nature sustainability

ANALYSIS

https://doi.org/10.1038/s41893-022-00874-z

Monitoring global development aid with machine learning

Malte Toetzke[1]✉, Nicolas Banholzer[2] and Stefan Feuerriegel[3]

Monitoring global development aid provides important evidence for policymakers financing the Sustainable Development Goals (SDGs). To overcome the limitations of existing monitoring, we develop a machine learning framework that enables a comprehensive and granular categorization of development aid activities based on their textual descriptions. Specifically, we cluster the descriptions of ~3.2 million aid activities conducted between 2000 and 2019 totalling US$2.8 trillion. As a result, we generated 173 activity clusters representing the topics of underlying aid activities. Among them, 70 activity clusters cover topics that have not yet been analysed empirically (for example, greenhouse gas emissions reduction and maternal health care). On the basis of our activity clusters, global development aid can be monitored for new topics and at new levels of granularity, allowing the identification of unexplored spatio-temporal disparities. Our framework can be adopted by development finance and policy institutions to promote evidence-based decisions targeting the SDGs.

The Sustainable Development Goals (SDGs) provide a global framework for responding to current challenges affecting people and the planet[1,2]. In addition to defining a set of development goals, the SDGs formulate the guiding principle of 'leaving no one behind' and thereby emphasize the necessity to improve conditions for all parts of society and across all countries[3]. In this context, the SDGs explicitly ask the international community for the provision of development aid[4].

Development aid plays a crucial role in helping developing countries meet the targets of the SDGs[5,6]. It supports developing countries through so-called aid activities. Aid activities represent different types of contribution, such as development projects, financial assistance, material supply, technological support and training. They are funded by various donor institutions worldwide (for example, national development agencies, multilateral development banks) and thereby pool substantial financial resources[6].

Given the financial magnitude of development aid, it is important to monitor how development aid is distributed across recipient countries and what topics are funded through aid activities (for example, disease prevention, electrification). On the one hand, this helps donors coordinate their global activities. On the other hand, it helps other stakeholders assess whether overall financial volumes are distributed reasonably and effectively[7–11]. For this, the topics of aid activities should be monitored with high granularity (for example, solar energy instead of energy) and without substantial delay. This enables donors and policymakers to respond to current disparities with coherent policy responses[12].

Yet, existing monitoring of global development aid has considerable limitations, especially in that it is extremely difficult and expensive[8,13–16]. A major reason for this is the highly fragmented structure of the global aid system in which various donors fund large numbers of aid activities across different recipient countries[16]. For example, the most widely used monitoring system is the creditor reporting system (CRS) of the Organization of Economic Cooperation and Development (OECD)[17,18]. The CRS draws on a decentralized reporting system to distribute the burden of data collection and categorization. It thus requires donors to manually categorize conducted aid activities based on a standardized set of labels that are predetermined by the Development Assistance Committee of the OECD. However, these labels are defined before donors start reporting aid activities for a given period and due to the bureaucratic complexity, these labels are rarely adjusted over the years[19]. Consequently, in some areas, the CRS suffers from a poor fit between conducted aid activities and assigned labels and fails to capture new trends within aid activities[13]. Furthermore, it produces excessive transaction costs to the entire development aid system[8,16]. These limitations are discussed in detail in Supplementary Discussion 1.

In this study, we develop a machine learning framework to generate a comprehensive and granular categorization of global aid activities based on the textual descriptions from donors provided in the CRS. Our data comprise textual descriptions of ~3.2 million aid activities. These activities were conducted between 2000 and 2019 and were financed by 757 different donor organizations. The total disbursement amounted to US$3.2 trillion. We clustered the textual descriptions of these activities into 173 *activity clusters* representing different topics of aid activities. Examples of activity clusters are PRIMARY EDUCATION, BASIC NUTRITION and BIODIVERSITY. In the following, we highlight names of activity clusters in CAPITAL LETTERS.

Our machine learning framework follows a bottom-up approach, where all textual descriptions of conducted aid activities are processed before generating a categorization. This is in contrast to the top-down approaches in existing monitoring where standardized labels must be predetermined ex ante. The bottom-up approach of our machine learning framework brings three substantial advantages. First, our framework ensures an accurate representation of aid topics and makes the categorization adaptive to emerging topics (for example, YOUTH EMPOWERMENT, MICROFINANCE). Second, our framework categorizes millions of individual aid activities algorithmically, which ensures consistency in the interpretation of aid activities. Third, our framework yields a high granularity of aid topics due to the bottom-up approach and thus allows for a more precise disaggregation of development aid.

[1]Group for Sustainability and Technology, ETH Zurich, Zurich, Switzerland. [2]Department of Management, Technology and Economics, ETH Zurich, Zurich, Switzerland. [3]LMU Munich School of Management, LMU Munich, Munich, Germany. ✉e-mail: mtoetzke@ethz.ch

This study makes an important step in advancing data-driven analyses in the area of sustainable development. Monitoring development aid through our machine learning framework allows for more cost-effective, granular and comprehensive monitoring of global aid activities. While previous literature used machine learning primarily to monitor prevailing needs in developing countries[20–25], our study uses machine learning to monitor aid activities conducted in developing countries. Therefore, our study complements existing work on the demand for aid (where is aid needed the most?)[26] with new evidence on its supply (where is it allocated?) to assist donors and policymakers in making evidence-based decisions to 'leave no one behind'[3].

Activity clusters generated by machine learning framework

Our machine learning framework categorized ~3.2 million aid activities by mapping their textual descriptions onto 173 activity clusters. The framework follows a bottom-up approach where labels are generated ex post after processing the entire corpus of activity descriptions. This is in contrast to existing top-down approaches where labels are defined ex ante and then assigned to new aid activities. As illustrated in Fig. 1a, this is achieved through three main steps; first input are the textual descriptions of all aid activities, which are embedded into a 200-dimensional vector space (*document embedding*). In this space, activity descriptions with similar topics are close to each other while dissimilar ones are more distant. Second, the document embeddings of the activity descriptions are clustered to group activities with similar topics into the same category (*clustering*). Finally, the content of each cluster is analysed, and clusters are provided with names (details on the machine learning framework and validation procedures are described in Methods).

The categorization through activity clusters produces a large variety of different topics at high granularity. As shown in Fig. 1b, the categorization differentiates among similar, yet distinct, topics of aid activities such as specific energy technologies (for example, SOLAR ENERGY, COAL-FIRED ELECTRIC POWER PLANTS). Activity clusters were also formed when comparably few observations were available, allowing for the categorization of less frequently conducted types of aid activity such as BIOFUEL-FIRED POWER PLANTS. To provide additional information regarding the topics of activity clusters, we present the characteristic words of activity clusters in Fig. 1c.

Our categorization assigns each aid activity to only one activity cluster. This ensures that the financial disbursements of aid activities can be directly attributed to activity clusters. Otherwise, if aid activities could belong to multiple activity clusters, it would be unclear how to split and attribute financial disbursements of activities between activity clusters. Still, activity clusters can also represent cross-cutting topics. Activity descriptions that combine multiple topics are located separately in the embedding space. Hence, they can form separate activity clusters. For example, this applies to activity descriptions that refer to energy generation from mixed sources (for example, ENERGY GENERATION (RENEWABLE SOURCES)) or research activities across different domains (for example, MEDICAL RESEARCH; Supplementary Fig. 6).

