Raval data commons (RDC)

RAVAL
DATA COMMONS

Manual for users and stewards

A project *developed by*

eticas foundation

with the *support of*

Ajuntament de Barcelona

February 2018, Barcelona.

1-Introduction

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.

This manual is aimed at **guiding potential RDC community members** to become part of this community and to share and exchange open data with its members. Before starting to read this manual, you should know that it is mainly targeted to those individuals in charge of managing data within their organizations (stewards), which means that a certain level of digital skills are required in order to understand it and implement it or that readers would be able to acquire the needed knowledge. Most of the organizations in the Raval do not have someone in charge of managing data. In fact, our stewards will be people that are either doing research, project management or carrying out direction tasks.

An **accurate selection, analysis and sharing of data will guarantee that the information uploaded to this system is legally compliant, proportionate and that protects the privacy rights of third persons**. The guidelines included in this manual cover technical aspects to be considered in order to efficiently exploit the functionalities of the RDC portal, as well as the most relevant references in terms of data protection and copyright.

In order to address these issues and provide useful methodological guidelines for data management within the RDC system, the manual is divided into four sections. The first part introduces the system and its general goals, the second part explains how to share data within the RDC, the third part shortly describes other ways to exploit the RDC platform and the last section includes a glossary of terms. In addition, at the end of the document you can find references and useful sources of information.

Presentation of Raval Data Commons

The **Raval Data Commons** is an online platform that will allow organizations in the Raval to upload, share and use (open) data provided by other citizens, organizations and also by the Barcelona City Council. By uploading data regarding the different areas on which the platform focuses, organizations will be participating in the online ecosystem and will help to create an innovative environment. Besides, they will be providing the inhabitants of the neighborhood with valuable insights.

The platform's four areas of interest -housing, insecurity, commerce and culture- were chosen due to their special importance in the Raval now. The neighborhood is currently undergoing a housing crisis since the Spanish housing bubble burst and is now subject to real-estate speculation, gentrification and an increase in the amount of touristic flats. These issues have in turn affected the level of insecurity, with an increase of theft due to the influx of tourists, *narcopisos*, drugs and drug-related crimes. Commerce have also evolved in the past years, adapting to the new environment in the Raval. Furthermore, the Raval is a multicultural neighborhood with multiple cultural facilities and with active social organizations in the arts and heritage domains. The **information stored in RDC will help neighbors in multiples ways**, including the elaboration of more informed and targeted public policies, and day to day issues such as finding housing, identifying threats for the community or accessing updated information on specific services or products provided by cultural groups or local stores currently active in the neighborhood.

What is Open Data? And what is Open Data about Raval?

The idea of data commons is to empower citizens to access and profitably use the database owned by public authorities in order to generate social, cultural and economic value for the community. The data stored by city administrations contain **information that can be used by policy makers, researchers and citizens** for developing a richer and more accurate representation of urban problems and potentialities. Moreover, by making it easier to access open data, the idea behind "data commons" is to trigger citizens-lead projects and initiatives that might lead to the re-evaluation and increased engagement with the urban spaces. Being aware of the amount of data that increasingly characterizes cities and having access to it would promote a model of open data grounded on responsibility and participation rather than on ownership.

Given these premises, the project is grounded on two conceptual pillars: "open data" and "commons". On the one hand, the concept of Open Data consists of giving free and public access to the databases containing information about the Raval. On the other hand, the older concept of commons, which points to the sharing of resources within the neighborhood in pursuit of a common or public good, which draws from the work of Ostrom (2015) on the potential of communities to cooperate and achieve shared goals. By providing open data from these areas, more insight can be gained to tackle the most challenging issues facing the Raval.

In line with this philosophy, the RDC system integrates a concept that goes beyond the usual open government data systems in two different registers. On the one hand, actively engage the community organizations, which can participate in the open data ecosystem not only as users but also as data contributors. Moreover, in line with this, the platform concept looks forward to establishing spaces of convergence outside the web. On the other hand, the RDC platform will be able to administer personal data with specific purposes and under strict security protocols. This is done by establishing a

scheme of data management that allows to segment data access according to the sensitivity of different types of data. Such approach promotes that researchers and policy makers can further contribute to the development of the neighborhood based on data commons.

As part of the RDC community, you can both share data with different members of the platform and with the general public. Besides, you can access data provided by other users and by the City Council.

Roles and responsibilities in Raval Data Commons

Stewards and users will contribute to the repository in different ways:

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

For an Open Data repository to be a useful resource for the community, ongoing work has to take place to ensure that the data and the platform are in compliance with agreed standards. These tasks will be conducted by the Data Management Organization[1] 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 that will enable RDC members, including authorities or researchers, to access files that are not open to the public (red and yellow data). The expected purposes of this access can be: either **scientific research or policy making (find detailed information of the access conditions in section 2 of the Controllers Manual).**

[1] 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.

