

A Practical Guide to Data Portability

Seize the opportunities offered by the new right to personal data portability.

This booklet focuses on the right to data portability, mandated by the General Data Protection Regulation (GDPR), which comes into effect on May 25, 2018. It clarifies which data fall within the scope of portability and explains how to implement the right inside your organization.

This booklet is the fruit of a joint effort between two complementary Fing project domains: the data literacy project 'Do you speak Data?', and the 'MesInfos' personal data and data privacy project.

The joint working group was led by Guillaume Jacquart (MesInfos), with the close collaboration of Armelle Gilliard and Charles Népote (Do you speak Data?). Core group members also included Hugues Roumez and François Herlent (MAIF), Jean-Marc Raibaud (Orange), and Régis Chatellier (CNIL). Other working group participants included Érik Arnaud (Maif), Emmanuel Malkor (Carrefour), Pauline Kullmann (GRDF) and Amine Essafoui (GS1). This booklet was written under the direction of Guillaume Jacquart.



“Do you speak data?” - Infolab **“Data Potential shared, at last”**

Since 2013, Fing has been exploring the ways that data literacy can better enable people and organizations to use data. The 2017 'Do you speak Data?' operational program explored the real-world deployment of data literacy practices within organizations, through a combination of workshops, best practices sharing and in-situ operational deployments.

<http://fing.org/?Infolab>



MesInfos

“If I (an organization) can use your data, you can too... however you please!”

Since 2012, the Fing MesInfos project has been exploring Self Data — 'the collection, use and sharing of personal data by and for individuals, under their complete control, and to fulfil their own needs and aspirations'. The G29 guidelines on the right to data portability cite the MesInfos project as one of the pioneering demonstrators of the potential for innovation that this legislative advance represents.

<http://mesinfos.fing.org/english/>

INTRODUCTION

Article 20 of the General Data Protection Regulation (GDPR) grants individuals the right to recover and reuse the personal data that organizations gather about them: it protects EU citizens' right to personal data portability. The data must be made readily available, 'in a structured, commonly-used, machine-readable format'. Individuals, known as data subjects in the new legislation, can choose to download their data to a hard drive, or transfer them directly to a third party.

Unlike the right of access (GDPR Art. 15) — which grants individuals the right to verify what organizations know about them and, in so doing, lays the groundwork for more transparency and trust — the new right to portability seeks to redress the balance between data holding organizations (data controllers) and individuals (data subjects) in terms of the value that can be derived from individuals' personal data. Organizations have continued to develop more and more powerful tools and collected more and more personal data in recent years to their great financial benefit. The right to portability intends for individuals to gain proper and full access to the data they personally generate and allows them to use these data for their own benefit.

The idea is to publicly offer individuals the same quality of data services that organizations have internally. Nothing more (it is not a question of collecting new kinds of personal data) nothing less (the data must be as usable by individuals as they are for the data holding organizations). Quick, easy, regular recovery of usable data without complex procedures, and provided for a nominal fee (if a fee is even necessary).

In order to map out this new legislative terrain, our working group has been focussing its efforts on two areas: the data covered by the law, and how best to ensure and be able to demonstrate that processing is performed in accordance with this Regulation.

Which data are covered by the law?

Unlike the right of access, the right to data portability concerns **a specific subset** of the personal data held by a controller. To help controllers and users more easily draw a perimeter around the data covered by the right to portability, the WG29 (European Data Protection Working Party) has proposed a tool in the form of five simple questions.

We will now take a closer look at these five questions, point out a few grey areas and examine some special cases. Let's start with the rule.

For any data, if the controller responds :

- 'Yes' to each of the five questions, then the data is portable;
- 'No' to at least one question, the data is not portable; or
- 'Yes, but...!', the data is likely portable, but will require some additional measures to make it truly portable.

In later sections, we will present some of these special cases.

1 - Do the data concern an individual data subject?

☒ yes

- Yes, but : the data subject is a service user who is not the service contract holder (e.g., mobile phone, mailbox, etc.)
- Yes but : several individual data subjects use a service collectively (simultaneously or not).
- Yes but : it is difficult to automatically link the individual with his or her personal data.

2 - Is data processing carried out by automated means?

☒ yes ☐ no

- Yes : Paper processing is not affected by portability legislation.
- No : Even archived data (difficult to access) is covered by the right to portability. These data may justify a longer restitution period. (However, is it justifiable even to keep such forms of data for long periods?)

