Raval Data Commons (RDC)

RAVAL
DATA COMMONS

Manual for the *RDC Data Management Organization*

A project *developed by*

eticas foundation

with the *support of*

Ajuntament de Barcelona

February 2018, Barcelona.

1. Introduction

This manual details the technical and human protocols to be followed when administering the RCD system. Besides describing its features and functionalities, it explains the **normative and operational criteria that data managers** (the Data Management Organization within the Barcelona City Council) **have to consider and follow** when developing their respective tasks. On this basis, it proposes concrete guidelines with this purpose.

It must be noted that this Manual focuses on the **management of data provided by the RDC stewards (Raval social organizations, universities, etc.)**, while processes for opening data owned by the Barcelona local administration (data on hands of the City Council) are already standardized within the City Council, and must follow the guidelines established for electronic administration in the Spanish State[1].

Still, the governance scheme currently applied to the Barcelona Open Data will need to be slightly adapted in order to be integrated to the RDC. This means that the security and technical protocols to be followed by the stewards of the City Council for opening data to be integrated in the RDC, will also need to be in line with the governance model suggested below. For instance, differently than the data management model applied within the Open Data Barcelona, **the RDC distinguishes data according to their level of sensitivity**. Therefore, the protocols and security standards included in this Manual will have to be considered by the Data Management Organization when opening data from the local administration to the RDC (public stewards) or when receiving external data (social stewards) to be integrated into the system.

Defining Raval Data Commons

The present project aims to **pilot a methodological model of Data Commons (collaborative data) in the Raval neighborhood**, responsible for generating social and economic value within the local community. The opening of data on relevant issues for Raval citizens will foster their participation in public life and promote a more innovative ecosystem. With these aims in mind, the RDC platform has been structured by picking up the needs of the community and focusing on four crucial domains for the Raval neighborhood: security, culture, commerce and housing.

This has been achieved by a comprehensive examination of Raval sociodemographic information and fieldwork conducted with organizations in the neighborhood. The information collected during these processes has been translated into the

[1] Check in: https://administracionelectronica.gob.es/pae_Home/pae_Estrategias/pae_Interoperabilidad_Inicio/pae_Normas_tecnicas_de_interoperabilidad.html#REUTILIZACIONRECURSOS

functionalities of the RDC platform (you can check it here) and has been considered in its initial conceptual design.

The Raval Data Commons project has been carried out by the Eticas Foundation and is part of the projects of socioeconomic impulse of the territory backed by the Barcelona City Council in 2017. With this plan, the city seeks to position itself at the forefront of digital transformation. As reflected in this manual, the role of the City Council goes beyond its support to Raval Data Commons initiative since it is planned that the **municipality will become the manager and controller of the system** once it is implemented.

2. Raval Data Commons governance

As referred above, this Manual focuses on the administration of the RDC system by the specific team to be established within the Barcelona City Council for this purpose. In this context, the data management processes considered in this Manual do not address the already developed technical and human protocols established by the City Council for opening the data of its different Departments, in the context of the its Open Data system. Nevertheless, even though some organizational changes will have to be achieved, the general protocols and security measures proposed in this Manual should be compatible with the open data management in operation.

Inspired by the CUSP Data Hub and the City of Seattle's open data policy models of governance, the RDC system is managed by a management team integrated by:

- the **Data Protection Officer;**
- the **IT Office;**
- the **Open Data Team,** integrated by:
    - **the data curator(s)**
    - **the Confidentiality Officer;**
    - **the Departmental Stewards.**

This team makes up the **Data Management Organization (DMO)**. Local organizations and individuals who want to open their data about Raval will directly interact with this team, which will support them during this process and afterwards in case of need. The legal requirements framing these data administration processes, the general distribution of roles and responsibilities established for data management and the policy principles guiding the whole project will be detailed below.

It should be mentioned that, aside from the above components of the Data Management Organization, stewards and users will contribute to the repository in different ways (see Users and Stewards Manual for further details):

- **Stewards**: Each Local Council's Department will have a steward who will be responsible for their department's participation in RDC. In addition, the organizations from the private sector that collaborate with the project must appoint a steward. They will be in touch with the DMO to set priorities and oversee the publication and ongoing management from the datasets in their Departments. They also keep the open data set inventory up to date.

- **Users**: They are citizens or researchers that use the RDC database to stay informed, carry out research or to develop initiatives in the benefit of the community.

2.1 Legal requirements around RDC

The General Data Protection Regulation (GDPR) will be applicable as long as personal data are collected, stored and processed as part of the Raval Data Commons project. This means that if personal data are not subjected to a successful anonymization process, they will be liable to the demands laid down in GDPR. That implies that public and private institutions must anonymize their data in order to freely release it to the public without accounting for GDPR dispositions. The following scenarios present a certain set of characteristics that makes it vital to consider the GDPR when dealing with them in the context of an Open Data portal:

1) RDC will allow users to **share personal data in concrete cases**, even though this information will not be opened to the general public.
2) Data provided by the RDC contributors to be integrated to the system as open data **may be wrongly anonymized**, even though safeguards in place in the users' manual and those integrated to the system are in place.
3) **Personal data having to do with the Raval are currently in the hands of the local government and** will have to be curated in order to be released as Open Data.
4) Besides, the data management organization of the City Council will act as data controller providing access to information that will be subjected to **different levels of security**. The City Council will have to define secure conditions to share personal data with service providers who will be legally considered as processors.

Hence, the requirements defined in GDPR mostly apply to the different processes leading to the filtering of data but also to the security mechanisms developed by the controller and processor(s) in order to ensure purpose limitation, privacy and data quality. As it was mentioned before, in the RDC governance scheme, the Local Council would be assigned the role of data controller, which would imply that it should ensure compliance with the following relevant duties:

Articles	Controller duty
Demonstrate compliance (Rec. 74, Art. 24)	The **data controller** must take a **proactive approach to compliance with privacy** and data protection legislation.
Data protection by design and by default (Rec. 78, Art. 25)	This would imply that considerations regarding privacy rights and data protection should be **embedded from the beginning into the platforms** developed within the scope of the project.
Keep record of processing activities (Rec. 82, 89 Art. 30)	Controller must keep a record of processing operations that should be disclosed to DPA's upon request.
Data security (Rec. 83, Art. 32)	The mandatory security measures will vary according to the nature of the processing operations being carried out. However, the general principle is that the data controller must **put in place adequate measures to ensure the integrity and security of the data**.
Report Data Breaches to DPA's (Art. 33)	In the **event of a data breach,** the data controller must give notice in the following 72 hours of becoming aware of it.
Ensure quality and accuracy (Article 5.1 d)	Official data generally enjoy what could be defined as a ''presumption of accuracy''. However, **data collected by private organisations and NGOs is likely to be closely examined.** According to article 5.1 d), measures must be put in place in order to ensure that data are as accurate as possible. Data subjects have the right to contest the accuracy of the personal data being processed (article 18.1 a).[2]

Source: own elaboration on the basis of GDPR.

