

Survey data of "Mapping Research output to the SDGs"



Table of Contents

Metadata.....	2
Dataset Content:.....	3
Dataset Takeaways:	4
License and Attribution, Acknowledgements and how to Cite	5
Acknowledgements.....	5
Please cite this data set as follows:	5
License Attribution when reusing this data:	5
Example Usage, some results:	5
Contributors, full list	7

IN PARTNERSHIP WITH:



Metadata

meta	data
description	Survey data of "Mapping Research Output to the SDGs" presented here is the comprehensive raw, aggregated and processed output of a survey that ran from October 2019 till January 2020. The goal of the survey was to collect data from (senior) researchers, to add and reflect their conceptual knowledge of their specific research domain that contributes to the SDG (Sustainable Development Goals). The outcome of the survey data can be used for validating current and optimizing future SDG classification model for research output.
date	2020-04-23
authors	Maurice Vanderfeesten (VUA); Eike Spielberg (UDE); Yasin Gunes (VUA);
contributors	Alessandro Arienzo (UNA); Roberto delle Donne (UNA); Ignasi Salvadó Estivill (URV); José Luis González Ugarte (URV); Didier Vercueil (UGA); Nykohla Strong (UAB); Eike Spielberg (UDE); Felix Schmidt (UDE); Linda Hasse (UDE); Ane Sesma (UEA); Baldvin Zariok (UIC); Friedrich Gaigg (UIN); René Otten (VUA); Nicolien van der Grijp (VUA); Yasin Gunes (VUA); Maurice Vanderfeesten (VUA)
related material	Aurora SDG queries version 4
file formats	data files are all in comma separated files (csv) in UTF-8 encoding.
Software used	Data collector: Aurora Data Collector , Survey tool: Survey Gizmo , Data aggregation: Elastic Search + Kibana , Data wrangling: Python , Jupyter Notebook , Elastic Painless , KNIME analytics platform

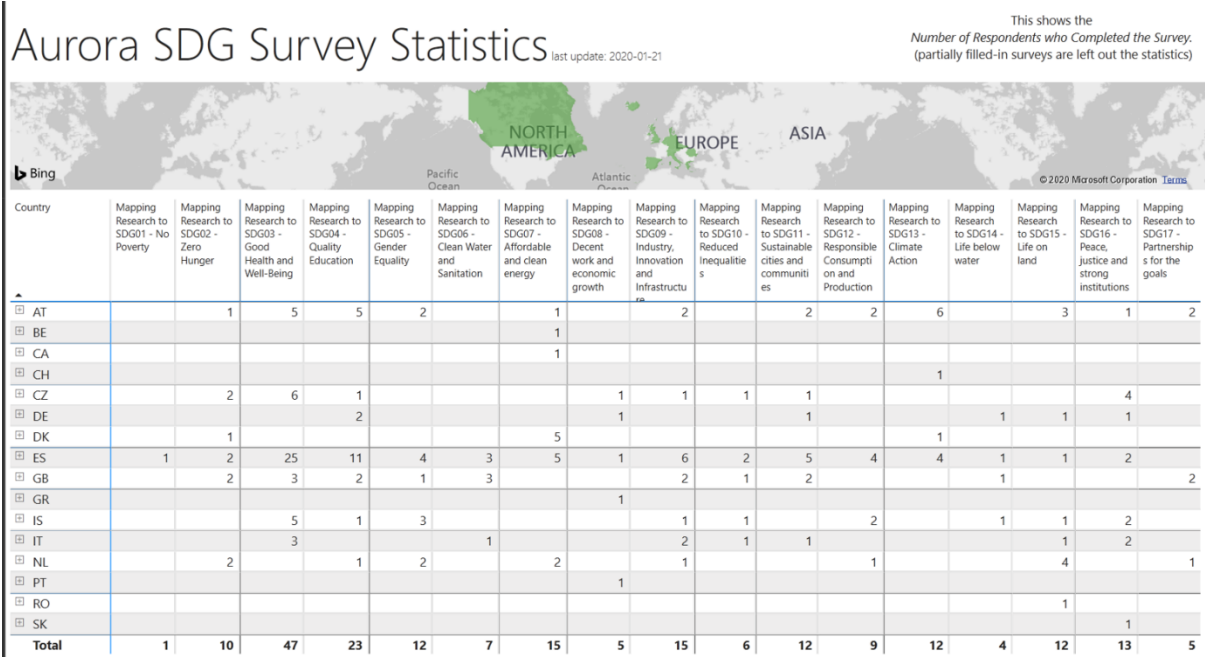
Dataset Content:

Directories and files. Privacy sensitive data is removed in public release version of the data. All sub-directories contain [README.md](#) files explaining what data you see and how to interpret.

