



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement num. 872396



ALLINTERACT

D1.1. Raw Data

April 30th, 2021



Project Number: 872396

Project Acronym: ALLINTERACT

Project title: Widening and diversifying citizen engagement in science



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement num. 872396



D1.1. Raw Data

ALLINTERACT project is committed to Open Science and, therefore, the deliverables produced within the project are part of the Open Research Data Pilot (OPDP). According to the GA (p.13) ALLINTERACT Raw Data will be submitted on M7 (April 2021) in open access. Following ALLINTERACT Data Management Plan (D7.2), this deliverable has been uploaded to Zenodo, an Open Access Repository that meets the requirements of FAIR Data of the project.

This document, which will be submitted on M7, will be periodically updated throughout the project in order to add new raw data that will be collected in further tasks of the project. The following table compiles all the information regarding the versioning and content of this deliverable, as well as the persistent identifiers generated in order to make data findable, accessible, interoperable and reusable.



Table. Versions of ALLINTERACT Raw Data to be uploaded

Version	Title	Content	Date	doi
1.0.0	D1.1.ALLINTERACT_RawData	- Raw Data obtained from Social Media Analytics in Twitter, Facebook, Instagram and Reddit	April 30 th 2021	https://doi.org/10.5281/zenodo.4729725
2.0.0	D1.1.ALLINTERACT_RawData	- Raw Data obtained from Social Media Analytics in Twitter, Facebook, Instagram and Reddit - Raw Data obtained from the Social Media Communicative Observation	To be collected on May 2021	
3.0.0	D1.1.ALLINTERACT_RawData	- Raw Data obtained from Social Media Analytics in Twitter, Facebook, Instagram and Reddit - Raw Data obtained from the Social Media Communicative Observation - Raw Data (anonymized transcript) obtained from the Focus Groups	To be collected on June 2021	
4.0.0	D1.1.ALLINTERACT_RawData	- Raw Data obtained from Social Media Analytics in Twitter, Facebook, Instagram and Reddit - Raw Data obtained from the Social Media Communicative Observation - Raw Data (anonymized transcript) obtained from the Focus Groups - Raw Data obtained from the survey	To be collected on November 2021	
5.0.0	D1.1.ALLINTERACT_RawData	- Raw Data obtained from Social Media Analytics in Twitter, Facebook, Instagram and Reddit - Raw Data obtained from the Social Media Communicative Observation - Raw Data (anonymized transcript) obtained from the Focus Groups - Raw Data obtained from the survey - Raw Data (anonymized transcript) obtained from the post-test Focus Groups	To be collected on September 2022	



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement num. 872396



On April 30th, 2021 the version 1.0.0 of the D1.1.ALLINTERACT_RawData has been uploaded. This dataset contains the raw data obtained from **social media interactions** (Twitter, Facebook, Instagram and Reddit) among citizens about **citizen participation in science** and **research with social impact** related to two **Sustainable Development Goals: Quality Education** and **Gender Equality**. The data collection has followed a twofold strategy 1) Top-Down, in which researchers identified and selected relevant Twitter and Instagram hashtags and Facebook and Reddit pages and 2) Bottom-Up, in which Twitter hashtags were selected based on daily Trending Topics.

The data was collected between March 9th and March 16th 2021 and has been obtained, cleaned and anonymized following Allinteract - Social Media Analytics Protocol (Flecha & Pulido, 2021). This dataset is part of EC **Horizon 2020** project ALLINTERACT Widening and diversifying citizen engagement in science (872396).

We provide five Excel files (one for each social network explored). Each file contains the main information of the extracted messages, however the information extracted in each case is slightly different.

1. **Twitter:** Row ID, Tweet ID, Tweet, Time, Tweet Type, Retweeted By, Number of Retweets, Hashtags, Number of Tweets, Number of Followers, Number Following
2. **Facebook:** Row ID, Post ID, Post, Link, Link Name, Link Caption, Link Description, Video, Type, Likes, Created Time, Updated Time, Comment ID, Comment Text, Comment Likes, Comment Time, Page Likes
3. **Instagram:** Url, content, likes, comments, date
4. **Reddit:** Row ID, sub_id, sub_title, sub_text, sub_score, sub_date, sub_link, comment_id, comment_body, comment_score, comment_date, comment_link

Funding: We acknowledge support of this work by the project "*ALLINTERACT Widening and diversifying citizen engagement in science*" (872396) from the European Commission Horizon 2020 programme.

References

Flecha, R., & Pulido, C. (2021). Allinteract - Social Media Analytics Protocol is licensed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License is available in https://archive.org/details/@crea_research

How to cite

Soler-Gallart, M. (2021). D1.1.Allinteract Raw Data is licensed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License