Some additional legal requirements apply to open data systems at the state level. In the framework of Law 19/2013 on transparency, access to public information and good governance, some limits to data opening by public administrations are delimited (articles 14 and 15). For our purposes, the two most important aspects are most likely **data protection/privacy rights and intellectual property (which constitutes an additional limit to the public releasing of data)**. Besides that, article 8 of 37/2007 l indicates that the reuse of public information could be subjected to the obligation to prohibit processes of de-anonymization from taking place in case that the data contains traceable elements that could enable the

[2] According to the recommendations of the EU data portal data publishers should: link to the data from their website, update the data regularly if it changes, and commit to continue to make the data available. See in: https://www.europeandataportal.eu/elearning/en/module5/#/id/co-01

identification of individuals (personal data). Another area of interest can be found in article 3.3, where a list of the exceptions to which this law will not apply is provided. This is particularly relevant since the general outlook of the law favors transparency and the extension of data reutilization, which is demonstrated by article 3.2, in which it is stated that the law will be applied to all documents elaborated by or under the custody of the Administration unless they have been explicitly exempted from it. In this list, there are a total of twelve exceptions. However, for our purposes the most relevant ones are the following:

- Documents that are protected by intellectual property rights (letter e).
- Documents to which access is limited by virtue of data protection legislation (letter j): Obviously, the pivotal piece of legislation in regard to the protection of personal data is GDPR. Hence, this letter could be currently interpreted as a referral to GDPR.

Besides, article 3.4 poses further limitations on the scope of application of this law. Concretely, the information referred to in articles 5.3 and 15 of law 19/2013 will not be fit for reutilization unless the dissociation process mentioned in article 15.4 of the same law is performed on the data. These protected categories of data are the ones listed down below:

- **Especially protected categories of data** (15.1): The categories of data that are considered especially protected at the national level are defined in article 7.2 of the Organic law 15/1999. In order to disclose these data, an authorization of the affected data subject will be mandatory.
- In other cases, when the data cannot be included into the ''especially protected categories of data'' label, the public entity to which the request of information has been forwarded should ponder how the right to data protection and the right to access to public information on the basis of a set of criteria that are included in the article.

The 1495/2011 Royal Decree, developing the 37/2007 law (amended by law 18/2015) dives deeper into the questions that were introduced in the previous piece of legislation. The reuse of public information is confined to non-personal data (article 11), which is in alignment with the data protection and privacy regulations at the European level (mainly GDPR). In order for personal data to be made available, it will have to be anonymized. The disclosure of the information is subjected to a small set of general limitations (article 7) and could be subjected to further conditions under exceptional circumstances (article 8). The ''*Administración General del Estado*'' has opted for a unified and comprehensive approach to the publication of public data, which translates its **willingness to not only make data available but to do so in a standardized and user friendly manner** (articles 4 and 5).

However, the law references some limitations to the general principle that information should be made available upon request. Those exceptions can be found in article 2.1, which redirects to article 3.3 from the 18/2015 law. Probably the most significant exception from a data protection perspective is contained in subsection j), which refers to those data to which access is limited by virtue of data protection regulations.

On the other hand, the Royal Decree 1/1996 on intellectual property becomes relevant in the case of the **commoning of information subject to intellectual property law**. Law 37/2007 establishes a set of exceptions to which that law does not apply in article 3.3. As was pointed out beforehand, personal data and data that are protected by personal data regulations are excluded (subsection j) but that is also the case for those data upon which property rights are upheld. This means the Public Administration will have to consider intellectual property before releasing data to the public.

At the regional level, the Law 19/2014 is particularly relevant since it establishes which information has to be published by the public sector and which are the enforceable limits. In its article 8 the categories of information that must be published by virtue of the transparency principle are laid down. Among these typologies we can find reports and studies (letter h), statistical information (letter j) and a general clause that encompasses any issue of public interest and the most solicited types of information via the exercise of the right to access public information. Article 16, on its part, allows for the reutilization of public information for any legitimate purpose, with a particular emphasis in application that will generate added value.

Concerning the legal limits to the processing of this information in terms of copyrights, article 17.1 makes it clear that it could be the case that certain public information would be assigned a creative commons license by regulation. The spirit of the Law is **predominantly strongly in favor of an extensive interpretation of the right to access public information**, which is enshrined in article 20.2, where it is said that limitations to this rights have to be justified case by case in a restrictive manner.

Finally, some legal provisions and requirements for reusing information of the public sector must be considered at Spanish level. In particular, the Decree 1495/2011, and its normative updating Law 18/2015, will be followed. These pieces of legislation promote the reuse of the information held by the public sector under the above privacy and intellectual property conditions. Expressly, article 5 of Law 18/2005 indicates the following:

> “(1) Administrations and agencies of the public sector will promote the publishing of documents for reuse as well as processing of requests for reuse are made by electronic means and through multi-channel platforms when this is compatible with the technical means they have at their disposal.
>
> (2) Administrations and bodies of the public sector shall facilitate their documents in any format or pre-existing language, but they will also try, when possible and appropriate, to provide them in an open and machine readable format, according to the provisions of the previous section and jointly with its metadata, with the highest levels of precision and disaggregation. Both the format and the metadata, as far as possible, must comply with open formal standards and norms. This does not imply that the Administrations and public sector bodies are obliged to create documents, adapt them or provide extracts of documents, when this entails a proportionate effort that entails more than a simple manipulation.”

2.2 RDC data scheme and data controller organization duties

A member of the Barcelona City Council will be appointed as Data Controller of RDC having the above detailed responsibilities and duties. As part of the data controller organization, the Data Management Organization **must manage all the data received by the system, monitor and provide users’ credentials and filter the data to be published on the platform, among other tasks.**

Following the CUSP model and in order to comply with the above-explained data protection by design and by default (Rec. 78, Art. 25 of GDPR) and data security (Rec. 83, Art. 32) requirements, data will be classified according to their sensitiveness in three categories. Security protocols will become more and more demanding as data get more private and potentially damaging for the individual’s privacy, property or integrity rights. In particular, the chart down below can serve as an example of how to classify the different typologies of data according to their potential risk for privacy and integrity:

GREEN DATA	❑ Open data on Raval housing, insecurity, culture and commerce ❑ Data that do not allow for the personal identification of individuals (anonymized data)[3] ❑ Public data from other Open Data portals(including the City Council)

[3] According to recital 26 GDPR, anonymized information will be that which ‘’*does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable*’’.