- **IT Officer 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.
- **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.

Overall, the Data Management Organization will be responsible for the regular management of the RDC, including the above tasks and others 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 DMO will also produce annual auditing and statistic materials on the performance of RDC for the City Council major, which will be accessible in the portal.

The portal

The RDC interface has been designed to minimize the time and effort that users have to dedicate to be familiarized with its features and functionalities. When we first open the website, we encounter the following elements:

- Header: By clicking on the header, we will be able to go back to the home page at any moment.
- Brief description of RDC: Right below the header, we can find a brief description of the website and the project.
- Sections: On the right side of the header we can find the following areas:

  -Datasets: By clicking on this section, users will be able to freely explore all the datasets that are currently available to the public by searching for the relevant keywords. On the left side of the web information is provided regarding various

aspects having to do with the data sets (organizations, groups, tags, formats, licenses), which is intended to help with the browsing.

-Organizations: In this section, information will be collected on the different organizations that actively collaborate with the Raval Data Commons project. Besides, users will be able to filter data sets according to the organization that uploaded them.

-Groups: In this section, the data are organized according to the issues they are relevant for. Currently the categories are commerce, culture, housing, security and other data.

-About: A brief summary of the nature and aims of the projects.

- Maps and visualizations

- Search bar: Users can do a direct search by using the search bar, which is located right next to the ''about'' section.

- Log in/Register: These two buttons are situated right above the search bar and allow users to either create an account (register) or to log in once they have created an account.

Finally, the data repository is available in Catalan, Spanish and English. In order to change language, click on the menu available on the right down corner.

2- How to Open up Data

For the Open Data repository to be truly useful for the citizenry and for researchers, certain **standards and criteria have to be followed**. Thus, it is not enough with simply dumping datasets into the platform regardless of their utility or their compliance with basic ethical and legal standards. Therefore, we have come up with a systematic guide that takes collaborators through all the questions to be considered when publishing datasets. **It must be noted that the references included in this document correspond with a manual approach to releasing the dataset**. There are automatic and programmatic ways to do it, which are more scalable. Once implemented, it is expected that the RDC will be scaled up**.**

Firstly, we shed some light on the following question: **should I make public this specific dataset at all?** Even if preventative measures are taken, publishing a dataset

involves certain risks for the rights and freedoms of data subjects, which makes it necessary to weigh up whether the utility of the data justifies assuming those.

Secondly, even if the dataset is deemed as useful, it is necessary to ensure that its **release is done in accordance with the legal framework**. For our purposes, GDPR will lay down the basic norms to be considered, although we will also comment on laws with a more limited scope.

In order to share Open Data in a way that is compliant with GDPR, **some procedures will have to be performed on data to prevent them from constituting a threat for the rights and freedoms of the people represented in the datasets**. Those techniques usually fall under the label of anonymization techniques, the basis of which can be found later in this document. Moreover, in the manual the reader will find specific guidance on how to safely share personal data through the RDC system in some specific conditions and circumstances.

To conclude, we provide guidance on how to upload the dataset to the platform once it has been processed in a way that makes it useful, ethical and legally compliant.

Step 1: Dataset selection

As it was already mentioned, the process of publishing a dataset is more complex than simply uploading information to the platform without paying attention to certain standards and criteria. Instead, it requires work at the moment of the uploading and on an ongoing basis to standardize the data, make it machine readable, write good metadata and take preventative measures to shield privacy and secure the data. In order for the process to be as seamless and effective as possible, the stewards from each organization or Department of the Local Council will have to pave the way for the Data Management Organization.

1.1: Data set format:

A **dataset is a set or "basic unit" of information**, composed of records with associated attributes. It can be published in multiple supports such as Word or other machine-readable formats when possible (PDF and other non-readable formats when there is no other option), or as a spreadsheet format (Excel) with records (rows) and attributes (columns). The rows often relate to an individual, an object, or any other entity and the columns relate to certain aspects of that entity (example below).

Name	Age	Postcode	Sport played
Anna	34	09365	Basketball
William	32	75236	Tennis

1.2: Is it a good idea to release this data set?

As it has been stated in the introduction, more data is not always better. **The first criteria that must be used to ponder if a dataset should be made public is utility**. The end goal of an Open Data repository is to be useful and allow for the transformation of raw data into public value. Such a process can only take place if the community has relevant and adequately curated data at its disposal. Subsequently, down below we provide insight into what is a dataset and how to make decisions regarding their utility for the community.

Now we understand what a dataset is and why it may be useful for the community, we have to quickly identify if the information managed by your organization fits RDC aims:

- **Examine if the available data of your organization fits within the domains of Raval Data Commons**: **housing, security, commerce and culture in Raval.** Check the definitions established for each of these domains in the platform.
- **Consider one dataset** at a time.

