RAYUELA
a fun way to fight cybercrime

Deliverable Report

D4.10 Open Data Considerations Report
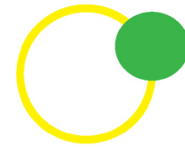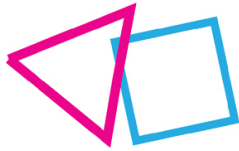
Deliverable No.	4.9	Work Package No.	WP 4	Task/s No.	T4.4
Work Package Title	ONLINE PRIVACY, DATA SECURITY AND ETHICS				
Linked Task/s Title	Open Data considerations for the project's research and development framework				
Status	Final	(Draft/Draft Final/Final)			
Dissemination level	PU-Public	(PU-Public, PP, RE-Restricted, CO-Confidential)			
Due date deliverable	30/09/2023	Submission date		30/09/2023	
Deliverable version	V2.0				

Document Contributors

Deliverable responsible			
Contributors	Organization	Reviewers	Organization
Roberts Rozis	Tilde	Lokesh Sharma	NEC
Aivars Bērziņš	Tilde	Manuel Álvarez-Campana	UPM
Jaime Pérez	COMILLAS	Sonia Solera	UPM
		Abel Muñiz	ZABALA
		Violeta Vázquez	ZABALA
		Gregorio López	COMILLAS

Document History

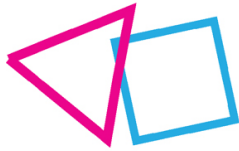| Version	Date	Comment
v1.0	20.09.2023	First draft sent to UPM and NEC teams
v2.0	21.09.2023	Tilde has addressed the provided comments with this update

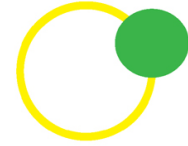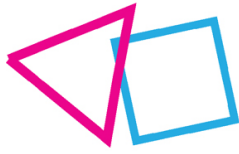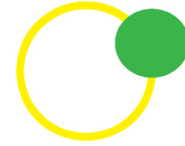# Table of contents

List of Abbreviations

Abbreviation	Description
CC	Creative Commons
CSV	Comma-separated values
DMP	Data Management Plan
EU	European Union
GDPR	General Data Protection Regulation
IPR	Intellectual property rights
MIME	Multipurpose Internet Mail Extensions
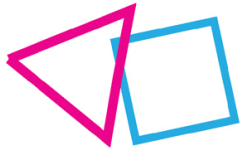
Executive Summary

This deliverable outlines crucial considerations when it comes to publishing data for potential reuse. The central element in this context is the availability of data in a machine-readable format, a paramount requirement without which any subsequent research or data modelling efforts would face significant challenges.

Despite not being obligated to disclose datasets, the consortium has taken proactive measures to release an anonymised dataset derived from the RAYUELA pilots. The dataset contains information gathered during the pilots conducted across various European Union member states as part of the RAYUELA project. Importantly, these pilots involved minors, with supervised sessions conducted in educational institutions and summer camps, ensuring strict adherence to safety and ethical standards.

For the convenience of potential dataset users, the dataset has been made accessible through Zenodo[1], a platform dedicated to promoting open access to scientific and research data.

By providing access to this data, the consortium not only supports future research efforts but also upholds transparency and fosters collaborative knowledge-sharing within the broader academic and research community.

This deliverable provides details about the dataset, including its origin and the rationale for making the data open (Section 3). Section 4 outlines the procedure for publishing data on Zenodo, while Section 5 delineates the steps taken to comply with GDPR regulations. Finally, Section 6 elucidates how the FAIR principle is being implemented.

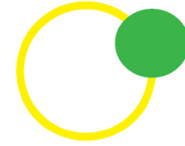# 1. Data resulting from WP6 Activities

1.1　WP6 – Data from RAYUELA pilots

This dataset contains the data collected from minors who have participated in the pilots of the RAYUELA project. From each participant we have collected demographic data (Age, Gender, Sexual Orientation, Migratory Background, Location, School Type, Daily number of hours spent using the Internet for leisure.
), the results of some psychological tests validated by the scientific community (social and family support, Big Five personality traits model, self-esteem and previous cyber-victimization/aggression) and finally the gameplay data (selected answers). No information is collected without the player's consent.