3 - Is the legal basis for data collection consent or contract?

☒ yes ☐ no

- Consent = Yes
- Contract = Yes
- Legal obligation = No
- The exercise of a public duty = No
- Official authority = No
- Legitimate interest = No

4 - Have the data been provided by the data subject?

☒ yes ☐ no

- Yes : All data explicitly transmitted by the individual.
- Yes : Also, all data resulting from direct actions on the part of the individual (transactions, use traces, etc.)
- No : This excludes (from the mandatory perimeter) inferred data, e.g. profiling, recommendations, etc.
- And YES! The key is to focus on reuse, and not just on portability.

5 - Will portability have no adverse effects on the rights and freedoms of others?

☒ oui ☐ non

- Yes but : will implementation of data portability respect employee regulations?
- Yes but : are the data linked to several people?

Yes, but: special cases

1 - Do the data concern an individual data subject?

Yes, but the service user is not the service contract holder, the service is collective, the service can be used by more than one person, the data relates to more than one person, etc.

Some examples :

*The service user is not the service contract holder: the call history for a mobile phone, when the contract is in the data subject's parents' name.
The service is collective: a supermarket loyalty card program.*

In the above cases, the problem is that the organization's information system does not explicitly link the data it gathers with specific data subjects. The data are still personal, and can be used to distinguish one data subject from another, but they cannot be automatically linked to a specific individual as the service now stands.

GDPR portability rights and obligations have introduced new constraints that throw contract models, data models, and even real-world ser-

vice processes into question. Even though the kinds of data gathering practices listed above have worked up to now, they have never been adequately addressed by organizations. Not only are these practices non-compliant with the GDPR, they also impair the quality and reusability of the data gathered. Situations such as these will require the implementation of significant changes to fully satisfy the law.

5 - Will portability adversely affect the rights and freedoms of others?

Portability is a right, but it doesn't take priority over other rights, such as privacy for example. Certain kinds of data and certain scenarios will force a compromise between enforcing the right to portability and respecting other rights. But this is nothing new — the right of access raised the same questions. Similarly, in terms of volume, when personal information is returned on paper — such as detailed telephone records — it also goes through a variety of verifications and annotations before being transmitted to customers. But the range and amount of personal data at stake here suggests that it will take a lot of work to properly identify, anticipate and respond to data requests like these.

'Yes, but' does not mean 'No'

Whatever the technical or organizational difficulties associated with GDPR compliance may be, and despite the apparent novelty of the process, portability is the users' right. The solution cannot be, 'We do not know how to make the data portable, so we'll just continue as before'.

When an organization requests to process an individual's personal data, whether by contract or with consent, it also assumes the responsibilities associated with that request, including compliance with the right to portability. If the organization is not able to fulfill its responsibilities, there is no need to process that data at all. To comply with the law, all the organization needs to do is purge that data from their servers.

Portable and reusable data

Why the new law?

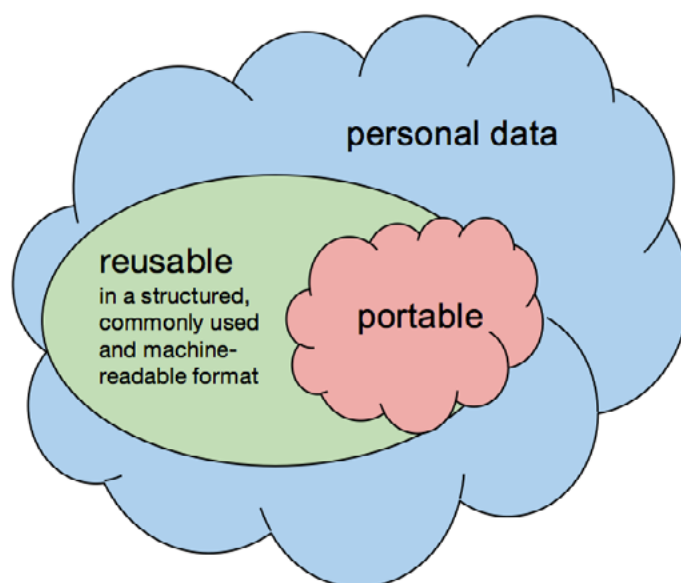
GDPR - Article 20: Right to Data Portability : *"The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller..."*

G29 recommendations!: *"The purpose of this new right is to empower the data subject and give him/her more control over the personal data concerning him or her... It will facilitate switching between different service providers, and will therefore foster the development of new services in the context of the digital single market strategy."*

Portability is a new right for individuals. It helps to redress the balance of power between organization and individuals. As the G29 guidelines point out, it also creates an opportunity for companies to **offer new services** or otherwise **gain a competitive edge**.