Decide whether the dataset you want to upload is useful for the community; does it help to understand the situation in the Raval? If after having considered the above points, you don't think the dataset will be useful for understanding the reality of the neighborhood and will not help citizens or policy makers to create solutions for it, **it may not be worth it to upload the dataset.** Its publishing may present with unnecessary privacy risks if the data is not anonymized correctly. This judgement is especially difficult, as the utility of a dataset may not be obvious at first glance. Besides, the utility of a dataset can change over time. Thus, the two main criteria to be used at this stage are the following:

- Utility of the data set for the community: Will the data set be of use in the short or medium term? Whom is it going to be useful for?
- Risks for individuals: How could the disclosure of the data to the public affect the individuals whose data are in the data set? Are the data anonymized or not?

The decision to upload or not upload the data set will result from pondering the pros and cons of each decision for both the community and individuals. Therefore, the judgement will necessarily have a contextual nature. The set of questions down below can assist stewards in making a decision in that regard:

- What data does your department use for internal performance and trend analysis?

- What data populates your department use for internal performance and trend analysis?
- What information is published as a performance metric?
- What information do you report to regional or state agencies?
- What information do you share with other city departments?
- What information do you share with external partners?
- What information does the public repeatedly request?
- What kinds of open data are similar agencies across the country publishing?

1.3: Are your data green, yellow or red?

When assessing data sharing you also need to consider that, **in some specific cases, personal data can be shared with the RDC community**. As shown in the Table below, the RDC system will store and manage red, yellow and green datasets, which are categorized according to their level of sensitivity. Hence, besides defining the relevance of your datasets for the RDC community you should also consider, at this stage, how your data fits within these categories.

GREEN DATA	❏ Open data on Raval housing, insecurity and commerce ❏ Data that do not allow for personal re-identification (anonymized data)[2] ❏ Public data from other Open Data portals(including the one currently in place)
YELLOW DATA	❏ Potentially re-identifiable information (including metadata).[3] For instance, information about crime with a certain level of granularity. ❏ Data subject to copyright (licenses that are more restrictive than open data license)
RED DATA	❏ Sensitive personal information (sensitive attributes, such as race, sexual orientation, religious or philosophical beliefs…)[4] ❏ Internal/Operational data of the City Council

[2] 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*''.
[3] 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.
[4] Red data corresponds with the notion of ''special categories of data'' as it is defined in article 9.1 GDPR.

Source: Own elaboration based on CUSP (Center for Urban Science + Progress, 2016).

Step 2: Check and confirm legal compliance

Once the utility of a dataset has been established, it is time to **take into consideration if it could be shared through an Open Data portal** in a way that is respectful of the basic legal requirements imposed by the legal framework. As it was said above, the core of the legal framework can be found in the General Data Protection Regulation, which recently entered into force in the whole of the European Union. We have boiled down all the complicated legal jargon to the guidelines listed down below in order to aid legal compliance:

- **Privacy:** Depending on the level of sensitivity of a dataset, the Raval Data Commons platform will only allow a restricted access to that dataset. Thus, the following rules apply:

    - Fully Open Data (**GREEN** data) must not contain personal data that allows for the identification of an individual. According to the GDPR (Article 4), personal data "'means any information relating to an **identified or identifiable natural person ('data subject')**; an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;" You should note that **biometric personal identifiers** (such as facial images, fingerprints or DNA) are defined as sensitive information, which requires specific safeguards for its treatment.
    - **Restricted access**: Data can be identifiable and potentially disclosed (**RED** and **YELLOW** data).

- **Intellectual Property**: Open Data is subjected to an open license which means that it can be downloaded, modified and re-used for free (**GREEN** data):
    - **Licensing[5]:** However, data may not be re-used unless it is licensed in a way that will provide the legal basis for its free processing and reuse. Therefore, you should ask yourself: "Are data currently published?" If so, it can provide a good starting point if this publication was done using an open data license. **If you cannot apply an open data license to the dataset, it may not be published on the platform**. There are different licenses, such as the Creative Commons Licenses[6], and **different open**

[5] Further recourses on this issue can be found in: https://www.europeandataportal.eu/en/content/show-license
[6] https://creativecommons.org/licences/publicdomain/

data licenses that you can select in the platform.[7] In order to be considered open, the chosen data license will have to follow the criteria established by Open Knowledge International here:https://opendefinition.org/od/2.1/en/ **We recommend to apply the Open Data Commons Attribution License (ODC-By)[8] to your dataset** (**GREEN** data).

- It could be the case that you want to apply some restrictions to your data in terms of reproduction, distribution, derivative works, sublicensing or use of patent claims. In that case, you can select other open data license and your data will have to be categorized as YELLOW data.

As summarized in the table below, before proceeding with step(s) 3 you need to confirm if the information you want to share fits the limited and specific conditions and circumstances under which personal data can be shared within the RDC.