At the beginning of the session the participants had to register on a website where they are assigned a username and password that will allow them to play later. During this registration, the validated psychological tests are also carried out, apart from the cyber-victimization/aggression questionnaire, in an attempt not to bias the answers given by the minors to the questions posed by the game.

In the event that participants are unable to finish the game completely, they can access the final quiz at any time. This feature was implemented for the last pilot phase, so there are participants who did not answer the final questionnaire.

---

[1] https://zenodo.org/
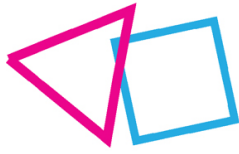
1.2   The origin of the data

The dataset has been generated from data collected during pilot projects conducted by the RAYUELA project across various European Union countries. These pilot projects involved minors, and the game sessions took place in schools and summer camps under careful supervision.

As far as possible, an attempt was made to homogenize the playing conditions of the participants, although some factors were not under the control of the RAYUELA members (e.g., the time that the schools could devote to the game sessions).

1.3   Reason for opening data

We believe that the data gathered could be of great use to the scientific community. Motivating other researchers to explore the dataset and try to validate or discuss the conclusions obtained. Obtaining as a result a better understanding of the social problems addressed in RAYUELA and a significant example of open and honest science. Which could translate into better prevention and mitigation policies.

2   Publication of Dataset(s) on Zenodo

Zenodo is one of the recommended repositories in line with the "Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020".

Within the framework of European Union Horizon 2020 projects, Zenodo operates as an expansive digital repository, comparable to a highly regarded online archive. This platform provides space for researchers and consortia to securely archive and distribute their research outputs, thus contributing significantly to the collective knowledge base within the scientific community. Zenodo represents a structured and easily accessible resource, purpose-built to facilitate the seamless sharing of a wide range of research materials, encompassing research papers, datasets, and associated scholarly content.

In the context of EU Horizon 2020 projects, Zenodo assumes a central and pivotal role, actively promoting the principles of open and collaborative knowledge exchange. It is aligned with the overarching mission to advance transparency and nurture innovation within the dynamic landscape of European research.

2.1   Data Set Metadata – Recommended Fields

Table 1 presents a comprehensive summary of the requisite metadata details essential for the publication of data within the Zenodo repository. This table enumerates the specific requirements, offers explanations for each mandatory metadata element, and provides corresponding values. Following the provision of all the necessary metadata, the dataset has been successfully published on Zenodo, as depicted in Figure 1. It is now readily accessible to the research community, facilitating its subsequent utilization for further studies.

Figure 1 Depict of the RAYUELA project dataset on Zenodo repository

Table 1 Metadata fields required by the Zenodo repository.

Metadata field	Explanation	Value
Communities	Specify communities which you wish your upload to appear in. The owner of the community will be notified and can either accept or reject	European Commission Funded Research (OpenAIRE)

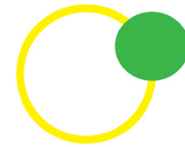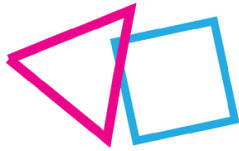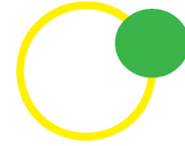|		
	your request. Please make sure your record complies with the content policy of the communities you add; reported abuse will be followed by account inactivation.	
Upload type		Dataset
Publication date		September 15, 2023
Title		RAYUELA - Open Data - Data collected through a serious game created to identify patterns and profiles of young potential victims/perpetrators of cybercrimes.
Authors	Name and Family name, Affiliation, ORCID	Jaime Pérez Sánchez (Universidad Pontificia Comillas, ORCID: 0000-0001-6044-0022) Gabriel Antonio Valverde Castilla (Universidad Pontificia Comillas, ORCID: 0000-0002-2521-4852) Gregorio López (Universidad Pontificia Comillas, ORCID: 0000-0001-9954-3504) Mario Castro (Universidad Pontificia Comillas, ORCID: 0000-0003-3288-6144) María Reneses Botija (Universidad Pontificia Comillas, ORCID: 0000-0002-9708-6896) María Riberas-Gutiérrez (Universidad Pontificia Comillas, ORCID: 0000-0003-2030-0310) Nereida Bueno-Guerra (Universidad Pontificia Comillas, ORCID: 0000-0003-1442-7905)
Description		The data of this dataset have been collected in the pilots carried out by the RAYUELA project in different countries of the European Union. The participants are minors and the game sessions have been carried out in schools and summer camps in a supervised way.
Version	Optional. Mostly relevant for software and dataset uploads. Any string will be accepted, but semantically-versioned tag is recommended. See semver.org for more information on semantic versioning.	V1.0
Language	Optional. Primary language of the record. Start by typing the language's common name in English, or its	ENG