Our aim here is to take advantage of the momentum created by the new portability legislation to illustrate how an organization can share the data it uses every day with the individuals concerned in ways that simplify the individuals' lives as well! It will be important for organizations to find ways to **make the data reusable**, and not just seek to avoid legal difficulty.

ILLUSTRATION



G29 guidelines on the right to data portability

On one hand, the notion of reuse can be used to frame **the perimeter** of the personal data to be returned — to be in compliance, a data controller must propose a data set comprising all available reusable personal data, but **nothing is preventing controllers from widening the perimeter** of what that set might include! And on the other hand, the lens of reuse can also be applied to data formatting and any associated documentation — these should also be intelligible and **reusable** by another data controller.

Reusability for individuals – to do what, exactly?

“Our very first question was, “How might our members potentially reuse these data?” Answering that took priority over any questions about legal obligations” . (MAIF, French insurance provider)

We invite you to explore your organization's data reuse potential from an entirely different angle and across a variety of uses that empower the individuals concerned. The **following three reuse cases** introduce a new reference point for the 'My' in 'My data'. Today, 'My' refers to organizations, who use data as a tool to drive everyday operations. But from another perspective, 'My' can also refer to the data subject (individual, client, citizen, etc.) who puts these data to multiple, self-directed uses!

The goal here is not to pre-define every single way that personal data might be reused -- because **most uses will only be revealed and understood after the data becomes available**. Nevertheless, this exploration will enable you to identify some potential priorities, and also help you start to think about metadata, and the supplementary information and tools you will provide alongside the datasets so that reuse is possible.

My data is a classic customer relationship management use case.

Customer call logs, sales histories, transaction

histories, etc: these data are useful to an organization as they facilitate more efficient exchanges with clients. Likewise, data subjects can use these data to better handle active queries on their end (documents sent, case/query/file numbers, names of contact persons, etc.).

My data has some proven individual uses.

Examples :

- Keeping track of my bank account balance, making a budget based on bank transaction history
- Using consumption data to perform impact analyses, for example to see just how much money 'sustainable' household habits save
- ...

My data offers the potential for tangible new uses by individuals.

- My geolocation data allow me to retrace my steps, remember timelines and sequences of events, see the patterns in the choices I make, etc.
- My telco data allows me to know retroactively when and how long a call was.
- ...



What can we do with our data? MesInfos : <http://mesinfos.fing.org/english/>

Reusability: *does My data have meaning?*

Will My data mean something to the client? To every client?

Two cases :

- Are the data too old? (It is important to look closely at the conservation limits of these data, and not to confuse them with data that must be saved but not exploited.)
- Low quality data....?

Two strategies to address these cases:

1. One line of reasoning is, 'I am not sure that the client will find certain data useful, BUT when in doubt I share, since it won't cost me anything'.
2. Another might be, 'These data are more trouble than they are worth, so let's add them to the list of data we purge in compliance with the GDPR'.

Reusability: *are My data intelligible?*

Will the data that organizations return be in “**human-readable**” format, on top of their mandated machine-readability? Will they be intelligible to others outside the original data controller organization?

Ease of (human) use is not a specific component of the right to portability, which compels data controllers to return data in “machine-readable” formats that others can make use of. The right to data access (GDPR Art. 15) gives individuals the right to acquire a copy of the data they have generated regularly, upon request. Portability takes that ease one step further: the right guarantees that a data controller must not only make the data available, but they must do so in a format that empowers individuals to make use of it, most likely with the help of dedicated tools. The value of the data will truly reside in how easily it enables others to lift the blockade on individual use(s) — not in the data's mere machine-readability. Thus, rather than just ensuring that the data

are readable, why not make the data **easy to interpret for reuse**, including via third parties?

Is My personal data intelligible or interpretable to others outside the organization?