Category	Privacy	Intellectual property
GREEN data	No personal data (anonymized)	Open Data Commons Attribution License
YELLOW data	Pseudonymised data on certain public events: Information or metadata (information that helps to understand and interpret the data set) on events that have occurred in so-called "hot zones" which could potentially lead to the re-identification of individuals. For instance, re-identification is particularly likely to take place with datasets that include information on insecurity or housing, which that can be useful for scientific analysis or policy-making but that, due its	Other Open Data licenses

[7] You can find more information on these licenses in these links: https://creativecommons.org/share-your-work/licensing-types-examples/
https://es.wikipedia.org/wiki/Licencia_Abierta_de_Bases_de_Datos
[8] https://opendatacommons.org/licenses/by/

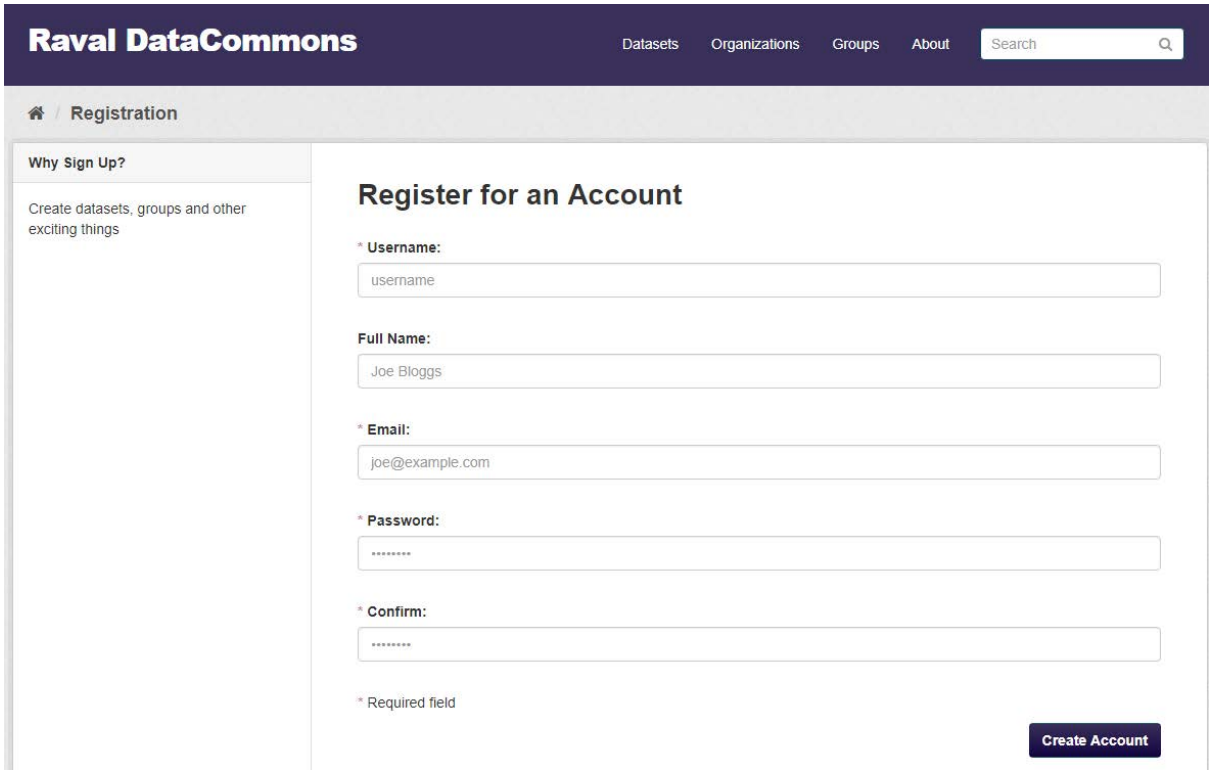
	characteristics, cannot be kept completely anonymous. Open Research Data. It can include personal information. It will have to follow the special requirements of science, such as being obtained through systematic methodologies. In the case of identifiable data, explicit consent of the data subject to use and exchange her/his data will be mandatory to guarantee compliance with data protection rules.	
RED data	Personal data	Open or not open licenses

If some of RED cases apply to your dataset, you should go directly to step 3.2.

