

---

## Plan Overview

*A Data Management Plan created using DMPTuuli*

**Title:** AIDA: Action and Intersubjectivity in the Digital Age

**Creator:** Pentti Haddington

**Principal Investigator:** Pentti Haddington

**Data Manager:** Pentti Haddington

**Affiliation:** University of Oulu

**Funder:** The Research Council of Finland (former The Academy of Finland)

**Template:** Academy of Finland data management plan guidelines

**ORCID iD:** 0000-0003-4378-4450

### Project abstract:

Social interaction is vital in the digital age. Emerging digital technologies present new environments for real-time and multimodal interaction, such as videoconferencing systems and immersive virtual reality (VR). These environments feature dimensions that have a fundamental impact on how social actors build common ground in real-time digital interactions. We know that, for example, building common ground, organising turn-taking or building a joint focus on a referent is hard in digital interactions, but we do not know how and why. AIDA asks the fundamental question: How is social action organized and common ground established in digital settings and interactions? AIDA uses the analytic mentality of Conversation Analysis (CA), an influential research field in pragmatics. It will be used to study the sequential and temporal organisation of digital action and interactions in distributed digital settings. It will also explore how people change their interactional conduct in digital interactions, indicating learning or adaptation to use new digital communicative tools. AIDA will initiate, underpin and provide a model for a digital turn in CA and pragmatics by building an extensive video corpus, creating ground-breaking video-based methods, and making a path-breaking theoretical and conceptual contribution to the research of intersubjective and digital social action. It will also have broader impact. By highlighting the overlooked social and interactionist features of meaning making in digital interactions, it can impact the existing individualistic and cognitivist theories and mindsets that currently underpin the development of digital technologies. AIDA will identify and describe problems that computer scientists are dealing with or cannot even imagine. Solving those problems will pave way for better digital communicative technologies (e.g., Social VR). Finally, AIDA will spearhead the digital transition in Europe by creating knowledge that can advance the building of more socially and ethically sustainable digital futures with stronger individuals and interactions.

**ID:** 22196

**Start date:** 01-09-2023

**End date:** 31-08-2027

**Last modified:** 08-06-2023

**Grant number / URL:** 356964

### Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

# AIDA: Action and Intersubjectivity in the Digital Age

---

## 1. General description of data

### 1.1 What kinds of data is your research based on? What data will be collected, produced or reused? What file formats will the data be in? Additionally, give a rough estimate of the size of the data produced/collected.

The description and principles of data uses in the AIDA project, as they are outlined below, apply to both existing data and new data collected in the project. The data used in AIDA are video recordings of people interacting in video-mediated, hybrid and immersive virtual reality contexts. The video data can include families, couples, friends, and colleagues, from different social backgrounds, age groups and working-life settings. Volunteers will be recruited through advertisements and/or through word of mouth in local communities or by using existing contacts and connections. There are no inclusion or exclusion criteria.

The procedures of collecting the recordings follow established methods and ethical principles in conversation analysis. The collected audio and video will be of high quality, making possible detailed analyses of participants' talk and multimodal interaction. AIDA will create, test, and implement new methods for collecting video data from digital interactions. More specifically, first, the video data come from video-mediated interactions that take place, for example, in Teams or Zoom. The recorded video data includes the participants' talk (audio) and embodied and multimodal actions (video), but in many cases will also cover each social actor's private perspective to the joint interaction (i.e., the screens and displays) and the physical multimodal environment they inhabit. The recordings are made with digital video cameras and built-in and/or external (wireless) microphones, as well as with screen-capture technology from computer displays (e.g., ScreenPal).

Second, the video recordings are made in interactions in immersive virtual reality (e.g., such VR environments as Rec Room, VR Chat or Sansar). Accessing the social actor's perspective in immersive virtual reality requires special technical solutions. To access an individual participant's talk and actions, we will capture their 'first-person perspective', i.e., the visual view as the participants see it through the VR glasses. This will be done by using screen capture technology. Additionally, we will also have a researcher-observer (acting as a "roving" camera) in the VR following the interactions. We will capture the view from their VR glasses to obtain a recording of the interactions from a 'third-person perspective'. Each of these recordings will also provide an audio recording. Finally, we will record the events in the physical space that the participants inhabit while in VR. This provides analysts access to the participants' private (usually embodied) actions that are important for the analysis. The data will be in English and Finnish.

