



“Building the Next Generation Personal Data Platforms”

G.A. n. 871370

PIMCity Whitepaper

H2020-EU-2.1.1: **PIMCity**

Project No. 871370

Start date of project: 01-12-2019

Duration: 33 months

Revision: 03

Authors: KULeuven



Dissemination Level

Project co-funded by the EC within the H2020 Programme		
PU	Public	

Approvals

	Name	Entity	Date	Visa
Author	Enzo Marquet, Peggy Valcke	KUL	31/08/2022	
WP Leader	Enzo Marquet	KUL	31/08/2022	
Coordinator	Marco Mellia	POLITO	31/08/2022	

Document history

Revision	Date	Modification
Version 1	16-08-2022	Draft version
Version 2	22-08-2022	Reviewed version
Version 3	31-08-2022	Final version



Disclaimer

The information, documentation and figures available in this deliverable are written by the PIMCity Consortium partners under EC co-financing and does not necessarily reflect the view of the European Commission. The information in this document is provided "as is", and no guarantee or warranty is given that the information is fitting any particular purpose. The user uses the information at its sole risk and liability.



Index

Index.....	4
1 Introduction.....	6
1.1 Description of PIMS	6
2 Goals of the White Paper	6
3 User empowerment	7
3.1 Consent Management	7
3.2 How to gather consent	8
3.3 Other lawful processing grounds.....	8
3.3.1 Contract.....	8
3.3.2 Legitimate interest.....	8
3.4 Criticism	9
3.5 PIMCity implementation.....	9
3.5.1 Uploading of personal data	9
3.5.2 Sharing personal data	10
3.5.3 How consent is gathered	11
4 Data marketplaces.....	13
4.1 Transparency.....	13
4.2 Connecting users to data buyers	15
4.2.1 Aggregation.....	15
4.2.2 Anonymity	15
4.2.3 K-anonymity	16
4.2.4 Differential privacy.....	17
4.3 Conclusion	18
4.4 PIMCity implementation.....	18
4.4.1 Transparency	18
4.4.2 Identity disclosure	20
4.4.3 Services implementing PIMS.....	20
4.4.4 Regular users	21
5 Roles of the parties	23
5.1 Capacities.....	23



5.1.1	Controllers	23
5.1.1.1	Determining the purpose and means.....	23
5.1.2	Processor	24
5.1.3	Joint-Controller	25
5.1.4	Case law evolution	25
5.2	Future views	27
5.3	PIMCity implementation.....	27
5.3.1	User environment.....	27
5.3.2	Data offering side	28
5.3.3	Determination of the means.....	29
5.3.4	Determination of the purpose	30
5.3.5	(Joint-)controllers?	30
5.3.6	PIMCity as processor	30
5.3.7	PIMCity as a separate controller.....	30
5.3.8	PIMCity as joint-controller.....	31
6	Concluding remarks	33
7	Sources	34



1 Introduction

1.1 Description of PIMS

The Web economy has been revolutionized by unprecedented possibility of collecting massive amounts of personal data, which led the web to become the largest data market and created the biggest companies in our history. This change has deep consequences for users, who, deprived of any negotiation power, are compelled to blindly provide their data for free access to services. Data collection is opaque, fragmented and disharmonic, so that users have no control over their personal data, and, thus, on their privacy.

The purpose of Personal Information Management Systems (PIMS) is to help give individuals more control over their personal data. PIMS allow individuals to manage their personal data in secure, local or online storage systems and share them when and with whom they choose. Providers of online services and advertisers will need to interact with the PIMS if they plan to process individuals' data. This can enable a human centric approach to personal information and new business models.¹

However, so far, PIMS have failed to reach business maturity and sizeable user bases. PIMCity offers tools to change this scenario through a PIMS development kit (PDK) to commoditize the complexity of creating PIMS. This lowers the barriers for companies and SME to enter the web data market. Novel mechanisms were introduced to increase users' awareness such as Transparency Tags, Personal Data Avatars.

2 Goals of the White Paper

This White Paper focuses on tackling legal uncertainties regarding the use of PIMS and how to align them with the current data protection legislation, mainly the GDPR and ePrivacy Directive. Three topics will be discussed as these are central to a functioning PIMS, from either a data subject's or a business' view. The topics go to the core of what a PIMS is and how a PIMS interacts with data protection legislation, which is essential for any project/stakeholder wishing to develop a PIMS. Best practices will be highlighted as well as voids in the legislation and potential developments in case law. Each topic follows the same structure: first, the relevant legislation will be discussed. Afterwards, the PIMCity implementation will be assessed, namely how a PIMS fits in the legislation and what the potential pitfalls are, possibly how they were addressed by PIMCity.

The first topic is user empowerment in a PIMS environment with a focus on personal consent management, how consent is gathered and how this approach can increase user empowerment.

Secondly, the data marketplace is discussed, mainly how it deals with the issue of transparency and identity (non-)disclosure. This part is strongly connected to the first topic.

¹ Definition used by the European Data Protection Supervision, [Personal Information Management System | European Data Protection Supervisor \(europa.eu\)](https://www.europa.europa.eu/press/pr/2016/06/16_06_16_en).



Thirdly, and lastly, current legislation and case law on the capacity of parties (controller, processor and joint-controller) is laid out and examined. Afterwards, different scenarios with different capacities of parties are presented with a description of their likelihood as well as their drawbacks.

3 User empowerment

One of the central ideas of a PIMS is that individuals are in control of their data. Data subjects (users) of a PIMS are in charge of what happens to their data, and decide what gets stored, what gets shared and with whom. However, to achieve this, users must fully understand the effects of sharing their data, who the data buyers are, what their purposes are, and so forth. This approach is in line with best practices in data protection by design and by default.

To place users in full control of their data, their choices must be informed and they should be fully aware of the consequences. The main idea of PIMCity is thus to put users in charge by giving them granular control over their data and by informing them, clearly and concisely, about the offers they receive from parties interested in accessing their data.

The following subsections describe how consent should be managed, as well as other processing grounds.

3.1 Consent Management

Consent is the corner piece of a PIMS. It is one of the six legal grounds on the basis of which personal data can be processed under the GDPR. For the data to be processed on the basis of consent, the latter must be freely given, specific, informed and unambiguous. Also, consent should be in line with the ePrivacy Directive as this allows the storage of and access to information stored in the terminal equipment of the user. The user should be provided with clear and comprehensive information about the purposes of the processing, and should be offered the right to refuse such processing.

Providing clear and comprehensive information is equivalent to transparent information. Without informing users about how their data is treated, they cannot make an informed decision and thus consent cannot be a valid legal basis as one of its requirements is not fulfilled. Transparency is discussed below in 3. Data Marketplaces.

All of this is dealt with in the personal consent manager (PCM) of PIMCity. In there, the user defines their preferences about which data a service is allowed to collect, process, or which can be shared with third parties. The PCM is meant to increase the users' awareness regarding the data they provide, generate and potentially share. It's the key to make informed decisions and adhere to the GDPR. A PCM provides following information:

- Easy to understand information on the nature of the service requested by the user;
- High-level information to be understood by regular users as well as more technical information for the more technically advanced users;
- Summary and integration of both information actively provided by data buyers and information collected using automatic methodologies, possibly highlighting contradictions;



The primary objective is to give the users the transparency and control over their data as well as being GDPR compliant. That is, give users the possibility to decide which data can be uploaded and stored in the platform, as well as how (raw, extracted or aggregated) data can be shared with data buyers in exchange for incentives when the opportunity arises.

Consent management, in a technical sense, over the internet is still relatively new as regulations define new rights and responsibilities. Therefore, there is no standard way of gathering and providing consents to store and use data. PIMCity thus focuses on implementing a simple, GDPR compliant solution.

3.2 How to gather consent

As stated, consent must be freely given, specific, informed and an unambiguous indication of wishes. The 'freely given' relates to an actual choice of the data subjects. There should be no negative consequences if consent is refused. Consent also cannot be bundled with other options nor tied with the provision of a contract.

3.3 Other lawful processing grounds

For any processing to be lawful, a legal basis as defined by the GDPR must be relied upon. As stated above, consent as a legal basis is the corner stone for PIMCity to function, empowering the users in their usage of the platform.

Generally consent is required to process *any* (not just personal) data in a PIMS. The criteria under the ePrivacy Directive provide a stricter interpretation of which lawful processing grounds are possible to use when dealing with processing data from the terminal equipment of the user. As data is gathered from the user by browsing, or by accessing the equipment (e.g. a voluntary upload), it falls within ePrivacy Directive's scope.