YELLOW DATA	❑ Potentially re-identifiable information (including metadata).[4] For instance, information on crime with addresses ❑ Copyright data (more restrictive open data license)
RED DATA	❑ Sensitive personal information (scientific research purposes) (sensitive attributes, such as race, sexual orientation, religious or philosophical beliefs…)[5] ❑ Internal/Operational data of the City Council

Source: Own elaboration on the basis of CUSP.

Green data includes both anonymized data provided by citizens and social organizations (stewards) and all the information that the City Council is currently allowed and able to make public or has already made public. Once the Open Data Team verifies its condition of public information, the focus is on quality controls more than on security mechanisms for guaranteeing privacy or nondisclosure. Instead, the open data system City Council is not currently able to facilitate access to other forms of data, such as sensitive data protected from disclosure, or to make under secure protocols and mechanisms. RDC will manage some specific information (although non-confidential), to be provided by both citizens and the local administration. Such information will be used only to support research or policy making as part of the Yellow and Red categories. **This is one of the added values of RDC**.

Besides the Data Protection Officer, a group of experts will manage the data received and processed by the Data Commons system. These tasks will be conducted by the Data Management Organization[6] according to the following distribution of roles, tasks and responsibilities:

- **Data Protection Officer (DPO)**: The DPO is responsible for the overall security of the data and also is in charge of examining potential members requests and authorizing members' credentials. In the same vein, he/she will oversee the process of approval of procedures, which will enable RDC members, including authorities or researchers, to access files that are not

[4] According to article 4.1 GDPR, personal data means "*any information relating to an identified or identifiable natural perso*n". Yellow data overlaps with the concept of "personal data" in GDPR.

[5] Red data corresponds with the notion of "special categories of data" as it is defined in article 9.1 GDPR.

[6] All the members of the team will have to be notified about their responsibilities and competencies. Experts in charge of yellow and red information will receive training on data protection beforehand and will have to take part in annual training updates in this field.

open to the public (red and yellow data). The expected purposes of this access can be either **scientific research or policymaking.**

- **IT Office managers**: These experts are ultimately responsible for the technical maintenance of the system, including the monitoring of security protocols and status. The IT Managers will also assign/manage passwords of RDC contributors/requesters based on the decisions made by the Data Protection Officer.

The Open Data team

- **Data curator(s)**: While the IT managers are in charge of the operational aspects of the system, data curators will be in charge of the system content and data quality control. These experts will ensure a correct data quality and organization within the platform. With this aim, the curators will approve or reject requests for uploading information, checking data and metadata quality (according the criteria/protocol detailed below) and legal compliance in terms of data protection.

- **Confidentiality Officer**: He/she evaluates official requests submitted by researchers or authorities for accessing yellow or red data. This Officer must assess the motivations and purposes alleged by scientists or civil servants to access sensitive data. If he/she does not see any threat to confidentiality contained in the research proposal, it will be referred to the IT manager who will verify that the facilities have all that is needed for the research to be carried out. Finally, the Data Protection Officer will approve or deny access in view of all the previous events.

- **The Departmental stewards**: each agency or department of the City Council contributing with data to the RDC will have to appoint a steward. This person will be in charge of the selection, preprocessing and provision of relevant data to the Data Management Organization according to the protocols and requirements included the users and Stewards Manual.

Overall, this team of the Barcelona City Council will be responsible for the regular management of the RDC, including the above tasks and other related to the promotion and enhancing of the system such as training, communication with the community and related activities including hackathons and other synergies. Among these tasks, the Management Organization will also produce **annual auditing and statistic materials** on the performance of the RDC for the City Council major, which will be accessible in the Data Commons portal.

Data governance and levels of confidentiality

As mentioned above, inspired in CUSP data governance model, each typology of data managed by RDC is subjected to different levels of confidentiality. These levels of confidentiality therefore correspond to different data access restrictions as follows:

Type of data	Access subjects and level
Red Data that includes personal identifiers.	❑ Access to this data is restricted to registered individuals (including researchers, civil servants or LEAs) under special authorization of the DPO (or to LEAs under judiciary request). A specific request must be submitted to the DPO with this purpose in all cases and scenarios. ❑ Researchers and civil servants must sign a non-disclosure agreement in order to get the information. ❑ This access can only be made at the facilities of the City Council and information will not can be exported or downloaded to external machines.
Yellow Data that does not contain direct personal identifiers but includes sensitive or copyright information.	❑ Access to data in the yellow database can only be achieved by to registered users of RDC, including scientific researchers. ❑ Researchers must sign a non-disclosure agreement in order to get the information. ❑ This access can be made both at the facilities of the City Council or remotely. Some of yellow information will not be available for export.
Green (open data) Data that does not allow re-identification.	❑ Public data available for all citizenry through different means and in multiple supports. No restricted green data will be established by the system.

Source: Own elaboration.

2.3 Policy principles

The above-described Raval Data Commons team will administer the RDC system and its complex security scheme according to GDPR and national data protection requirements, as well as also strictly following these **principles**:

1) In accordance with GDPR, all RDC community members will have to provide their **explicit informed consent** in order to register and get access to the platform. The information provided to members at the stage will detail all the legal requirements and managerial options to which data provided is

subjected. The community must have to direct access to some basic definitions around the system (e.g. to know what personal data is) in order to take part of RDC.

2) **No personal or confidential data will be opened or shared without in advance informed consent of contributors and explicit authorization of the DPO**. Moreover, information on the organizations acting as data suppliers will not be released without informed consent.
3) Data managed by RDC system will not be exploited with commercial purposes or with **any other purposes than those established in the informed consent templates** and detailed on the project webpage.
4) The data controller organization, in this case the Barcelona City Council, will establish a set of measures to **ensure the safety of data**. This includes the above explained data system architecture and organizational mechanisms, as well as the data quality assessments and security measures explained below.
5) Tracking of data origin and sources is allowed as part of the management processes established for securing data quality. This is also aimed at **preventing any negative social externality related to data opening**. In the same line, analysis of data accuracy, completeness and representativeness of data will be carried out.