Descriptive overview

Activity clusters generated by our machine learning framework provide a new and meaningful categorization of aid activities through the topics of textual activity descriptions (Fig. 2a). The categorization can be used to conduct important analyses regarding the distribution of aid across different countries and topics. For example, for any activity cluster, one can analyse how aid disbursements are distributed across recipient countries (for example, showing that among all recipients, Kenya is a major recipient for maternal health care) or one can analyse what activity clusters are financed most extensively (for example, showing that the majority of aid disbursements allocated to South Africa are targeting sexually transmitted diseases control).

Activity clusters cover a large variety of different topics. To summarize the scope, we assigned activity clusters to sectors of aid based on official sector categories by the OECD[27] (for example, we assigned the activity cluster MATERNAL HEALTH CARE to the health sector and SOLAR ENERGY to the energy sector). Figure 2b–d illustrates the most prominent sectors of aid. This reveals considerable differences with regard to the number of activity clusters, the number of aid activities and the total disbursements across sectors. Results show that the highest number of activity clusters (28 activity clusters), the highest number of aid activities (636,305) and the highest total disbursements (US$341 billion) come from the sector of government and civil society. This sector includes activity clusters such as HUMAN RIGHTS, GENDER EQUALITY, MACROECONOMIC POLICY and ANTI-CORRUPTION. 'Others' summarizes 19 smaller sectors that in total comprise 35 additional activity clusters such as RURAL DEVELOPMENT or WASTE MANAGEMENT.

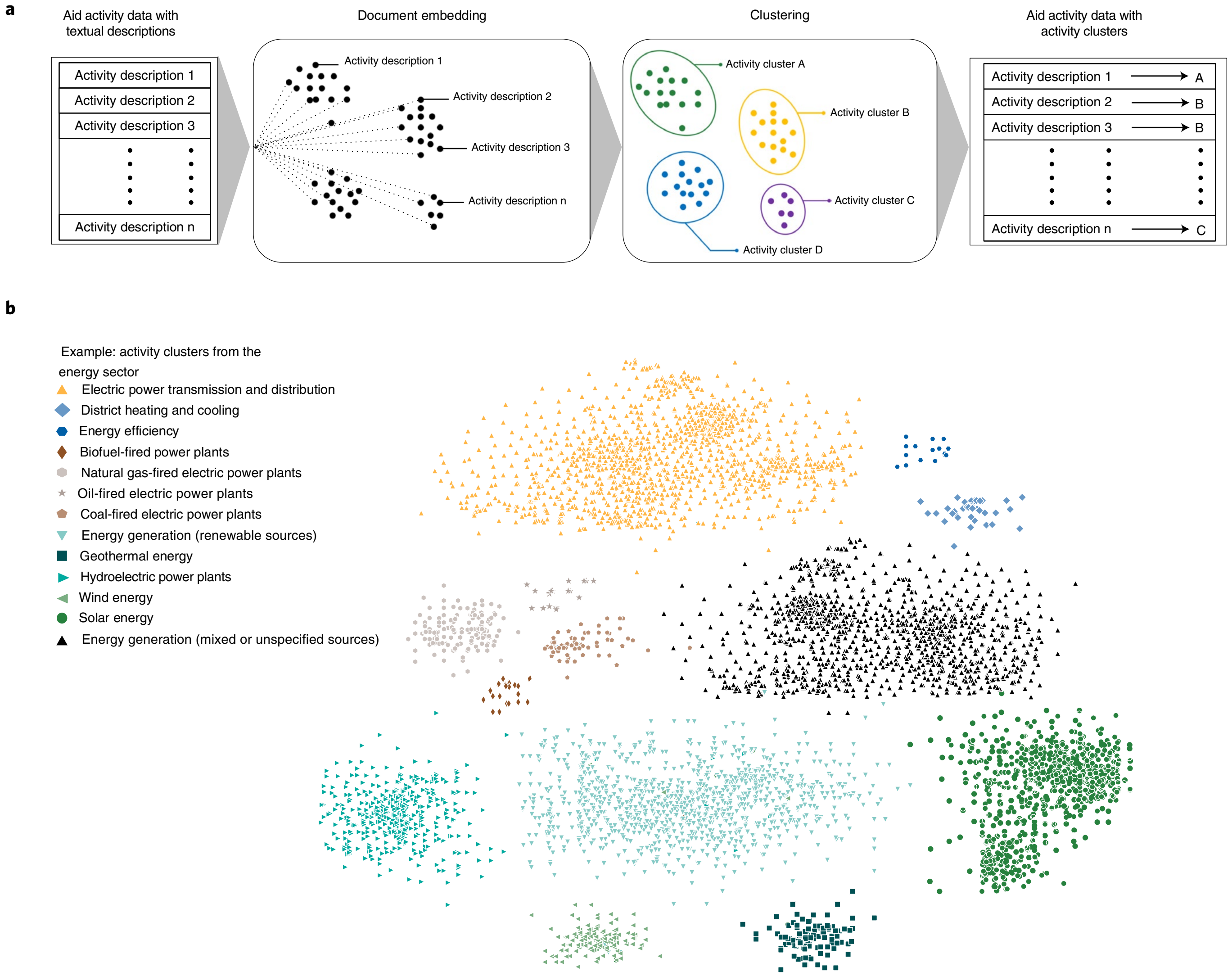
A comprehensive list of all activity clusters is provided in Supplementary Table 9 and the Supplementary Data.

Newly captured activity clusters beyond existing monitoring

In contrast to existing monitoring, the bottom-up approach of our machine learning framework allows ex post categorization of aid activities without using predefined labels. Activity clusters are generated by the framework by simultaneously analysing all 3.2 million descriptions of conducted aid activities, thereby enabling a systematic identification of topics. As a consequence, 70 out of 173 activity clusters have topics that have yet not been captured in existing monitoring (Supplementary Table 9). We refer to them as 'newly captured' activity clusters.

The occurrence of newly captured activity clusters can be explained as follows. First, some activity clusters capture topics at a higher level of granularity than existing monitoring. For example, unlike existing monitoring, activity clusters differentiate between CONSERVATION OF FORESTS, CONSERVATION OF WETLANDS and CONSERVATION OF COASTAL AREAS AND CORAL REEFS or between NEWBORN HEALTH CARE, REPRODUCTIVE HEALTH CARE and MATERNAL HEALTH CARE. Second, some activity clusters capture emerging topics that, initially, accounted for only a small number of aid activities but became popular later. In existing monitoring, such emerging topics are often not captured because labels are predefined years beforehand, and introducing new labels involves high bureaucratic costs. We observe that 35 of the 70 newly captured activity clusters had fewer than 100 annual aid activities in the first four years of our reporting period (2000–2003), but the number of aid activities increased by more than ten times when analysing the last four years (2016–2019). Examples of such activity clusters with emerging topics are YOUTH EMPOWERMENT and MICROFINANCE.

Between 2000 and 2019, newly captured activity clusters accounted for 784,500 aid activities and total disbursements of US$580 billion (Fig. 3a). The share of new activity clusters varies across sectors (Fig. 3b). Sectors related to technology and infrastructure (for example, energy, transportation and storage) have a comparably low share of new activity clusters. Activities within these sectors such as building streets or installing solar panels are relatively stable over time and thus more likely to be captured by existing monitoring. In contrast, sectors such as the business or the environment sector have a high share of newly captured activity clusters. This is presumably because they are strongly affected by political and societal trends, which change over time and are difficult to capture without updates to the categorization.