When data is strictly necessary for the performance of a contract (the requested service, being PIMCity), contract can be relied upon. When processing the data is necessary for a legitimate interest (such as fraud prevention), legitimate interest can be relied upon. Both contract and legitimate interest can only be relied upon in strict niche situations.

3.3.1 Contract

Contract is relied upon for the strict minimum of personal data necessary to use the PIMCity platform, being a name and e-mail address. The purpose of processing this data is thus limited to signing-up and being able to log in to the platform (e-mail address functions as log-in name). This is explained in the privacy policy and no other information (even voluntary) is requested at sign-up.

3.3.2 Legitimate interest

To rely on legitimate interests as the legal basis for processing personal data, a controller has to ensure that the processing activity does not override the fundamental rights of the data subject. The GDPR does not provide an exhaustive list of all contexts or processing activities where the legitimate interest lawful basis can apply. To support a



controller in such an evaluation, Article 29 Working Party (the predecessor of the European Data Protection Board) has clarified the examples of legitimate interests. Prevention of fraud, physical, IT and network security are legitimate interests to rely upon. After having done the necessary assessments and balancing tests, these legitimate interests can be relied upon.

3.4 Criticism

A centralised way of managing consent certainly assists users in expanding their knowledge and deepening the understanding of what happens to the user's personal data. This is not the end solution to solve all issues. Centralised consent management builds on the premise that users are interested in fully understanding what happens to their data, whilst also taking the time to read all information provided (e.g. through the privacy policy, in the interface).


However, certain structural information asymmetry and system barriers still remain. While an intuitive user face has been proven to combat these asymmetry and barriers, some users are not interested in learning about data sharing, or do not care about the impact of this. As such, structural deficiencies can't be fully solved by providing enough information presented in the right way.

3.5 PIMCity implementation

The consent management of PIMCity is tackled on different levels i) uploading of personal data and ii) sharing of personal data.

3.5.1 Uploading of personal data

The user is free to upload any information regarding themselves to the PIMCity platform. This could be sociodemographic information, contact information, browsing and location history. While there are (monetary) incentives to store more information and to share it afterwards, this is in no way obligatory. It is key to note that PIMCity functions as a personal data safe on the one hand, and a data marketplace on the other hand. As set out in the EDPB Guidelines 05/2020 on Consent, a user can always retract their consent to share the data with data buyers. This retraction does not impact any previously gained advantages. As such, it is not part of a contract and thus consent is the appropriate legal basis.

 **Here you can check and modify the data about you:**

- You can enter your demographic information which may be used by data buyers to filter audiences.
- Check your browsing history that allows the system to build a profile with your interests.
- Import the location history from your Google account so that data buyers could use it for their need.
- The sharing of which information data buyers can access, is controlled by the Consents you give. In the bottom part of the page you can find details about the amount of sites and services you browsed, the locations you visited, and the demographic information you gave to the platform.

Figure 1: Information on personal data safe



Users are informed about the possibility to upload their data and how this processing affects the sharing of their data. The purpose is clear and unambiguous, as well as informed and specific.

3.5.2 Sharing personal data

Users are able to upload personal data such as location data and browsing history to store in their personal data safe.

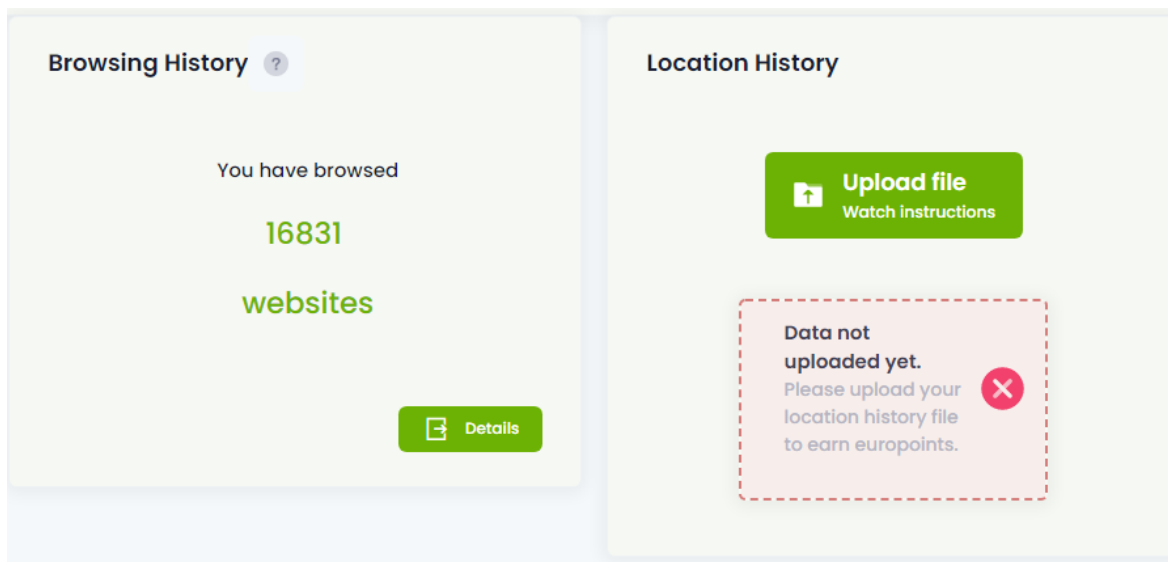


Figure 2: Browsing / Location History uploading

However, doing so is completely voluntary. If a user does not feel comfortable sharing this data, then they are under no obligation to do so. From this data, privacy preserving analytics are extracted to show the user's interests specific categories. These analytics are shared with the data buyers, but they do not disclose the identity or any personal data of the users.

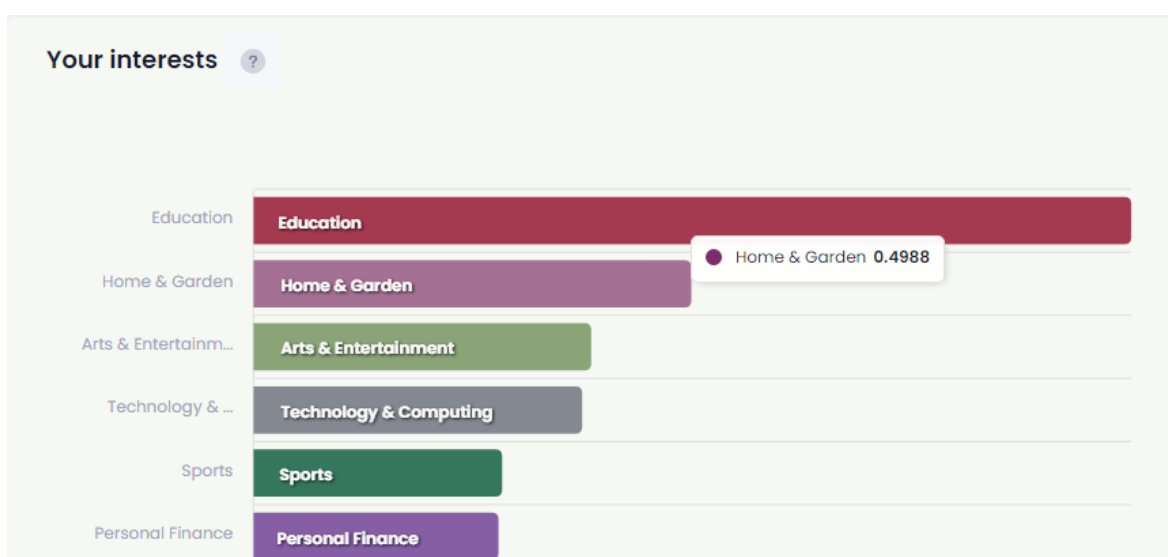


Figure 3: Interests deduced from browsing history



All of this is done locally and not shared with any other party. Only when the user consents to share data, interested parties will be able to provide data offers to users based on their interest through the EasyPIMS platform. However, those interested parties will never know the identity nor access the data before the user consents to this.

The user is able to contract their consent at any time and this does not impact any previously provided ‘advantages’.

3.5.3 How consent is gathered

To ensure PIMCity adequately gathers consent to share data, the users are first informed about the categories of purposes they can share their data for (e.g. commercial/research) and also what type of data they want to share (e.g. sociodemographic, contact, etc.), as shown in figure 1:

The screenshot shows a 'Preferences' section with a table of data categories and purposes. Below the table is a green button labeled 'Interest consents' with a sub-label 'Select interests to share'.