Step 3: Register in the platform

Fill in the **registration form** with the requested data and complete your organization's profile. Only the following information will be required from you:

- ❏ username,
- ❏ full name,
- ❏ email,
- ❏ affiliation: name of the organization (non exclusionary), + sector/domain,
- ❏ Raval/Others



Nickname and passwords: You need to choose your nickname (it must have less than 10 letters) and your password. It should include a combination of more than 8 numbers and symbols. **If you want to upload red or yellow information**, the system will also ask you to define a question + answer about you for future verification of identity. Keep this information stored in a safe place that is equipped with the right safeguards. It is also recommended to change your password regularly**.**

Step 4: Define format and layout of the file and modify its attributes if needed

At this point, you are ready to prepare the file to be shared. You might have to modify its format in order to publish the data. By following the five starts, open data plan[9] you can go beyond making your data available on the Web in whatever format (WORD, EXCEL or others). Data structuring should always be aimed at increasing the reusability and the linking of data (data augmentation).[10] For instance, if you are skilled enough, you can structure data in images as Excels, make data available in a non-proprietary open format such as CSV, link your data to other open data sources or even make data available through an API.

[9] https://5stardata.info/en/
[10] For further information on data structuring check: http://openrefine.org/#

- **Title the dataset in an accurate way.** As part of this process, you should correctly name your file. Try to be concise and clearly reflect its content. We recommend to include the following information in the title of the file:
    - Descriptive word(s) for content: e.g. *number of festivals in Raval*
    - Include project/institution name: e.g. *number of festivals in Raval (CCCB)*
    - Include dates following the ISO 8601 (YYYY-MM-DD): e.g. *number of festivals in Raval (CCCB), 2013- 2017 or 2017-04-09.*

Use lowercase letter only. Do not use special characters. We also recommend you to avoid words such as FINAL or similar.

NOW you have to choose between steps 5.1 if you want to upload GREEN data or 5.2 if you intend to share some type of personal information through the platform (YELLOW or RED data).

Step 5 (1): Open data: identify personal data and anonymize dataset

RDC has been devised for publishing and sharing data that cannot lead to the identification of an individual (anonymized data). Thus, a crucial step in the process of opening data is the identification of information related to specific individuals in the dataset. Here are the necessary steps to anonymize a dataset:

1. **Removal of direct identifiers**: No personal data that directly identifies an individual should appear in the dataset. Remove names, nicknames, ID and passport numbers, phone numbers, emails and addresses. Take into account that also other identifiers (that is, labels the identity of), such as club or associations ID, should be entirely removed.[11]

2. **Determine the quasi-identifiers**: Quasi-identifiers give information about an individual and does not directly identify them. ZIP code, gender, date of birth are common quasi-identifiers. Others are ethnicity, salary, and other demographic and socioeconomic information. Quasi-identifiers may lead to the re-identification of individuals if they appear in other databases that are linked with other identifiers, or when they are linked to one another within the same database. This must be addressed during the anonymization process.

3. **Identify other data that could lead to re-identification**: The remaining data can potentially still convey information about individuals. For instance, several aggregated statistics could allow for linking data belonging to an individual if correlated.

[11] For a comprehensive list of identifiers please check https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-are-identifiers-and-related-factors/

4. **Reduce the privacy risk**: Based on the previous considerations, apply the following anonymization techniques:
    a. **Disruption:**
        i. Permutation: switch values within the database.
        ii. Noise addition: Slightly change the value to make it less accurate. Example: Add two years to a person's age.
    b. **Generalization**: Generalize values by modifying the scale or order of magnitude. Example: Instead of giving a person's exact date of birth, give the year of birth only. Instead of giving an exact address, only give the first digits. The larger the scale, the harder it will be to single out an individual.
    c. **Suppression**: Remove cells.

Step 5 (2): Dataset containing personal data: check legal compliance and classify them before submission

As mentioned above, RDC will allow you to **share personal data only in specific cases and under special security conditions**. Organizations should only gather, store or use personal data if they have a clear objective for doing so. Following this criteria, we have identified the above-explained cases in which administering personal information could be beneficial for the community. This comprises:

- **Information derived from scientific research projects** containing personal information. These datasets are identified as red information (including pseudonymised data).
- **Pseudoanonimized information. For instance, data about "red zones"**[12] that could potentially lead to the identification of individuals involved in it. These datasets would be classified as yellow information.
- **Documents or sources** that are subjected to restrictive licenses, which limit the conditions under which they can be shared and reused.
- **Identified personal data**.

If your dataset includes one or more of the above kinds of personal information, you should classify it as yellow or red at the moment of submission. This will be further explained in the next step.

Step 6: Upload data to Raval Data Commons

Once data has been considered useful, legally compliant and has been anonymized (when required), it is time for it to be uploaded to the platform so it can be turned into public value by the members of the community and researchers. The necessary steps to do so are summarized down below.

