ATARCA

Project Deliverable 4.1 (D4.1)

Data Management Plan (DMP)

23.06.2022

[This is a revision of the original document submitted 30.06.2021]

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 964678.

ATARCA

Grant agreement number: 964678

Project acronym: ATARCA

Project full title: Accounting Technologies for Anti-Rival Coordination and Allocation

Deliverable number	D4.1
Revision history	First final document submitted 30 June 2021 Latest version revised based on feedback from the mid-review 23 June 2022
Deliverable responsible	Demos Research Institute Oy, DRI
Dissemination Level	Public document

Contributors: Streamr Network

Associacio Novact

Qbit Artifacts

Aalto University

Demos Research Institute

Table of Contents

1. Introduction

In ATARCA, we create cryptographically protected anti-rival tokens and test their applicability to governing industrial data markets and fostering cooperation in community-driven currencies. If successful, this technology will not only help to properly organize the markets for data and other digital goods but provide the structural fundamentals of a new type of economic growth. This will allow the societies at large to more widely explore structurally new incentives for systemic sustainability and scalable systemic intelligence.

For this purpose, ATARCA is conducting research, experiments and policy recommendations. This data management plan of ATARCA considers the aspects of data management, metadata generation, data preservation, and analysis. It includes guidelines for handling personal data. The document outlines how the project handles data both during research and after the project is completed. This ensures that data are well-managed in the present and prepared for future preservation.

2. Data Summary

2.1 Kinds of data ATARCA is based on

ATARCA will collect data and research material to address the objectives as specified in the ATARCA Grant Agreement. Specific measures will be carried out to obtain informed consent for data collection and processing, to protect personal data, to minimize the kinds of data to be collected, to avoid hurting or stigmatizing vulnerable groups or individuals, and to safeguard the rights of the data subjects under the GDPR. All data activities will follow the related national and EU ethics and legal requirements.

ATARCA will utilize three kinds of data:

1) **Qualitative and quantitative research material** on e.g. investigating the ecosystem design and operation, using methods common to social sciences and management studies.
    - Qualitative data include analysis of public policy documents, co-creation workshops and possibly interviews (interview recordings and transcript). Interview recordings will be erased when tapes are transcribed, and transcripts are anonymized and won't include personal or sensitive data. In addition, the ATARCA consortium will explore the potential of ethnographic research approaches, adapted to the virtual working environment (netnography), and will carefully consider whether some of the internal meetings, discussions, or messaging can be utilized as research data (e.g., internal discussions on project Slack channels, recorded Zoom virtual meetings, etc.). These data will only be utilized for research purposes after the affected individuals have been asked for an informed consent, based on a written document that specifies the material to be utilized, the purpose, practices, and personnel having access to the data, and the maximum period for storing and analyzing the data. The participants will retain the right to revoke their approval and retract/remove their personal data any time. This right will be nullified only in the case if the whole dataset will be completely anonymized.
2) Data generated by the participatory experiment via online questionnaires and interviews e.g. a) to define experiments and business model that prepares the innovation for its implementation beyond the experimental setting, b) to assess the appropriateness of the experimental design and calibration/redesign if needed, c) to generate knowledge regarding anti-rival, non-market community dynamics, and tokenization and d) to facilitate the social change in the community toward more sustainable outcomes.

- Users will be invited to take part in research activities that will lead to the design of a new token aligned with the project's field of study. Data collected and generated during the project will help us to identify gaps where further data are needed to address the pilot and prototype design and also will guide us to create a common methodology for the interpretation and use of the token concepts. The data will be anonymized and is described in more detail in chapter 1.2.

3) Publicly available data sources (such as company web pages, white papers, advertisement materials) to strengthen the explanatory power of the research. These data can improve the impact of our results by presenting and connecting the developed concepts and ideas to more relatable and previously known examples. In addition, ATARCA will utilize public transactional data stored on the blockchain in the pilots. This means transactions transferring cryptocurrency and transactions calling functions of marketplace smart contracts. Such data is completely open (as it is stored on the publicly accessible blockchain), pseudonymous (as it is connected to public, impersonal wallet addresses), and accessible to anyone interested in exploring the blockchain storing the data.

The use case of the Streamr platform collects a wide variety of data of the participants, including cryptographic wallet addresses, publicly available blockchain data, browser tracking cookies, social media usernames, and publicly available data of contributions. Personal data is collected both on the frontend and backend side of the system.

The use case of Food Futures Index collects of general data necessary for the App operation are means of user identification (email address), consumption interests and values communicated via the app, and consumption choices recorded via the app. This App data is stored by the platform operator and only has experimental value when users' actions are aggregated, and this is the data stored by Aalto University for the experiment. The second category of data collected relates to the participant feedback in three forms: i. notes from focus groups without any student names; ii. written essays stored without any student names; iii. Exit interviews, initially audio-recorded and transcribed (with originals deleted within 3 months) without names. The third category of data refers to the consent forms which are not used for the experiment and are held by Aalto University.

According to GDPR special categories are racial or ethnic origin, political opinions, religious or philosophical beliefs, and trade union membership, as well as the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health, or data concerning a person's sex life or sexual orientation. Sensitive personal data (such as data about the racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data, or other health or medical data) about the participants are not collected by the ATARCA Food Futures pilot use case.

The use case of Barcelona Green Shops will collect data about the participants' age, sex, educational level, employment status, and data about the participants' involvement in complementary currencies. Such data will only be collected with the participants' consent, and will be used in socio-economic analysis of the pilot experiment. The analysis will not require or use any information about the real-life identity of the participants, nor data which might be sufficient to identify any of the participants.

Moreover, the categories of personal data concerned in the study are as follows: user identification data (phone number, ID number and email address) and transactions of goods and services of users recorded via the App.

Article 35.1 of the GDPR establishes that it is the data controller who has the obligation to carry out the DPIA. Nevertheless, NOVACT prepares the present DPIA in order to comply with the requirements of the ATARCA consortium and to demonstrate its compliance with data protection regulations. The potential high-risk indicators identified in this section are possible given the characteristics and intended uses of the App.

1. Data processing that involves profiling: article 35.3 a) establishes that a DPIA referred to in paragraph 1 shall be required in particular in the following cases, in particular in the case of (a) systematic and comprehensive evaluation of personal aspects of natural persons which are based on automated processing, such as profiling.
2. Observation, monitoring, supervision, geolocation or control of the data subjects in a systematic and exhaustive way: following with the previous risk, the study focuses on observe the behavior of the participants which, necessarily, leads to an observation and supervision in a systematic and exhaustive way of the data subjects.
3. Use of new technologies or innovative use of established technologies: The GDPR makes it clear (Article 35(1), Recitals 89 and 91) that the use of a new technology defined "according to the level of technical expertise achieved" (Recital 91), may require the conduct of a DPIA.

Such data will be used in socio-economic analysis of the pilot experiment. The analysis will not require or use any information about the real-life identity of the participants, nor data which might be sufficient to identify any of the participants. The collection of data starts when the user has fully registered in the pilot application. This means that the participant has given its consent. In this sense, the data will be stored while the project is being carried out. In any case, the user has the right to revoke its consent at any time, without affecting the processing prior to the withdrawal of consent.

2.2 How will the consistency and quality of data be controlled?

ATARCA will not collect sensitive personal data. In the pilot use cases, the project collects a certain amount of personal data (depending on the pilots: participants' age, sex, educational level, employment status, and data about the participants' involvement in complementary currencies. ID number), transaction data, all with the users' informed consent.

When used in research inside the project the data will be aggregated and anonymized. ATARCA will apply pseudonymization to all personal data after data collection and before any further processing. Pseudonymization takes place before raw data is transferred from the user's phone to ATARCA's central repository for further analysis.

The interview data in the project will be collected mainly through 1-on-1 interviews with informants who will be selected based on purposive, theoretical sampling and who agree with our invitation to participate in an interview. The interviews will be video or audio-recorded and transcribed verbatim or documented into text notes. The interviewees' willingness to participate will be confirmed at the beginning of the interview, and they have the right to stop or cancel their participation at any time during or after the interview. The interview data, collected from consensual individuals, will be stored as written transcripts for no more than five years after the project, however with all possible identifying details removed. Also, the original audio or video recordings will be deleted once they have been transcribed and the work quality has been verified the pilot experiment has ended.

The data will be compiled and stored in partners' internal databases in compliance with this data management plan in generally accepted formats: transactions (permissioned blockchain) statistics, images, text messages, and excel - CSV, .txt, .docx, .xslx, XML, aac., mp3, and pdf format.

For comparative purposes, the project may reuse data collected in the previous project on Citizen Exchange System REC, carried out under the European project B-MINCOME (UIA). The collection of that data by NOVACT was in compliance with the legislation on protection of personal data (Regulation (EU) 2016/679 of the European Parliament and Council of April 27, 2016) on the protection of individuals with regard to the processing of personal data and the free movement of such data.

If a user who participated in the previous experiment participates also in the current project, he/she will be asked in the onboarding process if ATARCA can use the data that REC has previously collected. This choice is made via a checkbox in the consent form that is presented to the user. If the user decides not to give that consent, he/she can still participate in the pilot experiment. In such a case, only new data will be collected and any previously collected data will be ignored.

Data will be useful for ATARCA consortium, European Commission services, and European Agencies; and to the general public including the broader marketplace community.

The practices of data collections to ensure consistency and quality are described in more detail below.

2.2.1 Barcelona Green Shop

The pilot case focuses on creating a platform for professionals that increases the interactions of “Green shops of Barcelona” community members. The community, formed around sustainable products (hence the name: Green Shops), is managed by ATARCA collaborator Rezero, and consists of shopkeepers, their clients, and suppliers. ATARCA consortium member NOVACT, will work together with Rezero to coordinate the adoption and use of the new platform within the community. This new platform, in fact, an app mobile is the object of this DPIA (hereinafter referred to as "the App").

The case will make use of Barcelona’s social currency REC (originally promoted by NOVACT). REC is augmented by creating a medium of sharing that promotes coordination, information sharing, and community building among the shops. Moreover, the social currency will facilitate the consumers’ green and sustainable consumption, manifest the Green Shops’ sustainability impact, and build feedback channels (product demand and quality) among all members of the Barcelona Green Shop community. Regarding the tokenization of anti-rival value in the case, the idea is that clients of Green Shops will be rewarded with sntNFTs (non-transferable, but shareable NFTs) that will have smart functionalities and will be shareable among clients (within certain limitations).

Shopkeepers participating in the platform will, in turn, receive ntNFTs (non-transferable, shareable) for their contributions to the community of shopkeepers. As all these tokens are stored on a decentralized ledger, and the ownership of the tokens is recorded as a DLT-based wallet, the tokens' existence is not tied to the platform. Instead, token holders will be also able to prove ownerships of tokens outside of the platform, to enable engagement in various activities (e.g. as a proof of participation in the community).

Table 1. Examples of data that will be collected.

Type of data	Specific data
User identification data	- phone number - email address - picture of ID document - personal ID number (pseudonymization key) - for shop owners: documents on business ownership

Interaction data (actions in the platform)	- Interactions among shop-keepers (shop-keeper data). - Initiation of interesting public business-related conversations (counting of views & responses). - Participation in a topic (counting of likes received for participating) - Private conversations (counting likes received in a private conversation) - Posting/sharing external sources (number of clicks on the link) - Initiation of mentoring chats - Initiation of joint purchases chats. - Number of log-ins. - Invitations to new users - Number of times a certain number of likes is achieved - First time action actions are done in the platform - Interactions between shops and customers (user data) - Number of commercial transactions - Number of sustainability points (sustainability score). - Reviews by users - Posts/sharing of information about shops (Instagram/FB sharing, etc) - Good/service bought - Amount paid (euros) - Qualitative data (interviews and focus group meetings): - Personal data of subjects (name, age, sex). - interview recordings (video/audio depending on the medium, will be erased once the tapes are transcribed) - interview transcripts - written notes on the focus group meetings - Content shared on the platforms: sample of conversations and external content shared in the forum.
Content shared on the platforms	- Sample of conversations and external content shared in the forum (discourse analysis). Gathering conversations data from the discussions with more participation (once a month).
Focus groups with shopkeepers	- 3 focus groups with shop-keeper participants: 1 at the start, 1 during the experiment (mid-term), 1 at the end.

Survey to consumers	- Two surveys to consumers: once at the beginning, with first log in, and one at the end of the experiment. - Content: - Satisfaction question about the platform. - Socio-economic data (optional, only if user wants to disclose) - age - educational level - employment status - past experience with complementary currencies - neighborhood (mentioned in the consent form) - birthplace (mentioned in the consent form) - gender (mentioned in the consent form) - shopping habits (mentioned in the consent form) - type of profession (mentioned in the consent form)
Interviews (non-structured and non-scheduled):	- Short non-structured non-scheduled conversations that will help provide context of the interactions taking place in the network.

2.2.2. Food Futures Index

The case is designed to address the tragedy of the commons dilemma caused by individuals' negative externalities adversely impacting long-term ecological sustainability. It applies anti-rival community currencies in the form of non-transferable fungible (ntFTs) tokens and shareable non-transferable non-fungible tokens (sntNFTs) to recognize actors' contributions and actions with positive sum externalities.

The Food Futures (https://atarca.eu/food-futures-index/) use case is designed to encourage two forms of anti-rival goods generation: data sharing, and positive externalities inverting the tragedy of the commons. There are four types of actors operating in this ecosystem: vendors, customers, sponsors, and the platform operator. Vendors share enriched sustainability data regarding their products, in accordance with an indexed metric, with customers on the platform. Customers (study participants) share data regarding their sustainability values with vendors and the public (consensual, anonymized, aggregated, and GDPR compliant). Customers gain certificates in “sustainable consumption” for fulfilling initial tasks (ntNFT, non-transferable, non-fungible token), and gain “reputation tokens'' for every verified purchase (ntNFT, non-transferable non-fungible token). Reputation tokens confer governance rights once a “sustainable consumption” certificate is achieved. Customers also accrue “impact tokens'' when their acts contribute positive externalities affording sustainability of the environmental commons; these are augmented by a formula when individuals acting in concert have a joint constructive impact (sntNFT, shareable, non-transferable fungible token).

In Stage Two, ATARCA offers surplus goods (tax free items such as meal vouchers or items from the Aalto University shop) dedicated to individuals and members of groups who achieve positive impact according to indexed key performance indicators (KPIs). Impact tokens may be spent on available surplus goods. Surplus goods are made available via first-come-first-serve, a lottery system, or auctions, with the method set by a ballot system of governance open to community members. Thus, allocation of surplus goods, and their pricing, is established via decentralized governance by platform users (customers) who gain voting rights in accordance with their use history (recorded in reputation tokens). Thus ATARCA acts as a donor of surplus goods accruing shared indelible impact tokens demonstrating their support of maintaining a flourishing commons.

Food Futures pilot use case collects three types of data: (1) data necessary to operate the App, some given by the user, some generated by the user while using the App, and some given by the vendor and Aalto University acting in the capacity as donor of surplus goods; (2a) data necessary to give students credit for taking the associated university course; (2b) participant feedback on experimental design collected in i. focus group notes (no student names recorded) taken in three 1.5 hour sessions; ii written essays (with no student names when stored for the experiment (500 words per participant); iii. 30 minute exit interview (initially audio-recorded without student names, and then transcribed without student names); (3) data related to consenting to participating in the experiment which is stored by Aalto University and not used for experimental purposes.

Students provide the information for the App when they set up their accounts and interact with the App. The only data used for scientific purposes is aggregated information. Students provide feedback data which is stripped of any names when it is stored for use associated with the pilot study. Students provide their names when they sign the consent forms which are held by Aalto University.

Tables 2 and 3 specify all data collected in the use case, where it is stored, whether it is shared, and who owns it.

Table 2 Stage 1. Pilot experiment 15 March through 15 May 2022.

Context	Scope	Data	Owner	Storage
App	Not shared	User profile User id Email Data required for authentication (e.g. password hash; depends on the authentication method)	App (hosted for user); Aalto University	Flowa Oy hosted by Google's Firebase, stored on Google's Cloud Platform; for length of experiment—note that users can delete user profiles and their data will be deleted once the experimental analysis has been concluded in 2023; if users remain on the platform and its use continues, their data will remain active
App	Not shared	User's sustainable consumption values, which of 7 variables is of most interest: C02; water use; sustainable agriculture; animal welfare; distance of food source; nutrition	App (hosted for user); Aalto University	Flowa Oy, hosting and storage same as above.
App & ledger	Pseudonymized and possibly encrypted / public	User's meal choices (which main meal selected)	User (in user wallet & ledger) App has copy of that data; Aalto University	Flowa Oy has copy, but data is on a distributed ledger

App & ledger	Pseudonymized / public	Impact tokens (derived from meal choices and user's sustainable consumption values)	User (in user wallet & ledger) App has copy of that data; Aalto University	Flowa Oy has copy, but data is on a distributed ledger
App	Shared / public	Users' aggregated sustainable consumption values and meal choices (anonymized data, also aggregated by use group)	App (hosted for user); Aalto University	Data is aggregated so that a single user cannot be identified Data is public
App	Shared / public	Aggregated meal choices from restaurant (both app users and those who don't use app)	App (as data processor for restaurant)	Flowa Oy; Aalto University
App & Ledger	Partially public	Wallet address	User share this to the platform; App has data, Aalto University has data	Public can see contents of wallet, but wallet address (link user to wallet) data can be reconstructed by linking App user to actual person
Research & Not in app	Used in research article	Background data about users, who participate the research, saved as sampling data in Excel spreadsheet • Gender • Age bracket • Field of study	Researchers	Aalto University; data is deleted upon conclusion of the experimental analysis in 2023
Research & Not in app	Used in research article	Sampling data of restaurant customers; no names asked; recorded from surveys of samples; stored in Excel spreadsheet Gender Age bracket Field of study	Researchers	Aalto University; data is deleted upon conclusion of the experimental analysis in 2023

App	Shared / public	Vendor supplied sustainable consumption data for the app. This includes: • Meal information (name, description, etc.) • Metrics relating the meal (i.e. sustainable consumption values for the protein sources for the meal)	App (hosted for restaurants)	Flowa Oy, Aalto University
University Course	Not shared	• Student name • Student number • Grade for course • Grades for independent assignments	University of Helsinki; instructor (Co-PI)	University of Helsinki, subject to University of Helsinki storage rules; grades for independent assignments deleted after 3 years of course
University Course	Focus group (feedback on user experience); used for research—improving experimental design; reporting on user experience in publications Shared without student names	• Student feedback in written note form (without student names) • On paper, stored as PDF scan, no names	University of Helsinki and Aalto University	University of Helsinki and Aalto University; data deleted upon completing analysis of experiment in 2023

University Course	Exit interviews Data shared without any student names	• Student feedback with audio recordings and notes • No student names associated with files • Transcribed, audio file deleted within 3 months of transcription	University of Helsinki and Aalto University	Aalto University; data deleted upon completing analysis of experiment in 2023
University Course	Written essays	• PDF or word files • No student names	University of Helsinki and Aalto University	Aalto University; data deleted upon completing analysis of experiment in 2023
Experiment	Consent Forms	• Electronic copy	Aalto University	Aalto University; subject to Aalto University consent form policies

Table 2 Stage 2. (Note: has all of the above data; surplus goods to be supplied by Aalto in the form of tax-free items).

Context	Scope	Data	Owner	Storage
App & Ledger	Public	Voting history	User & App	Flowa Oy, Aalto University; outcome of collective voting public, individual voting action deleted upon completion of experimental analysis in 2023
App & Ledger	Public	Token transaction history	User & App	Distributed ledger, public record
App & Ledger	Public	Available surplus good data	Aalto University and shared in platform	Aalto University has copy of data, but it is also public in a distributed ledger

App & ledger	Public	Credentialing	Issuer (shared via app)	App is data processor for the credential issuer (e.g. a university); storage by Aalto University and University of Helsinki

2.2.3. Streamr Platform

In this pilot case, the Streamr platform (https://streamr.network) will be enhanced with anti-rival tokens - sntNFTs (shareable, non-transferable, non-fungible tokens). These will be linked to the existing real-time data ecosystem via Streamr community. Smart contracts with shareable NFT functionality will be developed.

sntNFTs are used to increase community engagement, the contributions of the community towards building the Streamr project, and to increasing its adoption. The research proposition is that by creating new types of tools for acknowledging the programming and non-programming contributions in the Streamr community, we can positively affect knowledge sharing and creation in the community. The assumption is also that the logic developed within the Streamr pilot case is repeatable and applicable to other Web3 communities.

From the research perspective, ATARCA is interested specifically in 1) recording and collecting textual metadata of the type of contribution e.g. reference to contribution such as hyperlinks, platform specific usernames or user handles, 2) details that define what was contributed, 3) pseudonymous wallet addresses of the contributors or the co-contributors, and 4) sntNFTs to contain reference to their origin (wallet address, or sntNFT) or originator (wallet address, smart contract address). All this data is collected with the consent and knowledge of the people who participate in the experiment.

An online platform will be created for sntNFT token creation, browsing and sharing purposes. It will be publicly accessible, and the maximum number of users is not limited. The current size of the Streamr community can be estimated to be around 7000 persons based on Streamr's channel in Discord which is a communication platform used by the community. Only a small subset of them will receive sntNFT tokens but the option to share tokens is available to all community members and even other internet users.

The contribution metadata collected by the sntNFT platform contains transaction data, e.g., cryptographic wallet addresses. As we are experimenting with a new type of tokens, the collection of the transaction data is essential for properly accounting for the legitimacy of transactions and thereby ensuring the integrity of the project. The data may also contain reference to the content that has earned the contribution - e.g. link to a blog. This may or may not be personally identifiable data depending on whether the user has revealed his/her identity connected to the wallet or in the content produced.

An invitation to an online survey related to the pilot experiment will be posted to all Streamr community members. Respondents of the survey will be asked to provide some basic user profile data such as age group, gender, country of residence, activities and interests in the Streamr community, and how long the person has been active in the community. A similar survey will be conducted at the end of the pilot experiment to detect what changes the pilot experiment has brought to the community.

A very small subset of the survey respondents, about 5-15 persons, will be invited to a further in-depth online interview (by asking their consent and contact details in the initial survey) which lasts about 30-60 minutes and focuses on further details about the person's activities and interests within the Streamr community. These interviews will be recorded provided that the person gives his/her consent to this.

No personal data is shared or transmitted between institutions or disclosed to third parties. Any research data shared between consortium members or analyzed or reported will be either aggregated or pseudonymous data which has been anonymized so that all identifying details have been removed. This applies also to socio-economic analysis (survey and interview data).

The analysis and research will be carried out in aggregated or pseudonymous fashion so that in no case can public research results data be attributed or traced back to identifiable individuals. Raw personal data (i.e. data which allows the identification of an individual) will not be accessible to the majority of researchers, developers or other personnel taking part in the project.

The platform will collect cryptographic wallet addresses, publicly available blockchain data, browser tracking cookies, social media usernames and publicly available data of contributions. Personal data is collected in several separate contexts and in summary as follows:

- Ledger: only the token id, wallet address and link to token specific metadata (in backend)
- Frontend: Website traffic data (Google Analytics), Privacy policy settings
- Backend: All specific metadata related to an individual token, with username (eg. Discord handle, or real name in case the person prefers to use it), category, title and link to contribution, and link to image

- Consent forms (related to receiving or resharing tokens): wallet address
- Contact information: needed for contacting token recipients to send consent forms
- Online surveys: Demographic data about respondents, information about their activities and opinions related to Streamr community
- Online interviews: Qualitative information about the interviewees' activities and opinions related to Streamr community

These different contexts (user research data, pilot data and consent data) are stored separately from each other. Only persons that could viably associate pilot participants to user research data are the named researchers on Streamr & TEX. All token metadata (with potentially several personally identifying details) is stored in a centralized backend instead of the ledger. The ledger does not contain any data that could be used to identify the person. This arrangement makes it possible for the recipients of sntNFT tokens to exercise their GDPR rights such as the right to be forgotten. In practice, this involves modifying or removing the metadata in the backend, resulting in the token metadata either changed or removed. The token metadata which is stored in the backend consists of publicly available data based on the contribution which the person has done to the community, eg. details of published texts or published code in Github.

Platform users will be invited to participate in user research consisting of two parts:

1) Online surveys (both at the beginning and end of the experiment) which are conducted to build an overall understanding of the community, its characteristics, behavior and opinions. As part of the surveys, respondents will be asked to provide some basic user profile data such as age group, gender, country of residence, activities and interests in the Streamr community, and how long the person has been active in the community.
2) In-depth interviews which are of 30-60 minutes of duration and conducted on the Zoom platform. An invitation to an in-depth interview is published as part of the survey (there is a question in which the respondent can provide his/her contact details and opt-in to be contacted by the ATARCA team for setting up the interview). Participation in the interview is voluntary and the respondent can opt-out any time even after giving the consent. Interviews are recorded for research purposes but only with the interviewee's consent. (Otherwise, the interviewers will take notes during the interview.) The interviews are transcribed based on the extracted audio recording, and with all identifying details (names etc.) removed. These transcriptions will be used for all analysis purposes.

While all results of our study will be presented with pseudonyms, the integrity of qualitative research requires that the researcher(s) maintain a chain of evidence between the qualitative data (e.g., interview quotes) and the informant. Yet, such information, (eg. The original interview recordings will never be shared between organizations and will only be managed by the leading researcher). Sensitive personal data such as data about the racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data, or other health and medical data of the participants will not be collected by ATARCA.

For the Streamr community pilot experiment, the data collected during the study will be removed within five years after the end of the project. Also, the pseudonymized interview data (stored as written transcripts) and the original audio or video recordings, collected from consensual individuals, will be stored for up to five years after the project, however with all possible identifying details removed.

Streamr sntNFT platform data will be stored as long as the platform remains operational. Data on the public blockchain (with wallet addresses but no identifiable details that would make it personal data) will be stored as long as the blockchain remains operational. The personal data collected as a part of ATARCA research project will be processed only in EU/EEA countries. While the sntNFT service is located in the EU area, on some instances it is possible that parts of the personal data will be also processed outside EU/EEA area (e.g., when a user located outside the EU/EEA area shares tokens on the platform or provides his/her data for a user research). However, these instances on handling personal data are not the result of the ATARCA project.

Participation in the Streamr community pilot case is completely voluntary. Selected Streamr community members will receive sntNFT tokens issued by Streamr as an acknowledgement for their contributions to the community. They also have the opportunity to share them further if they wish. The token recipients will be asked for their consent if they would like to receive this kind of token before it is minted and transferred to their wallets. (It is however unlikely that they would decline to receive the token as it is an award and highly valued by the community.) Also, the surveys and interviews are voluntary as community members can simply ignore the invitation to the survey and a possible in-depth interview.

As part of a survey or interview, community members give their informed consent to giving their personal data, and they will be informed about what data will be collected, how the data is used and processed, who is the data controller and the data processor(s), and what are the rights and freedoms of the user as far as data is concerned. A privacy notice and an information sheet are accessible on both the sntNFT platform website and on the project website.

Participants need to be adults. Minors who are unable to give informed consent will be excluded from the study. Sensitive personal data such as data about the racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data, other health and medical data of the participants, or data about their criminal convictions will not be collected in any experiment of the ATARCA project. The personal profile data collected from those who receive sntNFT tokens is minimized to include only the necessary details needed for issuing the token: contact information, username (or real name in case the person prefers to use it), wallet address, and reference to the publicly available information about the contribution.

Analytics data is collected from all users of the sntNFT platform with the aid of Google Analytics for research purposes (monitoring platform use during the experiment), and for software development purposes. GDPR compliant privacy preserving version 4 with IP anonymization enabled in Google Analytics will be utilized.

Those participating in the user research are also asked to submit some basic demographic data such as age group, gender, and country of residence, in addition to questions related to their activities and interests in the Streamr community, and how long they have been active in the community. The demographic data will be used in the project only to understand better the characteristics of the community members and will not be used for any other profiling. However, it is possible that some collected demographic information may contain or imply sensitive personal data indirectly (eg. the country of residence or username may imply ethnic origin). We assume that the described policies for storing and handling personal data such as pseudonymization and anonymization provide adequate security to also such indirect, sensitive personal data.

As part of the informed consent and privacy policy, as well as on the sntNFT platform and project website, participants are advised about their GDPR rights and how to exercise them if needed. A summary of the rights and practices is presented in Table 4.

Table 4: Data rights and practices.

Data right	Process to exercise data right
Access to personal data	A data subject may request a copy of their personal data in a structured and machine readable format at any time via email. The data subject may contact the Streamr team by email or the DPO either by email or via the data rights form on the project website to request the personal data to be sent to him or her by email.
Right to rectification	A data subject may contact the Streamr team by email or the DPO either by email or via the data rights form on the project website to request rectification (i.e. the correction of inaccurate personal data).
Right to be forgotten	A data subject may erase their personal data at any time by contacting the Streamr team by email or the DPO either by email or via the data rights form on the project website to request their personal data to be erased.
Right to restriction of processing	A data subject may withdraw their consent to data processing at any time via email. He or she may contact the Streamr team by email or the DPO either by email or via the data rights form on the project website to withdraw their consent.
Right to data portability	Once a data subject has requested and received a copy of their personal data, he or she is free to transit the data to another data controller. Right to object A data subject may withdraw their consent to data processing at any time via email. He or she may contact the Streamr team by email or the DPO either by email or via the data rights form on the project website to withdraw their consent.
Right not to be subject to automated decision making and profiling	A data subject may withdraw their consent to automated decision making and profiling at any time via email. He or she may contact the Streamr team by email or the DPO either by email or via the data rights form on the project website to withdraw their consent.

3. FAIR data

3.1 Making data findable, including provisions for metadata

The data collected in ATARCA will require careful anonymization and scrutinization via each of the project partners as described in the previous chapter with the support of ethics mentors for making the data openly accessible. In particular, due to the discrete nature of our research data, our priority is to secure the privacy of our study participants and informants. However, the project will implement policies to make data and results accessible, when possible, without compromising data security.

The consortium has made a Joint Controller Annex to the ATARCA Grant Agreement, specifying the roles and responsibilities of each ATARCA consortium member on handling personal data. Whenever possible, the data will be stored and processed within the EU area. The members of ATARCA will not transfer any of the project personal data outside the EU, except in accordance with the data protection legislation. The project will utilize communication tools and channels which take such requirements into account.

The data will be pseudonymized whenever possible, however the real identity of the persons can be revealed in some cases where a person has attached personally identifiable data to a public contribution residing outside of the experiment platform eg. a real name on a personal blog post. The user study will be planned in a way that personal data is minimized, eg. name and contact details are not asked and demographic data is entered in the survey with less precision (eg. age group instead of age). There are however cases in which the respondents can still be identified, such as when they give their contact details for setting up an interview, or if they voluntarily provide identifying details in their answers to open survey questions. However, whenever the answers are used as excerpts in publications, they will be first anonymized. The interviews will be analyzed using only the transcriptions, not original recordings. As part of the transcription work, all identifying details such as names (or usernames) of people or projects will be removed or pseudonymized. Interview recordings are stored only to maintain a chain of evidence, and they will be destroyed within five years after the end of the experiment.

The project will produce software artefacts and written reports that will be made openly available. We have selected the Github open data repository as our main channel for providing the software artefacts, found at https://github.com/atarca.

The written reports will be published in open access via SSRN repository for academic reports. We will focus our communications efforts to increase the publicity of these results. Following reports will be published:

- Report describing alternative incentive mechanisms capable of capturing some of the positive externalities arising from sharing digital goods with anti-rival properties.
- Report describing a set of anti-rival business model archetypes, documented as physical and virtual pattern cards, also published under CC-BY-SA 4.0 license.
- Documentation of the anti-rival toolkit to support developers.
- Report on the initial reception of the anti-rival business model design toolkit and the educational dissemination.
- Four observatory publications, policy recommendations and roadmap, final report, and report of participation.

Data transactions will be accessible on the blockchain to participants of the blockchain. Data collected for data marketplace will be available on the data marketplace:

- Data streams inside data products have user-defined metadata
- Data on a public blockchain is available to participants of this blockchain
- Research report metadata (published on ATARCA website)

3.2 Making data openly accessible

The main type of open data of ATARCA will be the technological artefacts that will be shared in the Github repository.

Once the data on the practical experimentation of the project has been successfully anonymised, we will carefully evaluate whether that data (or some elements of it) can be made openly available. Some of it could be protected with authentication inside the CKAN platform because it could contain sensitive data that cannot be opened. This protection will be decided depending on the type of data collected. A cautionary approach will be applied due to the discrete nature of the data.

All data on a public blockchain is publicly accessible. If project partners want to re-use personal data previously collected for a different purpose, they will first ask for the consent of the data subject before processing this data.

All data on the Streamer platform is publicly accessible. Creator of data products decides if he wants to offer data for free or for a specific price per hour: https://streamr.network/marketplace

Data collected in the Barcelona use case of Novact and Qbit will be available under data.rec.barcelona, which hosts a data management system (CKAN). The data will be accessible following the CKAN documentation under https://docs.ckan.org/en/2.9/ following open data foundation standards. CKAN allows following a clear versioning system for the data. Dome sensitive data sets will be restricted and will need to sign an NDA to be able to use it. NOVACT will manage the permissions to access this kind of data. Data sets can be accessed with an internet browser under data.rec.barcelona, and later downloaded in different formats like JSON, CSV, PDF, TXT, and other standard formats depending on the type of data (GeoJson). Also, CKAN allow accessing data using its API following CKAN documentation (https://docs.ckan.org/en/2.9/api/index.html).

In addition, the consortium aims to publish scientific research articles in the most prestigious journals that help to illustrate the practical relevance and theoretical implications of the research project. All journal publications will be published under the gold open access license. The publications in conference proceedings will be made accessible following the rules and policies of the conference organizing body (many times the conference proceedings do not support open access publishing). In addition, the publications and the metadata of data will be inserted to the consortium parties' electronic services, like in the Aalto University to the research information system ACRIS (acris.aalto.fi), whenever applicable. The data and the publications are publicly available via the open repositories, e.g., in Aalto university via the research.aalto.fi/en/, and via the other institutional or open access metadata catalogs, e.g. national services Etsin for the data and for publications the Virta as well via the repository of EU and national funded research OpenAIRE.

The data to be long-term stored will be made such that it will not need any additional information to be interpreted in the future, but it is applicable as such. All facilities to be used for managing, preservation, and sharing the data are free of charge and do not require additional help from experts.

3.3 Making data interoperable

The interoperability of the data results will be greatly improved by the results that support subsequent software development on anti-rival resources. ATARCA considers that the chosen emphasis on producing educational content, e.g., the planned, introductory-level MOOC on anti-rival business models can contribute to the interoperability of the project results and data. Such deliverables help to create common communication artefacts that facilitate shared understanding of this emerging topic and help to align future research efforts.

We will make NFT metadata publicly available. Data on NFT minting, sharing and other transactions are publicly available. Such data can be accessed either directly from the blockchain or via a performance oriented GraphQL interface.

2.4 Increase data re-use

If project partners want to re-use personal data previously collected for a different purpose, they will first ask for the consent of the data subject before processing this data.

Data collected in the practical experiments will be published under Creative Commons Attribution-NoDerivatives 4.0 International license. Some datasets could work with other licenses depending on the type of data and will be decided before publicly sharing.

4. Data security

The used and generated research data will not raise ethical or any other sensitivity issues, such as research carried out with human cell types or personal data.

However, the administrative materials will include personal data, and so in their handling, we will apply the EU's General Data Protection Regulation's (GDPR) requirements. Administrative material will be handled by Aalto University according to the Finnish Personal Data Act and to the Act on the Openness of Government Activities and Archives Act. The project administration is committed to following the guidelines issued by the Finnish Advisory Board on Research Integrity on good scientific practice. The project also follows the European Code of Conduct for Research Integrity by ALLEA.

ATARCA will use data storage systems that take into account the appropriate security level of the data in question and the needs born from our multidisciplinary collaboration. The relevant databases and systems are secured by passwords of sufficient length and complexity, and accessible only by authorized users involved in the research project. Further security measures will be put in place if vulnerabilities are detected in the risk analysis.

For data storage and publishing, we only consider solutions that fulfill the requirements above to safeguard all possible usage restrictions. The confidential industrial data will be stored in a secure environment agreed to and conforming to industrial partners' policy on data storage which will be available to the consortium. Permission should be sought prior to publishing any data provided to the consortium as part of the research.

Regarding the technical measures, ATARCA will use data storage systems with the appropriate security level for the data in question, taking into account the need for secure multidisciplinary collaboration. All consortium members hold high security standards and rely on professional IT services. The data is stored on secured servers. The relevant databases and systems are secured by passwords of sufficient length and complexity, and accessible only by authorized users involved in the research project. Database operations are restricted at CRUD level (Create Read Update Delete) and the access to sensitive data is managed by the responsible person. Laptops used by the researchers automatically encrypt data with Bitlocker or a similar system. Further security measures will be put in place if vulnerabilities are detected in the risk analysis.

The data will be stored primarily in the institutional servers' network drives of the organization, which has generated the data. The services include a snapshot feature and regular backups that make file versions automatically to recover from unwanted deletions; tape backups provide also system-level disaster recovery. The internal storage services are typically provided to internal users only:

- All consortium's sensitive and non-publicly available data is stored and is being worked with inside following services: Microsoft Teams, Dropbox, OneDrive (all these three are Aalto University's own private service instances). These tools provide storage, recovery, and transfer options. Aalto University is responsible for the consortium's joint workspaces.
- All DRI's non-publicly available data (excluding personal or sensitive data, which is stored to the consortium's common services mentioned above) is stored and being worked with inside the Google Workspace set of tools, which only DRI's employees have access.
- Some data is stored in Atlassian Cloud via the JIRA tool (responsible: Streamr).
- Data collected in the experiments by NOVACT and Qbit will be available under data.rec.barcelona, which hosts a data management system (CKAN).
- Laptops used by the researchers of Streamr dealing with pseudonymized transaction data are equipped with automatic data encryption with e.g. Bitlocker

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 964678.