		Sociodemographic information	Contact information	Browsing history	Location history	Interests
Commercial purpose	?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Research	?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Interest consents
Select interests to share

Figure 4: Personal Consent Manager

The users can select the kind of data they want to share with third parties, if any, and the purposes for which that data can be used. Moreover, a profile of the users will be automatically shown if they are uploading their browsing history. This allows users to filter the parts of their profile that can be used for both, data sharing and advertising.

Consent can also be withdrawn in the same way it is given. By unselecting the purposes and/or data to share (e.g. contact information, browsing history, etc.).

Additionally, users have the option to (de)select which interests they want to share with data buyers. This allows users to only receive offers based on interests of their choosing e.g. career or technology.



Interest name	Consent?
Arts & Entertainment	<input checked="" type="checkbox"/>
Automotive	<input checked="" type="checkbox"/>
Business	<input checked="" type="checkbox"/>
Careers	<input checked="" type="checkbox"/>

Figure 5: Consent per interest

It is clear that consent is given at a very granular and specific level. It can be withdrawn as easily as it is granted and thus users are in full control.

PIMCity tackles a part of the information asymmetry criticism by simplifying things for the users. While this plays exactly into one of the arguments (presenting enough information in the right way does not fix structural deficiencies), the implementation builds towards better personal consent management by introducing novel techniques and simplified transparency in line with best practices. More research should be done about how data can be shared in a way that users' personal data is not at risk, whilst also providing benefits to the users and interested parties.

As uploading personal data is entirely free (same for deleting or downloading it), and all other requirements for valid consent are met, consent is the legal basis for the functioning of the data storage of PIMCity.



4 Data marketplaces

PIMCity introduces a Data Marketplace, a platform (EasyPIMS) where companies can participate as Data Buyers in a huge data ecosystem, interacting directly with the end users (data subjects), without intermediaries. This marketplace integrates with PIMCity's Trading Engine, Consent Manager, Data Valuation Tools and many others. The marketplace is user centered in the following way: only users who have consented to share specific data, will have their data presented as a data point for which data buyers can make offers.

From the data buyers' point of view, they are able to acquire either raw, extracted or aggregated data, in a GDPR compliant way. Or on the other hand, make targeted offers to a specific niche of users without accessing their data.

Service providers implementing a PIMS are ensured that the data is stored in a safe way and that their users can trust the platform. Data can only be shared by the data subjects' consent.

Informing the end users of how their data is treated remains a difficult task. Some users are more privacy-sensitive than others, and convincing the first category that you process data in a privacy respecting way may not be an easy task. Some users will never consent to having their data shared regardless.

In this section, the following will be discussed: transparency towards users (data subjects, connected to consent management as discussed above) and the connection of users to data buyers i.e. how the data of users will be provided in an aggregated way (without disclosing the identity of the users, unless they chose to do so).

4.1 Transparency

Transparency is a major part of the GDPR and critical for any service wishing to process personal data. While being transparent about what you envision to do sounds easy, in practice, it has been shown over and over that getting your information through to the data subjects, in a clear yet concise way is not an easy task. It is the controller's duty to ensure data subjects are aware of the processing purposes.

There is no definition of transparency in the GDPR, but recital 39 as well as several articles mention as key aspects of transparent information:

- concise, transparent, intelligible and easily accessible;

In order to avoid information fatigue, any information provided should be easily digestible. This does not mean that details should be omitted; both concise and more elaborate information for the interested reader can be provided simultaneously.

A privacy dashboard is an adequate way to provide easily accessible and centralised information about the processing of a data subject's data, especially when the service provided extends to more than just one processing activity.

- clear and plain language;



Language should be simple and easily understood. The usage of vague terms (may, might, etc.) is to be avoided.

- the requirement for clear and plain language is of particular importance when providing information to children;
- in writing “or by other means, including where appropriate, by electronic means”;
- generally provided free of charge;
- provided at the commencement of the processing.

Of course, data subjects must be informed about the processing before it actually takes place, especially so if they want to provide their informed consent. To reduce information overload, information about the processing is often provided ‘just in time’, right before the actual processing would take place. This ensures the data subject has the necessary information about the processing at the right time whilst avoiding having to read an elaborate privacy policy at an earlier time e.g. sign-up.

Transparency comes in many ways, in essence, it enables the data subject to make an informed choice regarding the processing of personal data (in line with the consent management). Unfortunately, the relevant amount of work available in the literature which specifically focuses on solving the problem of delivering adequate and understandable privacy-related information to users is sparse. There are best practices such as layered privacy policies (e.g. [Juro Privacy Policy – Stefania Passera](#)) where the data subject is informed concisely at first and if they want to know more, they can click through the layers for a deeper explanation. However, text based privacy policies are often overly long and full of legal jargon. While they are factually correct, it is hard for regular data subjects to understand them. On top of that, most information pages are structured and built in a different way, meaning there is no uniformity between services. Simplifying things down is thus a road many services take. However, as pointed out above, providing more information does not solve a structural deficiency.

Some recent examples to provide transparent information are Google Play Store data safety information and Apple’s Privacy Labels. The latter has been introduced recently by Apple to classify apps downloadable from the App Store. Their objective is to give users more information about the impact on their privacy by letting them know which personal data is collected by app providers and how this data is collected. This initiative spurred a great noise in the media sphere, and most of users with a tech background acknowledged Apple reached the goal for which labels were introduced. However, some of them criticized Apple for not sufficiently advertising the labels, for placing labels in areas too difficult to find (privacy labels are placed almost at the bottom of the app page), and labels are informative up to a point, as they do not provide any information about how personal data is being used. From a design perspective they satisfy all basic UI requirements: they are indeed usable, easy to understand and visually enjoyable. Adequately informing data subjects remains a stringent task.

One of the most interesting features of Apple Privacy Labels is that they specify whether collected data is connected to the identity of the user or not. However, we have to highlight that information provided in Apple’s Privacy Labels are provided by app providers, and there is no real check of their claims to verify whether they correspond to the actual data collection practice. In this sense, Apple claims it will penalize or ban developers in case inconsistencies will emerge.



4.2 Connecting users to data buyers

The point of a data marketplace is to bridge the gap between users' data and parties interested in such data. Users can choose to share their data in a secure and private way with the data buyers; data buyers can make offers to aggregated groups of users without identifying them. The challenge lies in finding a balance between respecting the user's privacy and the interests of other parties. It is possible for these interests to co-exist.

To enable privacy respecting (data protection by design) data sharing, data of users can be aggregated in order to protect their identity, as we will discuss below.

4.2.1 Aggregation

Data aggregation techniques are used to bring together relevant pieces of information and provide a high-level overview of an entity in a scalable, efficient and reliable manner. However, the design and evaluation of such aggregation algorithms have not received the same level of attention that other basic operators, such as joins, have received in the literature. The nature and purpose of existing aggregation algorithms/techniques are also platform dependent in some cases, restricting their wider application across different scenarios. Moreover, the influence of Big Data techniques in various application domains is seen in the emergence of hybrid approaches for data aggregation.

Examples of hybrid techniques include the use of machine learning, predictive and clustering algorithms to perform data aggregation operations. In addition, some existing frameworks for data aggregation focus on combining data from multiple sources in real-time environments.

Aggregation attempts to achieve anonymous data. This is the process of modifying personal data in such a way that individuals cannot (reasonably) be re-identified, and no information about them can be learned. It is applied in several cases of data analysis. Most of the privacy anonymization models fall into two distinct categories. The first includes k-anonymity and its extensions, while the second consists of the notion of Differential Privacy, alongside with some variations. While k-anonymity is a mechanism that is applied on the data to make them a bit fuzzier to prevent revealing individuals, differential privacy is applied on the answers of the queries on the dataset to disclose private information/identities.

Identity disclosure can be prevented with a k-anonymous version of the initial dataset, that is, an original record cannot be distinguished within a group of k records sharing quasi-identifier values.

4.2.2 Anonymity

At this moment, what constitutes anonymous data is unclear. To clarify, anonymous data falls outside of the scope of the GDPR. Anonymous information does not relate to an identified or an identifiable natural person.

Whether data can be seen as anonymous is split into two approaches: the relative and the absolute approach. The relative approach is synonym to the risk based approach. With this



approach, if there is no reasonable risk (it is negligible) of (re)identification then the data is anonymous e.g. when the identification is a mere hypothetical possibility. To determine if this (re)identification is merely hypothetical, all means reasonably likely used by a controller must be taken into account such as cost, state-of-the-art technology, amount of time required, etc. It is only logical that the more sensitive personal data is (e.g. health data), the higher the perceived risk is and the higher the threshold.