- Some personal data are also part of a different (non-personal) database, internal to the company. Example: my Spotify playlist is a list of song references from Spotify's internal information system.
- Certain personal data also refer to the name of a product, a sales package, etc. For example, 'name of insurance contract: RA-QVAM', or 'contract option: BASE'.

In both cases, these kinds of data cannot be used without specific contextual elements that are often not made public!

In the first example (technical reference) it would be necessary either to make the reference table available openly, enrich the data with more easily exploitable bits of information (in the example of the Spotify playlist: track name, artist, date, etc.), or propose an identifier in a public repository, if it exists for the domain (similar to the barcodes in [OpenFoodFacts](#), for example).

The second case is more delicate because it leads to potential misinterpretation: the 'base' option might have the same name, but not the same characteristics from one competitor to another. This is where the explanation and documentation of the data makes sense. The risk is that, in the absence of such explicating measures, there will be misunderstanding, frustration and discontent among users seeking to reuse these data.

My personal data: how best to facilitate reuse?

- For example, if I use the INSEE municipality code (French administrative ID system) to identify a geographical area internally, is it not also in my interest to provide the name of the commune in plain language to streamline reuse?

- If data are stored using an internal coding system (see previous paragraph), should I replace a code number with its meaning in the data I return?
- What format should portable data take? What model should it adopt?

The more intelligible data is at first glance the easier it will be to understand and reuse. However, be careful not to further rob the data of meaning by giving them a more intelligible form. For example, according to the INSEE coding system, six municipalities in France have the name Montréal, 26 communes are called Saint-Martin, etc. Also, be careful not to burden data sets by drowning them in explanations.

As noted earlier, some of the data that an organization uses for a specific purpose internally can actually be used for a variety of purposes by others. In terms of the format and the model you choose to render personal data, two cases exist: **either there is one widely-used standard** (or several) for X type of data (the options Google offers for its [takeout](#) service are good examples of this); or, should no standardized format exist (yet), the simplest is probably **to stay as close to your internal model as possible** — even if that means making substantial changes down the road, when the predominant uses for these data or a standardized format emerges. In fact, although preexisting standards are not strictly necessary, establishing what they will be is a discussion you should have internally (or one you should join alongside others) as soon as you can, a) because having standards will help unlock the value of the data, and b) being part of the discussion will ensure that you do not have to submit to a standard imposed by competitors or new entrants.

Implementing the right to portability at your organization

Cross-departmental

Given the risk of damaging your organization's public image and (now) the threat of incurring financial penalties, implementing a portability initiative at your organization will require the close collaboration of several departments: legal (including a Data Protection Officer, or DPO), IT systems, data governance, and top management.

To transform the right to portability into an opportunity for value creation, securing the input of the professionals who process and use these data on a daily basis will also be important. Often they are the only ones to have accurate, up to date knowledge regarding the data, and the only ones able to maintain documentation and answer questions about the data. Communications will need to be mobilized both internally (to publicize the approach taken by the initiative and help drive it forward) and externally to enhance the organization's approach and reap the benefits for your efforts in the form of satisfied, loyal users.

Open externally

The right to portability requires organizations to reveal externally what has been kept behind closed (IT system) doors until now. On the one hand, this means **exposing** internal elements of your organization's operations, for example by translating internal naming systems.

Above all, however, it means **working alongside** third-party service providers and other data reusers as soon as possible. This will indeed be the fastest way for your organization to unlock the

value of these data, discover their potential uses, promote the development of future new services and, ultimately, experience the benefits offered by access to data — be it personal data, public data, or private sector data from other domains or your competitors — in ways that spur innovation.

Clear, concise internal communications

As we have seen, ensuring the right to portability will require collaboration among company stakeholders who have a variety of skill sets. Beyond this, providing data portability seems to go against the most basic principles that have long served internal company operations. *'Releasing' 'valuable', 'complex', 'sensitive' personal data to data subjects 'who don't even know the data exist', 'don't want to know' and 'won't even be able to do anything with them except steal them or give them away blithely to our direct competitor: it's heresy!'*

Beyond this initial reaction, the issue of portability will raise legitimate questions from every collaborator involved in the project. The risks of implementing portability are easy to discern: they are the risks that organizations run every day -- security risks, among others -- only more complex. And as this approach is a radical departure from long-standing internal protocols, the risk seems personal, individual.