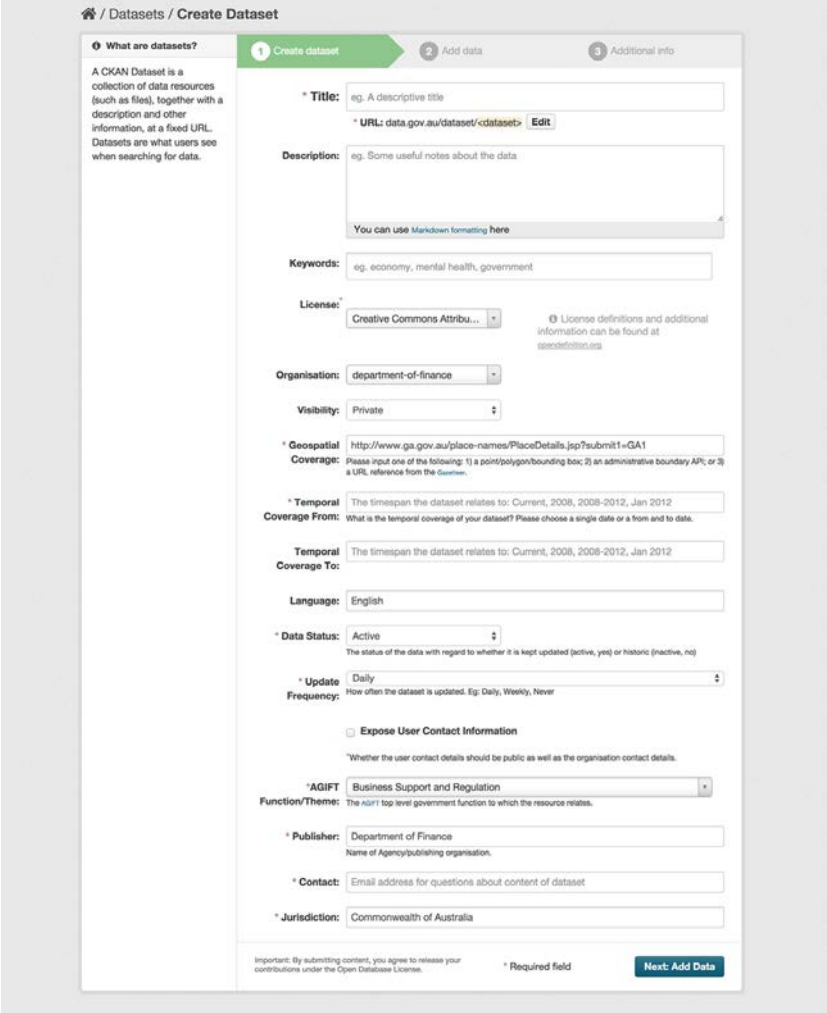
[12] Only streets and squares will be considered as red zones.

1. **Read and accept the informed consent sheet.** In this form, your rights and responsibilities in relation to data are all explained in depth, including your rights in relation to data rectification, erasure, restriction of and objection to processing. You must confirm that you have read this and the Privacy Policy before moving on to the next step.

2. **Fill the data transfer request:** Once the file is ready to be shared, you will have to complete the data transfer request including some complementary information. This information is aimed at checking the classification of the data you are sharing and make it more useful and searchable across the system.
    a) **Description of the data set**: This description must be concise (up to 100 words) and include all the keywords related to the information comprised in the dataset, such as the following: reasons for data collection, organization in charge of data collection and current/suggested exploitation.
    b) **Select the suggested classification** of dataset/file (red, yellow or green).
    c) Check and confirm the information on **data use agreement** in case **of red and yellow data;**
    d) **Select a theme for the dataset:** commerce, housing, culture or security/ others.

    2.1 Add tags: you can select up to five tags and attach them to the file by using the @. This will also contribute to enhance its usability.

    Example of registration of dataset in CKAN



3. **Once your data transfer request is sent, you will have to accept the *terms and conditions*. This message will include the following information:**

a) The RDC member (Data Contributor) has the authority to share the data with the data controller (Barcelona's Local Council).
b) The RDC member understands that the data provided will be added to the Raval Data Commons repository and will become available for the general public.
c) Data that are not public (yellow or red data), but whose existence will be made public in the RDC search engine will only be accessible to researchers who have been authorized by the City Council to use that data for approved research projects or with concrete policy making purposes. The Raval Data Commons Data Protection Officer oversees the research project approval process with the support of the Confidentiality Officer and the rest of the RDC team.
d) For yellow and red Data, the Raval Data Commons Data Facility and the Data Provider will sign specific data use agreements that contain non-disclosure clauses in addition to these terms of use.

e) In the case of sharing green data (open) through the RDC community, you need to accept the application of an Open Data Commons Attribution License (ODC-By)[13] to your dataset.
f) The Data Provider will, within reason, work with Raval Data Commons Data Facility researchers to validate newly developed metadata and data classification information.
g) The Raval Data Commons Data Facility will handle the data appropriately, in accordance with current legislation and our data management policies. The RDC team will classify the data you provided taking into account the specifications you provided as well as under the project policy and will handle the data accordingly. Moreover, the RDC team will share with the data provider (you) any newly created metadata or versions of the data, upon request.
h) The RDC data contributor has been given the option to Accept/Reject to be shown as part of the community (only logos and links will be published on the web).