On the other hand, the absolute approach defines anonymous data very narrowly. This approach is elaborated on by WP 216 and endorsed by the EDPB. The baseline is that even in 'anonymised datasets', three risks remain:

- Individuals can be singled out or isolated (this does not mean identified);
- Malicious parties can link the data from the anonymised dataset to another dataset and establish links between individuals/groups.
- Personal data can still be inferred based on an attribute from the value of a set of other attributes.

The absolute approach follows the zero-risk test, meaning that anonymisation of data means that identification is *never* possible. It is irreversible. This approach goes even further, stating that if the original dataset still exists, anonymisation techniques do not render data anonymous. The absolute approach endorses anonymisation as a good technical measure to protect data. Anonymisation renders data pseudonymous, which is part of a strong data protection practices.

With ever evolving technology, it is likely that current 'anonymous' dataset according to the absolute approach, will at some point in time, be reversible. Reidentification risks change over time as a result of technological advancement. As such, no anonymous technique will ever render data truly anonymous under the absolute approach.

4.2.3 K-anonymity

Many researches in the past years showed that removing the identifiers from a dataset is not a sufficient way to protect privacy: in 1997, Sweeney was able to identify the Massachusetts governor in a publicly listed set of medical records, using their birth date, its zip code and its gender. In 2006, Netflix published a dataset while launching a contest to improve its recommendation algorithm: the dataset consisted of rating provided by users, whose identities were not revealed. However, linking the information provided by Netflix with the one publicly available on sites like IMDB, Narayanan and Shmatikov were able to identify some of the Netflix users.

K-anonymity aims at reducing the probability for a user in a dataset to be identified down to $1/k$. The goal of k-anonymity is to group the users according to their quasi-identifiers (those attributes who are apparently harmless, but whose combination may lead to re-identification – such as zip code, birth date and so on) and enforce that each one of these groups (called *Equivalent Classes* – ECs) contains at least k rows of the dataset – i.e., k users: it is common to assume that each user only appears in a row. This property guarantees that if a user is linked with a sensitive information (such as a medical record, the income, and so on), it can hide himself among other k-1 users. K-anonymity is typically achieved through two methods: generalization and suppression. In the latter method, some rows may be removed from the published dataset in order to form ECs with appropriate size. With



generalization, instead, the single attributes may be modified in order to find k indistinguishable users. For numeric attributes, the actual number is replaced by a range that includes it. For categorical attributes, a common parent value in a hierarchy tree may replace the child ones (e.g., two values such as “French” and “Spanish” may be generalized as “European”).

The basic notion of the k -anonymity model and its extensions is to classify the attributes of a dataset into several non-disjoint types:

- Identifiers: These are the attributes in the initial dataset that explicitly identify the subject (e.g., passport number). The removal of the identifiers is a requirement to create an anonymized dataset.
- Quasi-identifiers: These are the attributes in the initial dataset that in combination can identify the subject (e.g., age, city of residence). Quasi-identifiers cannot be removed, since any attribute is potentially a quasi-identifier.
- Confidential attributes: These attributes contain sensitive information on the subject (e.g., salary, health condition).

K -anonymity does suffer from flaws. Through homogeneity attacks and background knowledge attacks, it is possible to weaken its protection. On top of that, each alteration to the raw data also diminishes the usefulness for analytical usage as it becomes less granular.

Regardless of the introduction of new privacy properties, k -anonymity, for its simplicity, is still regarded as the golden standard for anonymization. It prevents direct identification whilst rendering the data useful for statistical purposes.

4.2.4 Differential privacy

Differential privacy is the second category of privacy models, which aims to anonymize the answers to interactive queries submitted to a database. Differentially private data sets can be generated by creating a simulated dataset from a differentially private dataset or by adding noise to the attributes of the initial dataset. The main downside of differential privacy is that privacy guarantees deteriorate with repeated use and various techniques to address that need to be considered.

It aims at providing a formal definition of privacy. This definition claims that a user's privacy is protected if the knowledge an attacker has on a queried dataset is not too affected by whether the user is in the dataset or not. The “too affected” notion is parametrized by ϵ .

A common mechanism to obtain differential privacy is to apply the Laplace mechanism, which consists in adding a Laplace-distributed noise to the query result. The noise is extracted from a Laplace distribution which depends both on the ϵ value and on the sensitivity of the query – i.e., on how much the presence or absence of a user can affect the result. It is proven that the Laplace mechanism guarantees differential privacy.

However, differential privacy could also lead to inaccurate models and statistics as each alteration on the raw data deteriorates the accuracy of said data.



4.3 Conclusion

The above techniques are used on the one hand to protect the identity and personal data of users. On the other hand, the data must remain useful for statistical purposes. This is a difficult trade-off to make.

The applied techniques can be applied to severely reduce the likelihood that the users can be identified. When users are grouped into broader categories like interests, it becomes very hard to infer their identity or single them out.

The data presented to the data buyers can thus be seen as anonymous *from their side* e.g. users interested in politics between ages 28 and 35 located at city XYZ. On the side of the PIMS, they know which interests relates to which user. As such, the data is pseudonymous for the PIMS (they can match the original dataset to the aggregated dataset).

By operating in this split way, the GDPR remains applicable but the PIMS implements strong technical and organizational measures to protect the rights and freedoms of its users.

4.4 PIMCity implementation

4.4.1 Transparency

Transparency is implemented on multiple levels. The core is of course in the previously discussed consent management. A centralized privacy dashboard where the user has full control over their data. Information on data processing is provided in a layered and just-in-time approach. Only the relevant information is shown at the time it is necessary (e.g. information on the data required to sign-up, when uploading information) and layered by giving a concise and simple overview with the option to receive more information if the user wishes to know more. This satisfies the transparency responsibility of PIMCity as a controller under the GDPR.

To increase transparency and clarity among users, PIMCity introduced Transparency tags (TT). These can be compared with nutrition labels for food, but for data buyers requesting data of users. The TT has been introduced as an easy-to-understand way to provide the user with information about the nature of the web services contacted. For each web service (e.g., a website, a mobile app, a data buyer) the information exposed contains relevant information, such as its owner, its purpose, the personal data it collects, etc. Apart from providing all the details, the TT also summarizes such information in scores, automatically computed by the analytics behind the Privacy Metrics and revealing the potential security and privacy risks associated to that service. TT also attempts to describe how transparent the service under exam is towards the user, i.e., whether crucial information such as data processing purposes, data controllers, contacts, etc. is provided publicly.



Europa.eu

Browsing Not Available

Advocacy Groups & Trade Associations

Security ★☆☆☆☆

Privacy ★☆☆☆☆

Transparency ★☆☆☆☆

View

Figure 6: Transparency Tag template

The implementation of TT is focused around transparency towards the users, which is critical to empower them in their choices. Users will be informed on the performance of a company interested in their data, and allow/refuse to share their data with said company based on their judgement. The visuals of stars ranging from 0 to 5 out of 5 gives an easy indicator of how trustworthy a source is. This also encourages data buyers to improve their score.

Information provided by TBP	Webdata
Declared company name TBP	Company name Not available
Website TBP	Operates under europa.eu
Country TBP	Category Advocacy Groups & Trade Associations
Category TBP	Rank in category 54
Purpose of data collection -	Connected third-party services europa.eu
Data owner TBP	Third Party False
Data processor TBP	
View more	View more

Figure 7: Information on data buyer

Users wanting to learn more can do so by clicking further on the company’s page. However, at this point the data buyer is responsible for ensuring the data they provided is accurate and up to date. They are responsible for adequately informing the users about themselves. This comes down to the distribution of responsibilities of parties (controller, processor, joint controller) as discussed in 4. Roles of the parties. The controller is responsible for being complaint with the GDPR by, among others, providing the necessary information to the data subjects.

Transparency and consent go hand in hand. Without the necessary transparency, it is arguably impossible to give properly informed consent. Additionally, transparency is paramount to build trust with users. The PIMCity projects thus focuses hard on increasing transparency in PIMS by means of Transparency Tags in line with best practices.

4.4.2 Identity disclosure

In PIMCity, standard techniques were used to allow the data owners/providers to prepare their data to be shared in the PIMCity ecosystem. Such techniques are used for data aggregation to provide statistics like count, sum, average over specific data attributes such as location, age and time. Furthermore, these techniques will support the anonymization part of the data, supported also from anonymization algorithms to ensure identities are not disclosed. Even though the data will be aggregated, the results will still be useful for analytics.

. Finding appropriate ways to implement privacy principles in the big data business is the most efficient way to prevent a conflict between privacy and big data. Privacy by design is one of the fundamental mechanisms to address privacy risks from the beginning of the data processing, and apply necessary privacy preserving solutions.