	ISO 639 code (two or three-letter code). See ISO 639 language codes list for more information.	
Keywords		Serious Game, Educational game, Cyberbullying, Cybersecurity, Online Grooming, Psychological Questionnaire, Big Five Model
License	Required. Selected license applies to all of your files displayed on the top of the form. If you want to upload some of your files under different licenses, please do so in separate uploads. If you cannot find the license you're looking for, include a relevant LICENSE file in your record and choose one of the Other licenses available (Other (Open), Other (Attribution), etc.). The supported licenses in the list are harvested from opendefinition.org and spdx.org . If you think that a license is missing from the list, please contact us.	CC 4.0 International
Access right		Open Access
Funding	Grants	EC, EC GA: 882828, EMPOWERING AND EDUCATING YOUNG PEOPLE FOR THE INTERNET BY PLAYING
Related/altern ate identifiers	Specify identifiers of related publications and datasets. Supported identifiers include: DOI, Handle, ARK, PURL, ISSN, ISBN, PubMed ID, PubMed Central ID, ADS Bibliographic Code, arXiv, Life Science Identifiers (LSID), EAN-13, ISTC, URNs and URLs.	The H2020 project RAYUELA: A fun way to fight cybercrime http://dx.doi.org/10.18239/jornadas_2021.34.27 Virtual Harassment, Real Understanding: Using a Serious Game and Bayesian Networks to Study Cyberbullying https://doi.org/10.48550/arXiv.2309.08428

Format		CSV and txt
Subjects	Specify subjects from a taxonomy or controlled vocabulary. Each term must be uniquely identified (e.g. a URL). For free form text, use the keywords field in basic information section.	Cybersecurity-aware culture http://data.jrc.ec.europa.eu/ontology/cybersecurity/cyber security-aware_culture
Link to dataset		https://zenodo.org/record/8347722

The dataset was added to Zenodo on September 15, 2023. Since its publication, it has garnered attention, receiving 24 views and 18 downloads. According to Zenodo's statistics, it's worth highlighting that out of these 24 views, 15 are unique (first-time viewers), contributing to a total of 10 unique downloads. Figure 2 displays statistical data from Zenodo.

September 15, 2023

Dataset Open Access

RAYUELA - Open Data - Data collected through a serious game created to identify patterns and profiles of young potential victims/perpetrators of cybercrimes.

Jaime Pérez Sánchez; Gabriel Antonio Valverde Castilla; Gregorio López; Mario Castro; María Reneses Botija; María Riberas-Gutiérrez; Nereida Bueno-Guerra

The data of this dataset have been collected in the pilots carried out by the RAYUELA project in different countries of the European Union. The participants are minors and the game sessions have been carried out in schools and summer camps in a supervised way.

24 views	18 downloads

See more details...

	All versions	This version
Views	24	24
Downloads	18	18
Data volume	5.4 MB	5.4 MB
Unique views	15	15
Unique downloads	10	10
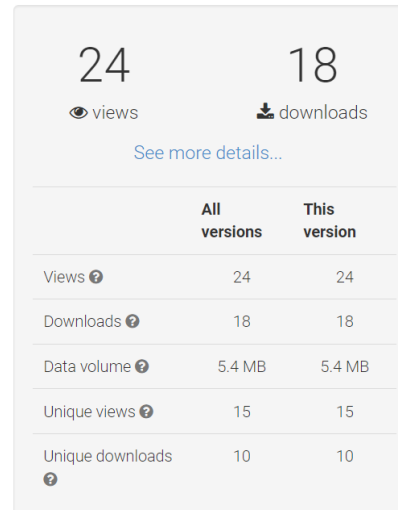
Figure 2. Depiction of Zenodo statistics.

