Plan Overview

A Data Management Plan created using DMPonline

Title: Does a workshop to create visual novels elicit bias regulation?

Creator:Dídac Jiménez Torras

Principal Investigator: Dídac Jiménez Torras

Data Manager: Dídac Jiménez Torras

Project Administrator: Dídac Jiménez Torras

Affiliation: Other

ORCID iD: 0009-0007-3565-437X

Template: DCC Template

Project abstract:

The aim of this study is to examine whether creating a visual novel (VN) video game in a dialogically mediated workshop contribute to bias awareness and self-regulation. A “visual novel” video game is like a digital comic, but interactive.

The study has already gone through a first edition, with a two group cohort in Barcelona. Three groups at Canòdrom civic centre, Joan Miró library, and Enti-UB university will participate in 10-session workshops on creating visual novels (individual, not with groups). My intention is to cover several steps to making a visual novel, with an emphasis on storytelling, identity, and iteration, encouraging students to give mutual feedback while learning about art, music, and interaction.

I plan to disseminate the results in a scientific article. This study is being conducted in Barcelona, in Catalan and Spanish.

ID: 183515

Start date: 07-10-2025

End date: 01-04-2027

Last modified: 23-07-2026

Copyright information:

Does a workshop to create visual novels elicit bias regulation?

Data Collection

What data will you collect or create?

Note that this is updated version [1] of the data management plan, and the original focused on a quantitative survey.

The main data results will come from three sources: Participant observation (qualitative), semi-structured interviews (mixed), and close reading of the participants' visual novels (qualitative). The data will be released in .xlsx, .csv, .ods, .doc, and/or .pdf files. The chosen formats allow versatility and interoperability and can be visualised in many repositories like OSF.

Some data already exist at the moment of this first DMP update [22-07-26].

Considering a total expected sample of 60 participants, there might be up to 55 original audios, and 20 visual novels to analyse, potentially including their .zips, and the observation of the different sessions, amounting to around 80 days of notes. Accounting for the filtered versions to maintain anonymity, the transcriptions, and the video .mp4 screen recordings of the analyses, the total amount of files may be over 200, with an expected size of 35 Gb, and an individual file size ranging from 0,1 Mb for transcriptions to 1 Gb for some videos/audios.

The chosen formats are easy to store, and although audio may be somewhat prone to corruption, the transcripts generated with Audacity's OpenVINO Whisper AI mode shall allow for long-term data maintenance.

How will the data be collected or created?

Interview recordings will occur in my Realme GT2 Android 15 device with Signal Messanger voice memos, or directly recorded using OBS on my desktop. The observations will be noted on my personal physical research diary and/or Notesnook notation app on my phone (ideally in airplane mode if I remember to). If the data is created on my phone, I will use the open source transfering app LocalSend to move it to my desktop. There onto a digital version at OnlyOffice or LibreOffice (note the first cohort sessions were annotated on my personal Google Docs but have since been downloaded, deleted, and backed up on Proton Drive).

Naming was initially planned not to be compressed into codes, and analysis files will still have self-explanatory descriptive titles, but the raw data will follow a structure that will be explained on a README txt file.

It is expected that the interview recordings, session diary notes, and VN analyses will have a structure to indicate the centre where they took place, the participant with their self-decided anonymity code, the day/number of the session, the cohort, and the absolute number in the case of interviews (for instance, as of now "13 10M Tecno1.txt" is a file that exists for interview number 13, from participant 10M of TecnoCampus edition 1). It is not planned yet what structure the data from this project will follow, but it is pressumed that it will be classified according to the different cohorts, and three data formats, plus a separate folder for the final analyses and summaries grouped together.

The days where the data gets analysed will be tracked on the documents themselves.

Versions will be tracked through the repositories (ResearchBox, and OSF). It is not expected that the data will get re-analysed after publishing, but in such a case, we will issue an update note and try to highlight it with ALL CAPS.

Documentation and Metadata

What documentation and metadata will accompany the data?

Only the titles, questions, and whether a question is reversely coded should be needed to understand the data. Dates, file descriptions, format, size, file language, keywords, and links to the original instruments or to the methodological references used following OSF's default metadata standard will also get included as part of the metadata. The programmes used to edit the files may be included. Information about the participants including age, name, nationality, gender, or job status shall be included inside the files themselves or as metadata, only upon decision by the participants themselves.

For information about the interview questions and final analyses documents, or regarding the titles of the spreadsheets/docs for the diary/VN rubrics, the files themselves will have explanatory comments, and the information may be repeated on a separate coding book. The thematic analysis codes will be represented on a correlations graph akin to Kraph et al's (https://doi.org/10.1016/j.chb.2021.106963) with a legend describing it. Data will get boldened, italicised, coloured, and highlighted, and the different modifications will get matched to a theme that will be explained at the beginning/end of the analyses documents. For quantitative data, like the aproximate score to the biases task on the interview, the results will be normalised on a scale from 0 to 10.The transcripts will have timestamps so that readers can locate what was said and when they said it even on the translated versions.

