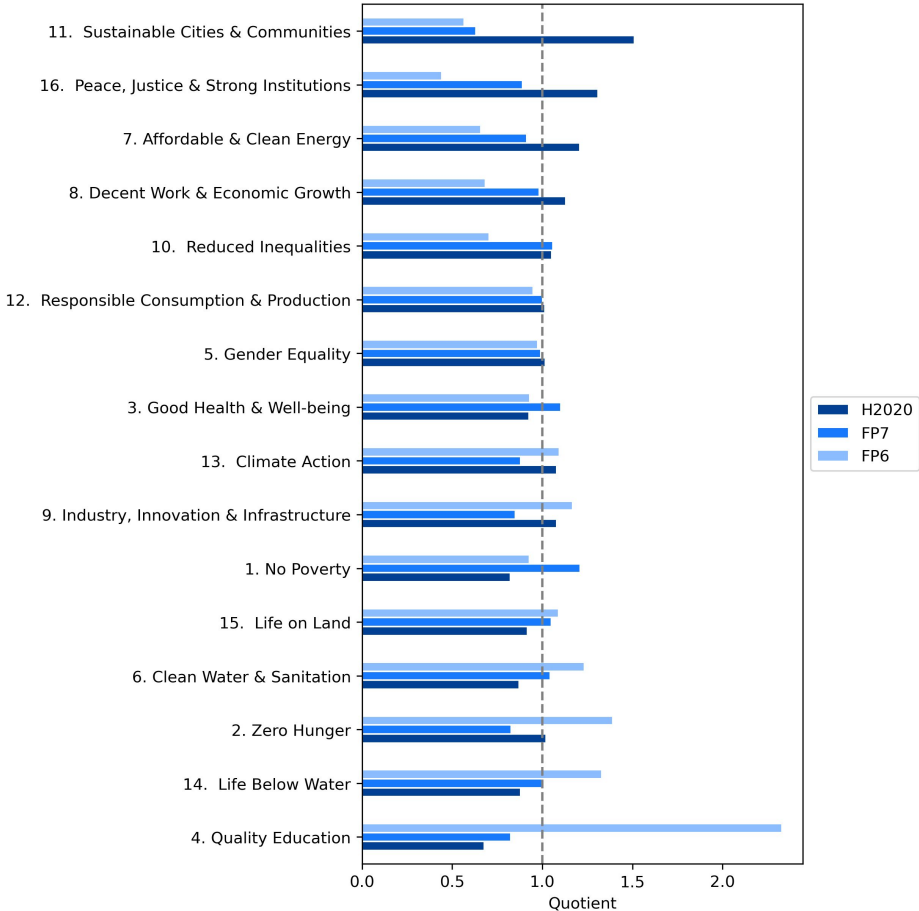


~100,000 project proposal abstracts

Processing time ~20 minutes



Deployment on R&I Abstracts



SDG research quotient for each framework programme (weighted by funding received) appears to suggest increasing attention to Goals 7, 8 11 and 16.

Suspiciously high funding activity for certain goals.

Inspecting the results

1. No Poverty

Improving loan quality and acceptance rates by predicting credit behavior through social mediadata.

Social innovation and urban revitalization in hyperdiverse local societies

Support to the implementation of the Strategic Research Agenda (SRA) of the Joint Programming Initiative on Cultural Heritage and Global Change (JPI CH)

Responsible Research and Innovation in Practice

Optimal Taxation of Wealth

13. Climate Action

Low-energy leak-proof double seat control valve based on a water hydraulic actuator system