/dir/ file.*	description
/00-survey-input/	This contains the survey questions for all the individual SDGs. It also contains lists of EIDs categorised to the SDGs we used to make randomized selections from to present to the respondents.
/01-raw-data/	This contains the raw survey output. (Excluding privacy sensitive information for public release.) This data needs to be combined with the data on the provisioning server to make sense.
/02-aggregated-data/	This data is where individual responses are aggregated. Also the survey data is combined with the provisioning server, of all sdg surveys combined, responses are aggregated, and split per question type.
/03-scripts/	This contains scripts to split data, and to add descriptive metadata for text analysis in a later stage.
/04-processed-data/	This is the main final result that can be used for further analysis. Data is split by SDG into subdirectories, in there you'll find files per question type containing the aggregated data of the respondents.
/images/	images of the results used in this document.
LICENSE.md	terms and conditions for reusing this data.

Dataset Takeaways:

- data collection period: Oct 2019 till Jan 2020
- respondents: 224 from 16 countries in Europe and North America
- questions asked for each SDG:
 - Suggest research papers that are relevant for this SDG (upload list)
 - Select research papers that are relevant for this SDG (radio buttons: accept, reject)
 - Select and Suggest Keywords related to SDG (checkboxes: accept | text field: suggestions)
 - Suggest SDG related glossaries with relevant keywords (text fields: url)
 - Select and Suggest Journals fully related to SDG (checkboxes: accept | text field: suggestions)
 - Suggest improvements for the current queries (text field: suggestions per target)



License and Attribution, Acknowledgements and how to Cite

Acknowledgements

We would like to thank the presidents of the Aurora Universities for their support and executive representation in this project. Also we would like to thank all researchers involved to have shared their expertise. Many people from within the Aurora University Network were involved making this survey possible.

If you want to tribute this hard work, please reuse this data to create or improve your services and share your outcomes.

Do so by respecting the license and attribute the contributors.

Please cite this data set as follows:

Survey data of "Mapping Research output to the SDGs" by Aurora Universities Network (AUR)
[doi:10.5281/zenodo.3798386](https://doi.org/10.5281/zenodo.3798386)

License for reuse:

[Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

License Attribution when reusing this data:

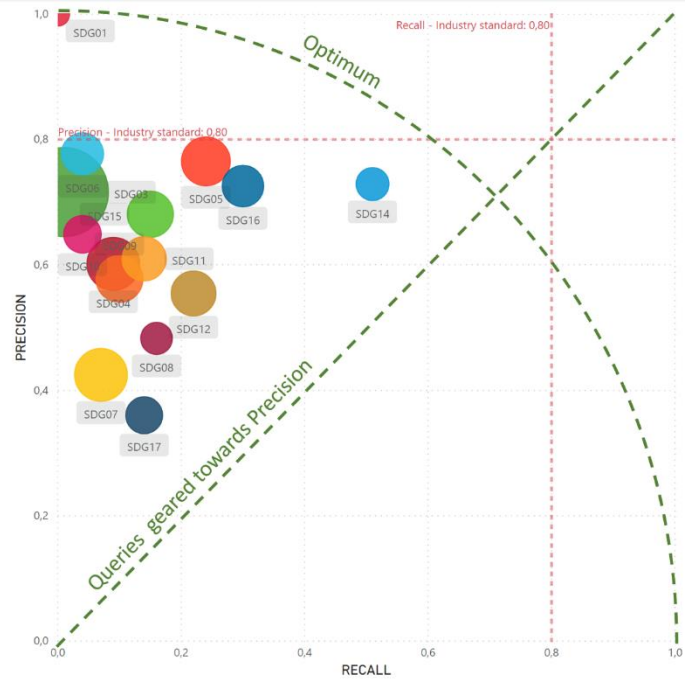
Survey data of "Mapping Research output to the SDGs" by Aurora Universities Network (AUR); Alessandro Arienzo (UNA); Roberto delle Donne (UNA); Ignasi Salvadó Estivill (URV); José Luis González Ugarte (URV); Didier Vercueil (UGA); Nykohla Strong (UAB); Eike Spielberg (UDE); Felix Schmidt (UDE); Linda Hasse (UDE); Ane Sesma (UEA); Baldvin Zarióh (UIC); Friedrich Gaigg (UIN); René Otten (VUA); Nicolien van der Grijp (VUA); Yasin Gunes (VUA); Peter van den Besselaar (VUA); Joeri Both (VUA); Maurice Vanderfeesten (VUA); is licensed under a Creative Commons Attribution 4.0 International License. <https://aurora-network.global/project/sdg-analysis-bibliometrics-relevance/>

Example Usage, some results:

We used the data for creating a baseline for the SDG classification model, and secondly to gather information on improving a next version.