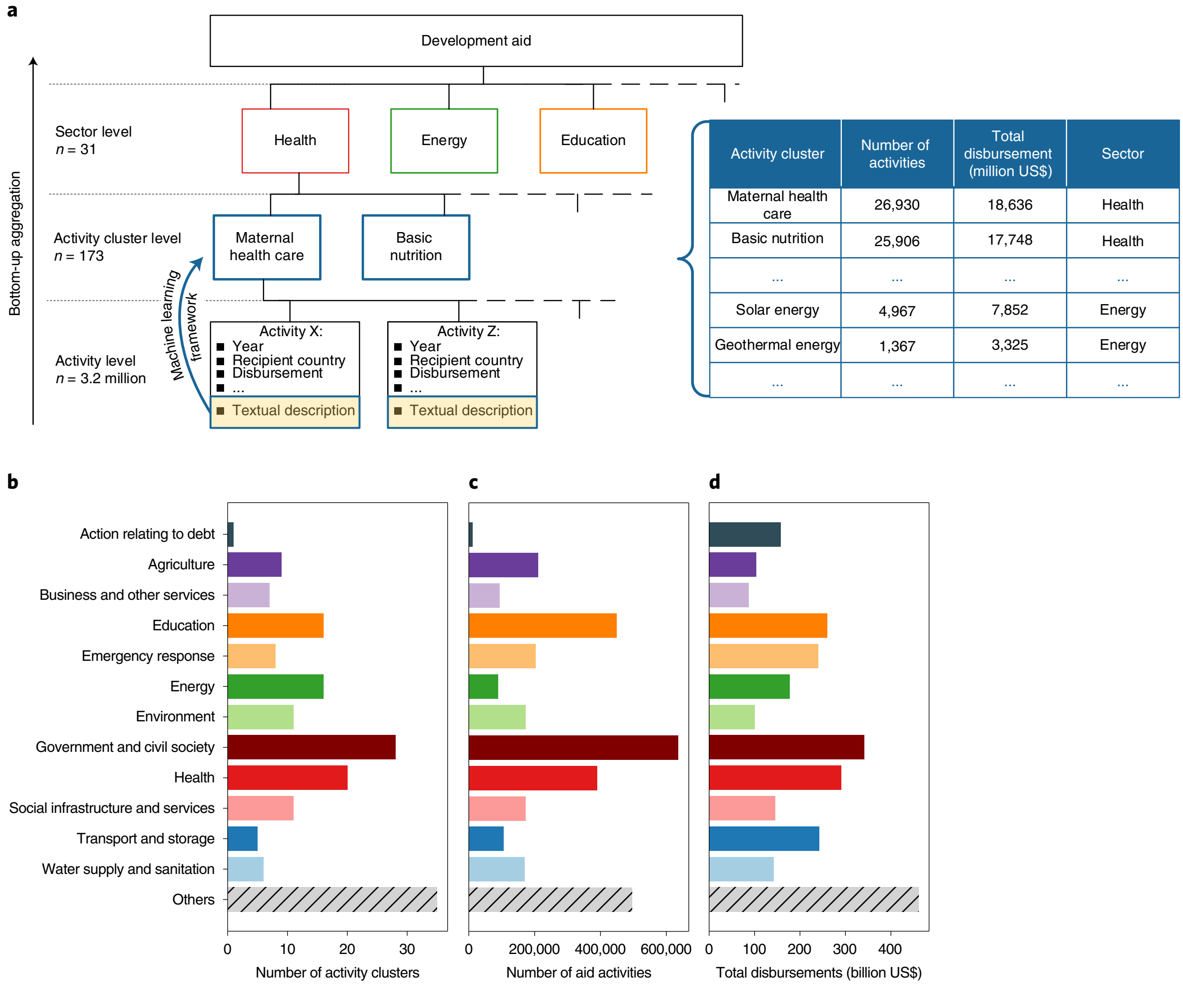

a
Aid activity data with textual descriptions
Activity description 1
Activity description 2
Activity description 3
Activity description n
Document embedding
Activity description 1
Activity description 2
Activity description 3
Activity description n
Clustering
Activity cluster A
Activity cluster B
Activity cluster C
Activity cluster D
Aid activity data with activity clusters
Activity description 1 A
Activity description 2 B
Activity description 3 B
Activity description n C
b
Example: activity clusters from the energy sector
Electric power transmission and distribution
District heating and cooling
Energy efficiency
Biofuel-fired power plants
Natural gas-fired electric power plants
Oil-fired electric power plants
Coal-fired electric power plants
Energy generation (renewable sources)
Geothermal energy
Hydroelectric power plants
Wind energy
Solar energy
Energy generation (mixed or unspecified sources)

c

Activity cluster from energy	Characteristic words
▲ Electric power transmission and distribution	kV transmission electricity grid substation substations electric voltage electrification electrical km circuit reliability smart interconnection
◆ District heating and cooling	Heat card central renewable cooling modernization pump stimulation emissions SECO smart USTDA DH inhabitants SDC
⬢ Energy efficiency	Efficient factor electricity building meter electrical utilities consumer inefficient loss TA turbine transform extent consumption
♦ Biofuel-fired power plants	Biofuels biofuel bioenergy biomass jatropha nut gas electricity retailers producers cassava mini fleet bio fire
⬢ Natural gas-fired electric power plants	Gas MW turbine fire electricity cycle thermal biogas emissions electric steam greenhouse loan kV pipeline
★ Oil-fired electric power plants	Oil fuel electricity steam heavy thermal turbine diesel generators biomass substations Wadi fire Moh expensive
⬟ Coal-fired electric power plants	Coal MW fire kV electricity electric tariff firm emission expensive emissions inefficient km Shanghai shortage
▼ Energy generation (renewable sources)	Renewable electricity energies clean electrification MW solar gas emissions wind greenhouse EE investments hydropower GHG
■ Geothermal energy	Geothermal MW electricity renewable exploration drill exploratory FS steam temperature surface GHG kV valley proximity
▶ Hydroelectric power plants	Hydropower hydroelectric MW hydro electricity transmission dam river kV renewable grid electric Himachal loan electrification
◀ Wind energy	Wind MW farm renewable turbine grid electricity PPA offshore Phu park transmission loan energies kW
● Solar energy	Solar photovoltaic renewable offgrid MW electricity electrification CSP households light install FS clean panel electric
▲ Energy generation (mixed or unspecified sources)	Electricity MW kV coal electric gas fire thermal renewable diesel reliability voltage fuel smart substation

Fig. 1 | Overview of machine learning framework to generate activity clusters. a, Schematic illustration of the machine learning framework for generating activity clusters from textual activity descriptions. **b**, Map of aid activities and corresponding activity clusters shown as distributed stochastic neighbour embedding[55] reducing the dimensionality of document vectors to two dimensions. To ensure visibility in two dimensions, we limited the sample of aid activities to 5,000 observations and included activity clusters only related to energy. Dots show the location of textual activity descriptions within the aforementioned two dimensions. Shape and colour of dots indicate activity clusters to which activities have been assigned. **c**, Characteristic words for activity clusters from energy, based on tf–idf scores (Methods includes further details regarding tf–idf).

Across sectors, we observe three major trends summarizing newly captured topics with particularly high growth rates (Supplementary Fig. 7): (1) 13 newly captured activity clusters promote inclusiveness, equal opportunities and empowerment of underrepresented groups (for example, INCLUSIVE LABOR MARKET DEVELOPMENT, GENDER EQUALITY, YOUTH EMPOWERMENT, ACCESS TO BASIC EDUCATION FOR DISADVANTAGED CHILDREN); (2) 11 newly captured activity



Activity cluster	Number of activities	Total disbursement (million US$)	Sector
Maternal health care	26,930	18,636	Health
Basic nutrition	25,906	17,748	Health
...	...	...	...
Solar energy	4,967	7,852	Energy
Geothermal energy	1,367	3,325	Energy
...	...	...	...