Energy efficient and sustainable building planning

Cyclone processes leading to extreme rainfall in the Mediterranean region

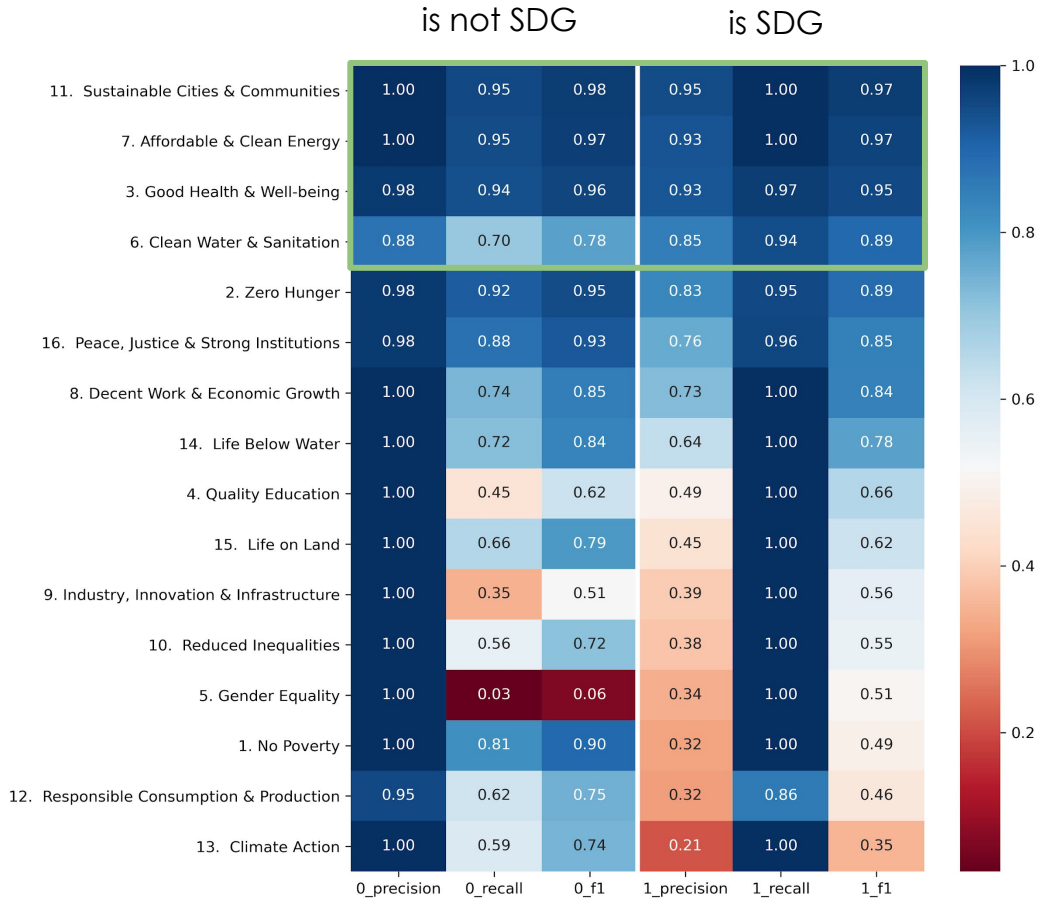
Closed-Loop Flow Control to Enhance Aerodynamic and Aeroacoustic Performance of Wind-Turbine Blades

Gas-Only internal combustion engines

Low precision: Very high false positive rate on several goals

Also leads to questions about recall

Rise of the machines humans



- Precision, recall and F1 scores for positive (1) and negative (0) SDG classifications, based on human verified data (support ~150). Sorted by positive F1 score.
- Only 3 SDGs have a precision & F1 of positive examples over 90% (5 over 80%).
- Imbalance in the data means there is little information on the nature of false negatives and the rate is likely underestimated.

Challenges

- **False positives**

- high semantic overlap between goals e.g. oceans and freshwater
- linguistic overlap between goals e.g. *fish* farming (falsely classified as Goal 14)

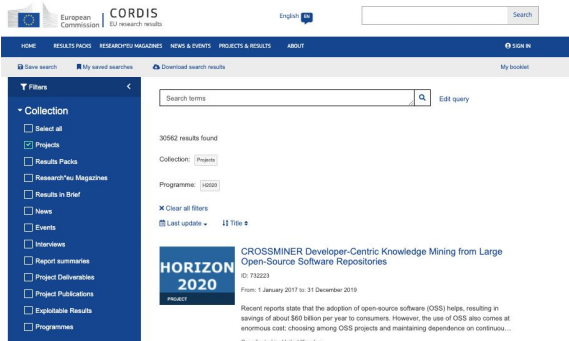
- **False negatives**

- reference to the goal is almost entirely implicit

- **Training data**

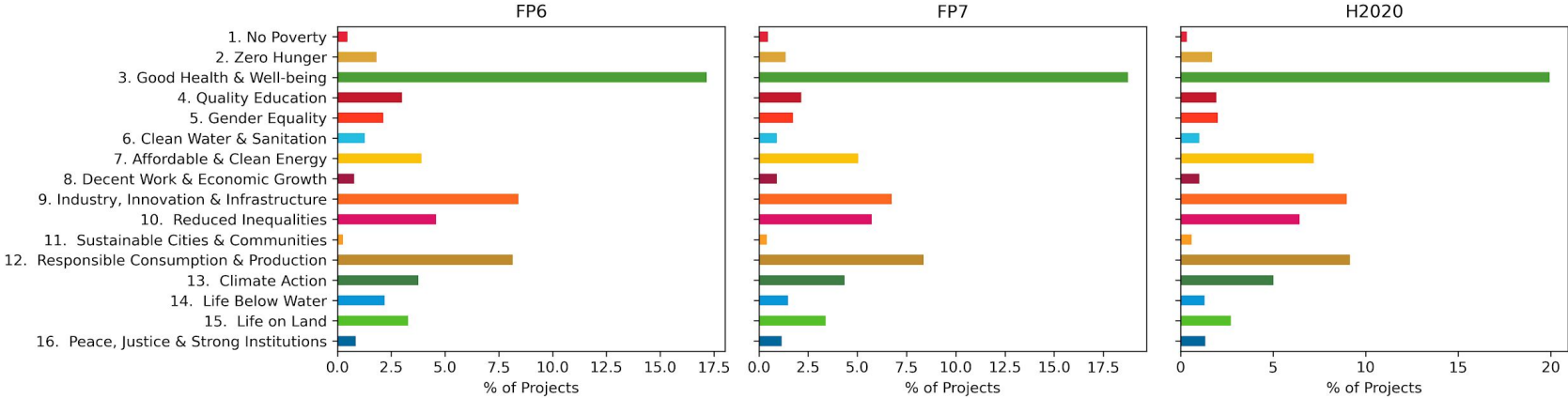
- over-zealous labelling with SDGs (e.g. many projects labelled with *both* 7. Clean and Affordable Energy & 13. Climate Action)
- large differences between news articles and research vocabulary