3. RDC computing environment

In this section, we will introduce the RDC system architecture and interoperability. Its data management technical mechanisms and its security measures will be explained as well, in the context of the RDC computing environment.

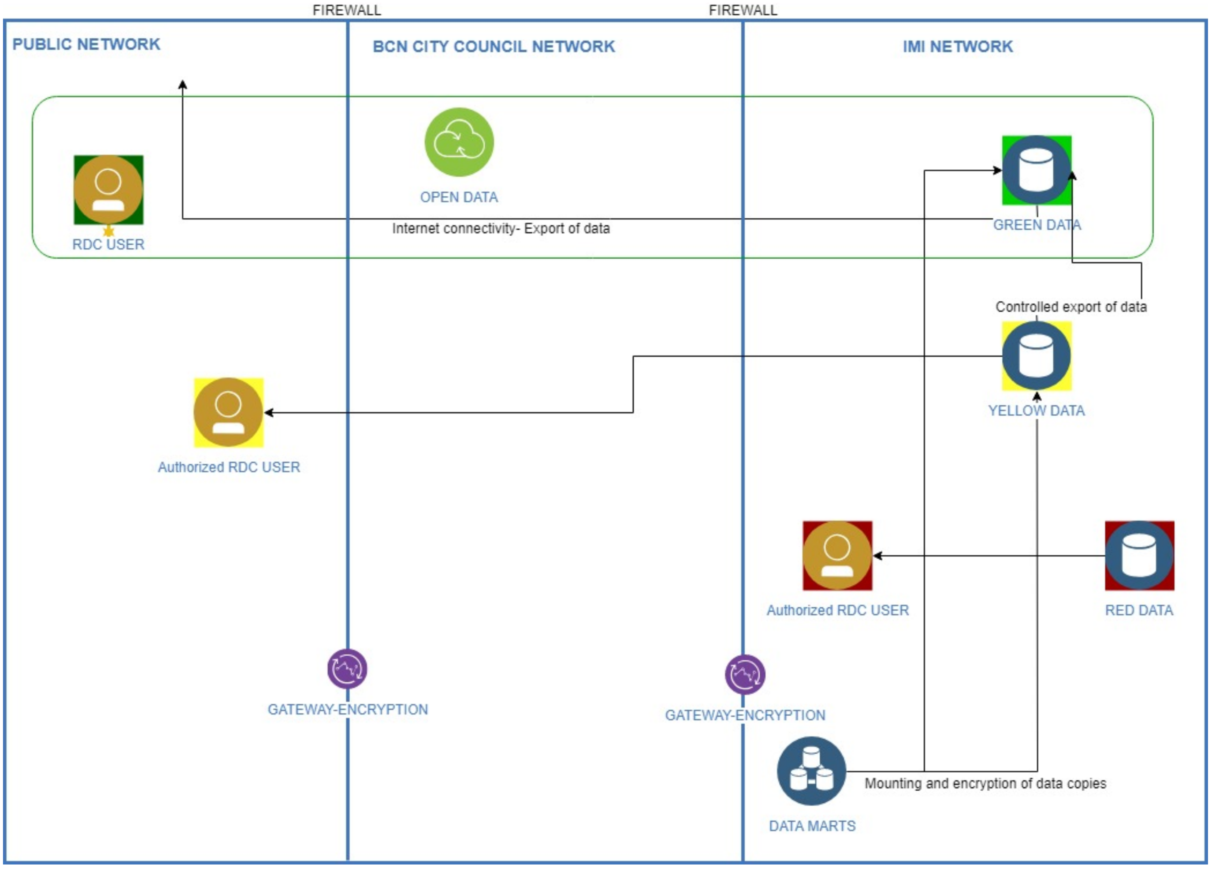
The **RDC system will be structured according to the already described (red, yellow and green) categorization of data**, which is established on the basis of their sensitivities. Three different databases are designed to store these different types of information as follows:

❏ The **green environment** will allow all RDC users to upload data and have remote access to all the information stored. Information stored in the green data environment can also be transferred without the DPO's authorization to other parts of the RDC system. However, after having been transferred, the information stored in the green database will only be further processed by the IT Managers and the Data Curators. Hence, stewards will only be able to edit the metadata and the descriptions once the data is stored in the database. This editing process will be performed according to the requirements and instructions included in the Users and Stewards' Manual.

❏ Yellow and red data will also be stored in specific databases, respectively. **Yellow and red servers have access restrictions**, which means that only some authorized RDC users will be able to access them. That is not the case for green servers. Following the CUSP model, the information stored in the yellow database will require further authentication mechanisms in order to be accessed and used. In this regard, the information stored in the **red database is only accessible on site and by individuals explicitly authorized by the Data Protection Officer**. In both cases, the servers are not connected to the internet and direct data export is not allowed.

Interoperability between these databases will be limited and subjected to a set of security mechanisms. Firstly, the **red database will be standalone** and will not be accessed via internet. Only anonymized information coming from the yellow area will be disclosed (transferred to the green database) under statistical disclosure control measures.

As shown in the image below, while green and yellow databases are interconnected, the red database is standalone and accessed within the *Oficina Municipal de Dades (IMI)* facilities.



As established by CUSP, communication across servers (External-BCN City Council and IMI) and also between the data vault used to store green and yellow data is encrypted. Moreover remote access is provided by using encrypted channel services (ssh tunnel <also encrypted> and Windows Terminal Server). **Hence, all the**

communications with data storage systems, with the RDC users/stewards and with the systems storing data copies will be encrypted[7].