Ethics and Legal Compliance

How will you manage any ethical issues?

In the previous and original DMP, an online questionnaire was the main data gathering instrument, and a question inside it asked for consent

for data sharing. Since then, the survey has been removed, and now the principal instrument is the interview. There are three interviews planned for every participant interested, and in the first one they are shown an infographic detailing data stewardship practices and blind spots, including the tools used to record the data, the devices where it be stored, and the repositories where it could be published openly once finished.

During the second interview, participants are asked about their anonymisation preference, and how they would prefer to appear in the transcripts. The script includes the following:

"Okay, I might start analyzing the interviews soon... When I do, how would you prefer me to cite you in the transcripts? There are 5 options: 1) As “a participant”, 2) With random letters and numbers “for example Q87”, 3) Semi-random “for example 7DID”, 4) With a nickname of your choice such as “for example Didacticus”, or 5) With your real name and demographic data that you feel like sharing “for example Dídac Jiménez Torras (23 years old, non-binary, researcher)”."

Lastly, during the third interview, participants will get asked about their data publication preferences, and how and whether they would be fine with their original interviews being published. The script mentions:

"Now in a few months I will start begin seeking to turn this into an open scientific publication, and let's see... I think that respectable science must be understandable and verifiable, and that is why I try to publish together with the explanation as much original data as possible, which includes this interview. So I was wondering if you would give your consent to publish this and the previous interviews with the supplementary material in the appendices? Before answering, if you prefer, I could anonymize your voice so that it is much more difficult to detect that you are speaking; I can give you an example." The voice changing process is to be conducted with free open RVC standard, in specific Tiger14n's GUI (https://github.com/Tiger14n/RVC-GUI/releases).

Now, before participating on the interviews, and for the analysis of the sessions and the visual novels, participants will be asked for written permission through an informed consent document based on Open University of Catalonia's (UOC) templates (https://campus.uoc.edu/estudiant/microsites/td_etica/en/nivell2_pmf/8_docs.html) and approved by the internal Ethics Committee with code "CE25-TE95" for third consecutive time.

Participants who do not adhere to it may appear only contextually referenced as part of the session diary annotations.

Throughout, I will know the names of the participants, and participants may disclose sensible data related to their political, work, religion, sexuality, gender, nationality, etc. beliefs and opinions. This is considered standard in educational participatory action research.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

I assume, I, Dídac Jiménez Torras, as the main researcher I am the "owner" of the analysed data whereby it came from my particular investigative efforts. Nevertheless, I have a record of not using "my" data for commercial purposes and I intend to keep it that way. The presentations of the workshops are openly and freely available from the first session, with the lastest version available on Figshare (https://doi.org/10.6084/m9.figshare.31563481). Still, the raw data will belong to the participants, and they can revoke access to it and ask for its elimination whenever.

An addition of the second edition [as of 22-07-26] is that data may be observed and analysed by a potential secondary co-researcher, Nicolás Alejandro Rojas Cajamarca. Their participation will have to strictly comply to the conditions laid herein and in my pre-registration over on OSF (https://doi.org/10.17605/OSF.IO/JP4DE) so long as its data stems from my consent documents and interviews.

Regarding publishing, the data will be licensed with the most open licence (CC0) I am allowed to provide once it is published in a journal. I don't put any restrictions in the re-use of data by third parties for research purposes, but data scrapping, AI, and corporate entities should avoid using the data for secondary analysis, specially for commercial purposes.

Data won't be postponed or restricted to seek patents, and participants may be invited to write, curate, and or even create their own workshops. The intent is to expand bias awareness and accessibility.

Storage and Backup

How will the data be stored and backed up during the research?

I have recently freed my local drives and currently [as of 22-07-26] have around 100Gb to analyse the data from the participants and over 600Gb for back ups, noticing audio files can be large, but not expected to reach over 1Gb.

Previously, the data was to be extracted from Microsoft Forms and uploaded to UOC's Google Drive to be analysed once once. Currently, with the focus on the interview, the privacy and safety have been enhanced by integrating an offline pipeline centered on analysing the files using LibreOffice and storing them locally only. This, nevertheless, leaves them more vulnerable to physical and targeted attacks or ecological adverse events, but those are considered very unlikely. As a result, the data will not be backed up until publishing, where it will be multiply synced over on ResearchBox, OSF, and possibly Zenodo.

I will check periodically (every 6 months) during the 5 years after publication whether the cloud services (OSF, ResearchBox, Figshare, and Zenodo) are still online, ensuring the links work, and finding apt replacement (Proton Drive, Zoho WorkDrive, or a journal's database appendices) if they don't.