4. Once your transfer request is accepted, you should upload the file!

Step 7: Fill the data quality self-assessment form

Once the file has been updated you will be asked to fill a short template about the quality of the information you are about to share. This is done in order to improve further data processing and data exploitation in the future. The template will request information on the following aspects:

- **Data accuracy**: Are you certain that the data resembles reality? Is there some source that can be consulted in order to assess that concordance?
- **Metadata**: Did you provide enough information about the data for users to really know how to make use of it? Metadata should cover questions such as the origin of the data, legal restrictions on its use and other contextual factors that help users understand how useful the data is and, therefore, the limitations that are known to you. In order to facilitate the process of uploading the data along with the necessary metadata, the RDC interface will ask stewards to fill up the following categories:

- Title: It will have to conform to the City's standards.[14]
- Description and business purpose: Further useful information can be added to this section.
- Keywords that people searching for the information are likely to search for.
- URL's that link to the website of the department or organization that is uploading the data set, as well as others that can be of interest.
- Primary contact information.
- Source and source type.

[13] https://opendatacommons.org/licenses/by/
[14] We recommend these guidelines be followed: https://centerforgov.gitbooks.io/open-data-metadata-guide/content/

- How is the data set extracted and prepared for publication.
- How the data set is published to RDC.
- How often this data set will be uploaded to RDC.

For each column in the data set you will document the:

- Column (field name).
- Data type (text, date, time, geocode, etc.).
- Sample value.
- Column metadata description (what the data in the column represents or means).

- **Is the data machine-readable**? This implies that can be directly processed and manipulated digitally, which means that information on formats such as PDF is not Open Data. Once we have established that the data is machine readable, can it be read in multiple formats?
- **Granularity**: Does your dataset have a sufficient degree of detail for it to be useful? Sometimes it is necessary to reduce the granularity of a dataset so it can be safely released. However, during that process its usability can go down to a level that renders it useless.
- **Would you consider that the dataset is interoperable**? Interoperability is usually achieved by consistency. That consistency can be internal (in relation to similar datasets) or external (in relation to other datasets uploaded by different providers). Example: if certain definitions are relevant for the dataset, is the definition being used consistent with more widely used definitions?

These questions' aim is exclusively to prompt the steward to consider the real utility of the dataset before publishing it. Nevertheless, the quality of the data and the privacy of individuals cannot be entrusted solely to stewards. That is why there is a second control instance constituted by the curators in the Open Data Organization who will review the data set in search of threats to privacy or data that does not meet the required quality standards. If threats to privacy are spotted, the case could be referred to the Data Protection Officer. On the other hand, if a lack of quality is detected in the data set, the reviewers will let the steward know in order to try to address the concerns or at least reveal the flaws detected in the description, so users are aware of the dataset's limitations.

The process described above will be divided into the two phases described down below:

- Publish privately: During this first stage, the data set will be labelled as ''private'' and will be accessed only by the publisher, selected testers and the Open Data Organization. The Open Data Organization will review the dataset and work with the actors involved in order to make the necessary changes. The IT manager will determine when the dataset is ready for publication.

- Open to the public: Once the X authorizes it, the data set will be released to the public.

3- I have already published my data, what should I do now?

Once the data set has been published, **the steward is responsible for keeping it up-to-date**. This also implies dealing with questions regarding the dataset that come from the public as well as working with the Open Data Organization to solve any issues that might arise.

Once you have shared information, you will also be able to update and revise the information that you have already made public through the platform. Therefore, it is also possible to modify your data and the data concerning your organization. To this end, the user will need to log into the system and get the corresponding authorization to introduce changes in metadata, descriptions and tags. At least once a quarter, users should review all the data sets published by their organization in order to ensure that they have been uploaded on the basis promised in the metadata. This applies both to data that has been published manually and to data that has been published making use of automated means.

Moreover, as the RDC is aimed to **commonify data but also to contribute to the dynamization of social relations in the neighborhood and its economic and social development**, you will find information provided by the City Council in the section "Other data" of the RDC platform and information about activities around the system in the section "Synergies". Different activities such as training on data management, GDPR compliance or hackathons will be announced in this section. It is also possible to confirm one's assistance through this section.

Finally, in the section called "design your data common" you will have the opportunity to suggest new topics or features, as well as any kind of activity that could improve the data commons ecosystem in the neighborhood.

4-Glossary

- **Spanish Agency for Data Protection (AEPD)**: in Spain, the agency responsible for ensuring compliance with the legislation on data protection is the Spanish Agency for Data Protection (AEPD). In Catalonia there is also the Catalan Data Protection Authority (APDCAT).
- **Algorithm**: a well-defined and ordered process or set of instructions or rules that allows an activity to be carried out through successive steps to be followed. In terms of programming, an algorithm is a sequence of logical steps that allow solving a problem.
- **Data storage**: process by which the data is recorded in the digital data storage medium, temporarily or permanently. These devices perform the read or write operations of the supports where the files of a computer system are logically and physically stored.