4.4.3 Services implementing PIMS

From the perspective of data owners (services wishing to implement a PIMS), they can import their data in an aggregated way through the following schematic:

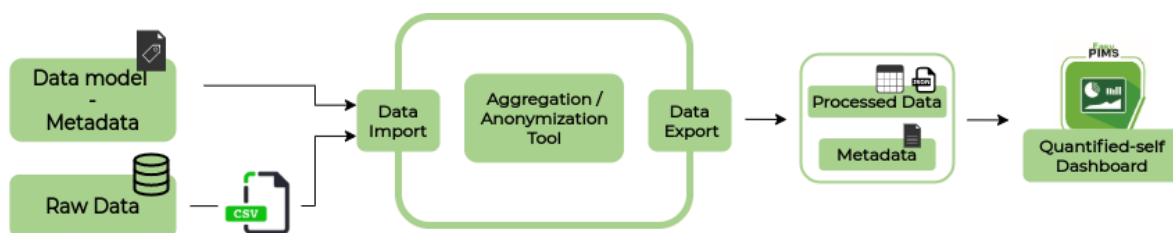


Figure 8: Aggregation/Anonymization schematic for imported data

The Data Aggregation (DA) tool enables data owners (for example an Internet Service Provider that holds a bulk of their users’ data) to perform two important processes on their data: aggregation and anonymization. Such processes enable data owners to share these data in a privacy-preserving way. The DA tool resides on the data owner’s side and its input is the raw data that is available through the initial sources (telco data, sensor data, etc.) and it is transformed in a predefined schema/metadata model. The user (data owner) is

responsible for preparing the data for processing (i.e., export from their initial source (internal database), clean them if needed, etc.). Afterwards, through the module, the data owner is able to choose the subset of the data to be aggregated / anonymized and set the related algorithmic parameters (for aggregation and anonymization). The output is the processed (aggregated / anonymized) data that can be exported to the PIMCity marketplace through an API that the module provides. The data resides on the data owner side and the interested party is able to retrieve them through this API.

Through this technique, data can be shared in a privacy respecting way between the parties. How the aggregation is done, is described below, from the perspective of regular users.

4.4.4 Regular users

From the perspective of regular users (data subjects), PIMCity provides several data aggregation/anonymization tool to prevent the identity disclosure of the users (data subjects) to the data buyers until the users have given their consent to share their data. The process can be summarized as follows:

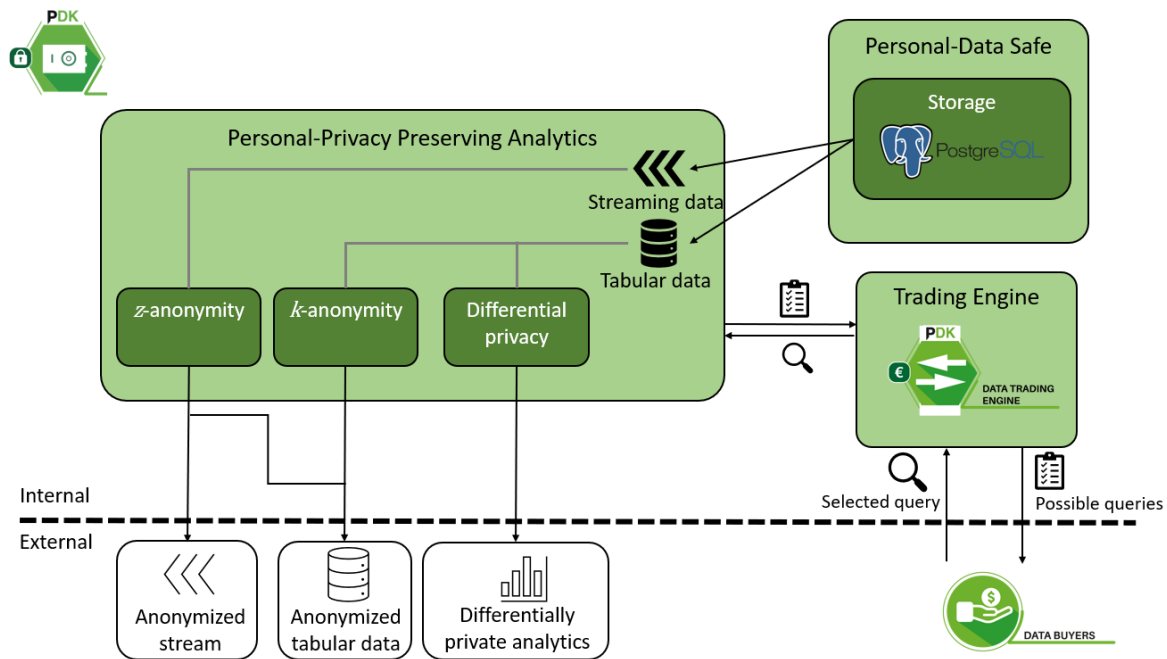


Figure 9: Personal-Privacy Preserving Analytics schematic

The user starts by uploading their own personal data into the platform. Here it is safely stored and nobody can access it. The EasyPIMS platform will run privacy preserving analytics (such as k-anonymity, z-anonymity, differential privacy) on it to determine meta data and tags useful for external services. These external services can make a request directly to the user to share their data for a certain incentive.

The request does not reveal the identity of the data subject, this is made through the PIMCity platform (from PIMCity to the data subject). The data buyer can only receive data from data



subjects once they have consented to sharing their data. The aggregated processing of data effectively protects the identity and identifiability of the data subjects.

As such, the identity of the users is adequately protected whilst the usability of the data sharing is not diminished. Interested parties can gain access to personal data through data offers, but only if they are sufficiently transparent and prove their trustworthiness to the users.

The implemented measures do not exonerate the application of the GDPR. PIMCity is obligated to implement strong technical and organizational measures to protect the rights of the users. This protection is exactly why strong aggregation/anonymisation protocol are used. The risk for users is severely lowered while the usability of the data does not suffer in a significant way. A win for all sides, in line with data protection legislation.



5 Roles of the parties

A major hurdle in the current legal landscape surrounding data marketplaces/PIMS and the GDPR is the qualification of parties. According to the GDPR they are either processors or controllers (or joint-controllers), which each have their own responsibilities. Controllers carry most obligations under the GDPR. They are those deciding on the purpose and means of processing of personal data. A processor carries out (a part of) the processing on behalf of the controller. Determining who has which role becomes increasingly complex when there are more parties and/or the processing takes place for different purposes. Joint controllership can also arise when two or more entities *jointly* participate in determining the purposes and/or means of the processing, where this processing is inextricably linked.

The relevant parties for PIMCity is the EasyPIMS platform on one hand, and the data buyers on the other hand. In the following chapter, the concepts of controller, processor and joint-controller and how to determine the capacity of parties will be discussed. Afterwards, this theory will be applied to the PIMCity project.

5.1 Capacities

5.1.1 Controllers

The (data) controller, being either a legal or natural person, an agency, a public authority, or any other body, determines the purposes for which and the means by which personal data is processed. The 'why' and 'how' data should be processed is decided by the controller. The controller is in control of how personal data is to be processed and for which reason. They carry the majority of responsibilities such as adequately informing the data subjects, implementing adequate technical and organisational measures to safeguard the personal data. They are also responsible for contacting the relevant authorities in case of a data breach, and so on. Part of these responsibilities can be delegated towards processors, but only in a limited way, otherwise the processor risks being labelled as a (joint-)controller itself for a (specific part of the) processing activity.

While the aim of this part of the white paper is not to give a long explanation on what constitutes a controller, it will expand on how to determine who is a controller as this is an unclear topic in a PIMS environment.

5.1.1.1 Determining the purpose and means

One of the key elements to determine which party is a controller, is its influence over the processing, its decision-making power. This analysis must be done on a factual, rather than a legal basis. If a data processing agreement stipulates that party X is a processor, but the factual elements say otherwise, then party X risks being labelled as a controller, including receiving all responsibilities that come with it. The contractual stipulations are to be taken into account, but they aren't decisive. As stated, the decisive influence on the purposes and means of processing is the deciding factor. The controller is empowered to give instructions to the processor(s) on how the processing should take place.



What can be understood under purposes and means of processing is the ‘why’ and the ‘how’ of the processing. What is the goal of the processing (why) and how will it be reached (how). A processor can never determine the purpose, as they are doing the processing *on behalf of* the controller, i.e. the controller contracts the processor for a certain predefined purpose. However, the processor can have *some* influence on the means of processing, because they are in a specialised position to deliver the service.

It must be noted that processing is a very broad term. Access to data is not necessary to fall under processing. Simply hosting the data on a virtual machine, fully encrypted with no way to access it falls under processing. Processing includes the collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction of personal data.