Recovery and/or backup in case of accidents an attempt will be made to restore the drives with a tool like iObit, and with the help of the technical and IT teams of UOC and/or the services providers, in a case-by-case basis, contacting them to ensure problems get resolved.

How will you manage access and security?

There are several risks to the pipeline described, particularly regarding transfer, since the raw files can be intercepted, and they may be vulnerable too to attacks on the science repositories. Internally, after analyses are done, for each cohort, the raw files will be file-encrypted with PeaZip, and the non-encrypted versions will be deleted. This reduces the vulnerability of participant voice data, even in case of

complete physical intervention of the main researcher's devices, as hackers would need to know what file is needed to decrypt the data.

To curtail data interception, I currently have Proton VPN over on my phone and use only either mobile 5G or my local private home Wi-Fi, but these may nevertheless be vulnerable to intrusion by the service provider (Orange), and the OS companies (Microsoft and Google). I am considering the purchase of a firewall-based open source VPN like Windscribe, as well as moving to Linux and Murena /e/OS respectively to increase privacy.

As per the data upload, files will be moved to OSF (https://osf.io/5hqem/) and ResearchBox (https://researchbox.org/4546) for publication.

The file formats to use have no known exploitable vulnerabilities, but nevertheless will have been previously revised to guarantee the anonymization asked by the participants without relying on special format features like censored text.

Observation rubrics will be stored in paper (and sometimes photos with no people but rather the whiteboard of the sessions), .pdf (final), .ods (spreadsheet) or .odt (docs) formats (less than 5Mb expected). The interviews and their transcripts will be stored in .txt and .wav format (400Mb average expected per file). The analyses recordings will be stored in .mp3 format (500Mb average expected per file). Infographics, and other secondary files not owing distinctly to the three methods described, like slides, may be stored in multiple formats, like .af (Affinity), or .png (for visualization).

The files will be maintained on the aforementioned open science repositories for as long as they exist, and I commit to periodically reviewing that they are still active once every 6 months (with an alarm), for the 5 years following the completion of the study.

Sharing between collaborators like Nicolás will be exclusively done under the conditions here described, and will ideally rely on first peer review access to OSF and ResearchBox, and second physical and offline pen-drive transfer where the special decryption file will be shared.

Participants will have my phone and my Signal contact if they need to relay any data or data deletion, standard queries may be answered over any platform, but documentation sharing will only happen over Signal.

Prospective researchers interested in the data will need to contact me first through phone at640017592 or email at djimeneztorr@uoc.edu, or didacticus@zohomail.eu, and every request of raw data will be parsed to ensure it is not intended to exploit sensible participant information. I find it hard to predict by then what service I will be using, but if the information is not already on the appendices, OSF, or ResearchBox, the decryption file will probably be sent through Signal, or else, a decryption password may be verbally shared during a call using Jitsi Meet or an analogous private meeting app.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

The data will be openly accessible for as long as the previously stated repositories to store it in exist, unless they get updated to a decentralised system, or otherwise prohibited for some reason.

The data of participants that haven't given permission to have their results published will be destroyed as soon as the article study with the results gets published. Any request for destruction will be undertaken over on the multiple platforms and files of that said participant, which is why I shall know their de-anonymised identities. I will do my best effort to remove and overwrite the raw files, copies, and half-deleted in-between versions with Eraser on Windows (https://eraser.heidi.ie/), and with Extirpater (https://f-droid.org/en/packages/us.spotco.extirpater/) on Android.

I plan to retain the data at least 5 years on my PC. I am currently holding the results of projects from 3 and 2 years ago, so I have that track record. The files may be ultimately deteled with the aforementioned tools afterwards. Longer lasting data includes my filtered analyses, which may be retained past the 5 years as will be still available on the repositories. Decryption passwords will be stored on a physical sheet inside my home. For phone interviews, the information may be deleted sooner, need it be for storage requirements.

Regulatory speaking, owing to UOC's ethics requirements: "Once the data has been dissociated, the files with the original data must be destroyed, i.e. they cannot be stored. The personal identifiers that are collected will be kept separate from the anonymized data in an encrypted password-protected archive.". I will try to do this, although I am still learning about tools like Amnesia. I am honest that this will be an on-going, and adaptative process.

The data will be used to adapt, promote or discard the creation of visual novels as an educational process/workshop, and to create a correlational mode of the affects that go into learning biases through VNs.

What is the long-term preservation plan for the dataset?

Once again, data will be stored in OSF and Research Box (public), and it could potentially be stored temporarily on my personal online Playbook drive with the caveat that it would be encrypted with an over 20 digits passwords kept in one of my physical notebooks as a means to guarantee future access for researchers who may be interested in validating the data in the age of AI fake papers. No monetary costs come associated with uploading the data and sharing it on these platforms.