Deployment on R&I Abstracts

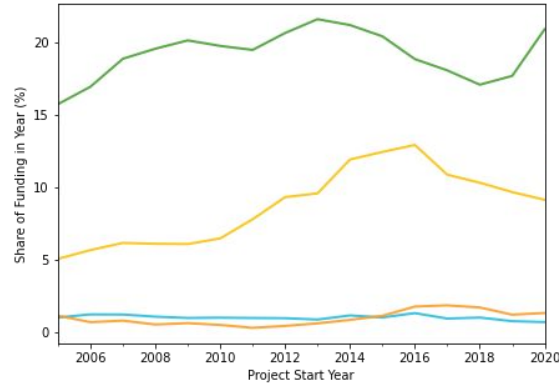
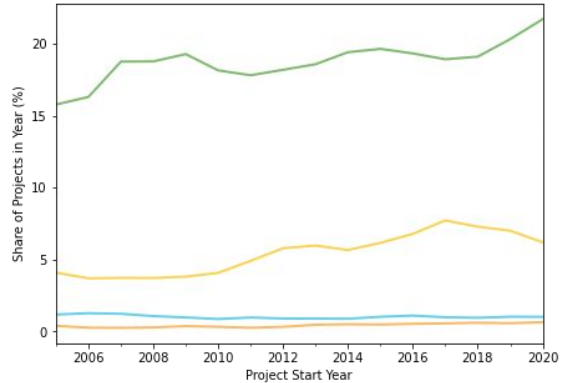
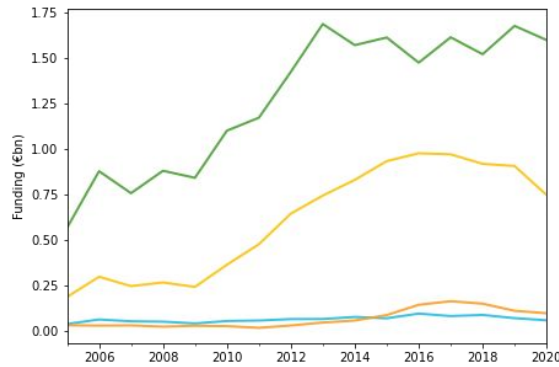
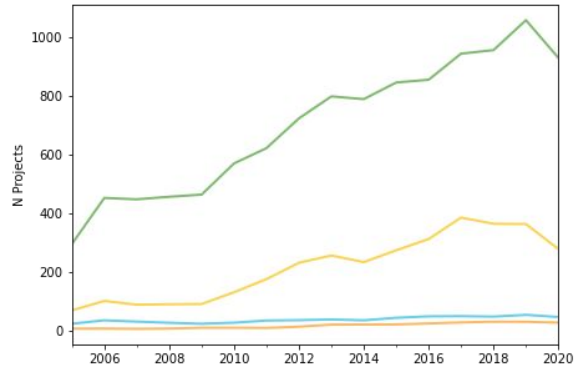