- precision and recall: how good is our SDG classification model (SDG queries v4)
- suggestions for improving the queries to version 5.

SDG-queries v4: Precision and Recall (and sample sizes)



SDG full name	PRECISION A/(R+A)	RECALL (SPRS/SP)	F1-SCORE 2* ((PRECISION*R ECALL)/(PRECIS ION+RECALL))	accepted (A)	rejected (R)	Number of matched suggested publications between 2009 and 2019 (SP)	Number of suggested publications in SDG result set (SPRS)
SDG-01 no poverty	1,00	0,00	0,00	9	0	4	0
SDG-02 zero hunger	0,31			186	414	60	
SDG-03 good health and well-being	0,72	0,01	0,02	1878	745	976	9
SDG-04 quality education	0,60	0,09	0,15	620	411	80	7
SDG-05 gender equality	0,77	0,24	0,36	456	140	318	76
SDG-06 clean water and sanitation	0,78	0,04	0,08	265	76	278	11
SDG-07 affordable and clean energy	0,43	0,07	0,12	357	483	281	20
SDG-08 decent work and economic growth	0,48	0,16	0,24	98	105	37	6
SDG-09 industry, innovation and infrastructure	0,58	0,10	0,18	362	265	193	20
SDG-10 reduce inequalities	0,65	0,04	0,07	253	137	55	2
SDG-11 sustainable cities and communities	0,61	0,14	0,23	373	239	86	12
SDG-12 responsible consumption and production	0,55	0,22	0,31	342	275	110	24
SDG-13 climate action	0,65			367	201	76	
SDG-14 life below water	0,73	0,51	0,60	164	61	49	25
SDG-15 life on land	0,68	0,15	0,24	403	189	188	28
SDG-16 peace, justice and strong institutions	0,73	0,30	0,42	386	146	57	17
SDG-17 partnerships for the goals	0,36	0,14	0,20	141	250	29	4

Contributors, full list

Below you'll see the full list of contributors to this project. Without them this was not possible.

Family Name, First Name	University	ORCID	role
Vanderfeesten, Maurice	(VUA) Vrije Universiteit Amsterdam	0000-0001-6397-4759	Project leader
Spielberg, Eike	(UDE) University of Duisburg-Essen	0000-0002-3333-5814	Work package leader
Gunes, Yassin	(VUA) Vrije Universiteit Amsterdam	...	Project member
Arienzo, Alessandro	(UNA) University Federico II	0000-0002-2867-5363	Project member
Delle Donne, Roberto	(UNA) University Federico II	0000-0001-8331-9436	Project member
Salvadó Estivill, Ignasi	(URV) Universitat Rovira i Virgili	...	Project member
González Ugarte, José Luis	(URV) Universitat Rovira i Virgili	...	Project member
Vercueil, Didier	(UGA) Université Grenoble Alpes	...	Project member
Strong, Nykohla	(UAB) University of Aberdeen	0000-0002-6137-591X	Project member
Schmidt, Felix	(UDE) University of Duisburg-Essen	...	Project member
Hasse, Linda	(UDE) University of Duisburg-Essen	...	Project member
Sesma, Ane	(UEA) University of East Anglia	0000-0003-3982-8932	Project member
Zaríoh, Baldvín	(UIC) University of Iceland	0000-0001-9317-2597	Project member
Gaigg, Friedrich	(UIN) Universität Innsbruck	...	Project member
Otten, René	(VUA) Vrije Universiteit Amsterdam	0000-0002-6485-8810	Project member
Grijp, Nicolien van der	(VUA) Vrije Universiteit Amsterdam	0000-0002-5119-3514	Project member

Besselaar, Peter van den	(VUA) Vrije Universiteit Amsterdam	0000-0002-8304-8565	Supervisor
Both, Joeri	(VUA) Vrije Universiteit Amsterdam	...	Supervisor
Kouwenaar, Kees	AURORA University Network	...	Sponsor
Beukering, Pieter van de	(VUA) Vrije Universiteit Amsterdam	...	Sponsor