For example, company X (controller) contracts a cloud service (processor) to host their customer data. The controller X does not decide how the cloud service works, nor how they set-up their technical environment. That is up to the cloud service. It is the controller’s duty to perform their due diligence by checking whether the processor complies with the GDPR (among other regulations).

Why the service is a processor:

- It’s a separate entity
- Which processes data on the *controller’s behalf*

The service receives instructions from company X, which it must follow. Company X decides the purpose (hosting customer data) and decides to outsource this hosting to the service.

This distinction can be split between deciding on the *essential* and *non-essential means*. The former being closely linked to the purpose of the processing while the latter deal with the practical side of the processing.

Separate controller to controller relationships also exist, where controllers share personal data but process it individually for their own purposes, and use its own means of processing. This happens when one controller sells a database to another controller. They both process personal data, but each for their individual purpose.

5.1.2 Processor

A processor is a separate entity from the controller, which does (part of) the processing activities on behalf of the controller, i.e. a service doing processing on behalf of and in the interest of their buyer. It is critical in a controller – processor relationship that processors cannot do any other processing on the controller’s personal data for its own purposes. If it does so (and thus goes beyond the controller’s instructions), then this is either a breach of contract and/or the processor risks being classified as a controller for its own purposes.

In terms of GDPR responsibilities, processors carry a light burden in comparison to controllers. They can still be held civilly liable for mistakes or errors on their parts, but this is more difficult under GDPR. For example, a cookie banner tool (processor) can be implemented in a non-compliant way by a business (controller) and the controller would risk



being fined. Some scholars note that part of the responsibility should also be on the processor, as they are the expert on their service and should thus only implement it in a compliant way.

5.1.3 Joint-Controller

A joint-controllership arises when parties *jointly* determine the purposes and means of data processing. Joint controllers have a *shared* purpose and agree upon the purpose and means of processing data together. However, this will not apply if the same data is being used for different reasons. As with controllership, the assessment whether parties jointly determine the purposes and means must be done on a factual, rather than a formal, basis.

It is key that a common decision (cfr. EDPB Guidelines 07/2020) is made, with converging purposes. All parties which have exerted influence over the means of processing can be joint controllers. If the processing is not possible without both parties (i.e. they are inextricably linked), then there is no joint controllership. The difference with a controller – processor relationship is that processors do not process the data for their own purposes.

However, as confirmed by the EU Court of Justice (CJEU) in the *Fashion ID* case, if controllers decide *solely* on the purposes and means of the processing, then it is a sole controller (this can create a controller to controller relationship).

As highlighted by the EDPB in their Guidelines on the concepts of controllers and processors, the mere existence of a mutual benefit arising from a processing activity does not automatically give rise to joint controllership. However, in the case of platforms, the CJEU ruled in the *Wirtschaftsakademie* case that joint controllership is possible when systems have been set up in a certain way by one party and then personal data can be used by the other party in a different way. It must be noted that this does not automatically lead to joint controllership if the processing is separate and done without intervention of the other party, or when they are a processor. If each party of a shared infrastructure/platform determines their processing purpose, then there is no joint controllership.

5.1.4 Case law evolution

There are a number of cases by the CJEU addressing controllership under the GDPR, and one prominent case by the Belgian Data Protection Authority which we will address.

The first case is *Wirtschaftsakademie* in which the CJEU decided that the creation of a Facebook fan page contributes to the processing of personal data of visitors and thus introduces joint-controllership for certain aspects of the processing. The facts that the statistics generated by the fan page were anonymous and that not every controller has access to the data were deemed irrelevant. Joint-controllership does not imply equal responsibility among the controllers. The ‘effective and complete protection’ of data subjects was an important point for this decision.

In *Jehovah’s Witnesses*, the term (joint-)controller was interpreted broadly, as any natural or legal person who, for their own purposes, exerts influence on the processing of personal data. An important point here was the access to personal data; even though a party had no access to the data, this had no effect on the capacity of that party being a (joint-)controller.



In the *Fashion ID* case, a 'like button' was implemented through a plug-in on a website. The CJEU ruled that even though the website had no influence on the processing of personal data, it facilitated its processing. The Court stated that processing may involve one, or a number of operations, each of which relates to one of the different stages. The plug-in's purpose was jointly determined by Facebook and Fashion ID and serves in both their economic interest, thus creating joint-controllership.

In a recent case by the Gegevensbeschermingsautoriteit (Belgian DPA) from 2 February 2022 involving IAB Europe, the role of its Transparency & Consent Framework (TCF) was up for discussion. Central to the TCF was a Consent Management Platform (CMP) for data subjects as well as an online ad-ecosystem for advertisers and agencies based on a real-time bidding (TRB) protocol.

The Belgian DPA (after consulting the other DPAs) ruled that IAB Europe is to be classified as a data controller for the use of the TCF. IAB Europe stated that they were neither controller nor joint-controller. As IAB Europe did not adhere to the obligations imposed by the GDPR on data controllers, claiming that they were neither controller nor joint-controller with regards to the use of the TCF, they were found in breach with several principles of the GDPR.

Relevant to the PIMCity project is how the DPA defines a controller broadly, in light of the legislator's objective to protect personal data as per the *Jehovah's Witnesses* case. It also states that IAB Europe played a decisive role in the processing (dissemination) of personal data, which may substantially affect the rights of the data subjects. The participating parties (other than IAB Europe), would not be able to reach the goals (purposes) set by IAB Europe without the TCF. IAB Europe set up a standardized approach and a technical protocol which their participants (such as adtech vendors and publishers) must adhere to before they can use the TCF.

Taken this all into account, among others, the DPA found that IAB Europe determines the purposes of the processing and is a controller for this.

To determine the means of the processing, IAB Europe monitors and defines the rules applicable to the TCF. Like in *Wirtschaftsakademie*, is it the entity which is responsible for laying down (and thus imposing) the setting in a process which determines the means and the purposes of the processing.

From this, among others, the DPA concludes that IAB Europe also defines the means of the processing.

The DPA goes further to determine the other parties involved should be regarded as joint-controllers for the processing of personal data as the purpose of the processing is then decided jointly (profiling, targeted advertising) and the processing wouldn't be possible without either party present.

This case is currently being appealed and the outcome will greatly affect the evolution of any PIMS.



5.2 Future views

It is clear that the CJEU and Data Protection Authorities aim for maximizing the protection of fundamental rights and freedoms of data subjects. The terms of controller and joint-controllership are being stretched beyond their original intention. It is only natural the legal landscape evolves, but uncertainty remains and operators are being branded as joint-controllers for the sake of protection of the data subjects. This approach is being criticized, as making everybody responsible will make nobody responsible in the end. A case could be made that the split between controller and processor is outdated in the quickly evolving technological landscape.

5.3 PIMCity implementation

From a technical perspective, PIMCity has three layers: the EasyPIMs environment consisting of the data marketplace and the login environment for the data buyers, the PDK which encapsulates the business rules (trading engine, data valuation and consent management) and lastly the user (data subject) environment.