Nevertheless, the process of generating, tagging, coding, and explaining the process to ensure it follows FAIR Open Access principles is estimated to take around 75 hours. 25 hours for preparation, 25 for analysis, and 25 for archival, revision, contacting, transfer, blinding, and other procedures. An additional 100 hours go to reviewing the pre-registration, maintaining the data management plan, seeking for

alternative open source apps, and creating infographics so that the information is conveyed accessibily.

Data Sharing

How will you share the data?

Data will be shared through online repositories (OSF, ResearchBox, Figshare, and possibly Zenodo) linked in the article with persistent identifiers (doi) along the research results. The data will also be available in the appendices. Potential users will be able to find the data in said purported published article.

During the process, one expected research collaborator, Nicolás that are to be determined as well as my three supervisors, Daniel Juarez Aranda (darandaj@uoc.edu), Joan Josep Pons López (jjpons@uoc.edu), and Joan Arnedo Moreno (jarnedo@uoc.edu) may partially see the raw data, to help me analyse it.

Potential users may also request to see the data through mail or phone, findable over on my articles and complimentary data.

Are any restrictions on data sharing required?

No restrictions or difficulties are expected beyond the possible hesitation of participants to openly share their information. If any participant sends me a message to remove its contributions, I will do that as soon as I see it.

No special system, algorithm, or biological sample/project is being undertaken here, so I deem non-disclosure and exclusivity agreements to be overstated. Arguably, participants taking part in the workshop may already be aware of its politicized character, so I find this to be an opportunity to raise their voices, instead of restrict them, and so far UOC has encouraged this participatory open action research approach.

I originally expected the survey analysis to take around 1 week. Nevertheless, the study no longer focuses on the (superficial), and instead focuses on the three methods, mainly the interview, so data will remain "exclusive" under my custody for about 8 months on deep DQA multi-round thematic analyses. I project to attend Digra Mx convention with some trial results during summer 2027, and I don't foresee the articles being ready before 2028.

Responsibilities and Resources

Who will be responsible for data management?

I, UOC grant PhD student, Dídac Jiménez Torras, am the sole main researcher in this study. Future updates may expand the scope, but [as of 22/07/26] this is not the case. My thesis on VN bias workshops is not meant as a large educational collaboration or a quantitative large sample test, but rather as an initial exploration of the potential and feasibility of workshops to have long lasting bias impacts. I am responsible for the implementation of this DMP, and my hope is that UOC's supervisors will keep me accountable to it, and remind me of its different parts.

For this reason, no CreDiT-like separation of functions like metadata production, backup, participant member checking, and data archiving or sharing is expected, I will take those roles myself. Backups are currently automated. Member checking is done via personal messaging, and archiving will occur as I publish articles about my thesis.

What resources will you require to deliver your plan?

I assume this point particularly relates to studies dealing with clinical or otherwise interview dense data, which can be heavy and difficult to maintain. Nevertheless, as I anticipated, this study won't be backed up in institutional university storage facilities.

I am promptly looking into training for myself as part of an on-going learning process, although I already participated in the past in courses related to FAIR data stewardship.

I will be using my desktop and phone, so I don't foresee special hardware being required, beyond the computer room and projector in which the workshop will take place, and some printed paper to make some dynamics work. No special software is required either, beyond the engine and tools to make the VN, although I am promptly looking into a recording app that allows you to encrypt and share encrypted voice files similar to Tella, and may update this DMP in the future if I find it.

I don't foresee needing experts for this particular project, but I may consult UOC's data experts if I can't find tools and processes myself.

Planned Research Outputs

Study registration - "OSF Registration for Does a workshop to create visual novels elicit bias regulation?"

OSF Pre-registration for the study.

Interactive resource - "Biases and Visual Novel Video Game Creation 2026 Workshop Participatory research Presentations"

ENG: These presentation slides are copyright-free CC0 and were used by researcher Dídac Jiménez Torras during 2026 participatory research workshops at multiple universities and civic centres. They are/were part of a thesis exploring effective ways to teach biases by creating visual novels and its intended public is learners from 21-60 years. The presentations were performed in 2-hour interactive sessions mixed with dialogical conversations, examples, and game development.

Planned research output details

Title	DOI	Type	Release date	Access level	Repository(ies)	File size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
OSF Registration for Does a workshop to create vis ...	10.17605/OSF.IO/JP4DE ...	Study registration	2025-09-27	Open	Open Science Framework		Creative Commons Zero v1.0 Universal	None specified	No	Yes
Biases and Visual Novel Video Game Creation 2026 W ...	10.6084/m9.figshare.31563481.v3 ...	Interactive resource	2026-03-07	Open	figshare		Creative Commons Zero v1.0 Universal	None specified	No	Yes