~100,000 project proposal abstracts

Processing time ~20 minutes

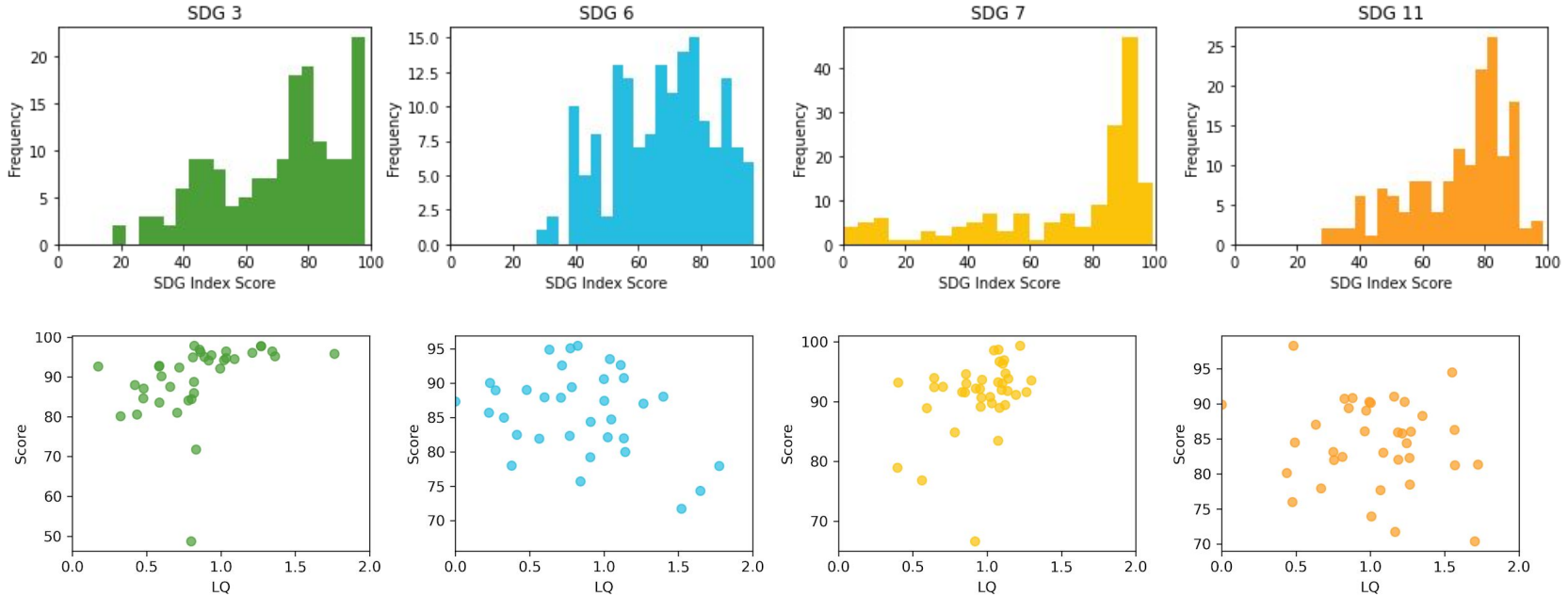


SDG Trends



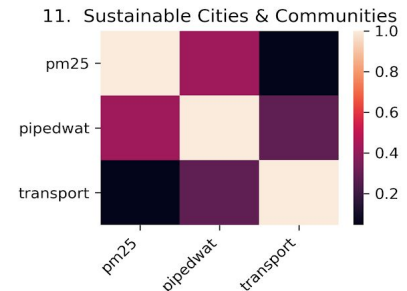
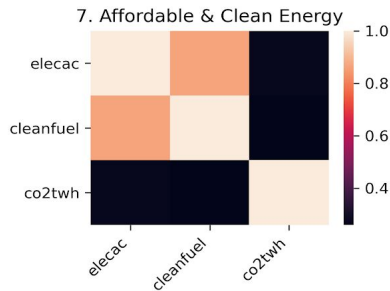
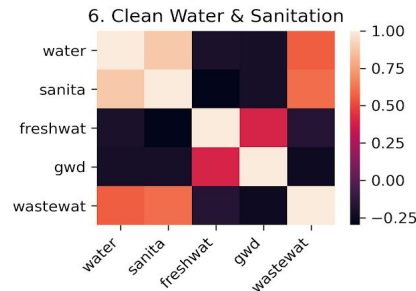
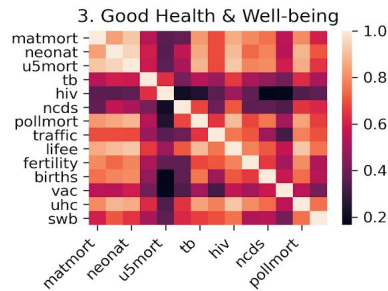
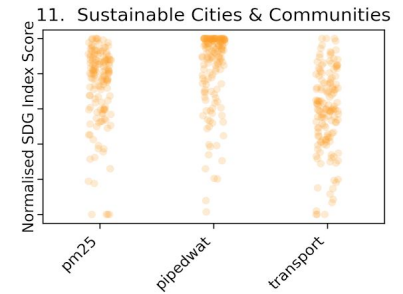
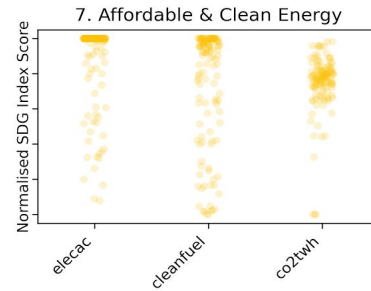
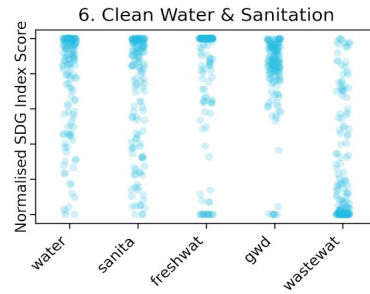
- Linear increase in proportion of projects for Goal 3. Goal 7 has received a significant boost towards end of FP7 and in H2020
- Trend towards larger projects for Goal 11 - number has not increased significantly but funding has. Now receives a larger share than Goal 6.

Research and National Performance in Europe



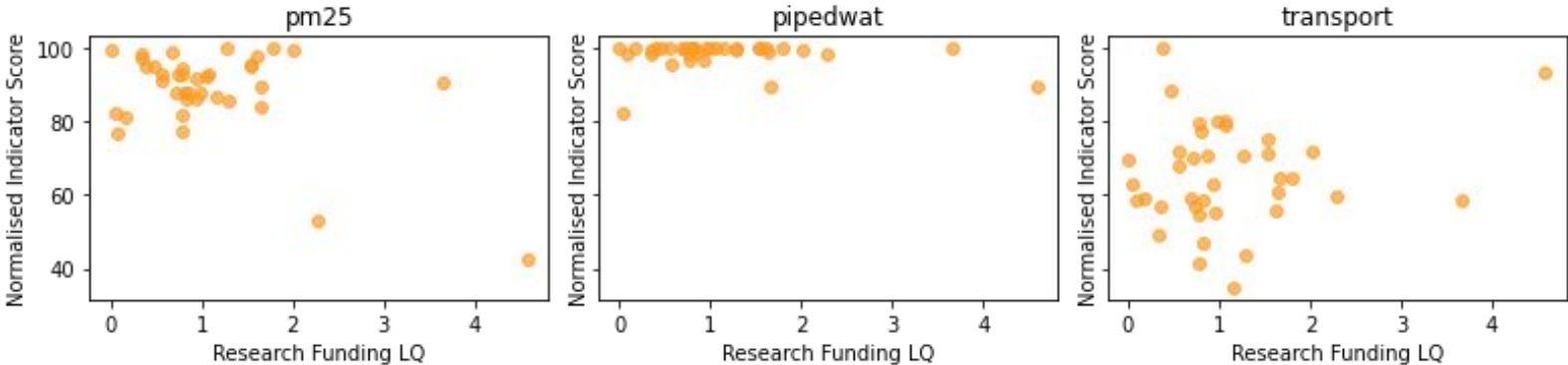
- Most European countries occupy the high end of the SDG Index 2019 score distributions
- The relationship between SDG related research specialisation and performance on the Index is a mix of positive and negative correlations of varying strength

