

Dataaccess



Data-responsible Enterprises

Les entreprises Dataresponsables

User Experience and Technical Specifications

V1 – February, 2018

A project led by



Since 2012, the MesInfos Project, led by Fing, explores and implements the Self Data concept in France: what would happen if organisations that gather personal data actually shared those datasets with the individuals concerned?
<http://mesinfos.fing.org/english/>

Licence





Content

Overview	1
Scenario 1: Download my Data	13
Scenario 2: Connect my Data	17
Technical Specifications for the "Connect" Scenario	22
Back-Office and Consent Management	27
Appendix 1: How to Select Which Data Are Portable	29
Appendix 2: Putting Data Portability Into Practice	32
Appendix 3: Backlog	37

1 Overview

Dataaccess project

V1 - February, 2018





Dataaccess project

Inspired by the US's "Green" and "Blue" Buttons + GDPR's Portability right.

Strategic goals :

To provide a common framework for **enabling personal data portability and turning it into an opportunity for innovation and customer relationship.**

Operational goals :

To **decrease the complexity** for companies through common specifications, guidelines and design,

To **ensure an efficient deployment** of this new right, making it usable by end-users, providing a common and consistent experience from one service to another, and preventing misuse or abuse,

And consequently, to **make it an opportunity** for innovation, customer relationship and trust.



2017 Objectives and Partners

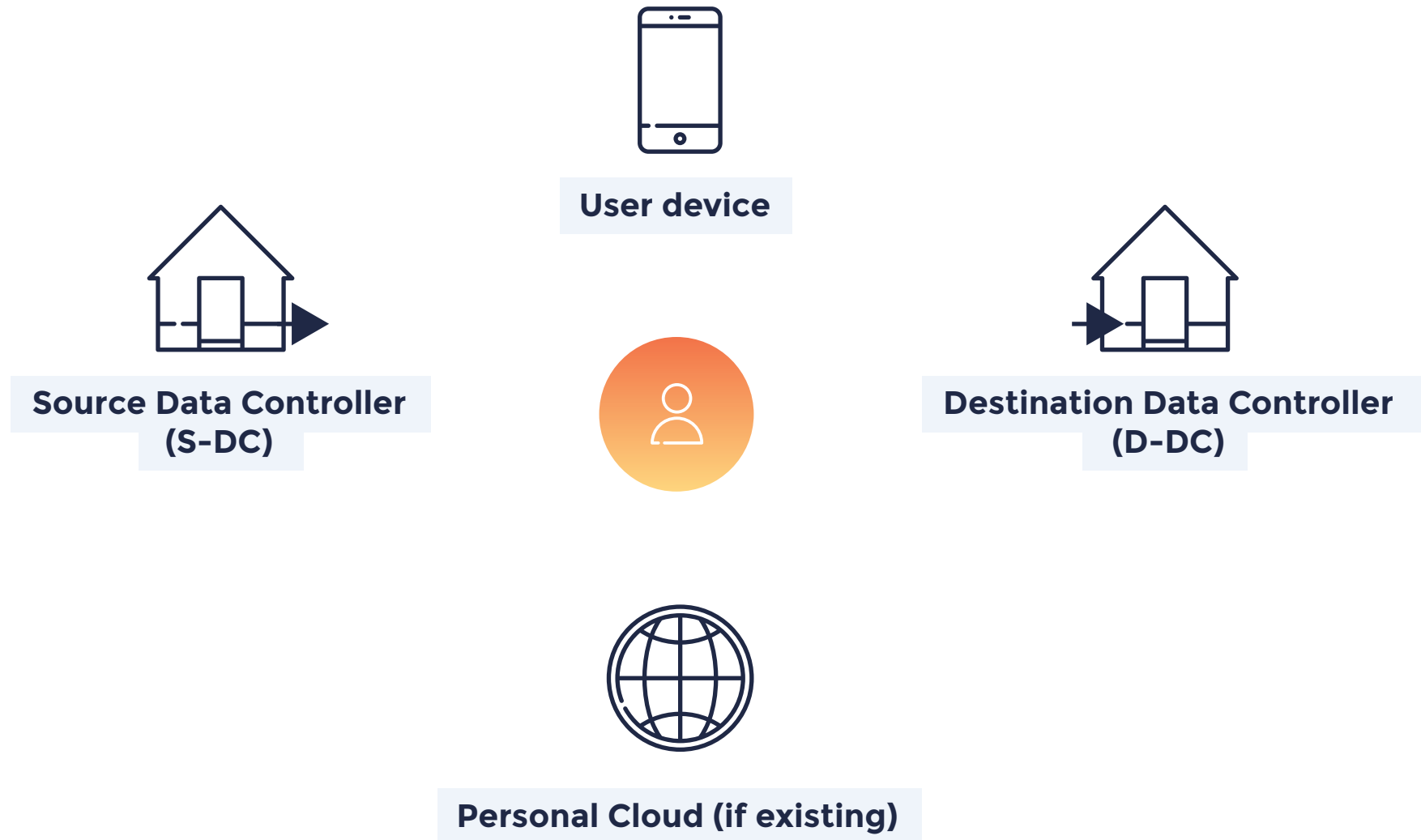
Goals:

- Design and specifications, from the point of view of users as well as of data controllers
- Prototyping and user-testing with 2 data controllers: Enedis (done), GRDF (in preparation)
- Publication of the specs, design elements, test conclusions, and other relevant material (will remain opensource)
- Choice of a name: Dataaccess
- Initial thoughts on governance (who can use the name and logo, specs review process, directories – if relevant –, etc.)

Partners:

- A project led by MesInfos and Orange
- An open working group of involved parties
 - ▶ Large data controllers: Enedis, Engie, GRDF, Maif, Mgen, Credit Cooperatif
 - ▶ Local communities: Lyon
 - ▶ Cnil (France's DPO)

Scenarios

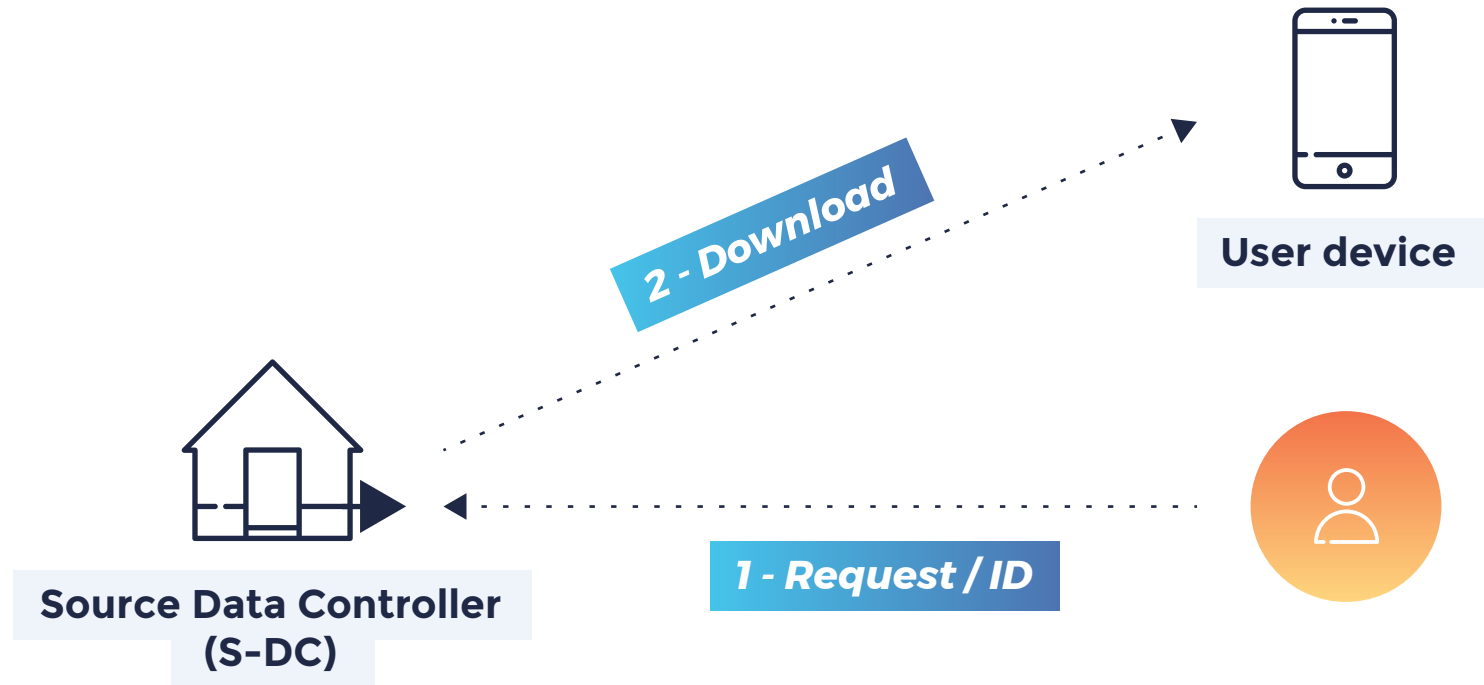