The audio and video materials will be stored as video MPEG-4 (.mp4) format. MPEG-4 is an International Standards Organization (ISO) specification, and this format is readable for most media players. Selected parts of the video files will be transcribed. The transcriptions will be stored together with the video files. Some of the video data will be supplemented with ethnographic notes that are stored along with the video. All in all, AIDA will collect approx. 30 hours of interactions from each setting (total 90 hours, approx. 1-2 TBs). With several video streams from one interaction (multi-source video), the total amount of video data will multiply. This is more than enough for a mostly qualitative research design that takes a micro-analytic perspective on social interaction.

### 1.2 How will the consistency and quality of data be controlled?

The consistency and quality of the data will be controlled and monitored by following the established and rigorous methods and protocols of collecting audio and video data in conversation analysis. These include the following principles:

1. The data are collected in naturally occurring situations and are not experimental, set up or controlled.
2. The talk and interaction in the recordings are transcribed using established methods and principles in conversation analysis.
3. The data collection follows ethical principles. Filming does not jeopardise the participants' integrity or put them in danger, not does it induce stress, anxiety or humiliation.

Quality and consistency of the data will also be ensured by using the technology, expertise and know-how in the LeaF infrastructure at the University of Oulu, which is specialised in the collection, handling and use of high-quality video recordings. LeaF hosts state-of-the-art video recording technology that will ensure the high quality of recordings. The same recording procedures will be followed in all instances of data collection. This applies to for example video file formats (mpeg-4) and methods of syncing audio and video. The project will also use established software to edit, view and analyse the recordings, ensuring that recordings remain usable for a long time. The research assistant in LeaF, who is an expert on the use of video in interactional research, will take part in the planning and implementation of the data collection.

Finally, the data will be named and archived so that using the data and making searches in it will be as easy as possible.

In addition to the video recordings, the researchers will participate in the recorded events. They will write ethnographic notes, which will help the researchers understand the events in the recordings and to support the analysis.

## 2. Ethical and legal compliance

### 2.1 What legal issues are related to your data management? (For example, GDPR and other legislation affecting data processing.)

The collection, use and managing of the project's data follow the ethical guidelines of the Academy of Finland and the National Advisory Board of Research ethics for human subject research ([http://www.tenk.fi/sites/tenk.fi/files/HTK\\_ohje\\_2012.pdf](http://www.tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf)), the European Code of Conduct for Research Integrity ([http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics\\_code-of-conduct\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-of-conduct_en.pdf)), and the national data protection act (based on EU General Data Protection Regulation).

As to the ethical aspects of data management, the following principles will be followed:

- 1) The video data and the field notes written during observation are used only for research purposes, including publications and conference presentations. In some cases, the data can be used for training purposes. This is regulated on a case-by-case basis and depends on the consent given by the recorded participants or the agreement signed with a possible collaborator.
- 2) The publicity of the data will depend on the specific data, the consent given by the recorded participants and agreements signed with possible collaborators. In order to ensure the privacy and anonymity of the participants and locations, some of the data (e.g., crisis management data) will not be shown publicly nor made open for other researchers. None of the video data will in any form be made available on the internet or in shared databases. This means that large parts of the corpus cannot and will not be made open for other researchers. Exceptions will be considered on a case-by-case basis and depending on the informed consent given by the participants or possible signed agreements.
- 3) As to the crisis management data, no demographic data will be collected or processed. As to other data, some demographic information may be collected for background information: age, gender, nationality and spoken languages (first language, second language, etc.).
- 4) As to the crisis management data, the anonymity of the participants is guaranteed. All details that may reveal the participants' identity (faces, names, name tags), rank and nationality (flags, military uniforms) will be removed or obscured in any representations of the data (e.g. transcriptions, image captures, illustrations, publications, conference presentations). As to other video data, anonymity is dependent on the details outlined in the consent forms. For example, non-anonymised data can be used in data sessions, seminars and workshops. In publications, participants can be anonymized in illustrations by blurring the image with photo editing software (e.g. Photoshop) or by drawing line sketches.
- 5) Before the recordings begin, all subjects are explained how the video data are used. Depending on the consent given the participants or possible collaborators, the data can be used for research teaching (and training) purposes, including publications and presentations. The use of some of the crisis management data is more restricted. The project does not specifically aim to study children, nor does it require that children appear on the recordings. However, since children are often a natural part of people's everyday activities, it is possible that children will appear on some of the recordings. In such a case, by considering their age, they are explained why the recording is taking place and how the recordings will be used. If the children and their parents agree to be filmed, parents are asked to give written consent on behalf of their children. Again, the subjects can refuse to sign. According to the rules of the Ethics Committee of Human Sciences at the University of Oulu, ethical clearance is not needed if the child and

their parents or carers give consent for the research.