Project Count by Country



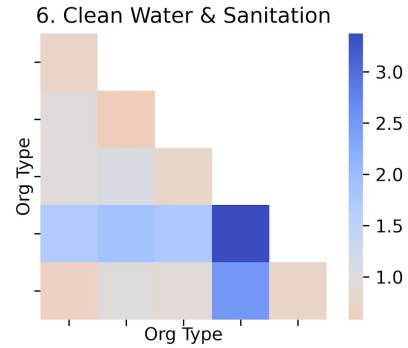
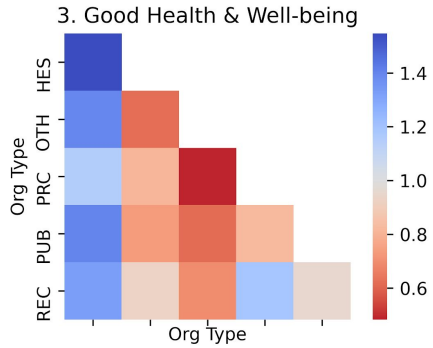
- Great variety between the distributions of individual indicators within each aggregate SDG Index score.
- Some of indicators are not correlated showing that they may not necessarily progress by the same mechanisms or at the same rate, or that they might in fact work in contradiction.

Focus on Sustainable Cities and Communities



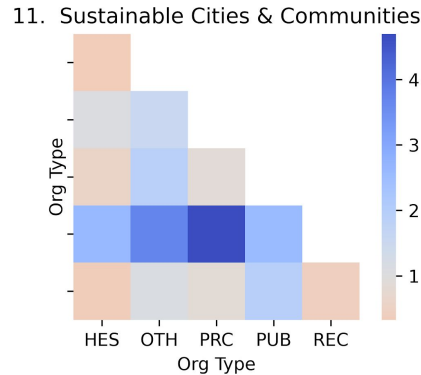
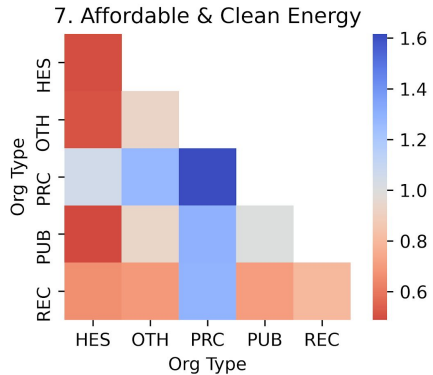
- Within the SDG Index score for a single goal, we find different relationships between R&D specialisation and the individual indicators

Collaboration Dynamics



Pairwise combinations of organisation type shows that particular collaborations are more highly represented in some goals.

Goal 3 - significant over-representation of higher education establishments collaborating with all other organisation types as well as collaborations between research and public institutions



Goal 3 - conversely a strong presence of private companies leads to over-representation of collaborations with them and much fewer with higher education establishments

Goals 6 and 11 - collaborations between public institutions and all other organisation types (particularly private companies) are highly represented.


Model v2.0

- Custom made dataset using manual annotation
- Annotating abstracts from CORDIS:
 - Research vocabulary
 - Spans fundamental and applied R&I in both STEM and social/policy
- Control over labelling process
 - A broad view of conceptual SDG relatedness
 - Strictly targeted to the targets and indicators of each goal

Distributed Annotation

README.md

SMART



build passing docs passing

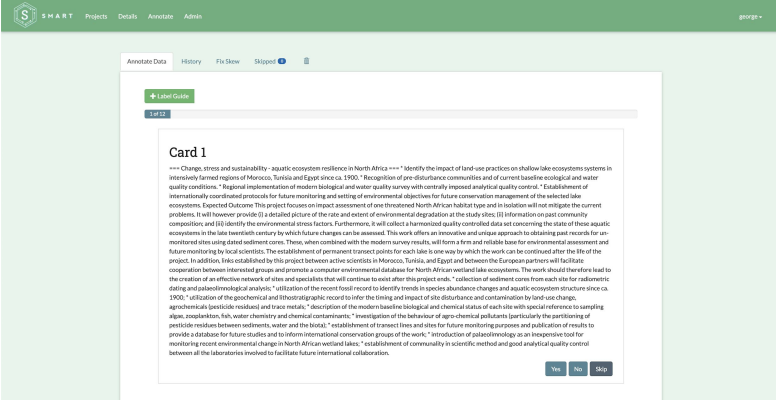
SMART is an open source application designed to help data scientists and research teams efficiently build labeled training datasets for supervised machine learning tasks.