● Scenario 1

Download



Scenarios 1 - Download



● Scenario 2

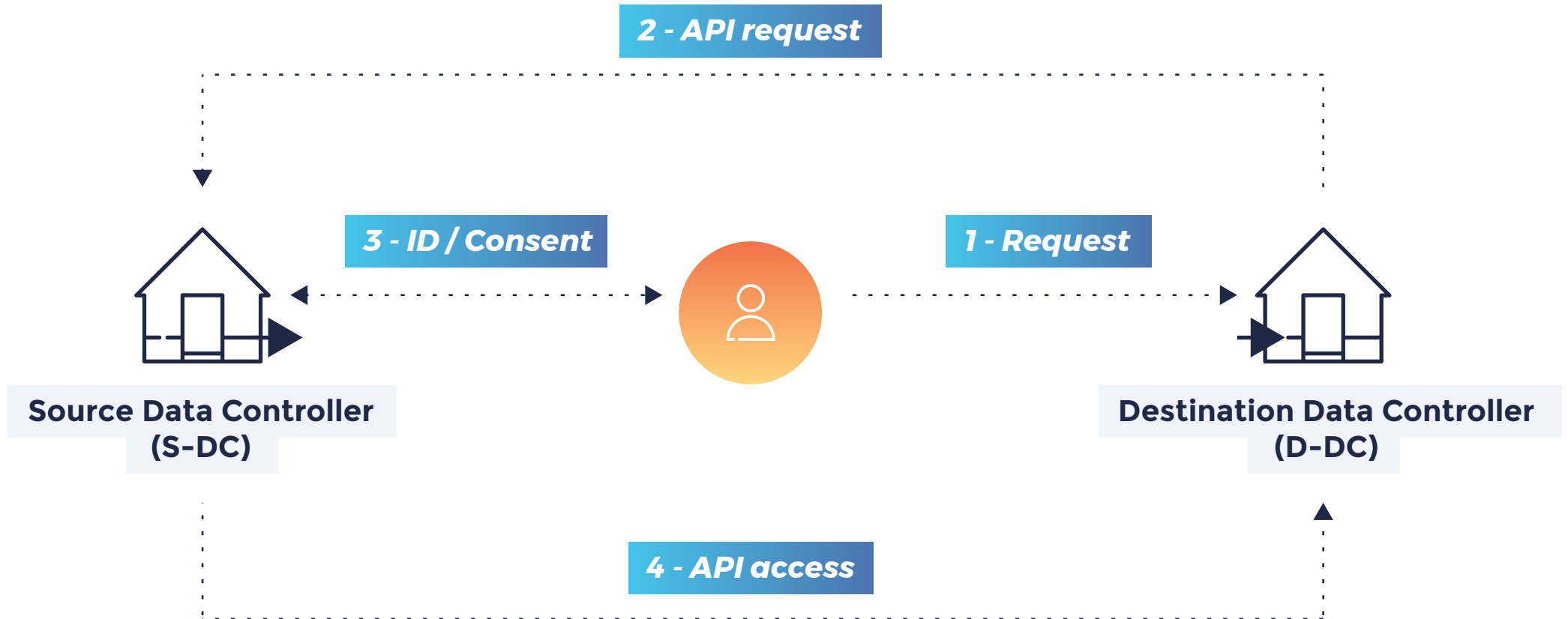
Connect

Likely to be the highest-volume scenario



Scenarios 2 - Connect

Likely to be the highest-volume scenario



● Scenario 3

Connect to Cloud

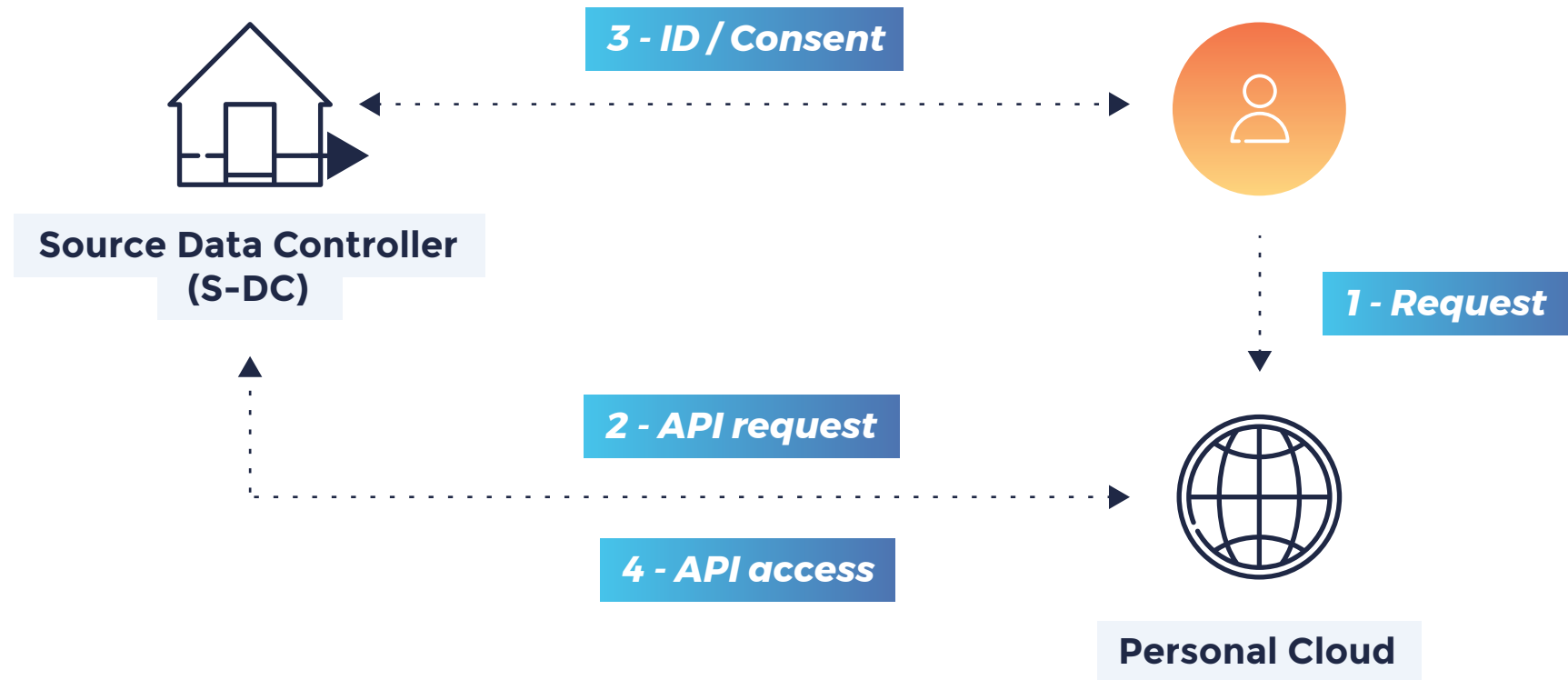
**Not specified as yet – see MesInfos project
for more details**

