Open Science Policy

Deliverable #1.2

Version: 1.0 Final

Date: 20/06/2023

Authors:

Nikos Deligiannis

CHIST-ERA 2021:
Foundations for Misbehaviour Detection and Mitigation Strategies
in Online Social Networks and Media (OSNEM)

Version History

Version	Date	Author	Description
V0.1	25/05/2023	Nikos Deligiannis	Initial version
V0.2	28/05/2023	Nikos Deligiannis and Jone Paesmans	Comments and suggestions by Jone Paesmans addressed by N. Deligiannis
V0.3	20/06/2023	Fintan Mc Gee and Nikos Deligiannis	Comments and suggestions by Fintan Mc Gee addressed by N. Deligiannis
V0.3	20/06/2023	Fintan Mc Gee	Marked as final version 1.0

Contents

1 Introduction

Online misbehaviour means we cannot take the information received via social media at face value. Sharing of misleading, out-of-context or simply false information, and coordination of behaviours ('brigading') distorts the quality of information we receive online, regardless of whether the misbehaviour is intentional. Additionally, by design, each user of modern internet products (e.g., search; social media) is exposed to a unique view of the information landscape without necessarily realizing that this is the case. Every user should be able to answer the questions: "How am I being misinformed?" and "How does my unique position shape the information I receive?" Leveraging a multilayer network approach to detect misbehaviour and describe users' positions in the dynamic information space, the CON-NET project will enable users to answer these questions for themselves or indeed any other network entity.

There are many different factors that make detection and mitigation of online misbehaviour a challenge. The complexity of the networks themselves, the role of automated agents (bots), and coordination to create misleading signals, the motivations behind misbehaviour and the types of misbehaviour all play significant roles. The medium through which misinformation can spread is also an important feature, adding further layers to the complexity. Doctored images and videos can spread through social media along with the ideas and memes of misinformation.

CON-NET will apply machine learning techniques to process the vast amounts of data and identify trends, signals, and suspicious behaviours as well as misbehaving entities, and we will further augment this with a visual analytics approach, including a human in the loop to provide context and human understanding of the concept of misinformation. We will address the complexity of online social media networks by taking a multilayer network approach to their analysis. The aspects for the network will characterise the different platforms, types of entities and interactions, language, media, as well as how they evolve over time. The multilayer network will be annotated via the machine learning and visualised via an online platform. This platform will push back the frontiers of multilayer network visualization to allow an end user to better understand the source and impact of online misbehaviour and misinformation.

This document describes the open science plan of the project and elaborates on its principles. With the aim of improving the quality, efficiency, and responsiveness of the conducted research, CON-NET will implement a comprehensive open science plan (OSP). The OSP Responsible for the CON-NET project is N. Deligiannis (VUB); in parallel, CON-NET's Management Support Team will follow-up during every network meeting on the adherence to the Findable, Accessible, Interoperable and Reusable (FAIR) principles of the generated data, reports, and publications.

2 General Principles

CON-NET's Open Science strategy is aimed at promoting transparent and collaborative research practices, facilitating knowledge reuse and data sharing.

CON-NET is committed to producing research findings, such as publications, datasets, and software, that can aid in the development of evidence-based policies, as well as provide training materials for building effective systems and processes. By linking data management to online information, CON-NET hopes to foster greater collaboration and transparency in detecting misinformation and make these resources available to the citizens of the EU. These efforts will be guided by best practices for data sharing and management.

The CON-NET project focuses on online social media data, which due to its nature and the identifiability of the users, is considered personal data.

The sharing of data and code will adhere to the Data Management Plan (DMP) plan and the Ethics policy of the CON-NET project, thereby respecting anonymity and avoiding identifying individuals or communities. These documents clearly indicate that data will not be shared with partners outside of the project without a formal agreement with the consortium.

In case of legal, ethical, copyright, confidentiality, or intellectual property considerations, researchers will need to contact the Project Coordinator and the OSP Responsible, who will advise on the policy to be followed with reference to the consortium agreement and the project's DMP. This includes data sharing in an open access data repository (after publication of the research results) under a specific license, which can restrict data re-use under certain conditions (e.g. CC-BY-NC) but not necessarily (e.g. CC-BY). Furthermore, metadata of restricted or closed access data will be shared in a FAIR data repository, where access to the data can only be granted after signing the appropriate agreements.

3 Open Access Publications

The CON-NET project is committed to open access to scientific publications, utilizing both the "Green" and "Gold" open-access models. All the scientific publications produced as a result of this project will be made open-access, meaning they will be fully accessible to all without readers needing to pay to access them. Under the "Green" model, pre-prints of manuscripts will be deposited in a selected repository before undergoing peer-review. To achieve this, the research portals of participating universities will be used.

Additionally, all project partners are required to provide bibliographic details and a copy of the peer-reviewed post-print of their publications and supporting data to the CON-NET Project Coordinator and OSP Responsible upon acceptance, at the latest. These publications will be deposited in a selected repository (such as Zenodo) and made available through open access.