- [SMART Landing Page](#)
- [SMART User Documentation](#)
- [SMART Publication](#)

If you use SMART for a research publication, please consider citing:

Chew, R., Wenger, M., Kery, C., Nance, J., Richards, K., Hadley, E., & Baumgartner, P. (2019). SMART: An Open Source Data Labeling Platform for Supervised Learning. *Journal of Machine Learning Research*, 20(82), 1–5.

User view



SMART Projects Details Annotate Admin george

Annotate Data History File View Skipped 0

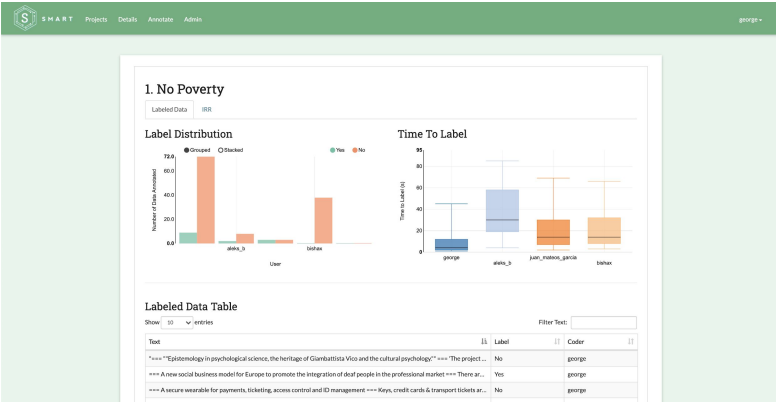
Label Data

Card 1

==== Change stress and sustainability - aquatic ecosystem resilience in North Africa ==== Identify the impact of land-use practices on shallow-lake ecosystems systems in intensively farmed regions of Morocco, Tunisia and Egypt since ca. 1900. * Recognition of pre-disturbance communities and of current baseline ecological and water quality conditions. * Regional implementation of modern biological and water quality survey with centrally imposed analytical quality control. * Establishment of internationally coordinated protocols for future monitoring and setting of environmental objectives for future conservation management of the selected lake ecosystem. * Expected Outcome This project focuses on impact assessment of one threatened North African habitat type and its isolation will not mitigate the current problems. It will however provide (i) a detailed picture of the size and extent of environmental degradation at the study sites, (ii) information on past community composition, and (iii) identify the environmental stress factors. Furthermore, it will collect a harmonized quality controlled data set concerning the state of these aquatic ecosystems in the late twentieth century by which future changes can be assessed. This work offers an innovative and unique approach to obtaining past records for environmental sites using dated sediment cores. These, when combined with the modern survey results, will form a firm and reliable base for environmental assessment and future monitoring by local scientists. The establishment of permanent transect points for each lake is one way by which the work can be continued after the life of the project. In addition, links established by this project between active scientists in Morocco, Tunisia, and Egypt and between the European partners will facilitate cooperation between interested groups and promote a computer environmental database for North African wetland lake ecosystems. The work should therefore lead to the creation of an effective network of sites and specialists that will continue to work after this project ends. * Collection of sediment cores from each site for radiometric dating and paleolimnological analysis. * Utilization of the recent fossil record to identify trends in species abundance changes and aquatic ecosystem structure since ca. 1900. * Utilization of the geochemical and lithostratigraphic record to infer the timing and impact of site disturbance and contamination by land-use change, agricultural landscape-related and trace metals. * Description of the modern baseline biological and chemical state of each site with special reference to sampling algae, zooplankton, fish, water chemistry and chemical contaminants. * Investigation of the behaviour of agro-chemical pollutants (particularly by the partitioning of pesticide residues between sediments, water and the biota). * Establishment of transect lines and sites for future monitoring purposes and publication of results to provide a database for future studies and to inform international conservation groups of the work. * Introduction of paleolimnology as an inexpensive tool for monitoring recent environmental change in North African wetland lakes. * Establishment of community in scientific method and good analytical quality control between all the laboratories involved to facilitate future international collaboration.

Yes No Skip

Admin view



SMART Projects Details Annotate Admin george

1. No Poverty

Labeled Data 100

Label Distribution

Time To Label

Labeled Data Table

Text	Label	Coder
==== "Epidemiology in psychological science, the heritage of Giambattista Vico and the cultural psychology" ==== The project...	No	george
==== A new social business model for Europe to promote the integration of deaf people in the professional market ==== There ar...	Yes	george
==== A secure wearable for payments, ticketing, access control and ID management ==== Keys, credit cards & transport tickets ar...	No	george
==== Accurate Income Measurement for the Assessment of Public Policies ==== This programme of activity will improve the com...	Yes	juan_matoso_garcia

SMART Data annotation tool

- Free and open source
- Easily deployable on EC2
- Includes features like active learning
- (a bit buggy)

Guided Annotation

Goal 1. No Poverty

End poverty in all its forms everywhere

Labelling Instructions

Key Concepts of Goal 1

To be labelled with Yes, a project should explicitly or implicitly refer to at least one of:

- Eradicating extreme poverty
- Reducing the number of people living in poverty
- Social protection and welfare systems
- Rights to economic resources such as land, property, technology and financial services
- Economic resilience of the poor and vulnerable to natural, social and economic shocks
- State spending on essential services (education, health and social protection)*
- State spending on sectors that benefit women, the poor and vulnerable*

*There are specific goals for education, health and gender equality. Please only label projects Yes if they mainly relate to these areas in the context of state spending.