4. Data Management

4.1 Registration and credentials

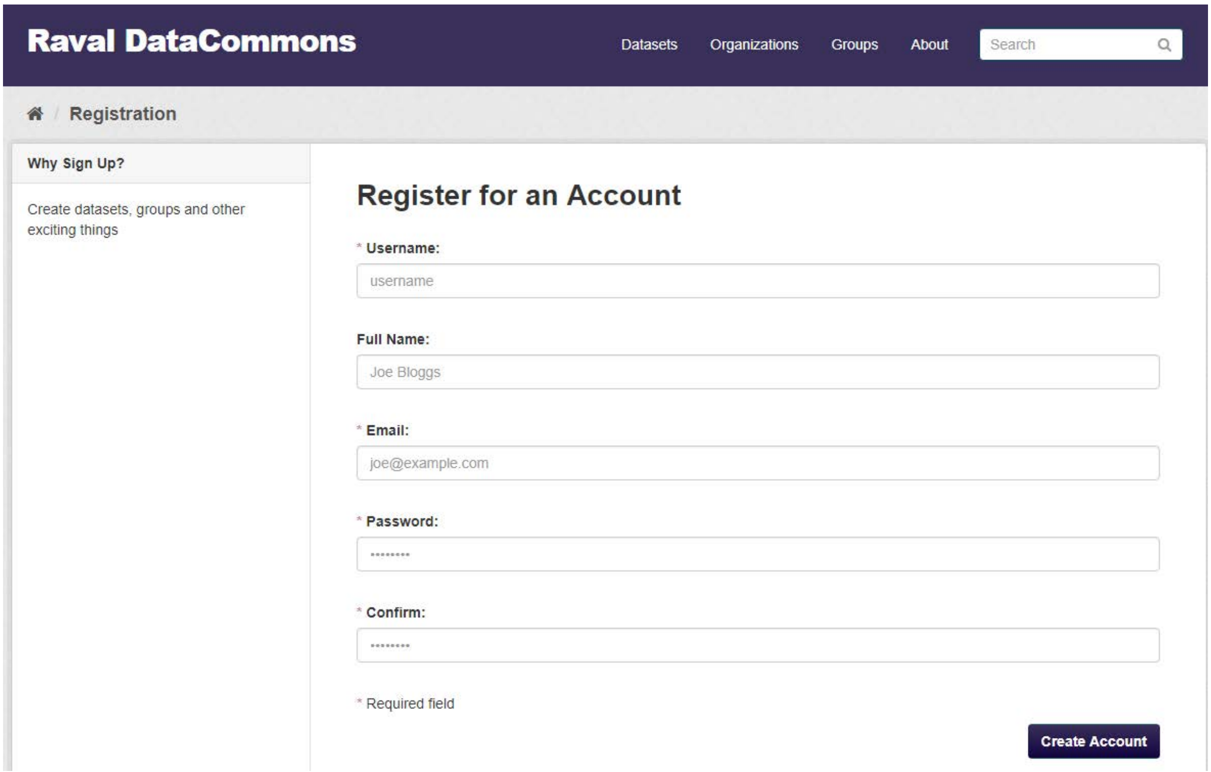
In order to become part of the RDC community individuals **must first register in the platform** via the Access Request Form, which includes the username and profile data. The following data will be requested when completing a contributor profile:

❑ username,
❑ full name,
❑ email,
❑ password,

This information may relate to:

❑ affiliation: name of the organization (non exclusionary), + sector/domain,
❑ Raval/Others

See below the registration form to be filled by the stewards:



Why Sign Up?

Create datasets, groups and other exciting things

Register for an Account

* Username: username

Full Name: Joe Bloggs

* Email: joe@example.com

* Password: ••••••••

* Confirm: ••••••••

* Required field

Once this registration request is completed, the IT Manager will provide the applicant with a unique **authentication password**. This password will allow the

[7] For further information on the security systems to be applied to the RDC see section 5.

steward to log in and access the green data platform, upload data and edit his/her profile as well as the metadata and description corresponding to the uploaded data.

In the case of stewards (including researchers and civil servants) who will upload red and yellow data, their data profile will be checked by the CO upon request of data submission or data access.

Besides these credentials and the ones corresponding to the Data Management Organization, only credentials to Raval Data Commons stewards from the City Council will be granted by the RDC Data Protection Officer. These credentials will have to be yearly renewed in all cases. Credentials will be modified in case the user changes his/her access status. Details on these procedures and about the mechanisms established to prevent unlawful access are explained in section 5.1.

❏ Credentials will can be reseted later at any time by contacting the RDC IT manager at ITmanager@RDC.com