Fig. 2 | Overview of activity clusters generated by the machine learning framework. a, Schematic illustration of bottom-up aggregation mapping aid activities onto activity clusters based on textual activity descriptions. To reduce complexity, only examples of sectors, activity clusters, and activity attributes are shown in the illustration. This simplification is highlighted by dashed lines and ellipses. **b**, Number of activity clusters by sector of aid. The category 'Others' includes 19 additional aid sectors of smaller sizes such as fishing. **c**, Number of global aid activities conducted between 2000 and 2019 by sector of aid. **d**, Total aid disbursements in billions of US$ spent on aid activities from 2000 to 2019 by sector of aid.

clusters address environmental sustainability and climate change (for example, GREENHOUSE GAS EMISSION REDUCTION, SUSTAINABLE LAND USE, ENERGY EFFICIENCY); and (3) five newly captured activity clusters focus on strengthening the private sector (for example, MICROFINANCE, BUSINESS DEVELOPMENT AND ENTREPRENEURSHIP, COMPETITIVE MARKET DEVELOPMENT). For the complete list of activity clusters assigned to these trends, see Supplementary Table 3.

Spatio-temporal differences in the global distribution of development aid

Our activity clusters reveal previously unexplored distribution patterns of development aid (Fig. 4). The reason is that distribution patterns can change substantially when disaggregating data with high granularity into activity clusters. For example, the global distribution of aid disbursements from the environment sector (Fig. 4a) differs considerably from disaggregated disbursements of the underlying activity clusters GREENHOUSE GAS EMISSIONS REDUCTION and CLIMATE CHANGE ADAPTATION (Fig. 4b,c) (Supplementary Fig. 9 includes an example from the health sector).

Thus, the granularity of activity clusters helps to generate new evidence on the spatio-temporal distribution of development aid. For example, between 2015 and 2019, Mexico and Brazil each received around 3% of global development aid disbursements related to the environment sector. When specifically assessing the subordinate activity cluster GREENHOUSE GAS EMISSIONS REDUCTION, the data show that Mexico received 21% and Brazil received 12% of the total global aid provided for activities aiming to reduce greenhouse gas emissions (Fig. 4b,c). Yet, for activities related to CLIMATE CHANGE ADAPTATION, both countries received less than 1%. Across recipient countries, upper–middle-income countries receive more than two-thirds of global development aid for reducing greenhouse gas emissions (68%), which also represent high emitters of greenhouse gases[28]. In contrast, development aid focusing on climate change adaptation is distributed more evenly and more targeted towards lower–middle-income countries (46%), least developed countries (34%) and also small island developing states (7%). Figure 4 shows that for example, the Philippines receive more than 7% of development aid for climate change adaptation but less than 0.1% of development aid for greenhouse gas emissions reduction.

Activity clusters also allow for contextualizing temporal changes in aid distributions with relevant multilateral agreements such as the Paris Agreement[29]. In the Paris Agreement, donor countries pledged substantial increases in global development aid related to climate change. Indeed, we see that aid disbursements for climate

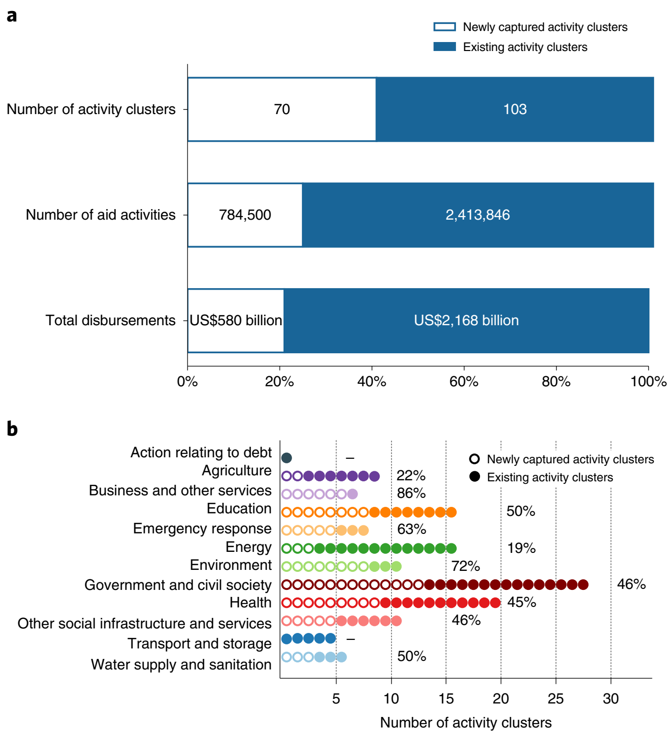


Fig. 3 | Overview of activity clusters representing 'newly captured' activity clusters. a, Summary statistics of newly captured versus existing aid activities (number of activity clusters, number of aid activities and total disbursements). **b**, Newly captured activity clusters grouped by sectors of aid. Percentages indicate the share of new activity clusters within each sector of aid.

change adaptation have more than doubled since the adoption of the Paris Agreement in 2015. Above all, this increase was channelled towards lower–middle-income countries, least developed countries and small island developing states—country groups that were designated in the Paris Agreement as extremely vulnerable to the impacts of climate change. In contrast, financial disbursements for climate change adaptation attributed to upper–middle-income countries have remained fairly constant. Notably, total aid disbursements for the whole environment sector, which, besides climate change adaptation, also includes other topics such as air pollution or biodiversity, even slightly decreased after adoption of the Paris Agreement in 2015 (Supplementary Fig. 10). On the one hand, this demonstrates that the importance of climate change adaptation has increased after the Paris Agreement. On the other hand, it shows that overall efforts by donors to protect the environment and the planet have been stagnating since the Paris Agreement and that donor countries may fail to meet their pledges.

Monitoring development aid based on activity clusters allows policymakers and researchers to analyse the global distribution of development aid in different contexts and promote coherent policy responses in cases where global distributions of aid are considered imbalanced revealing considerable disparities. The global distributions of aid from the remaining activity clusters of the environment sector are reported in Supplementary Fig. 10. An interactive map with all activity clusters can be accessed online: https://maltetoetzke.github.io/Monitoring-Global-Development-Aid/.

Discussion

We show how machine learning can enable a cost-effective alternative to manual categorizations with the aim of monitoring global development aid activities. Following a bottom-up approach, our machine learning framework helps to overcome important limitations in existing monitoring by providing larger flexibility and higher granularity for analysing aid activities. As such, our framework generates new and more timely empirical insights into the spatio-temporal distribution of development aid and thereby provides important evidence, which policymakers can use to locate and alleviate disparities in the global distribution of development aid. Furthermore, the generated data provide valuable input for further downstream analyses such as empirical studies investigating relations between development aid and outcomes of the SDGs[30–32].

The activity clusters generated by our machine learning framework show important topics of aid which are not yet captured by existing monitoring. This includes topics that only recently gained attention. By analysing these topics, trends in global development can be revealed over time. For example, many newly captured activity clusters are in line with the Paris Agreement as they address pressing issues related to climate change. Others follow the agenda on inequality of the SDGs by promoting inclusion and empowerment of underrepresented and vulnerable groups. Furthermore, we see a trend of newly captured activity clusters that aim to support the private sector in developing countries, for example, through microfinance, business development and entrepreneurship.