Labelling Tips

- Many projects will cover more than one Goal, but to be tagged as this Goal they must include at least one of the Key Concepts above. Further information on other goals can be found [here](#).
- The documents are research projects so it may not be clear whether they actually contribute to achieving a Goal. That is fine - we are interested work that either contributes to achieving a goal or generates knowledge directly related to the Goal.
- In some cases the distinction between Yes and No will not be 100% clear. That's fine. Just use your best judgement. You can always use the Skip button (although it is best to try to make a decision).

Key terms (non-exhaustive)

poverty reduction, social protection, welfare state, economic justice, low income, quality of life, extreme poverty, economic circumstances, economic crisis, material well-being, property rights, precariousness, poverty, economic exclusion, living standards, disaster resilience, financial inclusion, financial exclusion,

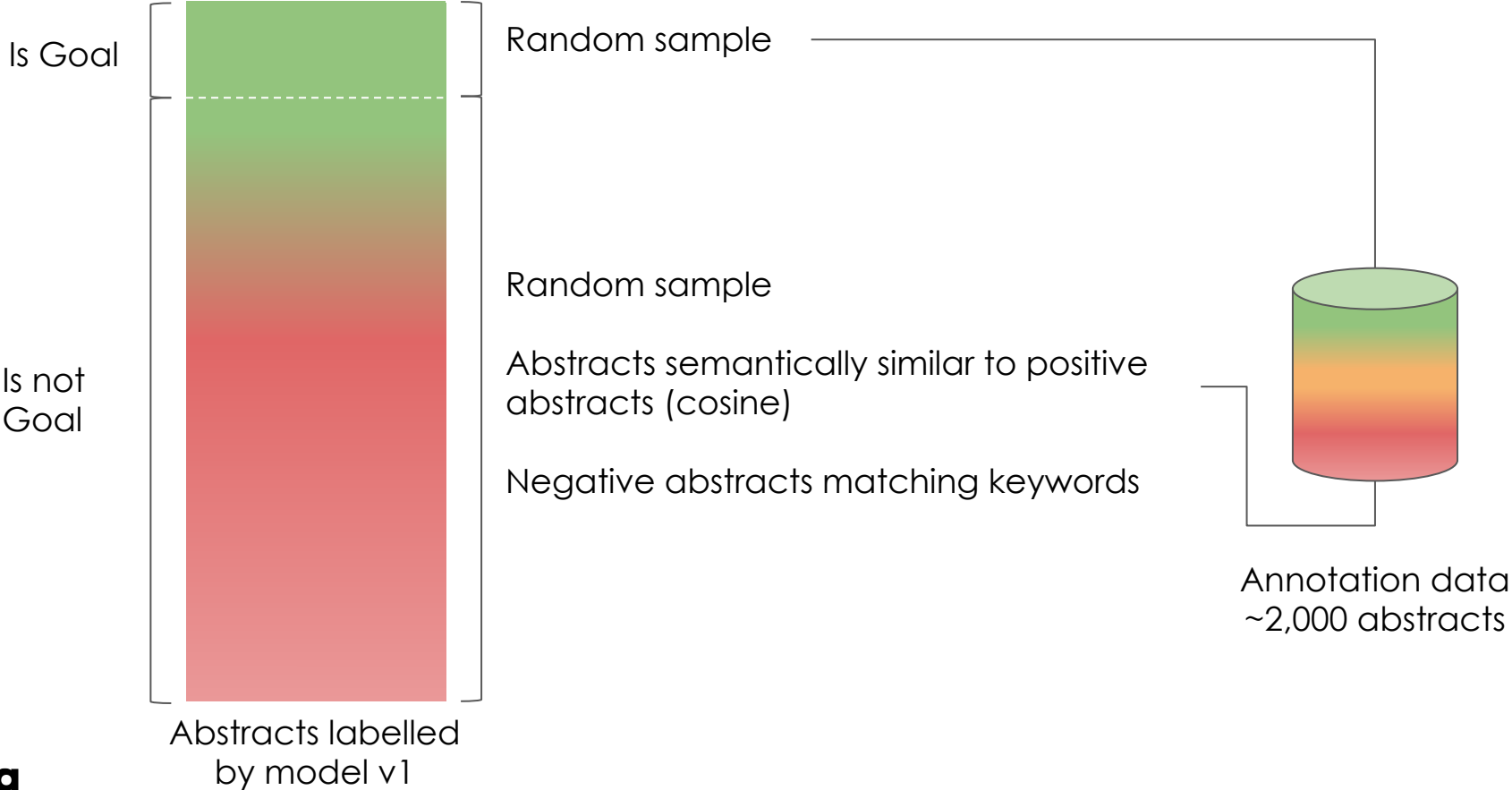
- Target 1.2: By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions
- Indicator 1.2.1: Proportion of population living below the national poverty line, by sex and age