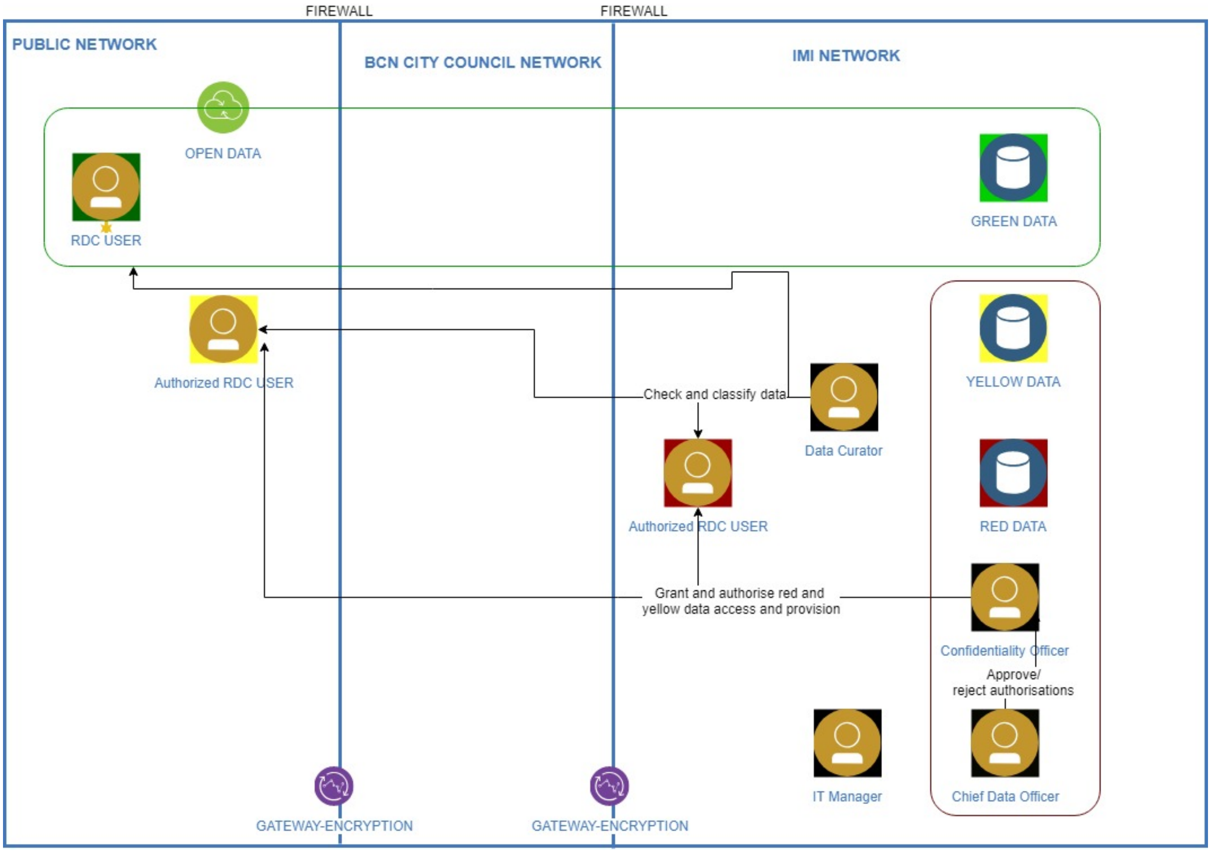
4.2 Management of data and roles

Once users have been correctly registered they can make requests to share and transfer data through the RDC system. Then the **Data Management Organization will carry out a set of activities oriented to (re)classify data, assess its quality and origin**, and assign different forms of access to different stewards. As mentioned above, yellow and red data will be disclosed to researchers or other users that claim a legitimate interest to access the data under controlled conditions. Thus, access to EACH yellow or red dataset should be requested individually so the proportionality of each request can be entirely assessed.

In this context, the IT managers and Data Curators will have to conduct the logistical, technical and content related aspects of data management, while the Data Protection Officer (DPO) and Confidentiality Officer (CO) will be in charge of monitoring and providing credentials, which will contribute to regulating access to datasets. These tasks can be classified in four main groups: management, analysis, self-assessment and training.

❏ **Manage and communicate**

Concerning data access, green data will be accessed by default by any individual without any restrictions as it was stated above. Conversely, the **Data Protection Officer (DPO)** will have to ultimately **authorize each access request to red and yellow datasets**. Once the user (data provider) has accepted the terms and conditions, his/her data transfer request will be checked and accepted/rejected by the CO. If the dataset is accepted, it will be stored in the RDC datahub. Before that, the **Confidentiality Officer (CO)** will assess if the requested data matches the purposes declared in the access request. Alongside that, it will check the type of information for which access is being requested. The CO has to analyze the proportionality of this authorization concerning the actors/agents involved, their institutional background as well as their data protection skills and credentials.



On his/her side, the **Data Curator will save the metadata** related to data transfer requests. As part of this process, the Data Curator will check the description, classification (green, yellow, red) and the quality of the data. After that, it will send a message to the corresponding steward suggesting the classification of the data set. If green data presents problems, the transfer request will be rejected by the Data Curator and a proposal for reclassification or improvement of the dataset will be submitted to the DPO and the steward involved.

It should be noted that, due to current problems in the domains of security and housing within the Raval, many of the data to be handled by the system could be sensitive. Considering this is very important in order to guarantee the security and integrity of both stewards and the citizens and for the correct management of the system by the data controller.

In line with this, user support is key for the system's functioning. Tasks conducted by the **Data Curators and IT managers** will include **contacts and communication with organizations or individuals** facilitating data. This has the purposes of clarifying doubts as well as enhancing data sharing and defining improved management protocols to be implemented in the future. Moreover, initiatives and requests coming from people in the neighborhood will be accounted for through ongoing communications with the different stakeholders and the media. Therefore, clear and effective communication with stewards must take place on a regular basis.

❏ **Analyze data quality and technical performance**

Data Curators will be in charge of checking **both correct anonymization in the case of Green data and overall data quality**. In order to assess if the anonymization has been successful the likelihood of re-identification of individuals within the data set will be gauged.

For the purposes of assessing the quality of the dataset, the data curators will conduct research in order to assess the quality of the data provided by stewards. On top of that, aspects such as the security measures put in place by the steward in the course of the processing of the data, the conditions in which the data were collected or its representativeness will also be taken into consideration. The self-assessment filled up by stewards during the final stage of the uploading process will constitute one basis for this process. Furthermore, advice from the IT department or statistic specialists from other departments of the City Council might be requested in case of need, always in this last case without releasing any personal data.

Data curators will also consider the level of sensitivity of the data when analyzing the three defined types of data. Red data are the most sensitive according to our protocols. Yellow data have a medium level of sensitivity. Finally, green data are the least sensitive out of all three categories, which is why they are readily available to anyone that wants to consult them.

Conversely, a priority criterion will serve as the basis for the releasing of the curated datasets. High priority data sets will be selected based on their importance relative to hot topics that are present in the public debate. In principle, given the current situation in the neighborhood, housing and security are likely to be prioritized.

The tasks conducted by the **IT manager** will also include the **analysis of the functionalities and outputs of the procedures conducted by the platform to assess its performance**. Statistics on each file's use, conditions and exploitation will be produced and analyzed by the IT office.

❏ **Self-assess data management**

Finally, a self-assessment will have to be conducted monthly to secure data quality and minimize any potential impact of data management on privacy or integrity of users and public in general. The assessment will include the analysis a set of variables around data management, including operational context and communication with RDC community members. It will cover the following process and variables:

- ❏ Was data source confirmed/contrasted?
- ❏ Was sensitiveness/typology of data correctly determined?
- ❏ Was data accuracy/completeness and up to data attributes checked?
- ❏ Were regular operational system security checks conducted?

A three-stage process will be carried out as part of this assessment. In the first place, the classification of the inputted data will be confirmed by analyzing anonymization and copyright issues. Secondly, classified data will be verified by checking data sources and data quality. Thirdly, conformity of the data to agreed standards will be assessed.

As far as data quality, data curators will use a template including the following criteria adapted from the Seattle Open Data Manual:

- ❏ **Good quality**: This category identifies datasets that are regularly exploited by users or stewards from the City Council, are understandable for the general public and contain accurate information.
- ❏ **Regular quality**: These categories defines data that is used by visitors of the platform and stewards from the City Council, but it is difficult to understand for non-owners/creators of the datasets or has accessibility/usability problems.
- ❏ **Poor quality:** This category includes datasets for which information about its exploitation is not available or is poorly exploited by stewards and users or has significant layout problems.

Training on RDC administration and use

Data curators and the DPO will organize data management training for the other members of the Data Management Organization and the external and internal stewards on a quarterly basis. This training course will be aimed at ensuring a standardized protocol for data management and that all technical and human aspects of the RDC system administration are covered consistently.