Our machine learning framework can be used by different stakeholders. On the one hand, donors can use our framework to review their portfolios of aid activities and coordinate activities among each other, thereby making more informed decisions about allocations of financial resources based on synthesized evidence. In this context, the high granularity of our activity clusters allow for the identification of developing countries that lack sufficient financial support for achieving specific targets of the SDGs. On the other hand, recipients can use our framework to gain a consistent overview of domestic and foreign activities related to sustainable development, thereby helping them to ensure that both funding streams are in line with each other. Our machine learning framework can also be used by international organizations or non-governmental organizations, which are not directly involved as donors or project partners in aid activities but ensure accountability (for example, OECD, AdaptationWatch). For such organizations, our framework offers new possibilities to evaluate global aid activities without reliance on donor-based categorizations, which also reduces cost and time needed for collecting data from multiple sources. Finally, our machine learning framework can be tailored to the needs of different organizations by providing more granular activity clusters for domain-specific organizations (for example, the Green Climate Fund or UNAIDS; Supplementary Discussion 4 provides details).

Our machine learning framework has limitations with regard to data and method. Current textual descriptions of aid activities provided by donor organizations vary in quality and length, which may represent a source of bias and limit accurate representations of aid activities. Nevertheless, a strength of the framework is that it can handle non-standardized texts of varying lengths. This allows donors to leverage existing texts such as internal project reports as input to the framework instead of compiling new activity descriptions that are dedicated specifically to a reporting system. Likewise, generic or ambiguous topics naturally impede high levels of topic granularity and thereby affect the interpretability of the resulting activity clusters (for example, electrification from unspecified sources versus solar energy). Furthermore, we acknowledge that the nature of our bottom-up approach precludes the use of 'ground truth' labels to train our framework. On the one hand, this is a major advantage offering flexibility and ex post categorization. On the other hand, it necessitates expert interpretation and thorough validations of outputs. Finally, these limitations emphasize the importance of high-quality input texts and human input and validation before deployment (Supplementary Discussion 4).

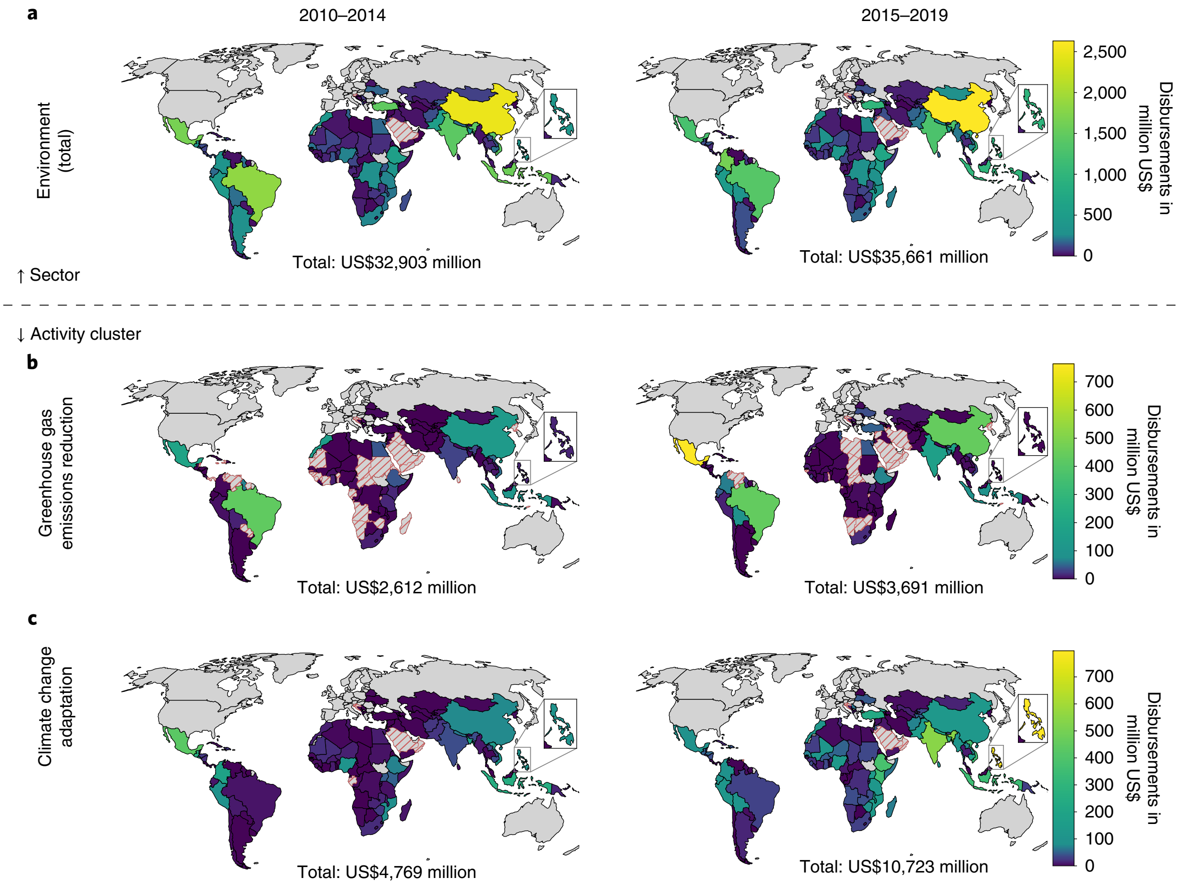


Fig. 4 | Global distribution of development aid allocated to recipient countries. a–c, Comparison of the global distribution of aid between aid data at the sector level and for underlying activity clusters based on the examples of the environment sector (**a**) and the two underlying activity clusters GREENHOUSE GAS EMISSIONS REDUCTION (**b**) and CLIMATE CHANGE ADAPTATION (**c**). Countries marked with red dashes are countries that were eligible for development aid but received no development aid for the respective activity cluster. The maps were created in Python3.6 using GeoPandas 0.7.0 (https://geopandas.org).

The use of machine learning offers the promise of automated updates as new data arrive. However, doing so is not straightforward, and there are two options which present a trade-off between adaptability to new topics and consistency. One option is to use the initially trained model and predict the assignment of activity clusters for textual descriptions of new aid activities. This represents a scalable option that can be operated in real time without human intervention. However, it holds the set of activity clusters fixed, thus limiting the ability to detect emerging topics (for example, COVID-19-related activities in 2020). Another option is to retrain the framework so that it is adaptive to emerging trends. However, the resulting activity clusters may not be fully consistent with the previous categorization. We expect that a combination of the two options is used in practice. That is, a new model is trained every five to ten years while predictions from the first options are used for annual updates.

Prominent examples[25,33–36] show how machine learning from research is increasingly put into practice by development institutions thus providing powerful and cost-effective approaches for evidence-based decision making. Here we add by offering a machine learning framework for monitoring global development aid.

Methods

Data. The underlying dataset for our study was provided by the Development Co-operation Directorate of the OECD and represents the raw data from the creditor reporting system (CRS), which is considered the most comprehensive data on global development aid[18]. The dataset is a collection of global development aid activities conducted between 2000 and 2019. Aid activities include ‘projects and programmes, cash transfers, deliveries of goods, training courses, research projects, debt relief operations and contributions to non-governmental organizations’[37]. Aid activities are reported by donor organizations at the end of each year in which they are conducted. In the reports, donor organizations provide attributes for each aid activity such as year, donor organization, recipient country and volume of aid disbursements in US dollars and a textual description of the conducted activity. If an aid activity proceeds over multiple years, it is reported each year with the corresponding aid disbursement of that year (details on reporting standards are given in[38]). We used aid activities from 2000–2017 for training and 2018–2019 for prediction to show the applicability of our framework to new data.