3 Steps taken to comply with GDPR

To ensure GDPR compliance and uphold intellectual property rights (IPR) neutrality for the dataset slated for publication, a following measures have been implemented:

- We have meticulously curated the dataset, retaining only those attributes that are directly pertinent to the scientific community's research needs and are free from any IPR constraints.
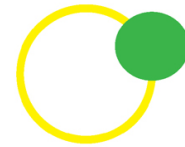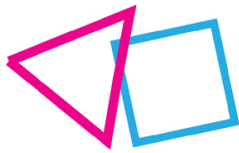
- We have undertaken specific actions to enhance privacy and data protection. This includes the removal of specific data points such as "Registration date" and "Player ID." Additionally, we have introduced randomization in the arrangement of dataset rows to prevent potential identification based on the timing of game sessions.

These steps have undergone thorough discussion and consensus within the team responsible for data collection and analysis, in conjunction with the legal experts from the RAYUELA project. This approach ensures not only GDPR compliance but also safeguards the dataset's IPR-free status in line with the principles of the European Union's Horizon 2020 program.

4 Compliance with FAIR Principles

The FAIR Guiding Principles[2] – FAIR—Findable, Accessible, Interoperable, Reusable – best describe the properties of Open Data. Applying these principles to the Data Set in focus will ensure excellent description and presentation of your Data Set from various criteria of metadata.

In the context of Open Data, it is considered best practice to construct Metadata with a thoughtful consideration of potential data re-users. Crafting metadata accurately not only facilitates the implementation of filtering and search capabilities, but also enhances the discoverability and accessibility of dataset tailored to the specific repository chosen for publication.
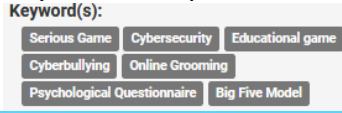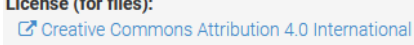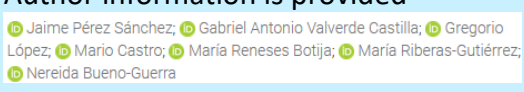
We've put together a checklist that lists the metadata elements that need to be specified, along with the methods and criteria for checking them. Table 2 gives you an overview of how well our dataset follows the FAIR principles when we publish it.

Table 2. FAIR principles validation

	Metadata Item	Validation Method, Criteria	RAYUELA - Open Data - Data collected through a serious game created to identify patterns and profiles of young potential victims/perpetrators of cybercrimes.
☒	Assignment of a Globally Unique and Persistent Identifier	Check the Identifier against its issuer	10.5281/zenodo.8347722
☒	Domain Assignment, domain classification scheme	Confirm the domain assigned	Domain classification is assigned: Subject(s): Cybersecurity-aware culture

---

[2] https://www.go-fair.org/fair-principles/

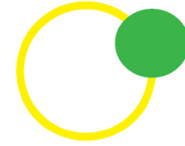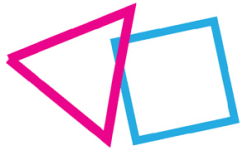| ⊠	Data Set Title – version in English present	Make sure that Title field is provided in English	Checked – Title field IS provided in English.
⊠	Data Set Description – version in English present Describe the background of the Data Set, how the data originated	Make sure that Description field is provided in English	Checked – Description field IS provided in English.
⊠	Data Set Type (Audio / Geospatial / Statistical / Text / Video / other)	Make sure that Data set Type field is provided	Zenodo repository does not contain a metadata field for Data set Type, but data Preview allows assessing Data Set Type.
⊠	Data set MIME Type	Make sure that a valid MIME Type is provided	Zenodo repository does not display MIME type, but CSV file type is unambiguously related to text/csv MIME Type
⊠	Keywords	Make sure that Keywords are provided	Keywords are provided. **Keyword(s):** Serious Game Cybersecurity Educational game Cyberbullying Online Grooming Psychological Questionnaire Big Five Model
⊠	Data set License Information	Make sure that a valid License Information is specified	A valid open CC license assigned. **License (for files):** ⧉ Creative Commons Attribution 4.0 International
⊠	Data set Author Information	Make sure that Author Information is provided	Author information is provided Jaime Pérez Sánchez; Gabriel Antonio Valverde Castilla; Gregorio López; Mario Castro; María Reneses Botija; María Riberas-Gutiérrez; Nereida Bueno-Guerra
⊠	Data set IPR holder Name / Institution / Address / Email / URL	Make sure that all possible Data set IPR holder attributes are provided	Content of IPR holder information is provided via Zenodo 'Cite as' field.