On the other hand, the opportunities that portability opens up will have an effect over the longer term and will have overarching consequences for the organization as well (and also each individual, the innovation ecosystem, ...). In any case, these developments will not seem exclusively individual.

How your organization ensures the right to portability is a choice, a new direction and a **long-term investment**, and must be **communicated as such**. The virtuous and societal aspect of the initiative should not be neglected, considering that each employee is also a data subject!

All these factors have shaped the **portability initiative framework** that follows. This proposal was developed in collaboration with stakeholders from large organizations, and based on the findings of field research carried out by the MesInfos team: MesInfos has been inviting organizations to return personal data to project participants since 2013, and is now conducting a pilot research project that began in 2016.



Reaching out to actors outside the organization

Training, coaching, assistance ³



Reaching out to actors outside the organization

A Venn diagram consisting of two overlapping circles. The left circle is labeled "3 Months" and the right circle is labeled "2 Weeks". The overlapping area in the center is shaded a darker blue than the non-overlapping areas.

Communication	3rd party services
Users	Top Management
Business Units	Data governance
IT Opt.	Legal, DPO

<http://infolabs.io/schema-portability>

1. Planning

In detail

Actors

Top Management	3rd party services	Communication
Business Units	Users	3rd party services
Data governance	Top Management	Users
IT Dpt.	Business Units	Business Units
Legal, DPO	Data governance	Data governance
	IT Dpt.	IT Dpt.
	Legal, DPO	Legal, DPO

Introducing data portability

MEETING

#risk
#opportunity
#schedule

Submitting a list of eligible data

MEETING

#mapping of personal data
#eligibility
#mapping of internal silos
#mapping of subcontractors

Analysing the list > Validation

PROJECT

#cost
#risk
#opportunity
#quality
#obsolescence
#scope of data according to applicant categories

Selecting the data transfer mechanisms

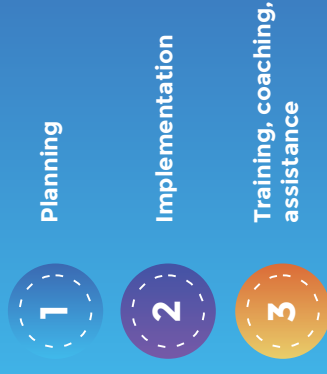
PROJECT

#direct download
#transmission to third party services
#privacy levels
#user interactions