Our data comprise 3,944,540 aid activities financed by 757 different donor organizations. Donors include 595 agencies from the 30 member states of the Development Assistance Committee (DAC) and 25 other non-DAC countries. Furthermore, the dataset comprises 162 multilateral donor agencies (for example, Global Environment Facility Trust Fund) and 28 private donors (for example, Bill & Melinda Gates Foundation). The largest share of aid activities is financed through official development assistance (ODA) grants (84 %). Our data also cover other financing sources such as ODA loans, private development finance and equity investments. According to the OECD DAC, recipient countries in the dataset are all countries that are eligible to receive ODA. The 160 recipient countries include low-income countries, lower–middle-income countries and upper–middle-income countries. An overview of aid donors, income groups of recipient countries and types of finance is presented in Supplementary Tables 1–3.

Aid activities are reported with textual descriptions from donor organizations. This includes a *ProjectTitle*, a *ShortDescription* and a *LongDescription*, which we concatenate to a single textual input string for our machine learning model. The textual descriptions are mostly written in the official language of the donor organization and they vary in length with a median length of 160 words and a maximum length of 3,752 words. Due to missing entries in the textual descriptions, we excluded 745,546 aid activities (18.9%) from our sample, leaving 3,198,994 aid activities as input to our machine learning framework. However, approximately two-thirds (60.8%) of excluded activities concerned the earlier period from 2000–2006. Exemplary textual descriptions of aid activities are provided in Supplementary Table 8.

Machine learning framework. We developed a machine learning framework that categorizes aid activities based on their reported textual descriptions. Due to the deployment of clustering techniques, the resulting labels are named 'activity clusters'.

The framework is presented in Supplementary Fig. 11. It consists of five steps. In step 1, we translate and preprocess the textual descriptions of aid activities (*translation and preprocessing*). In step 2, we train an embedding by converting textual descriptions into numerical representations in the form of document vectors (*document embedding*). In step 3, we cluster the document vectors (*clustering*). In step 4, we name activity clusters together with domain experts (*naming*). In step 5, we externally validate activity clusters through a user study and comparison with existing aid data (*validation*).

Step 1: Translation and preprocessing. In the dataset, 800,423 textual descriptions were written in a language different from English. These descriptions were translated into English as follows. First, the source language of each textual description was classified. To obtain robust estimates, an ensemble language classifier was constructed based on three common language classifiers that use different methods for language detection (spacy[39], nltk[40] and textcat[41]). Second, all texts from the classified source language were translated to English via the Google Translate API[42].

The following preprocessing steps from natural language processing were applied: we converted all words into lowercase, removed stop words and applied a lemmatization to reduce the size of the vocabulary. Furthermore, words referring to geographical regions were replaced by codes of the respective word–entity denotations ('gpe' for countries, cities and states; 'norp' for nationalities) to prevent document vectors from differentiating between them instead of differentiating between the topics of activity descriptions.

Step 2: Document embedding. We selected a paragraph vector model[43] to embed the textual descriptions of aid activities into document vectors. The model has been proven to perform exceptionally well in capturing semantic meaning when applied to documents of highly varying lengths including very short and very long texts[44,45] as in the case of the textual descriptions in our dataset.

The paragraph vector model fits a document vector for each input document, which can be interpreted as 'memory' capturing the content of the document. Each document vector provides a numerical representation of the document's content by locating it as a point in a higher dimensional vector space. Similar documents are closer together in the vector space, while dissimilar documents are distant. The paragraph vector model uses an unsupervised learning approach. It optimizes the prediction of the next words in a text using a sliding window. As in the word vector model proposed by Mikolov, Sutskever, Chen, Corrado and Dean[46], the concatenation of the word vectors is used as input features. In addition to n preceding word vectors, a document vector that provides the context information of the document is added to predict the next word. More formally, the paragraph vector model maximizes the following log probability

$$\frac{1}{T}\sum_{\mathrm{t=k}}^{T-\mathrm{k}}\log p(w_{\mathrm{t}}\mid w_{\mathrm{t-k}},\ldots,w_{\mathrm{t+k}},D), \qquad (1)$$

where T is the size of the sliding window and k half the size of the sliding window ($k=n/2$), p the probability of the word w_{t} at the position t in the sentence based on the sequence of words $w_{\mathrm{t\text{-}k}},\ldots,w_{T\text{-}k}$ and D which is the document vector of the document.

The following hyperparameters were used. We selected a vector size of 200 elements for document and word vectors and a sliding window of six words. Every word had to appear at least 30 times in all input texts to be embedded. We applied a hierarchical softmax[47] and stochastic gradient descent[48]. Word vectors and document vectors were fitted simultaneously as this has been proven to enhance performance[44]. Our paragraph vector model was pretrained on a large dataset (ten epochs on a Wikipedia training corpus of ~12 million texts). Subsequently, trained word vectors were kept while resulting document vectors from the Wikipedia corpus were discarded. Finally, we trained the paragraph vector model, including the pretrained word vectors, over 25 epochs on our aid dataset and extracted the resulting document vectors of each aid activity.

We also considered other possible models for document embedding. A discussion of these models can be found in Supplementary Discussion 4.

Step 3: Clustering. We clustered the document vectors of activity descriptions as follows. We selected a k-means clustering algorithm based on the specific characteristics of the clustering task (for example, high dimensionality of vectors, large number of clusters, large sample size). The performance of the k-means algorithm is dependent on the number of resulting clusters, k, as it affects the level of similarity between documents from the same cluster and the level of dissimilarity between documents from different clusters. We determined the number of clusters, k, based on the silhouette score[49], which compares the mean intracluster distance to the mean nearest-cluster distance for each data point. The range of possible k clusters was set between 150 and 250 to strike a balance between granularity and interpretability. Comparing the computed silhouette scores of each clustering with k clusters, the optimal number of clusters within our selected range resulted in $k=178$ (note that 10 activity clusters were merged in post-processing due to similar topics as described in 'Step 4: Naming').

We trained the k-means algorithm on the document vectors with a k-means++centroid initialization[50] and ten different centroid seed initializations. The distance metric was set to the cosine distance to address the problem of inflating distance in high-dimensional spaces[51].

Step 4: Naming. Each cluster was provided with a suitable name using the following systematic procedure: (1) for each cluster, we computed and evaluated 20 characteristic words, (2) we analysed n>500 random samples of textual descriptions per cluster and (3) we checked names of activity clusters with domain experts.

To extract the most characteristic words from each cluster, we applied the tf–idf (term frequency–inverse document frequency) statistic[52]. First, all textual descriptions of each cluster were concatenated to a single document. Second, the tf–idf statistic was computed to give a weight for every word in a given cluster. Each weight reflects the importance of the respective word in the given cluster by comparing how often the word appears in the cluster (term frequency, tf) to the number of clusters that include the word (document frequency, df). For a given term, t, in a document, d, the tf–idf statistic is defined as

$$\text{tf-idf}_{t,d}=\text{tf}_{t,d}\cdot\text{idf}_t \qquad \text{with} \quad \text{idf}_t=\log\frac{N}{1+\text{df}}. \qquad (2)$$

Third, the words of each cluster were ranked based on the corresponding tf–idf weights. Fourth, we extracted the 20 characteristic words with the highest tf–idf weights from each cluster as characteristic words.