☒	Data set Compiler / Producer Information	Make sure that Compiler / Producer Information is provided	Field N/A in Zenodo
☒	Data set Contact Person Information – name and email	Make sure that a valid Contact Person Information – name and email Information is provided	Field N/A in Zenodo
☒	Data Set size – quantity and unit. Multiple versions possible	Make sure that a correct quantity and unit information is provided	Data Set size in Zenodo repository is given in KB.
☒	Data Set Data Collection Date / Location	Make sure that a correct Data Set Date and Location information is provided	Field N/A in Zenodo
☒	Date of Data Set Creation	Make sure that a correct Data Set Date and Location information is provided	Date of Data Set Creation is provided

			Versions Version 1.0 Sep 15, 2023 10.5281/zeno do.8347722
☒	Data Set Funding sources	Make sure that a correct reference to Data Set Funding source is provided	Granting sources information is provided **Grants:** European Commission: • RAYUELA - EMPOWERING AND EDUCATING YOUNG PEOPLE FOR THE INTERNET BY PLAYING (882828)
☒	Version of Data Set	Make sure that a correct Data Set version is provided	Version of Data Set is provided Versions Version 1.0 Sep 15, 2023 10.5281/zeno do.8347722
☒	Metadata Assignment in relation and specific to the area of the Data Set[3]	Check the validity of Metadata Assignment	Metadata Assignment in accordance with Zenodo
☒	Information About Creation / Processing / Compilation of the Data Set	Make sure all possible information is provided about Creation / Processing / Compilation of the Data Set	Field N/A in Zenodo

---

[3] http://rd-alliance.github.io/metadata-directory/standards/

☒	Quick Data Set Content double-check	Download and unpack the data set files after publishing.	Data Set contents download and are readable
☒	Readability of files	Make sure all data files are readable and not distorted.	Data Set contents download and are readable
☒	Data Set Files and Metadata Files included	Make sure all data and metadata files are in the right places.	Data Set Files and Metadata Files are included in the Zenodo resource

## 4.1   Validation After Publishing

In our commitment to ensuring that the published dataset fully complies with the FAIR principles, we validated compliance after publishing data on Zenodo. This step aligns with the guidelines provided by the FAIR principles, emphasizing the critical importance of making data findable and accessible to the community. The validation process served as a quality check, confirming that our dataset met the criteria set by the FAIR principles. By conducting this post-publication validation, we could provide a level of assurance that the dataset adhered to the principles of Findability and Accessibility.

Table 3 Validation After Publishing

	Metadata Item	Validation Method, Criteria
☒	Metadata of the Data Set must be indexed and searchable	Perform search for the data set using predefined search keywords
☒	Data Set can be found by intended means of lookup and search	Make sure that Data Set can be found by intended means of lookup and search
☒	Data Set is accessible for free download to anonymous user	Make sure that Data Set is accessible for free download to anonymous user

5   Conclusions

The deliverable highlights essential considerations for publishing data for potential reuse, emphasizing the importance of machine-readable data availability. Despite no strict obligation to open any dataset resulting from the project activities, the consortium released a dataset from RAYUELA pilot projects. The dataset from RAYUELA includes demographic, psychological, and gameplay data, all collected during RAYUELA pilots across EU countries in schools and summer camps.
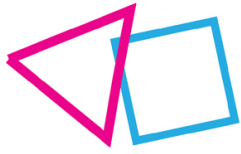
In alignment with the recommendations for EU Horizon 2020 projects, we have chosen Zenodo as the repository for publishing the resulting dataset. Zenodo's status as an Open Data repository, specifically recommended for EU Horizon 2020 projects, underscores its suitability for our data sharing endeavours. Tilde's in-house data experts conducted an assessment to ensure adherence to the FAIR principles during the publication of the dataset. The FAIR principles represent a set of guidelines designed to enhance data discoverability, accessibility, interoperability, and reusability. These efforts have resulted in a dataset that is accurately described, effectively presented, and compliant with the appropriate metadata standard. To streamline this process, we have developed a checklist that encompasses all the necessary metadata elements required to align with the FAIR principles.

In summary, this deliverable showcases the consortium's commitment to open and transparent science, facilitating data access and collaboration while adhering to essential regulatory and ethical considerations.