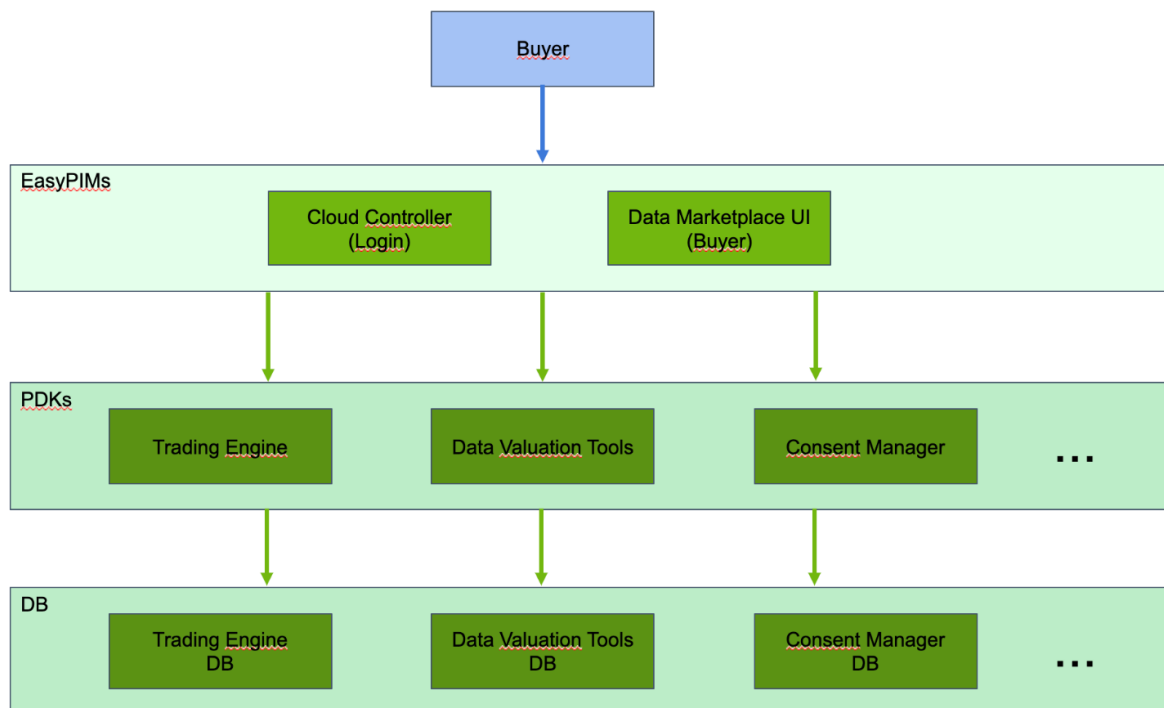


Figure 10: PIMCity high-level infrastructure

5.3.1 User environment

Starting with the bottom user layer. Here, all the data of the users is processed, stored and managed by solely PIMCity. This environment is set up and run by PIMCity and as such, they are the sole determinator of the means and purposes of the processing i.e. user sign-



up, possibility for users to upload data, consent management etc. The data is only processed for purposes decided by PIMCity.

While it is possible that *some parts* of such an environment are outsourced (e.g. cloud storage, data valuation, etc.), these are secondary and non-essential to the main processing. As such, any outsourced part would generally be constituted as a processor of PIMCity.

The users wanting to use the services of PIMCity contract the requested services directly. Their data is initially not processed for any other purpose, unless the user decided to share their data with a data buyer based on the offer received, which comes after the data offering as explained below.

It is clear that for this processing activity, PIMCity can only be seen as a controller, and the sole controller.

5.3.2 Data offering side

As the bottom layer contains all the data subjects' (users) data, the other layers can enable data buyers to access them. Based on aggregated data, data buyers can make a direct offer to users to share their data. It is important to note that this offers means that the data buyer will use the data for their own purposes, according to their own privacy policy and in their own system (means). To determine the capacity of the parties, we critically examine the determination of the means and the purpose. Afterwards, we examine which roles can be possible for the parties, with arguments pro and contra.

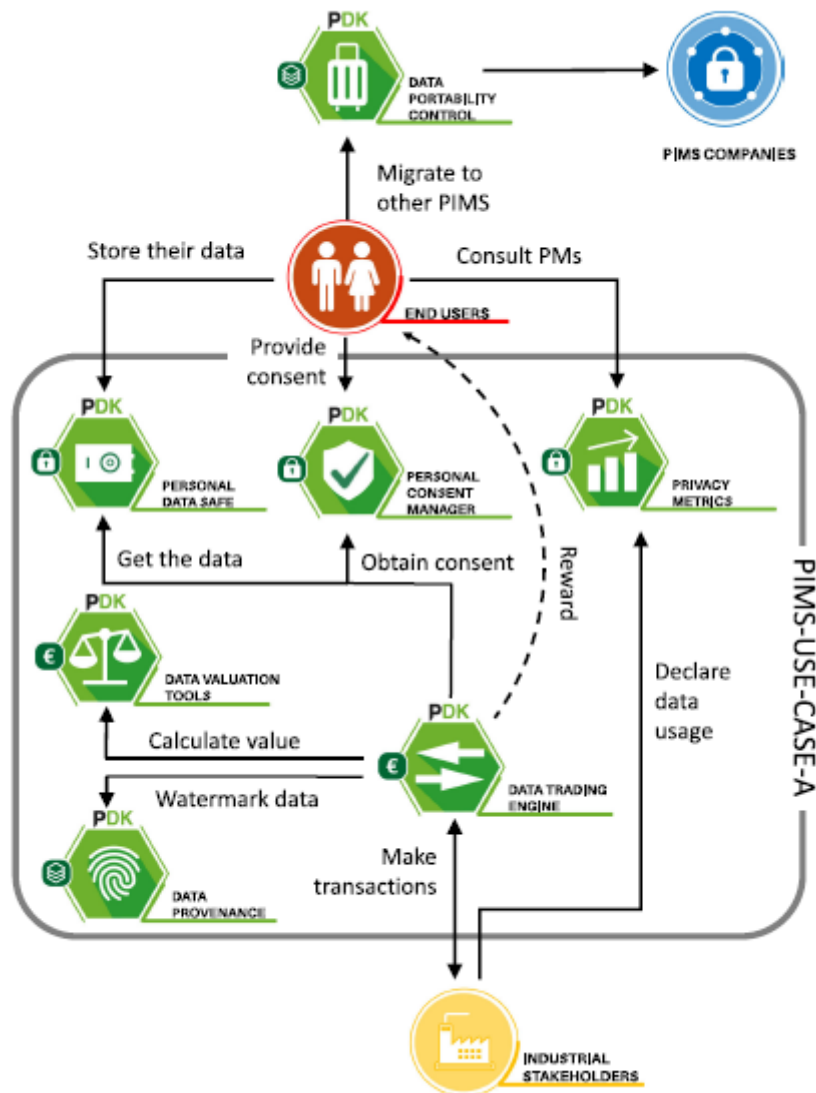


Figure 11: PIMCity personal data cycle

5.3.3 Determination of the means

It is clear that PIMCity determines the means of processing of the users' data at first. When a user consents to having their data shared and opens itself for data offering, aggregation techniques are applied to it to ensure the identity and personal data of individual users remains safe and confidential until they chooses to share their data. The creation of a Data Transaction presents the Data Buyer with the possibility to place a Data Offer in the Data Marketplace available to the users. When creating an offer, the data buyer can specify the audience, the type of data, the privacy policy and the budget to be spent, among other information. However, no personal data reaches data buyers unless a user accepts an offer to share their data. Users data is also not processed for sharing purposes if they did not consent to it.



Once this offer is accepted, the personal data is shared with the data buyer and it is at that moment that they can determine the means on how they will process it further (in accordance with their privacy policy and their offer). They become responsible for the processing of data.

5.3.4 Determination of the purpose

The processing purpose of PIMCity is equivalent to the service it provides, creation of a personal data safe and enabling data sharing. Collecting data subjects' data in the personal data safe and when the user consents to having their aggregated data shared in an anonymized way, allowing data buyers to make a data offer to those very users.

Here, several parties come together with different purposes. The purpose of PIMCity's processing does not change, but any data buyer can make an offer to a user for any purpose, which the data buyer themselves decided upon independently. PIMCity does not influence the decision-making of the data buyer, nor is it jointly determined. It must also be noted that no ad ecosystem is attached to PIMCity.

It is unclear if the purposes are converging, in other words: how inextricably linked the parties are at this point? This will also decidedly influence which relationship the parties have. As such, the inextricable link will be discussed per scenario below.

5.3.5 (Joint-)controllers?

A clear distinction must be made between the data buyers and a PIMS platform. For the most part, they are separate controllers, processing their own data for their own purposes. Yet there remains some legal ambiguity about the usage of a PIMS platform, specifically the trading engine. The data buyer will always be seen as a controller, the question is whether a PIMS platform is a processor, separate controller or joint-controller with the data buyer.

5.3.6 PIMCity as processor

This scenario is the least likely of all scenarios. While theoretically it can be seen that PIMCity does the processing of its data for the data buyers, PIMCity determines the means and purposes of its initial processing entirely by itself. As such, there being a separate processing activity and PIMCity being fully in control, PIMCity cannot be seen as a processor. PIMCity processes data for its own purposes and is thus a controller at least for this part.

5.3.7 PIMCity as a separate controller

It could be possible that the relation between PIMCity and a data buyer is a controller to controller, as each controller is processing the personal data for their own means and there is no common determination of the purposes and means, no converging purpose.

As stated above, PIMCity determines the means and purposes of the initial processing. Regardless of any data buyer being involved, the platform functions as a data storage. It is



only when a data buyer wants to make an offer to a user that a connection to the personal data of the user is made. However, PIMCity has no stake nor influence on the purposes of the data buyer. It is the data buyer's responsibility to inform the users adequately about their privacy policies and purposes. It is then up to the user to decide whether they want to share their data or not. PIMCity is not involved, except as a platform facilitating this transaction.