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



Scenarios 3 - Connect to Cloud

Not specified as yet - see MesInfos project for more details (<http://mesinfos.fing.org/english/>)





Common Choices and Specifications (valid for all scenarios)

- Each Source Data Controller (S-DC) decides which data are portable (scope) and organizes portable data into coherent subsets (e.g., customer coordinates, customer profile, contract, transactions, specific use category or service category...)
- User authentication must be dealt with, either by the S-DC, or by a commonly accepted and secure federated ID system.
- The interaction with the user must make him/her understand that the S-DC can not be held responsible for the future usage of his/her downloaded or transmitted data.



Name, Baseline, Logo

A creative workshop and a subsequent selection process has produced:

- The "values" that should be embedded in the name: Transparency, Innovation, Ease of access
- A "brand" name: Dataaccess
- A baseline: "Les entreprises dataresponsables" - rough translation: "Data-responsible enterprises"

Initial work on logo



Scenario 1

2 Download My Data

Data portability from a Source Data Controller (S-DC) to an end-user device

V1 - February, 2018



S1.1. General specifications for the “Download My Data” scenario

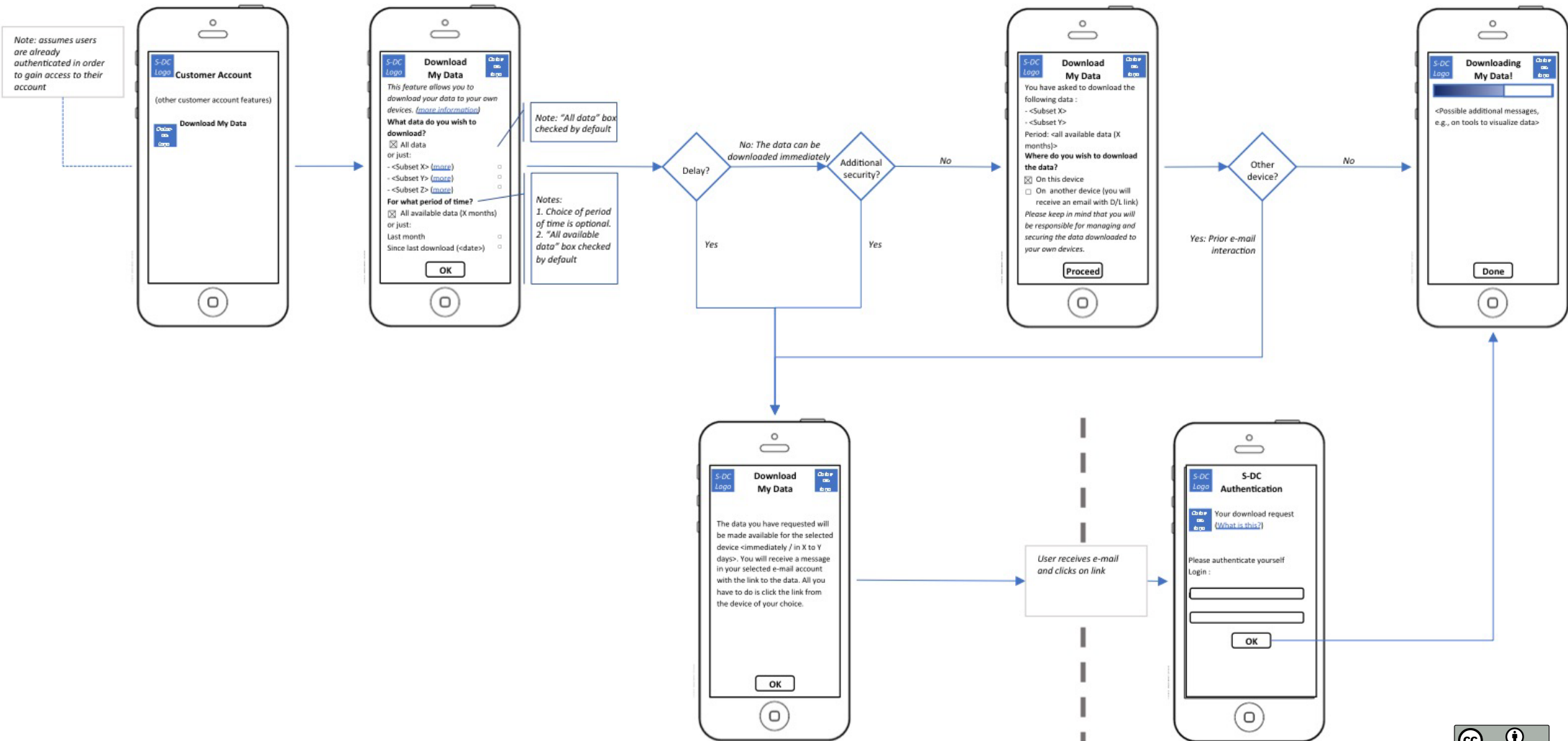
- Portability is initiated by the user from his/her customer account (thus ensuring user authentication). It provides a feature to download data into a structured, commonly used and readable format.
- Portability requests will apply to the whole set of "portable" data (by default), or to specific subsets (if the user chooses so).
- By default, the user will download all the available data within each subset available in the S-DC's active databases (excluding archives). However, several options may be offered:
 - ▶ All data / X months of data / all new data since last download (if applicable)
- The data can be made available in 2 different ways
 - ▶ Download starts immediately after user confirmation.
 - ▶ A one-time link to a download URL is provided via email to the user's registered address. Clicking the link will bring the user to an authentication page, and then to the download page. This option can be activated in 3 cases:
 - As an additional security measure decided by the S-DC. In this case, the email will be sent immediately after the user's confirmation.
 - If the user wished to download the data on a different device (typically: request made on mobile, download to PC). In this case, the email will be sent immediately after the user's confirmation.
 - If all or some of the data can not be made available in real time by the S-DC. In this case, the email will be sent when the data is available. Note: It is impractical for users to receive data in several batches. Therefore, even if only some of the requested data requires a delay (or if different subsets require different delays), the link will only be provided when all the requested data is made available.

S1.2. Recommended additional specifications

- The "Download my data" option should be made prominently visible in users' online customer accounts
 - ▶ If user account management includes a specific section on personal data and privacy, the "Download" option should be present in this section.
 - ▶ The option should be labelled "Download my Data" rather than "Portability". It can feature the Dataaccess logo.
- It is highly desirable to provide users with a way to understand and visualize their data.
 - ▶ This feature would:
 - Allow users to understand why the data were collected or produced in the first place
 - Make users perceive Dataaccess portability as a service provided by the S-DC, rather than just a legal obligation for the S-DC
 - Additionally, this can be the basis upon which the S-DC could provide user-centered data-based services in the future (e.g., analyzing consumption patterns and providing recommendations)
 - ▶ Such a feature could take several forms:
 - Simple documentation: "How to read and understand your data"
 - Spreadsheet templates providing tables and graphs
 - Or an online service, provided within the S-DC's customer account space
 - ▶ At this stage, Dataaccess will not provide common specifications for such a feature. This could be one of the future evolutions of the Dataaccess specifications.