- ❏ The training consists of five modules:
- ➔ Module 1: Introduction to the RDC system: overall architecture and functionalities
- ➔ Module 2: Goals and requirements of RDC: Open data and data commons, GDPR, open access.
- ➔ Module 3: Step by step data management for stewards
- ➔ Module 4: Step by step data management for members of the DMO
- ➔ Module 5: Communication and engagement in RDC

4.3 Step by step administration of access and data

4.3.1 Administration of credentials and accesses

- ❑ **Step 1: Process each steward registration request**

Registration requests are received by the IT manager, who provides stewards with their corresponding credentials.

- ❑ **Step 2. 1: Process each steward request for accessing yellow data**

After receiving this request, the data curator will send the corresponding information to the confidentiality officer. Access to yellow data will be processed by the confidentiality officer who will ask for the final approval to the DPO.

- ❑ **Step 2.2: Process each steward request for accessing red data**

After receiving this request, the data curator will send the corresponding information to the confidentiality officer. Access to yellow data will be processed by the confidentiality officer who will ask for the final approval to the DPO.

4.3.2 Administration of data

- ❑ **Step 1: Receive dataset and check user credentials**

Each dataset is received by the IT managers jointly with the data curators, who examine their sensitivity and quality.

- ❑ **Step 2: Administer dataset request /publish**
    - ❑ **Step 2.1 For green**

In case of green data, datasets can be published after the analysis conducted by the data curators with the support of IT managers.

- ❑ **Step 2.2 For yellow and red**

In the cases of red and yellow data, data curators must check datasets jointly with the confidentiality officer and request authorization to the DPO for its publication.

- ❑ **Step 3. Publish data**

IT managers jointly with the data curators can publish authorized data in the RDC platform.

- ❑ **Step 4. Audit, check credentials annually, check data updates**

The IT managers and data curators will conduct the above-explained regular audits concerning data quality, management and access.

5. Security control and regular auditing

The IT manager will regularly review system security status and credentials. These credentials can be adjusted or revoked in case that the user does not follow the guidelines established in the User's Manual. Membership will be automatically extinguished after two years of inactivity.

Part of the IT Officer Managers tasks will consist of assessing the results of the data anonymization processes conducted by the RDC members, including organizations and individuals. Since anonymization never completely frees a dataset from re-identification risks, it is essential to evaluate the effectiveness and impact on utility of anonymisation. Hence, IT manager who will be able to filter the received data.

Data access will also be regularly monitored. This monitoring will have to be proportionate to the type of data and their sensitiveness. Moreover, the system will provide an alert in case of repeated and failed attempts of login. **After five attempts, new access will have to be provided by the IT manager**.

In order to have relevant information on data security and prevent data breaches in line with GDPR, historic logs will be recorded and stored during two years. However, access to logs will have to be authorized by the DPO and the CO under audits or in the context of specific investigations. This is needed in order to guarantee the data protection rights of users, mainly considering the management of large amounts of data, in some cases of sensitive nature. At the same time, this is done in order to be compliant with the ARCO rights, and guarantee users' exercise of rights to access, rectification, cancellation and objection. Nevertheless, the management of the logs recording system is very complex and requires to sustain the custody of these logs until the DPO authorize their use under very specific conditions and security measures. It must be noted that, once the change of custody is broken, all the actors involved, including the IT managers in charge of accessing the logs will be responsible for their veracity.

The Data Management team of the City Council will regularly test the RDC systems and develop risk assessments on the basis of the project development. This will include the analysis of the data quality, system architecture and interoperability subsystems. Part of these tasks will be to confirm the correct authorization and provision of access to the different categories of information and the tracking of logs.

Concerning **storage,** the following mechanisms will be implemented:

- Device Management
- User cloud DRM
- User local DRM
- Admin lock management

Concerning **identification**, the following protocols are applied:

- Matching personal data with local government DB
- Validation mail with OTL

- Set first password
  Strength checker
- Integration with 2/3 factor
- Identification with something you…
  - know (password)
  - have (cookie, smartphone, etc.)
  - are (fingerprint, face recognition, etc.)

Regarding **authentication:**

- Confederation with 3rd party
- Avoiding password storage
- Avoiding phishing with best user practices
- No phone, no mail…
- Alerts from unusual login
- Unknown sources (block IP ranges)
- Failed attempt (slow down 2^n)
- IDS & IPS canary systems
- Renewing periodically and promoting OTP

Access and data breach

- Protected Project Work Space
- Monitored Data Export (red and yellow)
- Monitoring of access level
- Physical Access Controls to the City Council Facility
- Malware Protection
- Data Backup and Replication
- Unauthorized Acquisition and Agreement Violations (data breach GDPR)-Protocol

Below you will find a detailed description of the security protocols in RDC.

5.1 Identification and authentication

During the design of the identification and authentication policies for Raval Data Commons, the systems, applications, equipment and supports of the information system are evaluated. In this section, basic and advanced concepts are presented.

The policy validates that users have a **unique identifier** for their own personal and exclusive use. In addition, Raval Data Commons has chosen a technique to authenticate and verify identity claimed to a user.

Policies implemented

The policies that have been included guarantee the following safeguards:

- Prevention of **anonymous access**.
- Prevention of access through **credentials shared** by multiple users.
- Provide **complexity in passwords**. The commitment of maximum complexity should be associated with that which the user is capable of remembering.
- The **periodic renewal** of passwords, for treatment applications, is a necessary milestone to evaluate.
- **Limit wrong access** attempts. Alternatively, through thresholds, waiting times exponents of the nth errors.
- Waiting time for a new attempt = 2 ^ N seconds (where N is the number of errors committed).
- Through authentication in two steps, double factor of authentication or network blocking of the sources that access wrongly.
- **Authenticate in two steps and double factor**.
- A **password renewal** without the intervention of the administrator, through a *recover one-time link* sent to an alternative channel owned by the user.

Reduction of common risks

The above policies reduce the likelihood of the most common risks:

- Distribution of passwords: it ensures the method of how the secret key is delivered to the authorized user.
- Storage depends on the security of the database where keys reside.
- If control over the database were lost, the passwords would have been compromised and should be renewed.
- Method that does not force the user to memorize passwords. Implementing a password manager.

5.2 Access control

The **Access Control of Raval Data Commons manages grant permits to resources only for authorized users**. During the design, physical access controls to treatment centers and files are guaranteed, as well as logical access controls for applications, systems and networks.

During the final design process, the RDC will ensure that the Access Control covers the following functionalities:

1) Guarantee the lifting of confidentiality of information.
2) Minimize conflicts and problems arising from the assignment of permits.
3) Reduce the probability of errors of unskilled users.
4) Register the use of the services and detect cases of abuse.
5) Revoke permissions.