Each goal has a corresponding annotation project and a specific codebook

Codebook contents:

- Summary of the goal's Targets
- Clarifications and caveats
- Guidance for labelling research abstracts
- Non-exhaustive key terms
- Related goals
- Illustrative Positive and negative examples
- Full list of Targets and Indicators

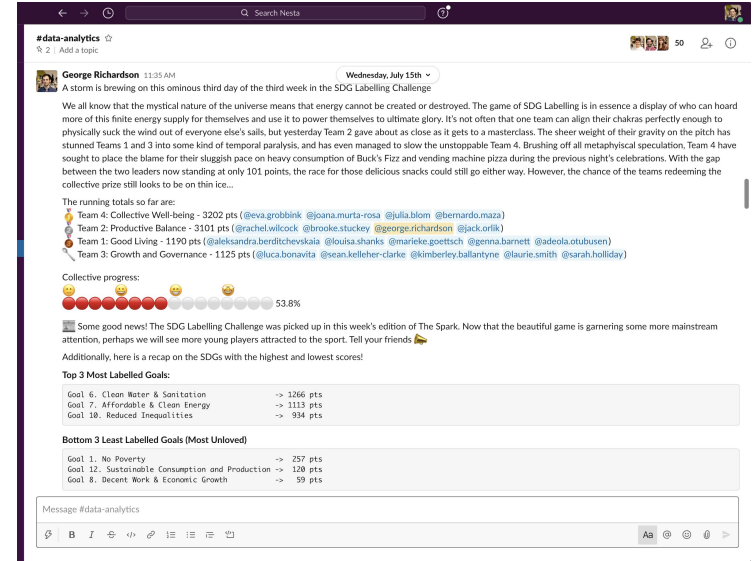
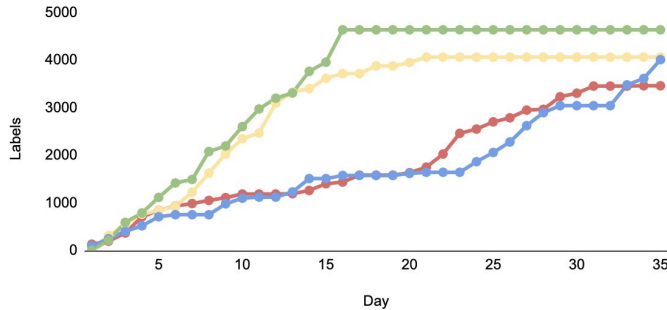
Annotation Data Preparation



It's A Team Sport

	A	B	C	D	E	F
1		Scores				Target
2		Team 1: Good Living	Team 2: Productive Balance	Team 3: Growth and Governance	Team 4: Collective Well-being	
		- 1. No Poverty - 5. Gender Equality - 12. Responsible Prod - 13. Climate Action	- 4. Quality Education - 9. Industry, Innovation and Inf - 10. Reduced Inequalities - 15. Life on Land	- 2. Zero Hunger - 8. Decent Work & Economic Grow - 14. Life Below Water - 16. Peace, Justice & Strong Institu	- 3. Good Health & Well-being - 6. Clean Water & Sanitation - 7. Affordable and Clean Energy - 11. Sustainable Cities & Comm	
3	1	136	33	99	7	1.71875
4	2	199	318	251	217	6.15625
5	3	372	555	405	595	12.04375
6	4	711	766	523	792	17.45
7	5	849	846	713	1119	22.04375
8	6	938	933	756	1422	25.30625
9	7	990	1229	756	1496	27.94375
10	8	1056	1630	756	2083	34.53125
11	9	1114	2031	985	2199	39.55625
12	10	1190	2350	1102	2606	45.3

- Team 1: Good Living
- Team 2: Productive Balance
- Team 3: Growth and Governance
- Team 4: Collective Well-being



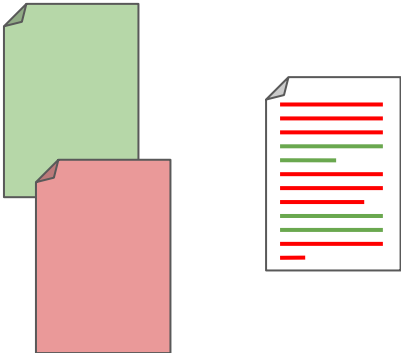
- ~1,000 labels per goal
- 4 teams of Nesta staff
- Daily 'game reports'
- 8 week duration with ~12 people

New Data, New Questions

Pro vs. anti-sustainable R&D



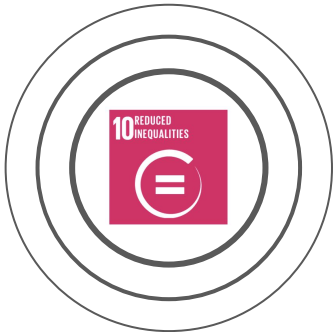
Binary or proportion



Knowledge generating vs



Direct vs conceptual



What's next?

- Develop new model with annotated data
- Validate and open source model and open source
- Establish firmer relationships between R&D and SDG Index particularly using European indicators
- Develop public crowdsourcing challenge to create a more robust training dataset and address different use cases