On the basis of the evaluations of the above 20 characteristic words and additional random samples of activity descriptions, we provided names to activity clusters. If the topic already existed as a category in other existing monitoring, we adopted the existing name. For all remaining clusters, we proposed a name for the activity cluster. Five times, we merged two clusters into one because after closer inspection, we found that they represented the same topic. These clusters were HUMAN RIGHTS, LABOR RIGHTS AND CONDITIONS, TRANSPARENT DEMOCRATIC INSTITUTIONS, RECREATION AND SPORT and INCLUSIVE GROWTH AND DEVELOPMENT.

Through a questionnaire, we asked nine domain experts whether they agreed with the names of activity clusters belonging to a specific sector of aid (for example, health, energy, education). Domain experts included professors, postdoctoral researchers and doctoral researchers from academia conducting research related to topics in the activity clusters they were asked to evaluate. For each activity cluster, they were provided the 20 characteristic words and five randomly drawn textual descriptions. On the basis of this information, they were asked whether the proposed name for the activity cluster matched the topic of corresponding aid activities and whether they would propose corrections. In total, we adjusted the names of six activity clusters from the sectors energy, education, water supply and sanitation, and agriculture based on answers from the questionnaire.

Step 5: Validation. We used a dual approach to externally validate activity clusters. Following best practice from related work[25,53], we (1) conducted a large-scale user study and (2) made a comparison with existing aid data at the sector level. Both validations demonstrate that the machine learning framework performs well in inferring topics from activity descriptions and in assigning aid activities to activity clusters. A direct verification of the overall accuracy of our machine learning framework is technically not possible because there exist no 'ground truth' labels with sufficient granularity.

In the user study, activity clusters were manually validated regarding the coherence of topics and the fit between textual descriptions and assigned activity clusters (details are explained in Supplementary Discussion 2). To validate topic coherence, human respondents were asked to identify an intruding word from a set of characteristic words for each activity cluster (*word intrusion*). To validate the assignment of activity clusters, respondents were provided activity descriptions and asked to identify the activity cluster assigned by the algorithm among other intruding activity clusters (*topic intrusion*; as we have only one topic assigned per document, we adapted the question of the topic intrusion test suggested by[53], accordingly). Each of the two tasks included 173 questions with one question per activity cluster. Results demonstrate a high coherence for the topics of activity clusters as in 93.6% of the tasks, respondents correctly identified the intruding word. Furthermore, results present a high accuracy for assignments of aid activity descriptions to activity clusters as in 94.0% of the tasks, respondents identified the correct activity clusters among intruding activity clusters (Supplementary Fig. 2).

Data of activity clusters were aggregated at the sector level (for example, health, energy, education) to compare them to sector data from the CRS[17]. This way, it was validated if our machine learning framework reflects a similar understanding of aid activities as reported by donors to the CRS (details are explained in Supplementary Discussion 3). Sector data of the CRS are classified and reported by donor agencies and represent a comprehensive collection of all donor-curated databases such as the Foreign Aid Explorer of USAIDS[54]. Results show that our machine learning framework reconstructs the sector distribution of aid activities from the CRS (Pearson correlation: $r = 0.98$ (aid activities), $r = 0.96$ (aid disbursements); Supplementary Fig. 3). Furthermore, the large majority of aid activities have been assigned to identical sectors (accuracy = 0.83; $F1$ score = 0.81, representing the harmonic mean of precision and recall). Some sectors account for a lower accuracy, which can be explained by their close relatedness to other sector categories (for example, between population policy and health; between business and industry). Thus, assignments between the CRS and the machine learning framework can be different while both being correct. Overlaps between sectors are discussed in Supplementary Discussion 3 and plotted in Supplementary Fig. 4. The validation included all aid activities conducted between 2015 and 2019, totalling to 1,093,065 aid activities.

Finally, we should mention that we explicitly refrained from mapping aid activities directly onto SDG targets for two reasons: first, compared with SDG targets (for example, eradicate extreme poverty), activity clusters (for example, SOCIAL PROTECTION POLICIES, HEALTH INSURANCE) are more specific, which helps policymakers and donors to turn evidence into concrete policies. Second, development aid and the SDGs are *not* subject to functional relationships. That is, one aid activity can address several SDG targets and, vice versa, one SDG target can be addressed through different activities.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Activity clusters can be monitored interactively via https://maltetoetzke.github.io/Monitoring-Global-Development-Aid/. The underlying data can be retrieved via https://github.com/MalteToetzke/Monitoring-Global-Development-Aid-With-Machine-Learning. For access to the raw data, please contact the DAC of the OECD.

Code availability

The scripts used for preprocessing the data and generating activity clusters can be retrieved via https://github.com/MalteToetzke/Monitoring-Global-Development-Aid-With-Machine-Learning. Analysis scripts are available on request from M.T.

Received: 19 November 2020; Accepted: 7 March 2022;
Published online: 11 April 2022