6) All participants are told that at any point after signing the consent form they are able to ask to see, hear, edit, or withdraw any portion of the recordings. All the recordings to which a permission is not granted or is denied afterwards are destroyed.

7) If the participating organisations or individuals deny the use of the data after they have been collected, the data will be destroyed.

## **2.2 How will you manage the rights of the data you use, produce and share?**

The research project will not use any data which is covered by the copyright, patents or any other similar legislation. The intellectual property of the video recordings made in the project remain with the researchers who have made the recordings.

As to the crisis management data, only the researchers in the project and who have undergone security screening performed by the Finnish Defence Forces have access to the video data. As to other data, certified researchers will have access to them.

Due to the nature of the data (video recordings made in crisis management exercises involving military staff and civilians), they cannot be shared, used by other researchers or made publicly available.

## **3. Documentation and metadata**

### **3. How will you document your data in order to make the data findable, accessible, interoperable and re-usable for you and others? What kind of metadata standards, README files or other documentation will you use to help others to understand and use your data?**

The data will be documented as follows:

1) Selected parts of the video files will be transcribed. The transcriptions will be stored together with the video files.

2) Other information and documentation, which includes tape logs and information about the participants and the setting, will be stored in electronic format together with the video files. Due to ethical and privacy reasons, access to the video data in the project will be restricted, and they will not be made accessible on open databases.

However, the metadata of the corpus will be made discoverable, identifiable and available online and on shared in the metadata tool Qvain (<https://qvain.fairdata.fi>). The metadata will include the following information:

1. Name(s) of the researcher(s) responsible for data collection
2. General description of the data
3. Where, when and why the data was collected (with possible restrictions)?
4. Which research methods are used?
5. How to refer to the corpus?
6. Who has access to and can use the corpus (including limitations for use)?
7. What is the structure, format and size of the data?

## **4. Storage and backup during the research project**

### **4.1 Where will your data be stored, and how will the data be backed up?**

The audio and video recordings and ethnographic notes will be stored and archived on a secure and password-secured servers at the University of Oulu. The data will also be stored securely on large encrypted external hard drives. The project's metadata will be stored at Qvain (<https://qvain.fairdata.fi>).

### **4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?**

The project's PI (Prof. Pentti Haddington) is responsible for controlling access to the data. Only certified members of the research project will have access to the data. As to some data, only researchers who have undergone security screening performed by the Finnish Defence Forces will have access to the data. Access to the data on secure servers will be carefully monitored and controlled. The data will be password-protected.

## **5. Opening, publishing and archiving the data after the research project**

### **5.1 What part of the data can be made openly available or published? Where and when will the data, or its metadata, be made available?**

Due to privacy and security reasons, the use of the data is restricted and the materials cannot and will not be made available for public use. The metadata of the video materials collected in the project will be made discoverable and available online on Qvain (<https://qvain.fairdata.fi>).

### **5.2 Where will data with long-term value be archived, and for how long?**

The audio and video recordings and ethnographic notes will be stored and archived on a secure and password-secured servers at the University of Oulu. The data will also be stored securely on large encrypted external hard drives. After the project ends, the data will be stored on encrypted hard drives if there is no funding for storing them on password-secured servers at the University of Oulu. The project's metadata will be stored at Qvain (<https://qvain.fairdata.fi>).

## **6. Data management responsibilities and resources**

### **6.1 Who (for example role, position, and institution) will be responsible for data management?**

The project's PI (Prof. Pentti Haddington) is responsible for controlling access to the data. Only certified members of the research project will have access to the data. In some cases, this involves undergoing security screening performed by the Finnish Defence Forces. Access to the data on secure servers will be carefully monitored and controlled. The data will be password-protected.

**6.2 What resources will be required for your data management procedures to ensure that the data can be opened and preserved according to FAIR principles (Findable, Accessible, Interoperable, Re-usable)?**

Handling and editing the video data will require additional services. The project will pay the salary of LeaF infrastructure's research assistant for 1-2 months per year during the project to help in this process. Archiving (preparing and storing on a secure server) the data will take several weeks. Only the project's metadata will be published. After the data has been edited and viewed, which will take several months, the writing of the metadata will take 2-3 days to produce. The video itself cannot be published and made accessible. Storing the video data on the University of Oulu's server will cost several hundreds of euros per year. The University of Oulu's data management specialists (library, computer services) will be consulted during the archiving process.