- **Anonymization:** process of converting the data in a way that does not identify the individuals and ensure that identification is not likely to occur.
- **Big Data**: extremely large data sets that can be analyzed computationally to reveal patterns, trends and associations, especially in relation to human behavior and interactions.
- **Data transfer**: the transfer of data is any consented transfer of information from one location to another through some method of communication.
- **Data cycle**: the data cycle or data life cycle is the sequence that follows a data or a set of data from its creation until its extinction. It is composed of the following phases: Creation and capture, Transmission, Storage and Security, Management and Access, Analysis and Exploitation, Elimination or Deletion.
- **Encrypt**: write a message in code through a system of signs formed by numbers, letters, symbols. Current encryption methods are based on complex mathematical formulas.
- **Connectivity:** is the ability to establish a connection, a communication, a link. In the context of computing, it is the ability of a device to be connected to another device or to a network.
- **Informed Consent**: is the act and the result of consenting or allowing something (the use of your data for example, or your participation in a study) after having all the necessary information to know exactly what you are giving your consent for and why . As a legal term, consent is understood as manifest will, and in this case written, as a contract.
- **Data:** A data is a symbolic representation (numerical, alphabetic, algorithmic, spatial, etc.) of an attribute or quantitative or qualitative variable. It is a value or reference that receives or generates the computer by different means. The data represent the information that the programmer manipulates in the construction of a solution or in the development of an algorithm. A data set by itself may not be information, but it is the processing (and its linkage to other data) of the data that gives us the information.
- **Personal data**: is any information about an identified or identifiable natural person, that is, any person whose identity can be determined directly or indirectly. For example, a name, an identification document number, location data, an online identifier, elements of the physical, physiological, genetic, psychic, economic, cultural or social identity of a person.
- **Sensitive data**: they are data that reveal the following information about an individual:
  • ethnic or racial origin
  • political opinions
  • religious or philosophical convictions
  • union membership
  • treatment of genetic data
  • biometric data
  • health-related data
  • data related to sex life
  • sexual orientation

  Sensitive data are also considered, data that may reveal circumstances of criminal record or police investigation, or cases of gender violence.
- **ARCO Rights**: the Organic Law 15/1999, of December 13, on the Protection of Personal Data, includes a series of fundamental rights of citizens: the right of Access, Rectification, Cancellation and Opposition (ARCO).

- **Algorithmic discrimination**: refers to the automated decisions made by the algorithms and especially in the event that you come to believe or prove that an algorithm has affected certain groups in a positive or negative way, discriminating against people according to age, race, religion, sex or sexual orientation.
- **Deletion of data:** it is the erasure of the content from storage devices, such as hard drives, flash memory, USB memory, etc.
- **The cloud**: the Internet cloud is a model for the use of computer equipment that transfers part of your files and programs to a set of servers that you can access through the Internet.
  It allows you to store your files on those servers, open them, use them or use programs and applications that are not on your computer, but on them. In English it is called cloud computing and hence it is known as the cloud.
- **Metadata**: many digital files contain additional information to their content. Simply open the file on the computer and check the properties to find information about the phone or camera used, identity of the person who created the file, identity of the person who has reviewed the document, etc.
- **Privacy policy**: a privacy policy is a legal document that states how an organization collects, stores, processes and deletes user or customer data. Each company or organization that manages personal data must have its own privacy policy available online in a visible and accessible place, which in clear and concise language explains how it manages and protects the data of its users. It is the responsibility of the user to read it, to have knowledge of the management of the privacy of the company or service that he is using.
- **Data protection**: refers to compliance with legislation on the right to data protection, which is a fundamental right of all people. It translates into having control over the use made of our personal data. This control allows us to avoid that, through the processing of our data, we may have access to information about us that affects our privacy and other fundamental rights and public freedoms.
- **GDPR:** it is the General Data Protection Regulation at European level, which will replace the current regulations in force. It entered into force on May 25, 2016 and will begin to be applied on May 25, 2018. The objective of the GDPR is to improve the protection of the privacy rights of people's data and to achieve the same standards throughout the EU. It forces organizations to take responsibility for ethical data management and to be transparent about the way they collect, store and use personal data.
- **Data processing**: any operation or set of operations performed on personal data or personal data sets, either by automated procedures or not, such as collection, registration, organization, structuring, conservation, adaptation or modification, extraction, consultation, use, communication by transmission, diffusion or any other form of authorization of access, collation or interconnection, limitation, suppression or destruction. "(GDPR).

5-References and relevant sources of information

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

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

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