Scenario 1

S1.3. User Experience – “Download”



Scenario 2

Connect My Data

Data portability from a Source Data Controller (S-DC) to a 3rd-party service (Destination Data Controller, D-DC)

V1 - February, 2018



S2.1. Conditions for access by 3rd party applications to S-DC portable data

- Some kind of contract may need to be signed between S-DC and D-DC
 - Commitments towards process, data protection, reasonable use of S-DC's IT infrastructure, etc.
 - Can it include monetary dimensions (eg, when request place a heavy load on S-DC's IT)?
- Possible (optional) conditions for access:

Qualification of D-DC	Data subset sensitivity	Result
Blacklisted (by S-DC or through a common blacklist; requires criteria, due process and motivation of decision)	Indifferent	No access
Unknown by S-DC	0-2	S-DC possible decisions depending on data subset sensitivity: • Access denied for lvl. 1 or 2 sensitive data (subject to legal verification) • Access granted with various levels of confirmation and disclaimer: - Lvl. 1: simple acknowledgement and disclaimer - Lvl. 2: acknowledgement + warning to the user on risks taken - Lvl. 3: acknowledgement + warning + request for reconfirmation
Known and qualified by S-DC	0-2	S-DC possible decisions depending on data subset sensitivity: • Access granted with various levels of confirmation and disclaimer

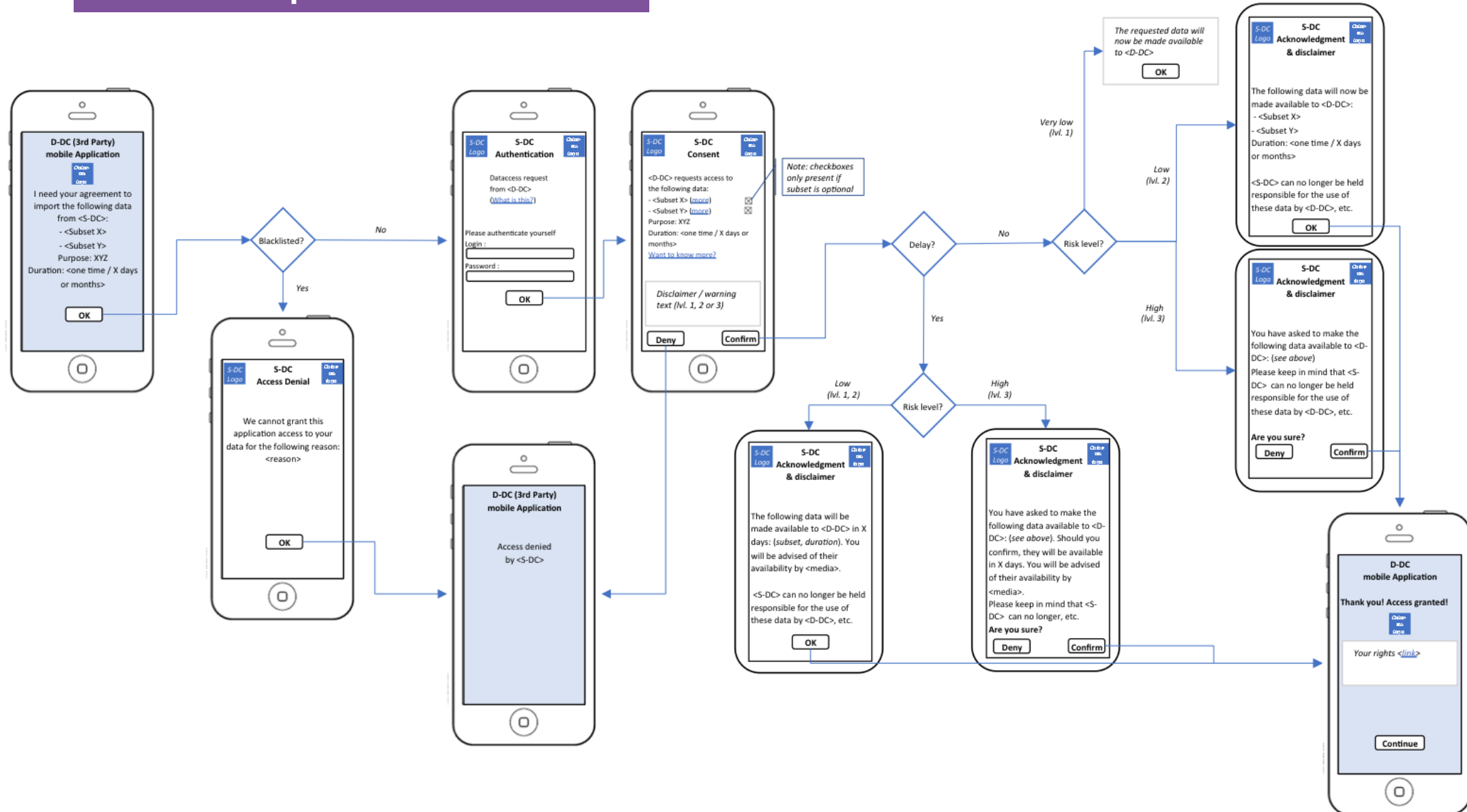
- However, need to remember that portability is a right that can not be restricted for reasons just based on the S-DC's commercial interest or goodwill
- A directory (either common, or operated by the S-DC) should allow 3rd parties to know and understand:
 - ▶ What the data are, how they are organized in subsets, each subset's "privacy sensitivity level" (0-Not sensitive to 2-Very sensitive)
 - ▶ The URL to each subset's API
 - ▶ The optional fields that 3rd parties can add in the requests (mandatory or non-mandatory subset, purpose of the request)
 - ▶ Possible conditions for access (see below)
 - ▶ Availability delay (default: data are provided immediately; otherwise, delay indicated in days/hours)
 - ▶ Duration of access token for one-shot requests and for repeat requests (if authorized)