The challenges that the access control system faces are:

1) Verify the identity of the users.
2) Verify the identity of the person authorizing the assignment of permits.
3) Verify that the user is requesting access to a certain service.
4) Integrate multiple levels of permissions for a specific user.
5) Determine quickly and reliably the level of user permissions.
6) Manage changes in user access requirements.
7) Restrict access permissions to unauthorized users.
8) Maintain an updated database of users and their rights.

Logical access control

Security measures for **logical access control**, which have been implemented during the design and by default, are indicated in the following list. In addition, it should be noted that the measures listed below have been verified in a test environment (if necessary, the policies could be extrapolated to new sites).

1) Verify the creation and management mechanisms of authorized users or credentials.
2) Verify that access control limits the privileges to be able to install programs by users.
3) Determine if the access control, despite avoiding the installation of programs, cannot prevent the execution of programs that the user would have obtained through alternative methods (Internet, a USB support, etc.).
4) Verify that an enhanced access control presents the security measures to ensure that only legitimate applications can be executed and visit the Internet sites necessary for the development of the functions of authorized users.

The logical access control in Raval Data Commons also includes security checks to the database (level data, applications, systems, servers and networks) to ensure confidentiality, integrity and availability. Thanks to the Access Control policies, the following most common technological problems are avoided:

1) **Excessive privileges**: databases that offer higher access permissions than those required by authorized users.
2) **Abuse of privileges**: credentials that abuse privileges for unauthorized purposes.
3) **Elevation of unauthorized privileges**: vulnerabilities that allow a low privilege credential to access with higher privileges. These vulnerabilities are avoided with intrusion prevention systems.
4) **Platform vulnerabilities**: Problems and errors in the systems that store the databases.

5) **SQL injection**: exploitation of existing vulnerabilities in the Web access layers to the databases.
6) **Weak Audits**: non-intrusive audits are often carried out to avoid a degradation in the performance of the database.
7) **Denial of Service**: due to *buffer overflows,* corrupted data, network flooding or exhaustion of computation.
8) **Vulnerabilities of the database protocols**: allows unauthorized access, data corruption or unauthorized accessibility.
9) **Weak passwords**: legitimate users of databases that suffer brute force password attacks and social engineering. The solution should be the tendency to double authentication factors.
10) **Exposure of backup copies**: theft of unencrypted backup copies.

During the choice of access control databases policies we have considered the following aspects:

a) **Registration processing personal data**: Recording access to personal data is an important measure, because users who have access or for whom access has been granted can be identified, but are not entitled to any specific circumstance. It also helps to detect if access control is failing. The record is also used to document. For example, what are the users who have accessed during the last week. The records must comply with security guarantees, because the records are also sensitive information.

b) **Encryption**: Most security regulations and guidelines require encryption of data. In addition, encryption can be used to enable an access control mechanism (for example for databases and backup copies).

c) **Pseudo-anonymization and anonymization**: databases process data to the minimum of sensitive information (such identifiable) possible. The possibilities of anonymization (depersonalization) and pseudo-anonymization (assigning aliases) include reducing the criticality of the database to a minimum residual risk and being able to relax the security measures.

Physical access control

On the other hand, policies for security measures in **physical access control** are listed below:

- All sites with computer processing systems or storage are protected from unauthorized access, using technologies to authenticate, monitor and record inputs and outputs.

- In addition, the third party operation has been physically separated from the controller access. If any third party activities needs access, this would be done by the controller-authorized person.
- Due to possible theft, vandalism and unauthorized use of information systems, access is restricted to areas considered safe.

The physical access control integrates different categories of staff:

a) Operators and users who work regularly in the safe areas (IT Managers from the RDC office, Data Curators, CO and authorized stewards).
b) Support staff that requires periodic access (other authorized IT Managers).
c) Others, who require access very rarely.

Here are some measures that will be implemented:

- Server rooms where deliberately robust access control is implemented.
- Server rooms are always closed (even with an automated system), and with an open door monitoring.
- Guardian locked treatment equipment.
- Communication cabinets with physical lock systems.
- Communications cabinet with visibility of its defined content.
- Reception of a treatment center, with surveillance point.

5.3 Register of accesses

The access registry of Raval Data Commons is responsible for ensuring that the permits are being properly filed. For this reason, **it is necessary to include monitoring and control of accesses by default and in the design of all activities**. Thanks to the feature, in case of detecting abuses, the situation can be documented, as an exception and sent to the IT Manager so that it can be solved. The review and audit of the access registry plays a fundamental role when detecting unauthorized accesses and in comparing them with the permissions that had been assigned.

If it is suspected that a user is violating access rules, making inappropriate use of resources or using data fraudulently, the access registry will provide evidence of data, time and even content to which the user has access in certain services.

The main measure of Raval Data Commons is that the access record will keep track of the logins that are made in the application. For example:

- The access record at the start of the system and access record at the start of the application (correct or incorrect).

The second measure would be to record access to information:

- The attempts to access information to which the user has allowed or denied access and access to the information were recorded.
- If a user could access, RDC should identify which registry modified. In this case, we would observe how the audit enabled in the previous section is insufficient and databases with granular audit should be used.

This measures guarantees:

- Veracity of records
- Biannual storage
- Monthly review

Version control

Raval Data Commons has a version control mechanism.

- A different version each time the information changes. This functionality allows having a tracking of the contents for each record.
- RDC will assess if this function is sufficient to meet the objective of having an access record.

Metadata

Access records can be included in the Metadata concept (the hidden data that defines other data). In addition, the design has been implemented to protect the security metadata. Metadata can respond to a record: who, what, when, where, why and how.

References

Ostrom, E. (2015). Governing the commons. [Place of publication not identified]: Cambridge Univ Press.

Center for Urban Science + Progress (2016). Data Governance and Confidentiality Policy. [online] Available at: https://datahub.cusp.nyu.edu/sites/default/files/documents/policies/Data_Governance.pdf [Accessed 12 Nov. 2018].

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

Other sources

National Institute of Standards and Technology (http://csrc.nist.gov/publications/PubsFIPS.html), Fibs 200: Minimum Security

Requirements for Federal Information and Information Systems (http://csrc.nist.gov/publications/fips/fips200/FIPS-200-final-march.pdf

https://opendatacommons.org/licenses/by/

https://centerforgov.gitbooks.io/open-data-metadata-guide/content/

https://ico.org.uk

https://creativecommons.org

http://opendatahandbook.org/guide/es/