IN-HOUSE PRESENTATION OF THE OPPORTUNITY

Opportunity & risks for each party - schedule

Steps



Duration



Reaching out to actors outside the organization

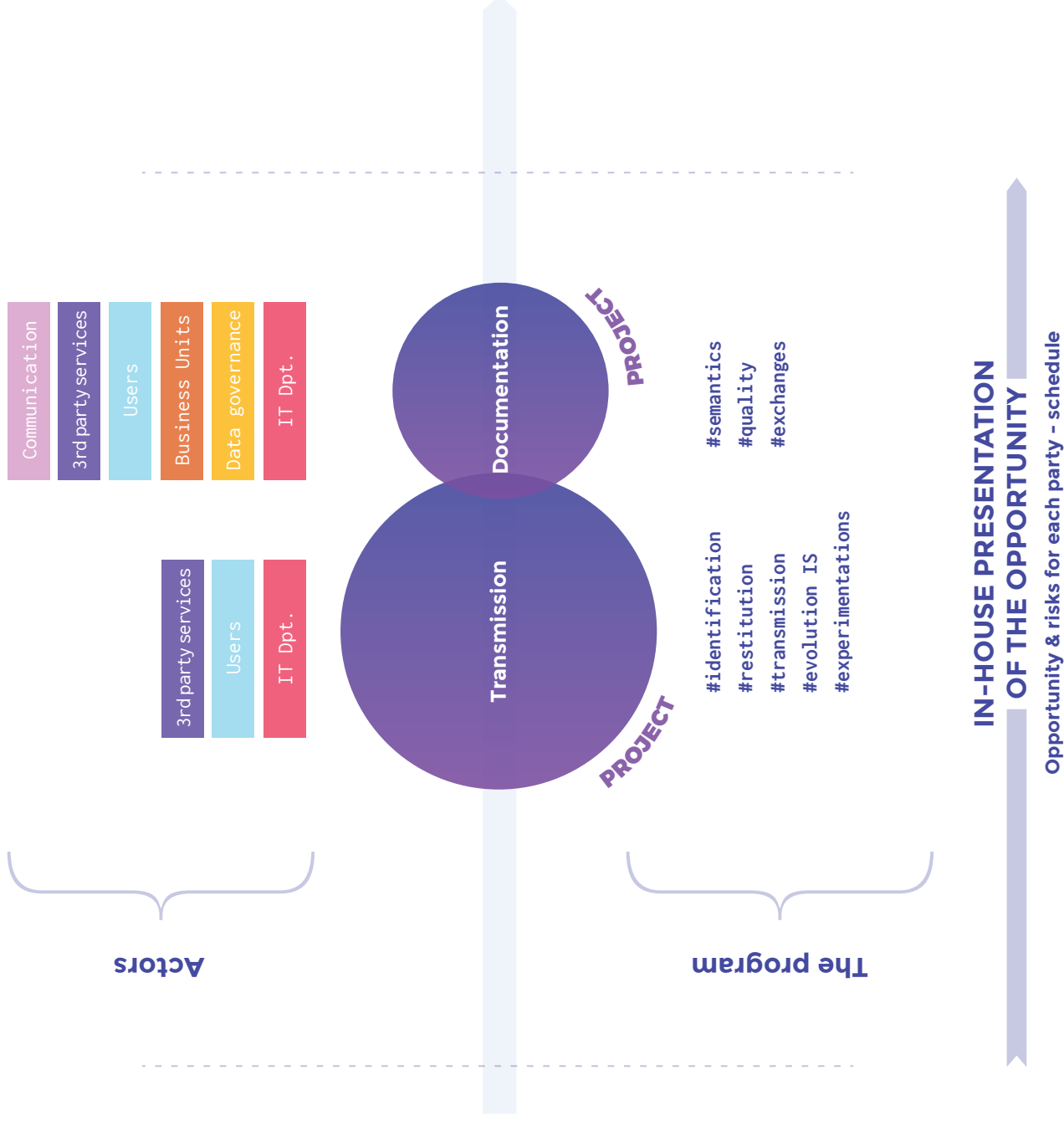


Licence



In detail

2. Implementation



Steps

- 1 Planning**
- 2 Implementation**
- 3 Training, coaching, assistance**

Duration



Reaching out to actors outside the organization

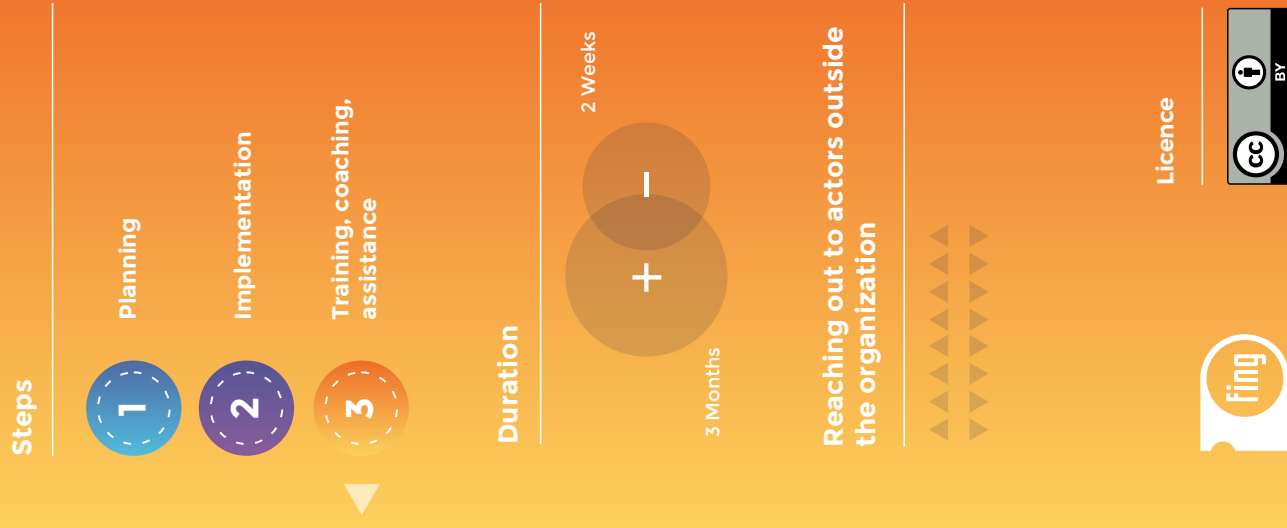