References

1. Liu, J. et al. Systems integration for global sustainability. *Science* **347**, (2015).
2. *Sustainable Development Goals: The Sustainable Development Agenda* (United Nations, 2015); https://www.un.org/sustainabledevelopment/development-agenda/
3. *The Sustainable Development Goals Report 2018* (United Nations, 2018); https://unstats.un.org/sdgs/report/2018/
4. Global Indicator Framework for the Sustainable Development Goals and Targets of the 2030 Agenda for Sustainable Development (United Nations, 2019); https://unstats.un.org/sdgs/indicators/indicators-list/
5. *World Investment Report 2014; Investing in the SDGs: An Action Plan* (United Nations, 2014); https://unctad.org/en/PublicationsLibrary/wir2014_en.pdf
6. *Development Co-operation Report 2018: Joining Forces to Leave No One Behind* (OECD, 2018); http://www.oecd.org/social/development-co-operation-report-20747721.htm
7. *Development Co-operation Report 2019: A Fairer, Greener, Safer Tomorrow* (OECD, 2019); http://www.oecd.org/dac/development-co-operation-report-20747721.htm
8. Nunnenkamp, P., Öhler, H. & Thiele, R. Donor coordination and specialization: did the Paris declaration make a difference? *Rev. World Econ.* **149**, 537–563 (2013).
9. Easterly, W. & Pfutze, T. Where does the money go? Best and worst practices in foreign aid. *J. Econ. Perspect.* **22**, 29–52 (2008).
10. Clemens, M. A., Kenny, C. J. & Moss, T. J. The trouble with the MDGs: confronting expectations of aid and development success. *World Dev.* **35**, 735–751 (2007).
11. Kenny, C. *What is effective aid? How would donors allocate it?* (World Bank, 2006).
12. Tierney, M. J. et al. More dollars than sense: refining our knowledge of development finance using AidData. *World Dev.* **39**, 1891–1906 (2011).
13. Pitt, C., Grollman, C., Martinez-Alvarez, M., Arregoces, L. & Borghi, J. Tracking aid for global health goals: a systematic comparison of four approaches applied to reproductive, maternal, newborn, and child health. *Lancet Glob. Health* **6**, 859–874 (2018).
14. *Toward Mutual Accountability: The 2015 Adaptation Finance Transparency Gap Report* (Adaptation Watch, 2015).
15. *State of Inequality: Reproductive Maternal Newborn and Child Health; Interactive Visualization of Health Data* (World Health Organization, 2015).
16. Flogstad, C. & Hagen, R. J. Aid dispersion: measurement in principle and practice. *World Dev.* **97**, 232–250 (2017).
17. Creditor reporting system 2019. OECD Statistics https://stats.oecd.org/DownloadFiles.aspx?DatasetCode=CRS1 (2022).
18. *Comparative Study of Data Reported to the OECD Creditor Reporting System (CRS) and to the Aid Management Platform (AMP)* (OECD, 2009).
19. *Purpose Codes: Sector Classification* (OECD, 2021); https://www.oecd.org/development/financing-sustainable-development/development-finance-standards/purposecodessectorclassification.htm
20. Burke, M., Driscoll, A., Lobell, D. B. & Ermon, S. Using satellite imagery to understand and promote sustainable development. *Science* **371**, (2021).
21. Kinyoki, D. K. Mapping child growth failure across low-and middle-income countries. *Nature* **577**, 231–234 (2020).
22. Local Burden of Disease Educational Attainment Collaborators Mapping disparities in education across low-and-middle-income countries. *Nature* **577**, 235–238 (2020).
23. Ricciardi, V. et al. A scoping review of research funding for small-scale farmers in water scarce regions. *Nat. Sustain* **3**, 836–844 (2020).
24. Xie, M., Jean, N., Burke, M., Lobell, D. & Ermon, S. Transfer learning from deep features for remote sensing and poverty mapping. In *Proc. 30th AAAI Conference on Artificial Intelligence* (AAAI Press, 2016).
25. Blumenstock, J., Cadamuro, G. & On, R. Predicting poverty and wealth from mobile phone metadata. *Science* **350**, 1073–1076 (2015).
26. Nature Editorial How science can put the Sustainable Development Goals back on track. *Nature* **589**, 329–330 (2021).
27. Glossary of statistical terms: sector of destination (of aid). OECD Statistics https://stats.oecd.org/glossary/detail.asp?ID=6808 (2005).
28. GHG data from UNFCCC. UNFCCC https://unfccc.int/process-and-meetings/transparency-and-reporting/greenhouse-gas-data/ghg-data-unfccc/ghg-data-from-unfccc (2021).
29. Adoption of the Paris Agreement FCCC/CP/2015/L.9/Rev.1 (UNFCCC, 2015).
30. Glennie, J. & Sumner, A. *Aid, Growth and Poverty* (Springer, 2016).
31. Qian, N. Making progress on foreign aid. *Annu. Rev. Econ.* **7**, 277–308 (2015).
32. Jakubik, J. & Feuerriegel, S. Data-driven allocation of development aid towards sustainable development goals: evidence from HIV/AIDS, *Production and Operations Management* (2022).
33. About us. *World Food Programme Innovation Accelerator* https://innovation.wfp.org/about-us (2021).
34. About givedirectly. *GiveDirectly* https://www.givedirectly.org/about/ (2021).
35. Adelman, M., Haimovich, F., Ham, A. & Vazquez, E. Predicting school dropout with administrative data: new evidence from Guatemala and Honduras. *Educ. Econ.* **26**, 356–372 (2018).
36. Calantropio, A., Chiabrando, F., Codastefano, M. & Bourke, E. Deep learning for automatic building damage assessment: application in post-disaster scenarios using UAV data. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **1**, 113–120 (2021).
37. Glossary of statistical terms: aid activities. OECD Statistics https://stats.oecd.org/glossary/detail.asp?ID=6807 (2005).
38. *Development Finance Standards* (OECD, 2020); http://www.oecd.org/dac/financing-sustainable-development/development-finance-standards/
39. spacy-langdetect (SpaCy, 2019); https://spacy.io/universe/project/spacy-langdetect
40. Natural language toolkit (NLTK, 2019); https://www.nltk.org/
41. Hornik, K., Rauch, J., Buchta, C. & Feinerer, I. textcat: N-Gram Based Text Categorization. R version 3.2.0 https://cran.r-project.org/web/packages/textcat/textcat.pdf (2018).
42. Cloud translation API (Google Cloud, 2019); https://cloud.google.com/translate/docs/reference/rest/
43. Le, Q. & Mikolov, T. Distributed representations of sentences and documents. *Proc. Mach. Learn. Res.* **32**, 1188–1196 (2014).
44. Dai, A. M., Olah, C. & Le, Q. V. Document embedding with paragraph vectors. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1507.07998 (2015).
45. Campr, M. & Ježek, K. in *International Conference on Text, Speech, and Dialogue* (eds. Král, P. & Matoušek, V.) 252–260 (Springer, 2015).
46. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems* (eds. Burges, C. J. C. et al.) 3111–3119 (Curran Associates, Inc., 2013).
47. Goodman, J. Classes for fast maximum entropy training. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*. 561–564 (IEEE, 2001).
48. Bottou, L. Large-scale machine learning with stochastic gradient descent. In *Proc. of COMPSTAT'2010* (eds. Lechevallier, Y. & Saporta, G.) 177–186 (Springer, 2010).
49. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
50. Arthur, D. & Vassilvitskii, S. *k-means++: The Advantages of Careful Seeding* (Stanford Univ., 2006).
51. Dhillon, I. S. & Modha, D. S. Concept decompositions for large sparse text data using clustering. *Mach. Learn.* **42**, 143–175 (2001).
52. Wu, H. C., Luk, R. W. P., Wong, K. F. & Kwok, K. L. Interpreting tf–idf term weights as making relevance decisions. *ACM Trans. Inf. Syst.* **26**, 1–37 (2008).

53. Chang, J., Boyd-Graber, J., Wang, C., Gerrish, S. & Blei, D. M. Reading tea leaves: how humans interpret topic models. *Adv. Neural Inf. Process. Syst.* **32**, 288–296 (2009).
54. *Foreign Aid Explorer* (USAID, 2021); https://explorer.usaid.gov/
55. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

Acknowledgements

We thank the SDG financing lab of the OECD for the provision of the raw data and the mutual exchange over the course of this study. Furthermore, we would like to thank all researchers from the Swiss Federal Institute of Technology (ETH Zurich) who helped us in evaluating and naming activity clusters.

Author contributions

M.T. performed data analysis and visualized the results. All the authors contributed to the conceptualization, interpretation of the results and the writing of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41893-022-00874-z.

Correspondence and requests for materials should be addressed to Malte Toetzke.

Peer review information *Nature Sustainability* thanks Max Callaghan, Lynn Kaack and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Stefan Feuerriegel

Last updated by author(s): Feb 22, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
□	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
□	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	□	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	□	A description of all covariates tested
☒	□	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
□	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	□	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	□	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	□	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
□	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used
Data analysis	As programming Language, python 3.6 was used. For language classification of textual descriptions the following open source packages were used: spacy-langdetect 0.1.2; nltk 3.4.5; textcat 1.0-6 (for RStudio Desktop 1.0.153). For text translation the Google Translate API was used: google-cloud-translate 2.0.0. For text preprocessing the following packages were used: nltk 3.4.5, spacy 2.2.1, RegEx (build in python 3.6). For document embedding Gensim was used: gensim 3.8.0. For clustering scikit-learn was used: scikit-learn 0.21.3

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Accession codes, unique identifiers, or web links for publicly available datasets made available upon publication

April 2020

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study uses machine learning (specifically natural language processing and clustering) to analyze what topics are addressed in global development aid activities and provide a comprehensive monitoring that maps where and when development aid was allocated for which topics.
Research sample	The study includes all aid activity descriptions collected by the OECD from the bilateral, multilateral and private donor organizations reporting to the DAC (757 organizations). Details can be found in the Methods.
Sampling strategy	NA
Data collection	Data was provided from a cooperation with the SDG Financing Lab of the OECD DAC. Details on the data collection by the OECD DAC are provided in the methods.
Timing	Data used in this study was collected from 2000--2019.
Data exclusions	We do not include aid activities with missing project description (745,546 aid activities which represent 18.9% of the data).
Non-participation	NA
Randomization	NA

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging