Impact on privacy and data protection of Citizen science projects

Introduction

1.1 **The processing of personal data in citizen science projects**

Citizen science is an inevitable trend in the context of a society that demands higher levels of participation in a growing number of fields. This trend is facilitated by the increase in the "digitisation" of social interactions and, in general, a greater presence of information and communication technologies (ICTs).

The extension of a vast network of computers and sensors enables outsourcing of tasks during research processes, leveraging existing resources through calls for distributed open collaboration. With the public increasingly equipped with individual devices (some of which, such as "smart" mobile phones, can accommodate up to 20 different sensors), and a growing culture of participation (some of which is a faithful reflection of the development of web 2.0), it is not surprising that scientific research projects involving citizens to varying degrees have emerged. This support is facilitated both by the availability of infrastructures for generating useful data and by the existence of extensive data flows that have already been created through these networks. However, the dynamization of digitised information flows is not without risks. On the one hand, due to the digitisation and interconnection of the devices themselves, increases the fragility of the information; on the other hand, due to the increase in the number of subjects involved, which means that part of the control and responsibility must be transferred to a large number of people.

Data management techniques in citizen science projects do not necessarily differ from those applied to traditional research. However, there are some specific issues arising from the involvement of volunteers during the process[1].

One of the key issues is the setting of clear objectives, in order to determine proportionately the information needs and the appropriate means. In the field of scientific research, the premise that a greater volume of information leads to better results must not fall into the trap of overloaded information; on the contrary, saturation of information can precisely obstruct the objectives and procedures of research. It is therefore important to obtain relevant and quality information in order to make useful analyses from which clear conclusions can be drawn. With regard to personal data, it is clear that individuals' specific identities are rarely relevant. However, some of the characteristics of such individuals (age, gender, etc.), which, alone or in combination with others, may reveal the unique identity of a subject, are. In the context of citizen science, conventional

[1] OCDE, 2013. Guidelines on the protection of privacy and transborder flow of personal data. Disponible en http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm#part2

relations between subject and object are altered. The delegation of tasks through collaborative networks of citizens implies in a certain way the "multiplication" of the scope that the subject of the research can have on a given object, facilitating its deepening in different aspects without having to carry out additional investments, in many cases difficult to assume.

Although sometimes the collaborator is at the subject-object boundary (since he is himself part of the research object and active or passive collaborator of it), his personal identity, and therefore the information associated with it, is in no case relevant to the research. The volunteer acts as a tool, provides a useful means to achieve an end (your camera, your smartphone, your shared notes, etc.), so your anonymity can and should be appreciated and respected. However, alongside the knowledge provided in quantitative terms by a collaborative network of citizens, there is the figure of the expert contributing knowledge in qualitative terms. In these cases, the quality of the expert is closely linked to a series of characteristics that support this condition, many times to the professional's own "name", so that in this case the identification of the subject is more justified.

An illustrative example of the collection of personal data in citizen science projects is provided by the BarcelonaLab's Office of Citizen Science. It has a form[2] that is made available to potential users to promote citizen participation in scientific research through collective experiments. This form allows volunteers to register in a system that makes it easier for project coordinators to contact potential experiment participants. To do this, they must fill in up to 10 information fields and accept the informed consent for the processing of their personal data. In general terms it is possible to see that the potential collection of information is excessive for the above-mentioned purpose ("to report on cultural activities organised by the Institute of Culture of Barcelona"). In the proposed form, participants are encouraged to give their first and last name (incentivizing anonymous registration), and optionally they are invited to give their telephone number and company name. Bearing in mind that this is a first approach to users for the recruitment of volunteers, the volume of information that can potentially be processed is above what is necessary.

In order to achieve good practices in terms of data protection in the field of citizen science, both organizational (including privacy policies) and technical measures can be taken. By proposing warnings and recommendations, and adopting rules of use, the volunteer is delegated responsibility for making wise use of the tools and, where appropriate, personal information about him/her. The use of technical resources to improve privacy complements the strengthening of privacy by avoiding the possibility that both managers and users of the tools have the possibility of making risk decisions. In any case, the first step towards systems that respect the privacy of individuals is to avoid unnecessary processing of personal data. On the other hand, citizen science project coordinators are obliged to provide volunteers with sufficient information on the consequences of the projects in which they participate, warning them of potential risks and threats.

[2] Bowser et al., 2014. “Sharing Data While Protecting Privacy in Citizen Science”. Interactions, XXI.1 January – February. Pp. 70-74.

1.2 Privacy and data protection

Personal data' means any information concerning an identified or identifiable natural person. They are those that allow to "identify a person", being able to acquire different formats (numerical, alphabetical, graphic, photographic and acoustic). The fundamental element to determine that this is personal data is that the information, by itself or in combination, allows inferring information about a specific person, either because it is directly identified through some data or because it can be identified by another means. Examples of personal data include: name, surname, date of birth, postal address or e-mail address, telephone number, marital status, bank account number, tax identification number, car registration number, fingerprint, DNA, photograph, social security number, etc.

The concept of privacy' does not necessarily correspond to that of data protection'. On the one hand, the idea of privacy covers a wide range of aspects, to the extent that some authors have come to identify up to seven categories of privacy. Among the dimensions that have been listed are privacy of information, communications, personal conduct and actions, group privacy, personal and family or household privacy[3]. Therefore,"privacy" encompasses a broad spectrum that goes beyond the mere right to data protection, as it refers not only to the processing of personal information but also to intrusions more closely linked to the individual himself or herself.

On the other hand, data protection goes beyond privacy, as it deals with any processing of personal data, regardless of whether they may compromise the privacy of an individual or not[4]. The importance of ensuring the protection of data by any organisation lies in a wide variety of reasons: ethical, legal, economic and organisational, etc. The protection of data by any organisation is important. The fundamental right to data protection is based on the right to informative self-determination. This premise promotes the protection of individual privacy and personal information, so that individuals can exercise control over their own personal data as much as possible. This implies that in the cases in which it is possible and relevant, they give them consciously and knowing the purpose of such collection, in addition to having the corresponding rights of access, rectification, cancellation and opposition. As stated by the Spanish Data Protection Agency:

The fundamental right to the protection of personal data derives directly from the Constitution and gives citizens a power of disposition over their data so that, on the basis of their consent, they may have access to them."[5]

Similarly, the fundamental right to data protection is recognised twice in the European Constitution, with Directive 95/46 governing the principles by which Member States' legislation must be guided. In Spain, Organic Law 15/1999 regulates this regulatory framework (the development of which is regulated by Royal Decree 1720/2007), with the Spanish Data Protection Agency in charge of protecting and guaranteeing this right.

[3] Based on: https://www.ipc.on.ca/images/Resources/7foundationalprinciples.pdf

[4] PIAF, p. 15. [EUROPEAN PROJECT]

[5] AGPD, 2014. Guía para una Evaluación de impacto en la protección de Datos personales. Available at https://www.agpd.es/portalwebAGPD/canaldocumentacion/publicaciones/common/Guias/Guia_EIPD.pdf

A reform in this area is currently under way in the European Union, which will lead to a broader, updated and standardised regulatory framework for all EU countries in the coming years, resulting in a new General Data Protection Regulation[6]. By means of the regulatory proposals that have taken place in recent years, the possibility has been suggested that under certain conditions organizations may have to submit themselves to IDPs (Data Protection Impact Assessments)[7]. On the other hand, the National Security Scheme indirectly already contemplates carrying out this type of analysis in the case of public administrations, through Royal Decree 3/2010, aimed at "ensuring access, integrity, availability, authenticity, confidentiality, traceability and conservation of data, information and services used in electronic media that they manage in the exercise of their competences"[8]

In the relevant regulation, it is relevant to mention exceptions concerning the retention of personal data for historical, statistical or scientific purposes. In any case, and in general, the data may only be kept for the purposes for which they have been collected and for the necessary time, since the principle of quality includes the minimum retention of personal data, which must be cancelled when they are no longer necessary or relevant to the purpose for which they were collected or recorded. If they wish to be kept for statistical and scientific purposes, it has to be verified whether there is specific legislation setting a time limit or specific requirements for it (e. g. the patient autonomy law or the Clinical Trials Register)[9]. At the end of the intended storage period, the data must be destroyed. They may be retained only if they are previously decoupled or exceptionally, on the basis of their historical, statistical or scientific value and in accordance with specific legislation[10], it is decided to maintain certain data in full. (art. 9.2 of RD 1720/2007). To this end, the controller must request it from the Spanish Data Protection Agency or, where appropriate, from the Devolved Regions' supervisory authorities, which may apply the procedure provided for in Articles 157 and 158 of Royal Decree 1720/2007. To this end, the controller must initiate a request through which he or she must[11]:

- Identify the data processing to which the retention is intended to apply
- Expressly state the reasons that would justify the authorization.
- Explain in detail the measures that the file manager intends to implement to guarantee citizens' rights.

[6] At the time of writing this report, the agreement reaches an advanced stage of negotiations: http://europa.eu/rapid/press-release_IP-15-6321_es.htm

[7] Proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Section 3, article 33. https://www.piawatch.eu/node/1023

[8] Royal Decree 3/2010, of January 8, which regulates the National Security Scheme in the field of Electronic Administration. Art. 1. http://noticias.juridicas.com/base_datos/Admin/rd3-2010.html#a1

[9] There is an exception reflected in RD 1720/2007, according to which "data may be preserved during the time in which may be required some kind of liability arising from a relationship or legal obligation or the execution of a contract or of the application of pre-contractual measures requested by the interested party ". https://www.agpd.es/portalwebAGPD/canalresponsable/conservacion_datos/index-ides-idphp.php

[10] In particular, the provisions of Law 12/1989, of May 9, regulating the Public Statistical Function, Law 16/1985, of June 25, of the Spanish Historical Heritage and Law 14/2011, of June 1, on Science, Technology and Innovation, and their respective development provisions, as well as the regional regulations in these matters.

[11] AGPD. Indications regarding data conservation: https://www.agpd.es/portalwebAGPD/canalresponsable/conservacion_datos/index-ides-idphp.php

- Provide as many documents and evidence as necessary to justify the existence of historical, statistical or scientific values.

The exceptional nature of the retention of personal data on historical, statistical and scientific grounds is evident, and the reasoning for considering such a situation must be emphasised.

In the common case of being able to dispense with personal data, especially if they are to be made available to the public by promoting the dissemination of results through open data initiatives, it is necessary to consider the decoupling and guarantee of non-reversible anonymity. The Article 29 Working Party has published a document presenting different techniques for anonymizing information[12].

In the field of scientific research, transparency, open participation and citizen collaboration are not at odds with privacy and risks should not be ruled out or underestimated. Irrespective of whether or not sensitive personal data[13] is used, which implies higher levels of protection (such as medical research), the processing of personal data must have supervisory mechanisms in place to avoid undesirable consequences, whether derived from intentional or unintentional errors. Trends towards open and collaborative science should not result in "total transparency" that could jeopardise the confidentiality of personal data, should they be collected.

The growing incorporation of ICTs in procedures implies, on the one hand, the digitalization of information and, on the other hand, the interconnection of devices. Scanning enables faster processing of information, more efficient storage and easier backups. It also facilitates the distribution and sharing of results, whether raw data or data subject to analysis procedures. However, digitised information also brings a number of additional risks arising from precisely the same features that bring benefits, such as increased ease of reproduction, the presence of metadata or differences in access recording and storage protocols.

The interconnection of devices facilitates the decentralization of tasks and the work in multidisciplinary and geographically dispersed teams, as well as the sharing and exchange of information. Similarly, these benefits are accompanied by a number of additional risks, such as the possibility of illegitimate access to information or involuntary sharing of data.

Concern for the privacy and data protection of individuals should not necessarily be based on plausible risks. The risk (s) may arise later, when it is already too late, or may not be perceived at all and some negative consequences may be directly apparent. In other words, the argument that there is no "nothing to hide" is totally wrong, as such a judgment is only entitled to be the right to be heard.

1.3 Data protection principles

[12] Ideally, more than one technique will be used. The document is available here: http://ec.europa.eu/newsroom/article29/news-overview.cfm

[13] Data revealing ideology, union affiliation, religion and beliefs; data that refer to racial origin, health or the sexual life; data relating to the commission of criminal or administrative offenses

In order to ensure the privacy of information, the recommendations described above will be based on the following list of data protection principles, inspired by OECD guidelines[14]:

1. **Principle of data collection limitation**. Promotes the limitation and minimization of data collection. Only the minimum data essential for the specified purpose (s) will be collected. This personal data will be obtained legitimately and legally, and as far as possible, after obtaining the data subject's informed consent.

2. **Data quality principle**. Personal data must be accurate, complete and up to date. The collection of information should be relevant and proportionate to the objectives and needs that motivated it.

3. **Principle of purpose**. The purpose of data collection must be communicated to the data subject and maintained during processing. Any change in purpose must be communicated and consented to.

4. **Principle of limitation of use**. Disclosure or publication of data or changes in the actual use thereof shall not be permitted unless agreed by the data subject or legitimized by a competent authority.

5. **Principle of safety**. Reasonable and proportionate security measures shall be taken to protect the collected data from risks such as unauthorised access, unwanted destruction, unlawful use, modification or disclosure.

6. **Principle of transparency**. It promotes the transparency of the rules and the development of data processing, including the communication of the existence and characteristics of the files created to the data protection authorities.

7. **Principle of individual participation**. It refers to respect for the rights of self-determination of information that favour the control of personal information by individuals concerned: rights of access, rectification, cancellation and opposition.

8. **Principle of legal responsibility**. Those responsible for the files must be legally accountable for complying with these principles through the established legal mechanisms.

For the specific case of Citizen Science projects, Bowser et al[15] refer to the four **privacy standards** proposed by the FTC's principles of good practice: notification, choice, access and security, and also provide a number of tips for protecting personal data in this area.

- **Notification**: Clearly inform participants about *any* personal data collection *before* it takes place. This obviously includes the cookies and user registration procedures involving the use of an e-mail address.

[14] OCDE, 2013. *Guidelines on the protection of privacy and transborder flow of personal data.* Accessible from http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm#part2

[15] Bowser et al., 2014. “Sharing Data While Protecting Privacy in Citizen Science”. Interactions, XXI.1 January – February. Pp. 70-74.

- **Choice**: Offer personal settings of privacy options to facilitate the possibility of participating more anonymously if the user so wishes. An example would be the possibility of not publishing certain data or distorting the location.
- **Access**: To promote transparency by facilitating the user's access to the data shared, as well as the modification and cancellation of these data. It will always be better if this happens through a direct interface than through intermediaries that delay or hinder the process.
- **Security**: Ensure the protection of information shared between participants and project managers, for example through encryption of shared information and the use of https protocols.

Specific **advice** in the field of citizen science for good practice includes:

- **Determine what data can be released** in terms of information accuracy, sharing and public visibility. Once decisions have been made, implement the relevant technologies (vague geolocation, anonymisation of identities, etc.).
- **Give sufficient notice of privacy options**. Explain the circumstances under which normal participation could pose a risk to privacy. Inform volunteers about who will review their data for quality control.
- **Give volunteers the option to hide some of the data** and locations in public interfaces, or the option to publish data anonymously.
- **To allow volunteers the possibility of modifying and deleting their data**, both personal information in a strict sense and information that contains data about the volunteer.
- **Require only minimal personal information about volunteers**. Demonstrate the value of the data collected, and explain who will have access to them. We recommend a multi-level access control system that takes into account the different roles and needs of the parties involved.

1.4 Privacy by Design and Privacy by Default

In order to guarantee individuals informational self-determination, developers of technological innovations may choose to implement Privacy by Design principles, and configure them based on the Privacy by Default parameters. Privacy by Design puts the privacy of users first, promoting their integration into the design of technologies (in a broad sense: hardware, software, network designs, etc.). Privacy by Default implies proposing by default the most restrictive options in terms of privacy when there are different options that can be configured for this purpose, in such a way that the active choice of the user when it comes to sharing data is reinforced.

Tools for mediating interfaces in Citizen Science projects, such as apps for mobile devices, can also achieve good practice in the field of data protection if they follow the 7 foundational principles of Privacy by Design (PbD)[16]:

1. **Proactive, not Reactive; Preventive, not Corrective**. The Privacy by Design approach is characterized by proactive rather than reactive measures. Anticipate and prevent privacy invasion

[16] Based on : https://www.ipc.on.ca/images/Resources/7foundationalprinciples.pdf

events before they occur. PbD does not wait for risks to materialize, nor does it offer remedies to resolve privacy breaches once they have occurred - its purpose is to prevent them from occurring. In short, Privacy by Design comes before the event, not after.

2. **Privacy as the Default Settings**. Privacy by Design seeks to deliver the highest degree of privacy by ensuring that personal data is automatically protected in any computer system or business. In the absence of intervention, the level of privacy remains intact. No action on the part of the individual is required to protect privacy - it is integrated into the system, as a default setting.

3. **Privacy Embedded in Design**. Privacy by Design is embedded in the design and architecture of computer systems and business. It is not added as a complement, a posteriori. The result is that privacy becomes an essential component of core functionality. Privacy is an integral part of the system, without diminishing its functionality.

4. **Total Functionality - positive addition versus zero sum play**. Privacy by Design seeks to accommodate all legitimate interests and objectives, without the benefits in one setting implying concessions on the part of another. Privacy by Design avoids the hypocrisy of false dualities, such as privacy versus security, demonstrating that it is possible to have both at the same time.

5. **Extreme-to-Extreme Safety - Full Life Cycle Protection**. Having been embedded in the system before any data has been collected, Privacy by Design maintains security conditions throughout the entire lifecycle of the data involved. Robust security measures are essential for privacy, from start to finish. This ensures that all data is retained and deleted safely at the end of the process, without delay. Therefore, Privacy by Design ensures secure end-to-end management of the information lifecycle, from start to finish, from end to end.

6. **Visibility and Transparency - Keep it Open**. Privacy by Design seeks to ensure that all relevant parties, regardless of the business or technology involved, are operating in accordance with the stated promises and objectives, and subject to independent verification. Components and operations remain visible and transparent to users and suppliers. Confidence but at the same time verification.

7. **Respect for Users' Privacy - User-Centered Approach**. Above all, Privacy by Design requires that architecture managers and operators maintain in a superior position the interests of the people, offering measures such as robust default security, proper notification, and user-friendly options that increase control. The aim is to maintain a user-centred approach.

2. Description of the procedure and methodology

In order to manage the possible risks related to the processing of personal data within an organization (considering the options of avoiding or eliminating risk, mitigating, transferring or accepting it), there are procedures that provide a **preventive framework** that helps to systematize the assessment of potential dangers. While ideally these procedures should be carried out **prior** to the design of a new technology or service (whether it involves software, hardware, or a combination of both), the incorporation of an analysis after implementation can be equally useful, especially if no significant incident has yet occurred.

There are several concepts to call such procedures, which are not necessarily interchangeable[17]: Privacy Impact Assessment, Privacy Risk Management and Data Protection Impact Assessment.

Due to the use of the latter in the context of the European proposal for a General Data Protection Regulation, it will be the one chosen to refer to such a procedure. However, it is important to note that currently, the DPIA is a **concept under construction** and that only recently are methodologies with a certain level of standardization being developed.

According to the Spanish Data Protection Agency (AEPD), a Data Protection Impact Assessment (DPIA) is an "analysis of the risks that a product or service may entail for the protection of data of those affected and, as a result of this analysis, the management of such risks by adopting the necessary measures to eliminate or mitigate them"[18]. The methodology used here is intended to go beyond mere legal compliance in the field of data protection and technical audits of information security. It aims to cover, from a multidisciplinary point of view, both legal aspects (legal warnings, compliance with procedures with data protection agencies), as well as technical aspects (interface security, structure of data storage), organisational aspects (relationship between members of an organisation, relationship with users), attitudinal aspects (perceptions, doubts, predispositions, etc.).

The methodology proposed by the AEPD proposes a comprehensive 8-step evaluation, ranging from the mere analysis of the need for an DPIA to the implementation of recommendations, which will then be subject to new review and feedback processes.

In order to guarantee both the quality and confidentiality of the data collected, an analysis of the **information flows** through the different stages facilitates the detection of specific risks at each moment. There are several approaches to defining the different stages of the data lifecycle.

In the case of **data processing for scientific research open to public participation**, the DataOne Public Participation in Scientific Research Working Group proposes a total of 8 steps: planning, collection, quality, description, storage, selection, integration, and analysis[19]:

- **Planning**: Description of the data to be collected and details of its processing.
- **Collection**: Manual or automated data collection and subsequent digitization.
- **Quality**: Checking the quality of the information obtained through reviews and inspections.
- **Description**: Labelling and classification of data through metadata.
- **Storage**: Sending data to a file for long-term storage (e. g. a data center).
- **Selection**: Location of potentially useful data collection.

[17] PIAF, p. 15. [European Project]

[18] AGPD, 2014. Guide for an Impact Assessment on Personal Data Protection. Available at: https://www.agpd.es/portalwebAGPD/canaldocumentacion/publicaciones/common/Guias/Guia_EIPD.pdf

[19] DataONE’s Primer on Data Management (Strasser, Cook, Michener & Budden, 2012). Available at: https://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf

- **Integration**: Combination of data from different sources to create homogeneous analyzable sets.
- **Analysis**: Study of the available data.

However, in this analysis we have tracked the data flow according to the following simplified structure of the information life cycle:

- data collection
- data storage
- Data processing, editing and analysis
- Relationship with third parties
- data deletion

The team has relied on the following procedures and tools for the implementation of the DPIA:

- **Analysis of web pages and applications**. Observation and navigation through the tools managed by citizen science projects in order to check any privacy related issues within the framework of usability and interaction with them.
- **Questionnaires**. Forms structured around life cycle phases to be completed by team members with direct access to the information required in each case. They provide a deeper insight into the project, with information on the design of the tools, their operation and possible future changes.
- **Legal analysis**. Observation of compliance with the legal requirements established in the field of data protection.
- **Technical analysis**. Observation of data exchange dynamics through the tools used, taking into account both weaknesses and good practices in information security issues.

3. Impact and overall assessment

While it is true that not all situations in which personal data processing is carried out entail the same risks to privacy (considered in both a quantitative and qualitative sense), the importance of data protection as a right to be observed and protected should not be underestimated regardless of the circumstances surrounding it.

A bank or hospital does not bear the same risks as a car dealer. The type of information handled and protection needs are different. However, individuals have **the same right** in both cases to have the personal data transferred handled with due protection. As this is a right, an organization cannot unilaterally judge the need to protect privacy. In fact, the actual "fragility" of a data and the privacy consequences that may result from malpractice in its processing can vary from one day to the next depending on the circumstances or specific situation of the individual.

It is a mistake to believe that transparency and participation are at odds with privacy and data protection. They are simply two issues that need to be harmonised in such a way that they are compatible. The accessibility of the results for both participants and the general public has great advantages in terms of transparency and bringing science closer to the citizenry and making knowledge accessible to everyone. However, the personal data of a person involved in a project, whether as a "participating subject" (intermediary), as a "participating object" or as a "participating subject" are generally not decisive or even relevant to the dissemination of results. There is one key issue in opening up science to the public, and that is the fact that not only closed results and conclusions about a particular study in which a scientific team has identified a number of objectives and approaches must be made available to the public. The key is to release the output in such a way that other teams can take advantage of the information obtained, and thus generate new results, cross and relate them to other sources or apply new approaches and interpretations. That is to say, make the datasets available in their least elaborate state possible. In this way, the context in which data processing in citizen science is framed is closely linked to the retention of data for statistical and scientific purposes. However, as mentioned above, the retention and publication of raw data does not justify the retention of personal data and therefore data sets must be subject to the relevant decoupling processes.

The potential diversity of projects both in terms of objectives and needs and means, implies great differences in terms of personal data protection. When it comes to carrying out scientific research, the deployment of means is totally different when observing the behavior of bacteria in Biology than when carrying out a life history in Sociology. This is why each area is unique and therefore impossible to establish a "magic formula" to guarantee good practices in data protection. It is doubtful whether it is possible to establish a series of axes that allow in some way to increase the degree of systematization in the analysis of the impact on privacy of a citizen science project (e. g., depending on the type of science involved, which in the case of multidisciplinary projects may be of little use).

The projects analysed thus respond to a variety of needs and objectives. However, it is possible to point out that data protection principles help to establish common guidelines, and in particular the aim of collecting the least amount of personal data necessary. Likewise, the projects present a very different compliance with the principles of protection of data and the relevant regulations, so that it is difficult to make a general assessment in one way or the other.

A common feature of projects is that none of them processes personal data by using files that are not in digital format or are not connected to a network (e. g. paper files). In this regard, it should be remembered that printing the contents of the database or creating copies through optical drives would involve reconsidering the privacy and security needs of the resulting files, to avoid both unwanted access and loss of information.

As regards the volume of data collected, their variability does not necessarily reflect the real need for collection. We also appreciate various efforts to anonymize the personal data collected. However, it is not sufficiently clear whether this is due to a diversity of objectives, which give different value to the real identity of the collaborators. One point to note is that not only the actual data processing currently taking place, but also the potential for collection and processing must be taken into account.

This is where a point of friction arises between transparency and data protection, due to the figure of the "expert", whose ability to validate results must be guaranteed.

In general, regulatory compliance is satisfactory, although there are differences in the visibility of the use of cookies and guarantees to obtain informed consent. One aspect in which there are also wide differences is how to deal with the communication of privacy policies (as well as the terms of use of the websites/applications). Although the legal minimums are respected, it is possible to go further and make a communication effort that guarantees not only the comprehension of the terms on which the data is processed, but also the acceptability of this processing. To this end, the legal texts can be complemented with simplified and illustrative schemes, highlighting the type of data to be collected and how they will be processed, especially in the previous moment of data collection.

With regard to security measures at a more purely technical level, it should be noted that the web pages used within the framework of the projects lack the https protocol to ensure safe navigation. It is also advisable to address the issue of password management, given the widespread practice of reusing passwords between different services. A highlight in this area would be metadata (data of the data). The participants in the projects collaborate with elements that contain information at two levels, one more explicit (e. g. an image) and the other that is collected in the background (e. g. date, camera model, etc.). Metadata, if collected on a passive scale, can lead to the generation of unique identifiers because they create combinations of variables that are difficult to match from one user to another, making it difficult to anonymize if they are crossed from different sets, when that same element or another with the same combination of metadata is linked to personal data.

Finally, the deletion of data is another issue where it is necessary to intervene to improve the protection of personal data. This is one of the most controversial aspects, since in research, the renunciation of data retention may be somewhat paradoxical. However, it is essential to establish a justified expiry date in which a decision must be taken between decoupling or deletion, since the long-term retention of files containing personal data must be fully justified on the basis of an obvious need. Both decoupled data and results can be retained, but it is important to avoid indefinite storage of personal data.

In summary, it should be noted that the two most important aspects to be taken into account with respect to the projects analysed are the minimisation of data collection in a justified, consensual and informed manner, and the provision of disposal procedures once their use is no longer necessary.

4 Project analysis

4.1 Bee-Path

Bee-Path is a tool that "allows studying human mobility, registering it through an application for mobile devices"[20]. The final objective of this project is to offer an **automatic analysis of the type of mobility carried out by the user** through mathematical models that explain the observed phenomena. In this way, it will be possible to make mobility forecasts in specific contexts. *Bee-*

[20] Web page of Bee Path: http://bee-path.net/?lang=es

Path involves the multidisciplinary work of physicists, biologists and artists within the Office of Citizen Science of the *BarcelonaLab*.

The experiments of the *Bee-Path* project have taken place on three different occasions. The first was held in 2012 as part of the Science and Technology Festival at the Ciutadella Park in Barcelona, organised by the Barcelona Institute of Culture, with an influx of around 12,000 people. The same experiment, but adding a further degree of complexity (the participants had to look for objects and the application registered the strategies used), was repeated in the 2013 edition of the Science Festival.

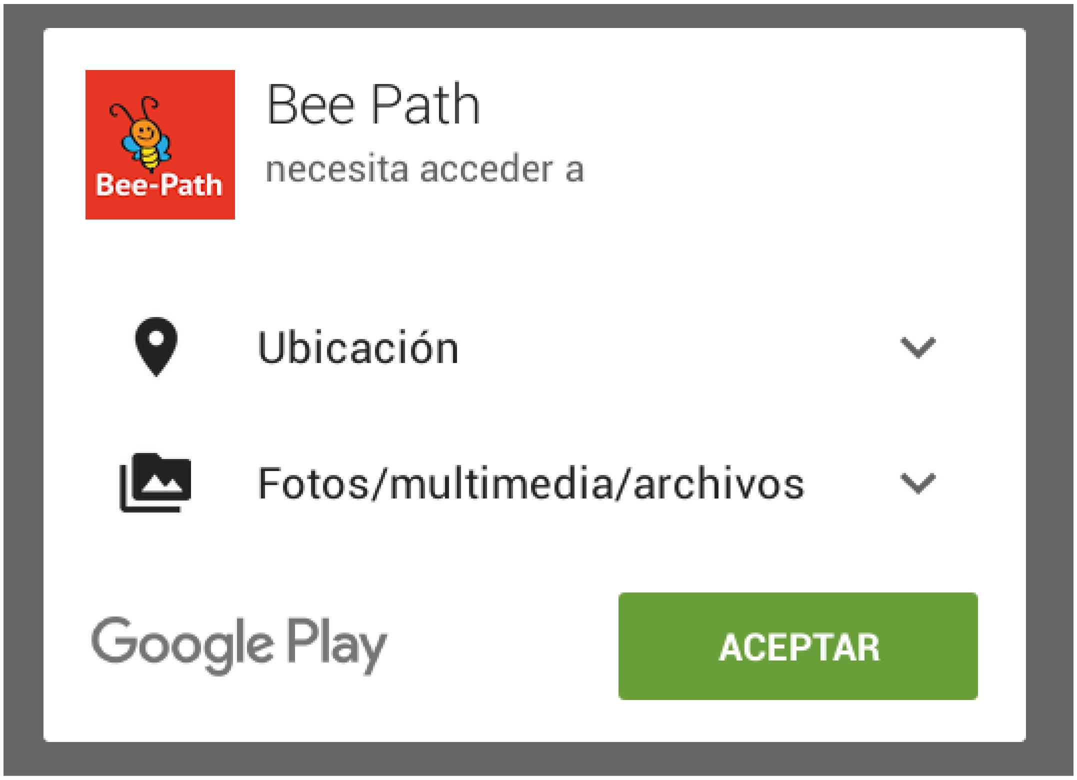
In 2014, a collaboration was established with the *Domestic Data Streamers* collective and *Counterest* company, to analyse visitor mobility patterns in a room of the Big Bang Data exhibition at the Barcelona Centre for Contemporary Culture (CCCB).

4.1.1 DATA COLLECTION

This project collects personal data from two main sources: online forms (available via the web-app and app), and sensors managed by the *Bee-Path* application. Through this application, volunteers submit information on mobility patterns via the GPS locator, and accelerometer use is planned for the future. Thus, the variables collected are as follows:

- User (ID)
- password
- email
- gender
- geolocation
- age group (0-25;25-40;40-65; + 65)
- how they perform the experiment (alone/accompanied)

When the application is downloaded, *Google Play* warns of limited access to user resources:



Bee Path necesita acceder a

The main strength of this project in terms of data collection lies in the fact that the application **collects data only in the relevant area**. This is why the idea of opening up spatio-temporal conditions for data collection increases the risks of re-identification and excessive data collection or non-relevant data. In this way, it is advisable to maintain the objective of the study in analyzing **how individuals move through a particular area**. The decision to focus the weight of the study on a specific area or on human mobility in general has different consequences in terms of privacy and data protection.

In short, it can be said that this collection is limited and proportional to the context and needs of the project, since it avoids the unnecessary collection of variables, avoids the collection of data outside the relevant context, avoids the easy identification of individuals and renounces the collection of personal data of special protection. However, it is important to create a study framework that considers *future* possibilities for data collection and data collection, as the possibility of including gender, age and type of mobility (alone/accompanied) in the analysis in the future is considered.

In this regard, it is necessary to highlight the risk of **interest in additional data** on the population studied. When carrying out studies on human mobility, it is possible to segment the patterns of human mobility according to variables of different types (e. g. age), especially if it is necessary to make the study profitable through the commercialization of results. The experiment itself opens the door to questioning differences in movements according to additional characteristics (age, gender, etc.), which would increase the possibilities of deanonymization. There is no specific number of variables from which identifiable profiles of individuals could be more easily inferred. However, it is possible to affirm that **as the number of variables increases**, although they do not reveal unique identities, the real anonymity of an individual is reduced. That is, for a given population, there are a limited number of people who meet a specific combination of a number of variables (e. g. only the combination of age, gender and address can shed light on a real identity). Therefore, the increase of variables is generating increasingly clearer profiles on specific identities. By Therefore, the increase in variables is generating increasingly clear profiles about identities concrete. In this sense, the use of broad categories can be recommended as opposed to collection of specific data, as has been the case with age, collected by groups of age instead of requesting more invasive data and that facilitate the identification of unique individuals, as is the date of birth.

On the other hand, it is important to point out the risks that would derive from the activation of the mobility analysis **outside the specific context of the study**. That is, it is important that the analysis be carried out with respect to a series of **anonymous** individuals at a given time and place. If it were decided to expand the geographical scope of the study and focus the study of mobility on **specific individuals**, there would be risks of unnecessary identification of subjects and excessive data collection. For example, despite the fact that the subjects were anonymous, an analysis of their mobility patterns in a broader context could shed light on their real identities as it can illustrate heatmaps that show places of common presence (home, workplace, etc.).). If, for example, we could reveal unique patterns of mobility that are repeated in different experiments, we would construct unique, differentiable profiles. Even if personal data of an individual is not collected, the way in which it moves, especially if it is subject to detectable routines, may reflect information that potentially helps identify the individual.

With regard to the **notification** of data collection, it is possible to state that it meets minimum criteria but can be improved. During the registration process in the app, the first screen displays the following warnings:

"Bee-Path: Experiments of human mobility"

In order to process your data correctly and so that you can consult the results of your experiment yourself, we need you to fill in the following fields. The username will be the same as that which will appear on the ranking screens with the best search times.

General conditions

We inform you that your personal data will be used exclusively for scientific purposes. The data obtained during the experiment will be anonymized, unlinked from your email and will remain open at https://github.com/bee-path/. Bee-path researchers undertake to use the collected data only for the scientific purpose mentioned above. The interested party declares to have full knowledge of the destination and use of personal data by reading this clause and the express approval of the clauses set forth.

The privacy policy stated in this DPIA questionnaire is as follows:

The personal data provided in the registration form and during the activity to be carried out as part of the Bee-Path research project will be incorporated into the "Research, development and innovation projects with basic personal data" file, owned by the University of Barcelona, with the final aim of to be treated in order to achieve the objectives established in the mentioned research project. The body responsible for the file is the General Secretary. In any case, you may exercise the access, rectification and cancellation rights by means of a written communication, enclosing *a photocopy of the ID card or other identifying document, addressed to the General Secretary of the University of Barcelona, Gran Via de les Corts Catalanes 585, 08007 Barcelona, or by electronic mail to the following address: secretaria.general@ub.edu.*

I declare that I have received the information.

The latter text should appear on the first registration page of the web-app before any data is entered. That is, including the details of the name of the corresponding file and the responsible for it.

However, it is important to point out a success: the warning that "the username will be the same as that which will appear on the ranking screens with the best search times". Non-identifying pseudonyms could be also suggested.

On the other hand, there is no warning on the website that informs the general public of the privacy policy and precautions taken in the project, which could be very useful, for example to attract the interest of potential participants. In other words, the advantages of limiting data collection in order to encourage greater participation are not sufficiently exploited and there is no specific section in which individuals can access the terms of use and privacy policy.

Finally, note that the website installs cookies from the following sites, which can be used to monitor the user:

- vimeo. com
- dropboxusercontent. com
- linkedin. com
- slideshare. net
- scorecardresearch. com

Both Slide Share and Vimeo use Google Analytics, which also collects information from users who visit *Bee-Path* through the videos and slides embedded in the page. This is why the use of cookies should be properly warned and accepted.

4.1.2 DATA STORAGE

With regard to data storage, it is possible to point out the accuracy involved in using differentiated databases that store on the one hand the most easily identifiable data (User ID; password; e-mail; gender; age range; solo/accompanied) and on the other hand, geo-positioning data (User ID, GPS; timestamp). With regard to the management of e-mail addresses and passwords, the most recommended option is to restrict access to information related to e-mail addresses, and manage passwords through the procedure of "salted password hashing"[21].

The possibilities for re-identification are case-by-case and are not easy to predict a priori, so the different procedures for decoupling data should be laid down. Although a user can register as 'Anonymous01 User' and have an e-mail address that is' 40132@yahoo.es', with very low possibilities of reidentification, can also do so with the username 'WilfredoBasterretxea1979' with an e-mail address that is' WilfredoBasterretxea1979@nameofthecompany.es', which increases the chances of identifying a person with certainty. Moreover, it is common for individuals to reuse passwords for different registration processes, so that a security breach in a project that may be apparently of little interest to an attacker could result in a customized attack with greater consequences. Appealing to individual responsibility for the use of data is not a safe option.

Likewise, the bet for an internal server (from the University of Barcelona, in this case) against the use of cloud services allows **greater control of data**. The consideration of cloud services in the future must be subject to the corresponding guarantees of security and data protection, so that the file manager has no difficulty in carrying out any request to enforce the ARCO rights of users, as well as the effective and definitive removal of data when necessary. It is also important to know the exact physical location of the servers in order to consider the specificities of the relevant legal frameworks (if located outside the EU). One possibility to consider would be to keep some of the data on the internal server (e. g. e-mails and passwords) and to outsource another part of the data (e. g. geopositioning and other characteristics of individuals less easily reidentified). In any case, it is important to consider limiting access to information through the separation of privileges, as well as controlling access through activity logs that reflect data management protocols (these may reveal unusual or illegitimate accesses).

4.1.3 DATA PROCESSING, EDITING AND ANALYSIS

Within the framework of the *Bee-Path* project, there are a number of initiatives aimed at increasing both transparency, participation and openness:

- Publication of the data created in open format, as well as the corresponding source codes.
- sending a personalized report to users on their mobility.

These procedures do not have to involve any significant risks as long as the relevant pre-anonymisation processes are carried out. In no case is it relevant, of course, the inclusion in the raw data packets of the passwords listings or the e-mail addresses of the users. The inclusion in the future of more complex data sets (including more variables) could pose risks to data protection, and possible increases in the ability to identify individuals should be assessed on a case-by-case basis. That is, if additional features are added to an anonymized or apparently anonymized profile,

[21] The procedure of Salted Password Hashing is detailed here: https://crackstation.net/hashing-security.htm

the chances of re-identification increase. Moreover, each space defines the probabilities of reidentification of an individual; there is not the same probability of randomness in a spatial-temporal context as the "Science and Technology Festival in the Ciutadella Park", that in a context such as "no holiday Monday at 10:00 a. m. in Room 506 of the Law School." The first case is not linked to specific identities, but in the second case there is a specific population to be expected, and therefore the criteria for considering the possibilities of re-identification must be tightened. Open spaces with the potential to accommodate a larger number of individuals are more conducive to respecting the anonymity of participants. To this must be added the desirability of correctly explaining to the public good practices in terms of data protection with regard to the publication of raw data and the generation of personalised reports. These reports should be made available to the user through the app and in no case by using e-mail services.

Since user names can be highly identifiable, it is advisable to use the numerical IDs obtained by order of registration for any procedure that will take place, including publication of data. During the information cleansing processes that take place to achieve quality data (users without experimenting or failing, etc.), it is important to note the importance of correctly and securely deleting data that will not be used, so that appropriate decisions are made about the data sets. If there are purified subsets, they must be subject to the same data protection measures as the previous ones.

With regard to data access, it can be said that the proportion of staff with access to data is not excessive. Although there are no systems for recording access to information and actions carried out, there are mechanisms to prevent risk practices and unauthorised access: the server cannot be accessed from outside the UB outside the periods in which experiments are carried out, and on the other hand, permission to receive data from new volunteers within the UB is regulated manually.

4.1.4 RELATIONSHIP WITH THIRD PARTIES

As regards the relationship with third parties, it is possible to point out the obvious absence of unnecessary risk practices. This is the case, for example, of the moderate relationship with social media. Registering users through your own interface, instead of using third parties such as Google or Facebook, avoids unnecessary sharing of personal data. Equally, it is a good thing that the settings for sharing results are limited to the project context, rather than making excessive use of others such as those based on social networks.

The transfer and manipulation of data by third parties is currently, in principle, very limited. The consideration of future experiments in the field of mobility should take into account that the same privacy guarantees are maintained in all phases of the data lifecycle. In other words, the involvement of third parties to support research work should leave no doubt as to the aspects outlined in this report: data minimisation, relevance, clear purpose and limitation of use, anonymisation, etc. The other parties involved must also commit themselves to these criteria. For example, in the case of the 2014 experiment at the Big Bang Data exhibition, it is crucial to ensure that Counterest maintains the same standards when collecting information through sensors and, on the other hand, to communicate the procedure clearly to visitors.

It is important to note in this respect that if the data are effectively anonymised, **no transfer of personal data to third parties takes place**. On the other hand, if there is personal data, it will be necessary to consider which entity carries out the collection in order to know who is legally responsible for the file.

Given the idea of generating platforms to share results, it is recommended to avoid publishing the content in open or social networks. In other words, for results that may contain more identifiable data, such as the user name, it is advisable to offer the information in an environment that requires prior registration of users.

In the case of networked or global initiatives (e. g., including other universities), in order to generate comparative studies or a global pool of data, it is desirable to establish a joint data protection plan based on the above-mentioned criteria. The compatibility and compliance with national data protection regulations will also have to be discussed with the relevant partners in the third countries concerned. In these cases, it is important to clearly define who exercises the role of the "data controller", which according to Spanish legislation (LOPD) is defined as the "natural or legal person of nature public or private, or administrative body, which decides on the purpose, content and use of the processing "[22].

4.1.5 DATA DELETION

The deletion of data is undoubtedly one of the most problematic issues, not only in this area, but in any organization that carries out the processing of personal data. In the case of scientific or statistical studies, it is common to foresee indefinite storage for such purposes. However, it is important to differentiate between completely anonymized and identifiable datasets. In the first case, there are no incompatibilities when storing information indefinitely. However, raw data that may contain identifiable information should be subject to some sort of expiration date. In order to carry out diachronic studies with single subjects, a specific period of time must be considered in order to obtain or renew the consent of the subjects whose data are being processed.

If the data is processed by third parties, the guarantees for the deletion of personal information must also be maintained once it is no longer used or at the request of those affected. It is essential that the subjects can maintain control of their personal information, even after having terminated the contractual relationship with third parties that carry out collaboration tasks. In other words, if, after a certain period of time, the external organisations processing personal data in the framework of the project no longer have any relationship with the organisation in charge of managing it, care must be taken to ensure that ARCO rights are exercised and that the data expires and is subsequently securely deleted.

In this case, the information is not expected to expire, nor are there any established protocols for its deletion. It would be advisable, therefore, to establish a scheme to define the protocols for the

[22] Organic Law 15/99 , 13 December, on the Protection of Personal Data

elimination of personal information once it is not necessary or when the affected subjects demand its elimination.

4.1.6 GENERAL AND SAFETY PROCEDURES

With regard to other procedures related to information security and data protection, there are no major risks. The testing of interfaces and systems is not carried out on a case-by-case basis, the same conditions of use and security levels of the websites/apps are guaranteed for all possible variants of use (e. g. in different language versions), and the division between personal and organisational tools is encouraged.

Although there has been no access to security protocols and policies[23], it is clear that the team has clear references to address in case of doubt, in this case the Legal Adviser to the UB's rectorate and Data Protection Coordinator.

However, it is essential to point out that the website does not support SSL and this means that the administrator's password is easily captured by an attack. It is currently recommended to enable HTTPS and change the password.

As for organisational and technological changes, they have been discussed previously throughout the previous sections. A reduction of the research team does not necessarily mean an increase in data protection risks.

4.1.7 CONCLUSIONS

In short, it is worth noting that one of the project's strong points in terms of data protection lies precisely in the restriction of data collection spaces and times, which prevent unconscious or non consented collection by users. The key of moderation in data collection lies in compartmentalising studies and linking them to a given space-time context, rather than linking them to specific subjects whose data are collected in different space-time contexts. Different approaches to the objectives and questions to be answered show different levels of risk. For example, if you want to answer the following questions:

- how does a particular individual move around a neighborhood?
- how does a concrete individual move through a city?
- how does a particular individual move, with defined age and gender characteristics, generally without geographical limits?
- How does a particular individual move for an hour?

[23] Although according to the answers of the questionnaire it is affirmed that there exists a security document and protocols or policies of security, copies that confirm it have not been attached.

- how does a particular individual move for 24 consecutive hours?
- how does a particular individual move for a period of 7 days?

Here it can be seen that the levels of intrusion are very different, and that the information gathered and profiles that can be drawn about an individual range from patterns of mobility in limited contexts to obtaining a person's movements over a more or less extended period of time. Therefore, higher levels of data accuracy in the study will require greater argumentation about collection needs, purpose definition, etc. By way of example, if an individual's mobility patterns are captured for periods of more than 24 hours, it is feasible not only to ascertain their address but also their time out of it. If the subject's e-mail address is used in other fields (i. e., it has not been created specifically for registration), and in some way it can be linked to a user with registered mobility patterns, a relatively detailed profile can be established about a specific individual, their schedules and origin-destination patterns.

Likewise, in order to prevent possible negative reactions from the volunteers, as well as to properly deal with their possible complaints, it is essential to inform, communicate and explain the characteristics of the study, indicating who is responsible for the treatment of the data, the purpose of the collection, the time during which they will be stored and of course, the possibility and means to exercise the rights of the data subjects. The voluntary and non-remunerated nature of participation does not make a significant difference in any of the above-mentioned aspects. This communication work can be carried out not only by means of the corresponding explanatory texts, which include the legal minimums required, but also by linking the different phases of the data flow through illustrative schemes.

A series of **recommendations** can lead to improvements in the project:

- Clear specification of the **objectives** for any campaign: In this way, you will get an idea of the **most relevant data** to obtain. To this end, it is possible to find answers to a series of questions on basic aspects of each specific study:

➢ What do you want to find out?

➢ Why and for what?

➢ Where will it take place? (specific area)

➢ What forecasts do you want to obtain?

➢ For how long will the information collected be needed?

- **Avoid additional data**: Taking into account the previous point, carefully evaluate the use of data not previously collected: Is it planned to use them? What real utility do they bring? What difference does it make in the results? Can new information be usefully exploited? What additional risks are involved?

- Limit data collection to the **specific context of the study**: Avoid time-consuming data collection for the same group of subjects.

- Clear **communication** of the collection: Facilitate the privacy policy on the website and explain to both app users and potential participants the best practices regarding data protection. Insist on the anonymization of openly shared data and highlight the flow of information through explanatory schemes.

- Submit the data to **decoupling procedures**. Both the username and the e-mail address can, depending on each case, reveal specific identities. Submit password management to' Salted Password Hashing' procedures.

- **Avoid external servers** as much as possible.

- When it comes to sharing the results, create **secure platforms** with anonymous data and not make raw data available that might reveal identities (e. g. usernames, e-mails, etc.). Avoid automated interconnection of results with external interfaces such as social networking services.

- Define the **expiration** date and the protocols for data **deletion**.

- Enable **secure browsing** via HTTPS.

4.2 Sea Observers ('*Observadores del Mar*')

Observers of the Sea "is a citizen science project that covers research into various phenomena related to the marine world. Through feedback from volunteer participants, it is possible to reflect information covering issues as diverse as overfishing, the occurrence of invasive species, pollution, changes in biodiversity, etc.

To this end, a web platform coordinated by the Barcelona Institute of Marine Sciences (CSIC) is made available to volunteers. A number of experts from different national and international research centres validate and comment on the observations received. In this way, the website is conceived as "a meeting point between citizens and scientists, which aims to create new knowledge together"[24]. The results are shown on a map that combines the multiple contributions, allowing visibility of the different phenomena and projects covered according to different variables and sub-variables.

[24] Web page of Observadores del Mar: http://www.observadoresdelmar.es/projectes.php

4.2.1 DATA COLLECTION

In order to reflect the observations whose contribution is decentralized through a network of volunteers who contribute through forms and experts that validate the contributions. The interface for data collection is the website itself, which first of all encourages user registration. The first phase of registration (http://www.observadoresdelmar.es/apuntat.php) urges volunteers to fill in 4 mandatory fields: name, surname, email and password, and offers two optional fields: language and type of user (individual/association).

After this first registration phase, ODM sends a confirmation e-mail to the user, including the password in plain text format. It is advisable to avoid this way of managing users' passwords, since the security of e-mail and webmail services is beyond the reach of the ODM. It is known that many times users reuse passwords, so it is not convenient to have messages inside the mailbox that reveal them easily.

Once the volunteer has registered, in your user area you will be able to access the "Personal Data" section, where you can modify and collect additional personal data. More specifically:

- Permission to publish the e-mail (yes/no)
- Postal address
- Population
- Province
- Country
- Biography
- Photo
- Web
- Facebook
- Twitter
- Google+
- Linked In

It also allows you to change the password and displays information about the registration date and IP.

It is important to stress here that no unambiguous informed consent is apparent. The privacy policy is located in a link on the website, but there is no box to fill in accepting the privacy policy before sending any personal data.

In addition to these variables, precise geo-positioning data (latitude and longitude) as well as the date and time of the sightings are included when entering observations. Assuming that the observers whose profile is collected on the web are those who have actually carried out the observations, in several cases it is obtained that a person' X' (with a relatively detailed identity defined in a relatively detailed way) has been in an exact' Y' place at the exact time' Z' This information is publicly available and traceable, and can reveal patterns of mobility.

At this point, a key question arises: the relationship between purpose and relevance and proportionality of personal data collection. Although some of the data collected are optional, their collection must also be justified in terms of the purpose to which they are linked. With regard to, for example, permission to publish e-mail, it would be advisable to disable it by default (opt-in mode). It has been proven that its publication on the web is affected by Google's indexing engines. Moreover, after later trying to hide the e-mail address of the Observers of the Sea website, (since it is not possible to hide it from the beginning) the Google search of the address "xxxxxxxxxxx@gmail.com" still redirects to the file originally created, which, however, no longer shows this address. It is equally important in any case to avoid the publication of e-mail addresses on web pages, due to the existence of robots that crawl the Internet in search of text strings under the format "xxxxx@xxxxx.xxx" in order to send spam.

The privacy policy shown on the website indicates that personal data is collected "for the purpose of **maintaining** and **managing** the relationship with the user". The questionnaire indicates that the purpose of the data collection is to "**connect with the user for the validation of data**, know if the data refers to the **area in which they live**, know their **profile** to know if they are an **expert** or not in any subject". On the other hand, there is no identified purpose for collecting the logging date, IP address, and date of the last visit (although these data may be useful for enhancing the security of user accounts).

We therefore have serious doubts about the purpose of collecting personal data. On the one hand, the most basic (name, surname, e-mail and password) are the ones required for managing and maintaining the account. The rest is for the validation of observations and the assessment of volunteers as experts. Overall, it is possible to generate very precise profiles (depending on the extent to which optional data are filled in), available not only to the people involved in the project, but also to anyone browsing on the Internet.

Here it is necessary to assess the importance and real need to know in detail the identity of the subjects who make the contributions. Is it in fact intended to create a "meeting point" where the different contributors of information form an interconnected network in which personal information is a useful contribution? In other words, do you really want to encourage the generation of some kind of' social network' for marine research? Or does the value of the information lie mainly in the observations made? Would contributions have the same effect if they were anonymous or the identification of more limited observers (e. g., would they be anonymous? that the coordinators had access to details of who provided the comments but these were not published on the web).

Depending on the actual objectives of the research and, as a consequence, on the real needs of identifying and defining volunteers, the potential for data collection as a whole can be considered excessive. In any case, the system does not facilitate or promote the anonymous contribution of observations, and the problem of this fact lies in the potential loss of control of data by the user.

One issue to highlight is the dubious contribution of the field "Mailing address", bearing in mind that there is an opportunity to provide information about the population and the province.

In the initial recommendations about observations, it may be possible to warn that the attached photos should confine their content to the observation itself, avoiding additional elements (persons/situations). On the other hand, it should be noted that the images shown on the page contain metadata. The metadata of the images themselves do not reveal personal information, but in the case of Observers of the Sea they are directly associated with the identity of the person who published the photograph. There is a risk that the combination of these metadata will eventually generate an identifier of the device that was used to obtain the image. If so, the same identifier could be used to desanonymize other images that have been obtained with the same device and published anonymously (for example, in other citizen science projects) thus obtaining additional information on the places where a subject has been found. To mitigate this risk, metadata in the images should be deleted before publication on page[25].

With regard to cookies, it is possible to indicate that they are properly informed of their use at the beginning of navigation and a link is provided that leads to a special section with the relevant information.

4.2.2 DATA STORAGE

While it is a good idea for the project to have its own server instead of using cloud services, it would be advisable to store passwords using the Salted Password Hashing procedure[26]. As noted above, sending passwords to the user after registration is a avoidance practice. On the other hand, it is recommended not to limit too much the maximum length of the password generated by the user (within reason).

Another noteworthy issue is access to information. According to the questionnaire, up to 15 team members have access to raw data containing personal data. The separation of access privileges is important for each to limit access to personal data. To this end, degrees of access to information can be defined according to real needs (administration and management, research, consultation, etc.). For example, a researcher does not have to need access to information about usernames and passwords. This way, it is convenient to decide within your team which people have access privileges to the different data segments.

To avoid unwanted access, it is also recommended that protocols be established to reflect the activity within the database by means of a registry.

4.2.3 DATA PROCESSING, EDITING AND ANALYSIS

[25] The complete list of images is accessible from: http://www.observadoresdelmar.es/fotos_observacions/

[26] Salted Password Hashing process is detailed here: https://crackstation.net/hashing-security.htm

The main concern here is the publication of open data, i. e. making the data available to the general public. This issue is closely related to the section on data collection and the definition of the objectives and purpose of collecting personal information.

The privacy policy shown on the website indicates that personal data is collected "for the purpose of maintaining and managing the relationship with the user". However, for the performance of these tasks it is not necessary to publish user data, and the real need to collect them should even be assessed. While many of the fields are optional, it is possible for observers to fill them in, generating very detailed fact sheets about themselves, which will then be made public. Publishing information on the Internet de facto means **losing control** over that information, as it can be captured and indexed by third parties automatically.

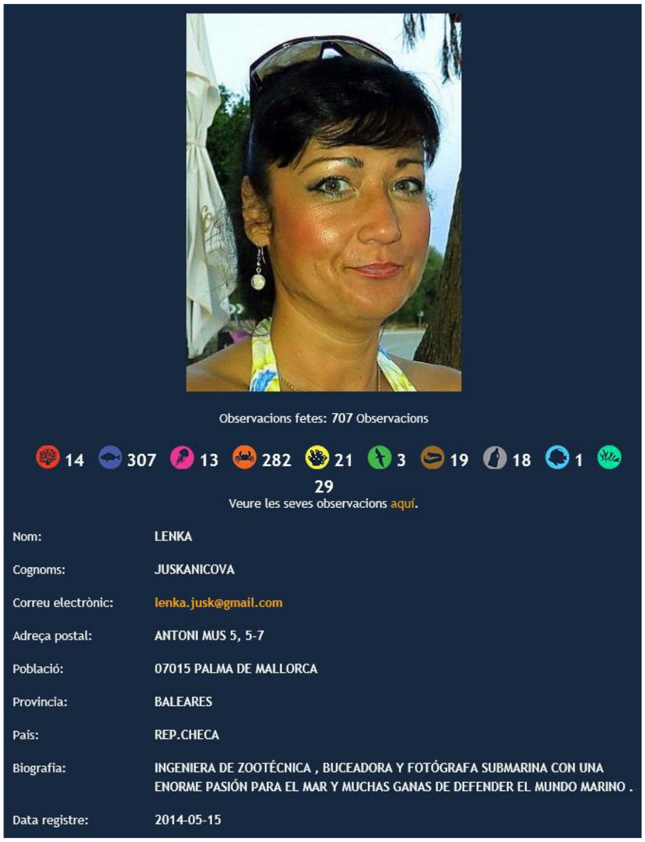
Internet Archive is dedicated to documenting and archiving different content in various formats because of its historical interest in creating a publicly accessible online 'library', and the *Internet Archive Wayback Machine* collects samples of web pages to document the history of the Internet. As an example, the *Internet Archive* has saved content from the Observadores del Mar website a total of 32 times since 2010[27]. This content is publicly accessible and contains personal data that is out of control[28]. Just as the *Internet Archive* collects this information automatically, others may collect it for other purposes.

Along with the profiles created - whether completed or not - Observers of the Sea allows its users to upload photographs located in time and space, associated with the identity of the person who provided the photograph. This information can be easily captured, processed and crossed by third parties. You can, for example, see where each person has been and when. It is therefore advisable to consider anonymizing user input to mitigate this risk.

For example, we have the case of Lenka Juskanicova. Its profile is publicly accessible, with photograph, full name, e-mail, address, town, province, country and biography. This would be Lenka Juskanicova's token:

[27] https://web.archive.org/web/*/observadoresdelmar.es

[28] https://web.archive.org/web/20140930154304/http://www.observadoresdelmar.es/participants.php



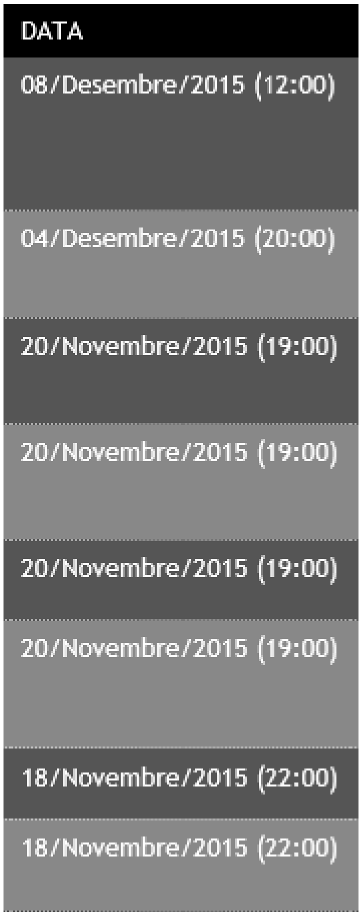
Observacions fetes: 707 Observacions

14 307 13 282 21 3 19 18 1
29

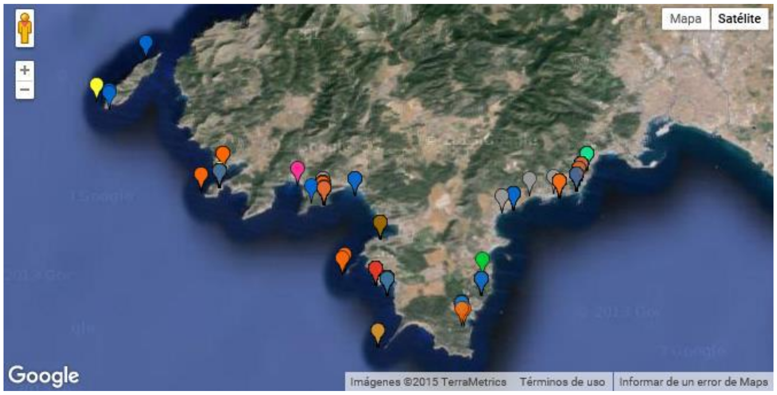
Veure les seves observacions aquí.

Nom:	LENKA
Cognoms:	JUSKANICOVA
Correu electrònic:	lenka.jusk@gmail.com
Adreça postal:	ANTONI MUS 5, 5-7
Població:	07015 PALMA DE MALLORCA
Provincia:	BALEARES
Pais:	REP.CHECA
Biografia:	INGENIERA DE ZOOTÉCNICA , BUCEADORA Y FOTÓGRAFA SUBMARINA CON UNA ENORME PASIÓN PARA EL MAR Y MUCHAS GANAS DE DEFENDER EL MUNDO MARINO .
Data registre:	2014-05-15

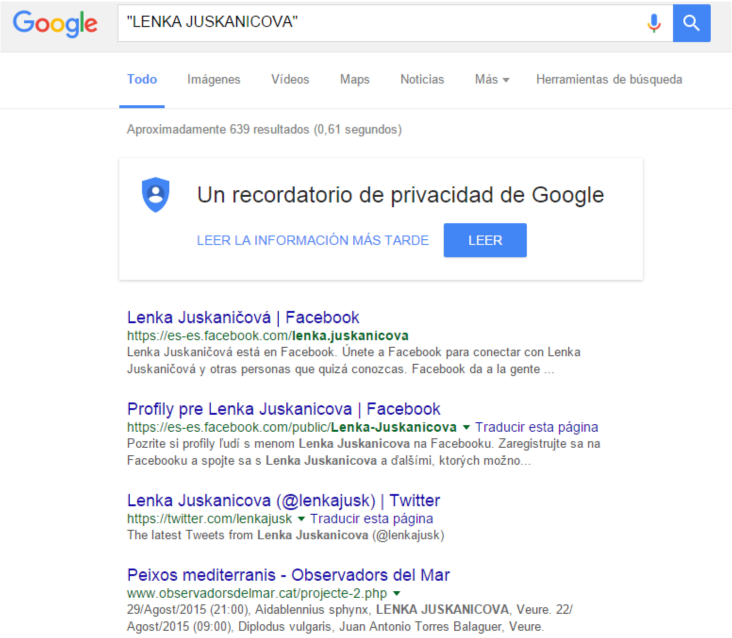
Next, it is possible to see a map with all the observations provided, which are geolocated. In other words, it shows a precise record of where this person was at each of these moments, providing a pattern with the places where this person moves to make the relevant observations:



DATA
08/Desembre/2015 (12:00)
04/Desembre/2015 (20:00)
20/Novembre/2015 (19:00)
20/Novembre/2015 (19:00)
20/Novembre/2015 (19:00)
20/Novembre/2015 (19:00)
18/Novembre/2015 (22:00)
18/Novembre/2015 (22:00)



A simple search in the search engine of Google ("Lenka Juskanicova") throws a link to the information published in the fourth place of the first page:



Aproximadamente 639 resultados (0,61 segundos)

Un recordatorio de privacidad de Google

LEER LA INFORMACIÓN MÁS TARDE LEER

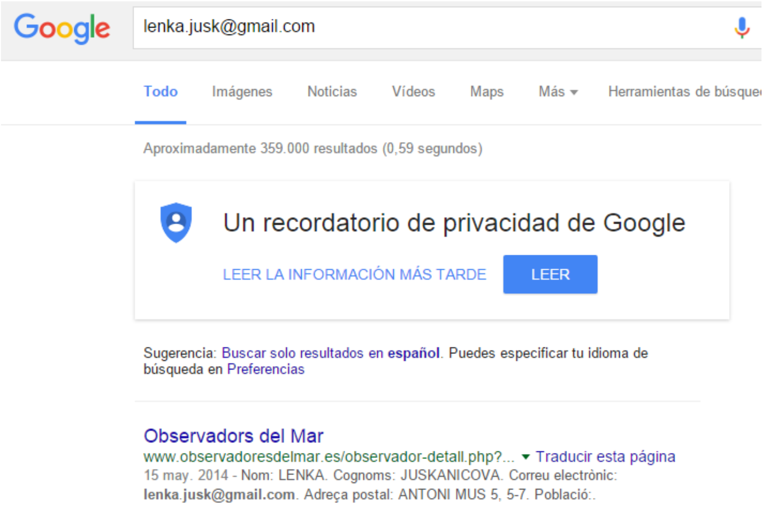
Lenka Juskaničová | Facebook
https://es-es.facebook.com/**lenka.juskanicova**
Lenka Juskaničová está en Facebook. Únete a Facebook para conectar con Lenka Juskaničová y otras personas que quizá conozcas. Facebook da a la gente ...

Profily pre Lenka Juskanicova | Facebook
https://es-es.facebook.com/public/**Lenka-Juskanicova** ▾ Traducir esta página
Pozrite si profily ľudí s menom **Lenka Juskanicova** na Facebooku. Zaregistrujte sa na Facebooku a spojte sa s **Lenka Juskanicova** a ďalšími, ktorých možno...

Lenka Juskanicova (@lenkajusk) | Twitter
https://twitter.com/lenkajusk ▾ Traducir esta página
The latest Tweets from **Lenka Juskanicova** (@lenkajusk)

Peixos mediterranis - Observadors del Mar
www.observadorsdelmar.cat/projecte-2.php ▾
29/Agost/2015 (21:00), Aidablennius sphynx, **LENKA JUSKANICOVA**, Veure. 22/ Agost/2015 (09:00), Diplodus vulgaris, Juan Antonio Torres Balaguer, Veure.

In other words, the information published is not only published in the relevant environment of Observadores del Mar, but is also indexed in Google and therefore its access is facilitated to anyone who searches for your name on the Internet. Moreover, assuming that this person uses their e-mail in different environments, we can see that it is not even necessary to search for the profile by entering your name and surname in the search engine. The Google search for information about Lenka Juskanicova, knowing only her e-mail address, firstly provides a link to Observers of the Sea and the relevant file.



Aproximadamente 359.000 resultados (0,59 segundos)

Un recordatorio de privacidad de Google

Sugerencia: Buscar solo resultados en **español**. Puedes especificar tu idioma de búsqueda en Preferencias

Observadors del Mar
www.observadoresdelmar.es/observador-detall.php?... ▾ Traducir esta página
15 may. 2014 - Nom: LENKA. Cognoms: JUSKANICOVA. Correu electrònic: **lenka.jusk@gmail.com**. Adreça postal: ANTONI MUS 5, 5-7. Població:.

That is to say, an individual who decides to inquire about Lenka Juskanicova, just needs his name and surname or e-mail address and will have quick access to a file containing his postal address and a list of places where he has been moving during his observations, with the corresponding precise data on latitude and longitude, date and time.

In this sense, it could be thought that if the individual has voluntarily surrendered the data, there should be no conflict of interest. However, if this person, for whatever reason, decided to remove this information from the Internet, he or she would have some difficulty. This is why it is not a question of assessing the appropriateness of a series of data being published and their access facilitated through search engines, but of the real possibility that individuals whose data have been collected and published have the capacity to exercise a right recognised as such, in this case, the right to data protection based on information self-determination.

A first solution would be to collect a smaller amount of data, restricting access to information among registered users. This would include anonymisation of observations to the public (these could be linked to specific identities of each to facilitate validation of the observations, with restricted access to those identities). In each observation, It could be included, openly to any person, a link to contact the person who has made the observation.

However, there is a chance that the possibility of facilitating contact and networking between volunteers and researchers themselves (e. g. in the same area) could be pursued as a 'marine research social network'. To this end, an intermediate solution would be the publication of information (e. g. fact sheets of commentators) in a restricted environment accessible only through pre-registration. In any case, it is essential to improve data collection warnings so that users are aware of what is happening with their personal information.

4.2.4 RELATIONSHIP WITH THIRD PARTIES

No major risks have been identified as regards the potential relationship with third parties. However, it is important to bear in mind that some of the comments reflected above take place in third countries (France, Italy, Turkey, etc.) and concern users living abroad. It is therefore essential that these users have full knowledge of the processing of personal data that will take place outside their territory, as they will also be stored on the same internal server, and therefore will be carried out in accordance with Spanish legislation (in this case organic law 15/1999).

The page installs Twitter cookies in the browser of its users, which are used (by Twitter) to monitor their activity inside and outside Observers of the Sea. In addition, the site uses Google services (*Google Maps, Google Analytics*) that also allow the monitoring of user activity within ODM by this company. However, as mentioned above, the cookies policy is correctly communicated to users.

4.2.5 DATA DELETION

The deletion of personal data is usually the most contentious issue and in this case it is not an exception. In Observers of the Sea, no provision is made for the deletion of data within the information life cycle scheme, and therefore there is no established protocol to which we can refer.

Moreover, by navigating through the user interface, it is intuitive that in order to unsubscribe a registered card, it will be necessary to contact the administrator so that they can carry it out manually, since there is no option to indicate such a possibility. This is problematic as long as a registered user will find it difficult to remove the shared information. At this point it is necessary to remember the expansion work carried out by search engines when disseminating information. In fact, the concealment of e-mail a posteriori (since it is shown by default), as previously pointed out, does not prevent the e-mail address from being linked in search engines to the corresponding ODM profiles.

4.2.6 GENERAL AND SAFETY PROCEDURES

In this section, the first thing to highlight would be the need to notify the relevant authority of the file created as a result of the Sea Observers project. Its creation is clearly stated in the data protection policy:

> “ In accordance with the provisions of Organic Law 15/99 of 13 December on the Protection of Personal Data (...), Observadores del Mar informs the user of the existence of an automated personal data file created by Observadores del Mar for the purpose of maintaining and managing the relationship with the user ".

However, according to the questionnaire responses, there is no data file notified to the Spanish Data Protection Agency. The data controller must notify this agency, detailing the characteristics of the file. Any modification of its characteristics must also be communicated[29].

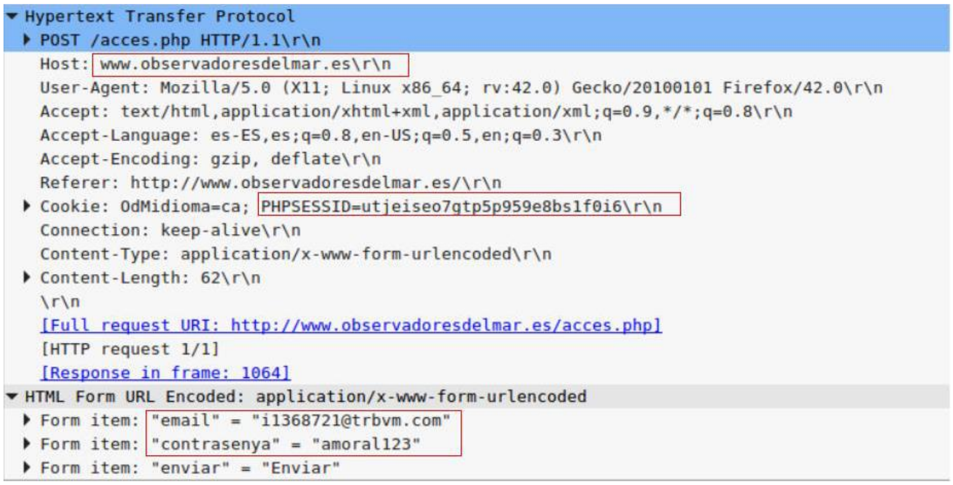
Once the previous recommendations concerning the collection, processing, editing and analysis of data have been put into effect, avoiding that the personal data of observers are not exposed to the public, it would be advisable for successive tests of interfaces and systems to be carried out with profiles created specifically for this purpose. That is to say, to dispense with the use of real cases for any kind of test related to the web and the database.

Finally, indicate that the website does not support HTTPS secure protocol. This means that both the information downloaded by users and the information sent by them is sent unencrypted. This poses a risk to users' privacy, as passwords and session cookies are liable to be passively captured by an attacker when sent to the server.

This attack would be made merely by listening to the communication between the user and the server, storing all the exchanged packets, without the attacker having to intervene at any time. If the attacker were to acquire any of these data, he could enter the page with the user's account, being able to supplant his identity or access non-public data. In addition, if the user reuses the password for other services (something that is commonplace) such as email management or social networking, they may also be compromised.

The hypothetical attacker described in the previous paragraph can also observe all interactions between the user and the server. Any person or institution that has access to any point in the connection between the two can make this attack. As an example, we show a capture of the information that can be obtained from an intercepted packet at the time of authentication of a user named Alberto Moral with an email "i1368721@trbvm.com" with the password "amoral123".

[29] More information about the registration of files:
aquí:https://www.agpd.es/portalwebAGPD/canalresponsable/inscripcion_ficheros/preguntas_frecuentes/cuestiones_generales/index-ides-idphp.php
https://www.agpd.es/portalwebAGPD/canalresponsable/inscripcion_ficheros/index-ides-idphp.php



```
▾ Hypertext Transfer Protocol
  ▸ POST /acces.php HTTP/1.1\r\n
    Host: www.observadoresdelmar.es\r\n
    User-Agent: Mozilla/5.0 (X11; Linux x86_64; rv:42.0) Gecko/20100101 Firefox/42.0\r\n
    Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8\r\n
    Accept-Language: es-ES,es;q=0.8,en-US;q=0.5,en;q=0.3\r\n
    Accept-Encoding: gzip, deflate\r\n
    Referer: http://www.observadoresdelmar.es/\r\n
  ▸ Cookie: OdMidioma=ca; PHPSESSID=utjeiseo7gtp5p959e8bs1f0i6\r\n
    Connection: keep-alive\r\n
    Content-Type: application/x-www-form-urlencoded\r\n
  ▸ Content-Length: 62\r\n
    \r\n
    [Full request URI: http://www.observadoresdelmar.es/acces.php]
    [HTTP request 1/1]
    [Response in frame: 1064]
▾ HTML Form URL Encoded: application/x-www-form-urlencoded
  ▸ Form item: "email" = "i1368721@trbvm.com"
  ▸ Form item: "contrasenya" = "amoral123"
  ▸ Form item: "enviar" = "Enviar"
```

The best way to mitigate this risk is to use HTTPS by default.

4.2.7 CONCLUSIONS

In general, certain data protection shortcomings have been identified and need to be addressed as soon as possible in order to ensure the control of personal information by users. Transparency and publication of results must not be at odds with the data protection of project partners.

The most noteworthy **recommendations** are as follows:

- **Clear definition of objectives and purpose of data collection**. The collection of personal data must have a clear and defined purpose field by field. Each personal data collected must be based on a clear need and linked to a specific contribution related to the established objectives.
- **Excess data**. There is currently an excess of data collection, which makes it impossible for users to remain anonymous (if they wish to do so), and reduces the possibility of information being controlled by those affected by data processing. Overall, the possibility of generating very precise profiles about the observers arises (e. g. if they complete the whole sheet and provide geolocated observations).
- **Sending passwords and publishing e-mail addresses**. During the registration process, it is strongly discouraged to send the user's password to the e-mail account by default. In principle, there should be no need to communicate this information to the registered user, as he or she has not requested it. If you wish to renew your password, you can be informed of one that should ideally be temporary. As far as the publication of the e-mail address is concerned, the best option would be to hide it by default, and it is advisable to avoid publishing it as such (under the recognisable format "xxxxx@xxxxx.xxx").
- **Storage of passwords and access privileges**. Use of the 'Salted Password Hashing' procedure for password management. Set different levels of access privileges. Limit the number of team members who can access personal information.
- **Publication of personal data on the web**. This is a key point in the review of Observers of the Sea's privacy strategies and policy. It is essential to review and, if necessary, rethink the objectives of the project and therefore assess the real need to collect and publish such a large amount of data. On the one hand there are the fact sheets with the details of the

observers, and on the other hand there are the observations themselves. It is precisely in the combination of both that a precise tracking of identified individuals is potentially generated. In order to avoid this situation, it is proposed, depending on real needs, to limit the collection of personal data, to anonymise observations or to publish them in restricted environments accessible by prior registration. Defining which actors have a real need for access to personal information about observers can provide clues on how to manage this situation.

- **Notification of the personal data file**. The existence of the data file should be reported to the Spanish DPA using the procedures available, indicating the characteristics of the data file and the security level established. In the future, any modification of the characteristics of the file must also be communicated, so if changes are made in the way in which data are collected and stored, it is advisable to carry out the procedure after their adoption.
- **Prevent data deletion**. Try to establish periods and phases for project management where personal information can be dissociated from observations made. While it is important to store observations in order to allow analysis to be carried out over long periods of time, it is questionable whether it is indispensable to link such observations to identified persons, especially once they have been validated.
- **Hypertext transfer protocol Secure (https)**. Any kind of web service that requires the sending of personal data and/or passwords should be protected by this protocol.

4.3 Catch the Tiger (*Atrapa el tigre*)

Based on the phenomenon generated by the expansion of the tiger mosquito (Aedes Albopictus), an "invasive" species that has established itself in urban areas lately, the **AtrapaelTigre.com** project proposes a citizen alert system to investigate, monitor and control its expansion. This initiative aims to make the discovery of adult tiger mosquitoes and breeding sites visible. The project relies mainly on an app called *Tigatrapp*, which allows, through a mobile phone or tablet, to report the location of adult tiger mosquitoes and breeding sites. This information will thus be made available to citizens, public administration, law enforcement officials and researchers.

To this end, the collaborators send photos of the specimens and breeding sites, which in turn are validated in network by other users. Both experts and citizens collaborate in the validation of photos sent with the *Tigatrapp app*. Once validated, they are incorporated into the map of the tiger mosquito.

Through the *app*, users also participate in different "missions" through which they focus their observations based on different objectives.

4.3.1 DATA COLLECTION

Conducting the data protection impact assessment for this project is not straightforward. On the one hand, according to the questionnaire responses, **the project would not collect personal data** through its observation and validation system. This is why the data protection impact analysis does not follow the same structure based on the life cycle of personal data. On the other hand, it is necessary to ensure the follow-up of the project on the basis of the recommendations of this report, i. e. to assess possible changes in the project that could incorporate data relating to identifiable persons, such as certain specific missions.

In general, it is possible to point out a whole series of successes in the design of the project and the citizen collaboration network that together facilitate the protection of privacy and personal data. Firstly, it is a relatively simple system that does not collect more information than necessary. The role of collaborating citizens is none other than to function as an information intermediary. You don't even need to register as a user to use the app and share your comments. While it should not be forgotten that the absence of a registration process does not automatically prevent any type of identification from taking place or generate a trace of identifiable users (through one of the identifiers of the device or the combination of different data provided). In this case, minimising the collection of data helps to avoid unnecessary collection of personal data.

Due to the characteristics of the photographs required, the risks are minimized, as they present very close planes, avoiding the inclusion of backgrounds with potentially user-identifying content. In photographs referring to breeding sites, the risk of including irrelevant elements in photos is slightly increased, but not significantly.

It should be noted here that the purpose of *background tracking* (background location of devices used by participants), although specified[30], is not sufficiently substantiated. Does it make a significant difference to assess which areas are covered by users who have different levels of activity and participation? (e. g. users who have downloaded the app but are not collaborating). In this regard, it would be interesting to assess whether there is redundancy in view of the fact that those who contribute reports already geolocate their contributions. Fortunately, the background geolocation system can be disabled. It is important that this option is disabled by default and can only work after activation by the user.

On the other hand, it is important to note that the distortion of the location is carried out properly, i. e. before being transmitted by the device. This would only call into question the actual input of data such as battery status and device model.

Knowing that data collection is practically the minimum necessary and is carried out in proportion to the needs of the project, avoiding the direct collection of personal data, it is possible to analyze the way in which this is communicated to volunteers and whether the information collected is capable of generating traces of data relating to individuals, which although not directly identified, are potentially identifiable.

4.3.2 COMMUNICATION STRATEGY

Carrying out a privacy and data protection policy that avoids the unnecessary collection of personal information is sometimes not enough if these efforts are not transmitted to individuals, so that the

[30] According to the privacy policy, "the approximate location of the volunteers is collected at random intervals in time (in order to evaluate the geographical areas covered by the volunteers)”. http://tigaserver.atrapaeltigre.com/en/privacy/

real effects of it are not reflected in the interaction between organization and users. The communication strategy of *Atrapaeltigre* towards the user is widely successful. The use of the app can only be made with prior acceptance of the terms of use and privacy policy, which consists of the following text[31]:

The AtrapaelTigre project collects data shared by volunteers and makes it available to third parties and the general public. In order to protect the privacy of its volunteers, the collection of information that can identify an individual is avoided. Thus, no names, addresses, passwords or other personal information is collected from volunteers. Volunteers' exact locations are collected during reporting, but this information is linked only to the report and a randomly generated user ID. It also collects the approximate location of volunteers at random intervals over time (to assess the geographical areas covered by volunteers), but this information is also linked only to a randomly generated user identifier. Volunteers may include notes and photographs in their reports, but it is their responsibility not to include in these notes and photographs any information they wish to keep private or any information that violates the privacy rights of others or is inconsistent with the agreement of the AtrapaelTigre user.

The detailed list of information collected is as follows:

- *Reports collected and sent by volunteers regarding breeding sites and sightings of tiger mosquitoes, as well as specific missions. This includes the precise location of volunteers when submitting the report, the location specified by volunteers on the report map and photographs, notes, or any other information the volunteer adds to the report.*
- *Information on the approximate location of volunteers, sampled at random time intervals. This information is collected to know the geographic areas explored by volunteers, and to improve knowledge about tiger mosquito pathways. To avoid accurate collection of volunteer locations, latitudes and longitudes are rounded to the nearest 0.05 degrees before transmitting the location from the device.*

Thus, it is impossible to determine the actual location within an area of 0.05 degrees latitude by 0.05 degrees longitude (about the size of a small town like Blanes).

- *Information about the phone model, battery, language settings, and the Tigatrapp installation (i. e., the installed version of Tigatrapp and the operating system and version in which it is installed). This information is used to help the research team improve the application."*

However, it is true that a certain paradox can be seen in the fact that personal data are not collected but that "the exact locations of volunteers are collected during the reporting process". In this case, it could be more appropriately indicated that the locations of *observations*, not of identified users, are collected.

Moreover, the effort beyond the legal imperative to ensure data protection is evident in asking the participants for their commitment in this regard:

[31] Tigatrapp privacy policy: http://tigaserver.atrapaeltigre.com/en/privacy/ Tigatrapp user agreement: http://tigaserver.atrapaeltigre.com/es/terms/

> *“ As your photos will be made public, don't photograph people or anything else that contains personal information or that you don't want to be public. Remember that your participation is always anonymous."*

In summary, it can be said that both the conditions of use and the privacy policy and additional indications are visible, accessible, comprehensible and available in various language versions.

4.3.3.3 ADDITIONAL RISKS

As regards the possibility that the information collected may be capable of generating *traces of data* relating to individuals, which although not directly identified, are potentially identifiable, a number of residual risks should be noted. It is true that under current conditions profiles are anonymous by default (linked to a numerical identity). However, if one or more personal data is included in the future, or if one or more of the profiles are included in a participant's personal data, one or more of the profiles may be anonymized.

In order to avoid these situations, it is important to analyse the new "missions" entrusted to volunteers: to check in each specific case that there is no risk of sharing or disclosing personal data. As far as comments on the photos are concerned, since they assume a free field, they could include any kind of information, including personal information. However, the website is already warning of the convenience of avoiding this type of data.

One point to highlight and commonly overlooked is the metadata of the photos provided. Some of the images contain metadata that could allow the creation of a unique identifier for the device from which the image was taken. In fact, images can be easily downloaded from the web server[32]. Metadata can also be quickly extracted and indexed into a data set. Once this set has been created, if it is possible to create an identifier for each device, it is possible to find out which images have been taken with the same camera and to cross these data with the geolocations of those images, which can be obtained by processing these files:

- http://tigaserver.atrapaeltigre.com/static/all_reports2015.json
- http://tigaserver.atrapaeltigre.com/static/all_reports2014.json

With this information, there is the hypothetical possibility of desanonimize users based on two options:

1. Crossing data: Device identifiers can be cross-referenced with other external datasets to find out who owns which devices.
2. With previous knowledge of some facts: Suppose Alice wants to know where Bob is going. Alice knows that Bob collaborates with "Trap the Tiger" and knows the make and model

[32] See http://tigaserver.atrapaeltigre.com/media/tigapics/

of his smartphone. Using the datasets mentioned above, Alice may be able to obtain information about the places Bob has visited, along with the date and time he visited those places.

Hence the need to rethink the real need to collect data such as the device model used. We also recommend reviewing the need to allow the full list of images to be publicly available and to remove metadata from the images before uploading them to the server.

4.3.4 RELATIONSHIP WITH THIRD PARTIES

As far as the relationship with **third parties** is concerned, it is possible to point out some data protection issues.

- **System of validation by the users**. The project is aided by the validation and classification of the photographs through a system that uses crowdsourcing logic to optimize resources: *Crowdcrafting*. This is a different environment to the one of the website or the *Atrapaeltigre* app, and has its own terms of use and privacy policy. It should be noted that although they allow the registration of users and encourage this procedure in order to receive the corresponding credits after validation, it also allows the validation processes to be carried out anonymously. The registration process would involve the collection of a limited amount of personal data: username, password, e-mail, photo/avatar (optional). It also uses the possibility of registering through third parties (*Facebook*, *Twitter*, *Google*), which would involve obtaining another set of data (public profile, email address). In short, an important part of the project is delegated to third parties, so volunteers should be aware that they are entering a different platform from that of *Atrapaeltigre*, with different conditions of use and privacy policies. It is true that this web platform has practices similarly designed to ensure the privacy of the user and with the corresponding guarantees of quality. However, legal warnings are still available in English only.
- **Share photos on social networks**. The website encourages volunteers to share photos on social networks such as Twitter or Facebook. It is important to note here that the warning "remember that your participation is always anonymous" may be invalidated. In other words, if the same geolocalized "anonymous" photo is published in social networks under identified profiles (i. e. without using pseudonyms), it will be possible to establish the data triangulation referred to previously. That is to say, the same photo (which could also share the same metadata), can be linked in one environment to a defined identity and in another, to a precise location. This exception to anonymous participation may be included in any of the warnings and/or privacy policy.
- **Cookies**. The website uses *Google Analytics*, a system that collects information about user activity on the website. The page installs cookies that allow the user to monitor the activity of the following domains:

➢ Youtube.com

➢ Highcharts.com

- Twitter.com

Although there is a specific section for the cookies policy, informed consent is not given before navigating the website.

- **Observations made outside Spain**. It is possible to include observations in places outside Spain, subject to legislation other than Organic Law 15/1999. In this case, it should only be noted that the different language versions are guaranteed, as has been the case, in order to ensure that the corresponding privacy policies and warnings are understood.

4.3.5 DATA DELETION

Finally, just to point out a brief note about the deletion of information: Although no personal data deletion protocols are required since they are neither required nor collected, it is important to consider the need to delete personal data in case they are accidentally collected. In this sense, it is possible to mention two specific cases: the removal of inappropriate photos when they are detected and the removal of metadata from published photos.

4.3.6 CONCLUSIONS

In summary, the project of *Atrapaeltigre* can be highlighted as a citizen science initiative that clearly watches over the protection of personal data from the very beginning. Although resources are used that can commonly generate direct risks in terms of privacy and data protection (images and geolocation), the possibility of participating anonymously is evident. The key to good practice in this area lies in minimising data collection, for example by avoiding unnecessary registration of users.

Risks are therefore relegated to a more subtle level related to potential data interconnection and the hypothetical introduction of changes and developments. These risks and improvements relate to the following aspects:

- **Social networks**. The use of social networking services is always a source of potential risks, while many users are registered under identifiable names.
- **Purpose of the information collected**. It is essential to assess the real need to carry out *background tracking* of the devices (possible redundancy of information), as well as to collect the phone model and battery level, considering the actual contributions for the project.
- **Data validation**. The platform through which the photos are validated is managed by a reliable organization but in any case external to *Atrapaeltigre*.
- **Missions**. The commissioning of new "missions" should be monitored against the recommendations and principles mentioned in this report.
- **Photo metadata**. Maintaining the metadata of the photos could lead to the generation of unique identifiers and therefore to the de-anonymization of individuals, their mobility and/or location patterns.

- **Cookies**. The user of the website must be informed and consent to the use of cookies before starting browsing, since the existence of a section with a cookie policy is not sufficient.