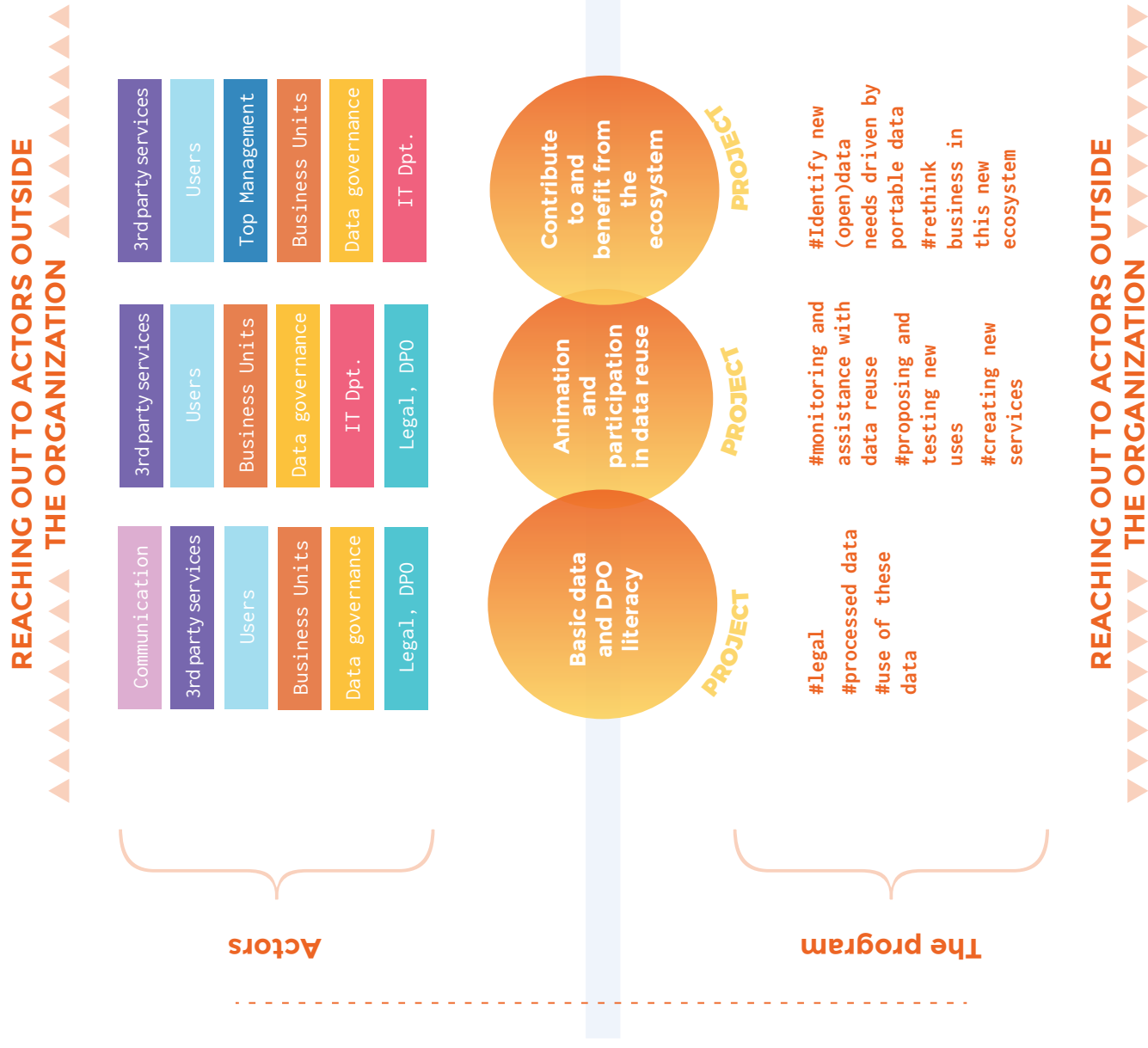


Licence



3. Training, coaching assistance

In detail



NOTEBOOK #4

A practical guide to data portability

<http://infolabs.io/dp17>



This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0).

This notebook was co-produced by Fing's Infolab Campaign and MesInfos project



and supported by MesInfos and Parlez-vous Data project partners



Fing receives support from its key partners