A data buyer will never process, in any way, personal data of users unless that user has consented to it. PIMCity is able to provide aggregated data to a data buyer about its users, but this is done in an anonymous way. No user can be identified. However, access to the data has been put aside as a determining factor by CJEU case law. As there is no ad ecosystem present, and PIMCity does not facilitate more than just sharing the data, it can be argued that PIMCity and data buyers are *not* inextricably linked.

Considering the above, there is a strong case to be made for a controller to controller relationship as both parties are entirely separate from each other, both fully in control of their own processing and determination of means and purposes. PIMCity acts merely as a platform to facilitate data transfers.

5.3.8 PIMCity as joint-controller

Given the evolution of case law, and the recent case against IAB Europe by the Belgian Data Protection Authority, it is not unlikely that the usage of PIMS platforms will be labeled as joint-controllership between the platform and the data buyers. PIMCity also sets up a standardized way to process data and data buyers have to conform before they can use it, similar to IAB Europe.

However, different to TCF, PIMCity provides considerably more freedom to data buyers to determine their processing purposes and how they process the shared data. On top of that, PIMCity works with data provided by the user directly and only when it chooses to do so. This is fundamentally different than the TCF which gathers user data by activity of the data subject. Additionally, users themselves are in control whether they want to share their data with the data buyers. So while PIMCity certainly facilitates the link between users and data buyers, they have no stake in the relationship between them.

When compared to the TCF, some similarities must be considered. First, PIMCity creates a standardized way of working. If the data buyers want to use PIMCity, then the data buyers must conform to the platform. This is in line with how IAB Europe decided the purpose and the means of the TCF. It *could* be stated that PIMCity plays a decisive role in the processing (dissemination) of personal data, which may substantially affect the rights of the data subjects. The participating parties (other than PIMCity), would not be able to reach the goals (purposes) set by PIMCity without the trading engine. PIMCity sets up a standardized approach and a technical protocol which their participants must adhere to before they can use the trading engine.

However, the PIMCity implementation is not as far reaching as the TCF. IAB Europe's implementation encompasses a whole ecosystem, bringing together adtech vendors and publishers alike). On the other hand, PIMCity serves as a data storage and a data marketplace (two separate purposes). It is in the end the user which remains in control of their data, whereas this is not the case within the TCF.



The elements at hand may lead to a joint-controllership relation, depending on how case law evolves. However, this would complicate the usage of any PIMS platform. In any case, legal clarity is necessary.



6 Concluding remarks

Personal Information Management Systems are an upcoming and promising tool to manage personal data of users, whilst also facilitating sharing of data with interested parties. Stakeholders interested in setting up a PIMS should head additional attention to the following issues, issues which are the fundamentals of PIMS; consent management, data marketplaces and the capacities of parties.

While the technological advancements are there to implement robust and functional PIMS, it is clear that the legal certainty surrounding the usage of PIMS is not.

Maximising the resources used to inform the user does not translate 1-to-1 in informed users. A structural imbalance is not solved by providing more information in a certain way. Some users simply do not care about their data.

Data marketplaces struggle with usable statistics and the protection of personal data, which often collide. However, by using privacy-preserving analytics, it becomes possible to provide useful information to data buyers. Additionally, the identity of users remains protected, although no solution is perfect on both sides. Anonymisation techniques are at minimum a good practice to implement adequate technical and organisational measures according to the GDPR. Whether these techniques truly anonymise the data is still up for debate.

Lastly, the capacity of parties is a real pain point. Depending on which role a party gets assigned (based on the concrete facts), their responsibilities drastically shift. It is important that this legal uncertainty is cleared up, either through case law or through clarification by legislation or interpretative bodies. Without the necessary stability, the usage of any PIMS remains risky and opens parties to administrative fines.



7 Sources

Article 29 Working Party Guidelines on transparency under Regulation 2016/679.

AUSLOOS, J., DEWITTE, P., NAUDTS, L., *Meaningful Transparency through Data Rights: A Multidimensional Analysis*. In: KOSTA, E., LEENES, R. and KAMARA, I. (eds), Research Handbook on EU data protection, Edward Elgar, 2022.

BECKER, R., THOROGOOD, A., BOVENBERG, J., MITCHELL, C., HALL, A., *Applying GDPR Roles and Responsibilities to Scientific Data Sharing* (May 01, 2021). Available at SSRN: <https://ssrn.com/abstract=3851128> or <http://dx.doi.org/10.2139/ssrn.3851128>.

BU-PASHA, S., *The controller's role in determining 'high risk' and data protection impact assessment (DPIA) in developing digital smart city*, Information & Communications Technology Law, 2020, 29:3, 391-402.

DUCUING, C., SCHROERS, J., *The recent case law of the CJEU on (joint) controllership: have we lost the purpose of 'purpose'?*, Computerrecht Tijdschrift voor Informatica, Telecommunicatie en Recht 2020(6), 424-429, 2020.

EDPB Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subject.

EDPB Guidelines 05/2020 on consent under Regulation 2016/679.

EDPB Guidelines 07/2020 on the concepts of controller and processor in the GDPR.

EDPS TechDispatch on Personal Information Management Systems, Issue 3 / 2020.

ELLIOT, M. et al, *Functional Anonymisation: Personal Data and the Data Environment*, Computer Law & Security Review (2018).

FERRETTI F, *Data Protection and the legitimate interest of data controllers: much ado about nothing or the winter of rights?*, Common Market Law Review 51: 843–868, 2014.

GRUSCHKA, N., VASILEIOS, M., KAMER, V., MEIKO, J., *Privacy Issues and Data Protection in Big Data: A Case Study Analysis under GDPR*. 5027-5033. 10.1109/BigData.2018.8622621.

HÖLZEL, J., *Differential Privacy and the GDPR*, EDPL, 2/2019.

JANSSEN, H., COBBE, J., SINGH, J., *Personal Information Management Systems: A User-Centric Privacy Utopia?*, Internet Policy Review 9, no. 4 (December 18, 2020).

JHA, N. et al., *A PIMS Development Kit for New Personal Data Platforms*, in IEEE Internet Computing, vol. 26, no. 3, pp. 79-84, 1 May-June 2022.

KELLEY, P. G., LUCIAN, C., BRESEE, J., and CRANOR, L. F., *Standardizing privacy notices: an online study of the nutrition label approach*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10). Association for Computing Machinery, New York, NY, USA, 2010, 1573–1582.



KNIGHT, A., STALLA-BOURDILLON, S., *Anonymous data v Personal data – A false debate: An EU perspective on anonymization, pseudonymization and personal data.*, Wisconsin International Law Journal, vol. 34, no.2.

MICHELE, F., PALLAS, F., *They Who Must Not Be Identified—Distinguishing Personal from Non-Personal Data under the GDPR.* International Data Privacy Law 10, no. 1, 2020, 26.

Ruddell, B. L., Cheng, D., Fournier, E., Pincetl, S., Potter, C., Rushforth, R., *Guidance on the Usability-Privacy Tradeoff for Utility Customer Data Aggregation*, Utilities Policy 67 (December 2020).

SAMARATI, P., *Protecting respondents identities in microdata release.* IEEE Trans. Knowl. Data Eng. 13(6), 1010–1027 (2001).

SANTOS, C. et al., *Consent Management Platforms under the GDPR: processors and/or controllers?*, ConPro '21, available via [2104.06861.pdf \(arxiv.org\)](https://arxiv.org/pdf/2104.06861.pdf) (last accessed 16/08/2022).

SOLOVE, Daniel J., *The Myth of the Privacy Paradox*, 89 Geo. Wash. L. REV. 1 (2021).

SWEENEY, L., *Achieving k-anonymity privacy protection using generalization and suppression*, Int. J. Uncertain. Fuzziness and Knowledge-Based Systems, 10(5):571–588, 2002.

SWEENEY, L., *k-anonymity: a model for protecting privacy.* International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002; 557-570, 2002.

SWEENEY, L., SAMARATI, P., *Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression.* Technical report, SRI International, 1998.

TRUTA, T. M., & VINA, B. *Privacy Protection: p-Sensitive k-Anonymity Property.*, 2006, 22nd International Conference on Data Engineering Workshops. Atlanta.

VADHAN, S., *The Complexity of Differential Privacy.* In: LINDELL, Y. (eds) *Tutorials on the Foundations of Cryptography, 2017, Information Security and Cryptography.* Springer, Cham.

VEALE, M., NOUWENS, M., and SANTOS, C. T., *Impossible asks: Can the Transparency and Consent Framework Ever Authorise Real-Time Bidding After the Belgian DPA Decision?*, Technology and Regulation, 2022, 12-22.

WACHTER, S., *The GDPR and the Internet of Things: a three-step transparency model*, Law, Innovation and Technology, 2018, 10:2, 266-294.