Selected publications will be made available through the highest standard of open access (Gold Open Access). Participants will leverage national funding agency schemes for open science, e.g., the open access fund of the FNR[1], to fund gold level publications, where possible. All other publications will be made available on the project's website, and OpenAIRE's Zenodo.org or more recent HuggingFace Spaces . By following these open access practices, CON-NET hopes to increase accessibility and dissemination of scientific findings, ultimately advancing scientific research and knowledge.

3.1 Early Access to CON-NET results

To increase the reach and impact of its research, CON-NET will make its results available early through informal publications to reputable repositories such as the Computing Research Repository (CoRR), which is a partnership of ACM, arXiv.org, and other organizations.

These repositories are indexed by many other repositories, including DBLP and Google Scholar, further increasing the visibility and accessibility of CON-NET's research. By utilizing these reputable repositories, CON-NET aims to promote the timely dissemination and open sharing of its findings, ultimately advancing the state of scientific knowledge in the field.

[1] https://www.fnr.lu/funding-instruments/open-access-fund/

3.2 Open Research Europe Publishing Platform

As part of its commitment to open access and transparent communication of its research, CON-NET partners will endeavour to publish articles in the Open Research Europe Publishing Platform (ORE).

By publishing in the ORE, CON-NET aims to make its research results widely accessible and increase its visibility and impact. The ORE is an open-access platform that supports the publication of a wide range of research outputs, including articles. Through its publications in the ORE, CON-NET hopes to contribute to the advancement of scientific knowledge and promote the adoption of open science practices.

3.3 Open Peer Review

CON-NET partners will also participate as much as possible in conferences with open peer reviews. This open review process will promote transparency, collaboration, and knowledge sharing among the scientific community. By engaging in open peer review, CON-NET aims to receive constructive feedback and improve the quality of its scientific outputs, while also contributing to the advancement of the field as a whole.

4 Open-Source Software

Given valorisation constraints, CON-NET aims to make some of the developments of its academic partners available as open-source software. These developments will be published on Zenodo and Github[2], subject to valorisation plans and restrictions. The valorisation plans and restrictions will be on a per partner basis, and set in coordination with the technology transfer offices of the participating institutions.

CON-NET will strive for open access for research software in platforms such as Zenodo, GitHub, and code ocean under proper license scheme (e.g., MIT license). The consortium partners already have a track record of Open-Source Software (OSS) contributions and experience, and their expertise will inform the selection of the appropriate license for the open-source software. VUB researchers can contact TechTransfer (techtransfer@vub.be) or us (dmp@vub.be) for advice on this.

By adopting an open-source strategy, CON-NET aims to promote collaboration and transparency in the scientific community, while also facilitating the dissemination and adoption of its research results. Through the availability of open-source software, CON-NET hopes to contribute to the development of innovative solutions and processes that can address pressing challenges in the field.

5 Open Data and FAIR Data

Given valorisation constraints as well as exceptions related to data confidentiality and sensitive personal data (see Data Management Plan and Ethics Policy), CON-NET aims to share its datasets as open data in various FAIR repositories and the OFair Data Marketplace. Platforms for open access for research data include Zenodo, GitHub, and IEEE DataPort. Moreover, CON-NET will strive to share its constructed AI models, and data analysis flows in open repositories like OpenML. Through these practices, CON-NET promotes collaboration, transparency, and the adoption of best practices for data management and analysis, ultimately contributing to the advancement of the field.

This will be done as much as possible subject to constraints related to valorization and data privacy, and ethics. The valorisation constraints will be set in coordination with the technology transfer

[2] GitHub links directly to Zenodo (Referencing and citing content - GitHub Docs). Using this it is possible to reference and quote your code as a PID is assigned to the code.

offices of the participating institutions. The data privacy and ethics constraints are discussed in the Data Management Plan and the Ethics Policy of the project (see associated documents).

It is important to mention here that, subject to the DMP and Ethics Policy documents, all data shared will undergo pseudonymisation. The contents of online social media post and messages will not be made publicly available as part of this project to help preserve anonymity.

The partners will also investigate using restricted access data repositories. This means that the metadata is available to everyone and a dataset can therefore be retrieved by a PID, but that you do not have immediate access to that dataset. As a researcher you will have to request access and after signing the necessary documents you will be able to access those data.

Moreover, data is gathered from social media platforms, following the terms and conditions of these platforms.

This will prevent sensitive data being published publicly and vulnerable groups form being targeted due to the project outputs. As part of our project infrastructure, we will provide a means by which an individual can request that their online social media data be removed from our data sets.

5.1 FAIR Data

Below the FAIR principles we will follow:

- F(indable): we will make data findable by considering PID for datasets, ORCID for researchers, ROR for research institutes, and by linking to project and publication
- A(ccesible): As mentioned above we will endeavour to make data as accessible as possible by investigating open and restricted data under appropriate licenses (having data restricted data is due to valorisation, ethical and anonymity constraints)
- I(nteroperable): we will consider interoperable formats for data (e.g. csv instead of xlsx)
- R(eusable): this includes data documentation, licenses, explanation of used software and models