S2.2. User Experience – “Connect” – Overview and Options

- The experience starts with the Destination Data Controller (D-DC), typically a 3rd-party app or service that needs the S-DC's data in order to provide the service
- The user's authentication and consent dialogues happen on the S-DC's servers
- The experience ends with the D-DC, after:
 - ▶ The user has been correctly authenticated and has approved the request to transfer data (default)
 - ▶ The user has been correctly authenticated, has approved the request to transfer data, and has been told that the data would be made available via a one-time use link that will be emailed to him/her (either immediately, or in X days if the data require a delay before they are made available by the S-DC)
 - ▶ The process has been aborted for one of several reasons: authentication error, consent denial, or the D-DC is blacklisted
- The Dataaccess logo and graphical elements are present on both the D-DC's and the S-DC's screens in order to provide a sense of continuity in the experience. There are 2 possible options on the S-DC's side:
 - ▶ The UX is clearly branded by the S-DC, using their graphic charter and only adding a few elements from the Dataaccess charter
 - ▶ The UX is more "neutral", mainly branded by Dataaccess, with the S-DC's logo added to it.

Scenario 2

S2.3. User Experience – “Connect”



S2.4. User Experience – “Connect” – What Happens Next

1. S-DC sends email to user:

CASE # 1: IMMEDIATE DELIVERY

On dd/mm/yyyy at hh:mm, upon your request, we have made the following data available to <D-DC>;

- <Subset X>*
- <Subset Y>*

Duration: <one time / X days or months>

<S-DC> can no longer be held responsible for the use of these data by <D-DC>, etc.

You can review and revoke authorizations given here.

CASE # 1: DELAYED DELIVERY

On dd/mm/yyyy at hh:mm, upon your request, we will make the following data available to <D-DC> within X days:

- <Subset X>*
- <Subset Y>*

Duration: <one time / X days or months>

You will be advised of their availability by <media>. Please note that after they data have been made available, <S-DC> can no longer be held responsible for their use by <D-DC>, etc.

You can review and revoke authorizations given here.

2. On S-DC's server, customers can:

- Access all authorizations given and revoke ongoing authorizations
- Access information on risks, rights, etc.

3. S-DC's customer service must be trained to answer requests and complaints regarding portability, and have access to authorizations given by customers (plus, if possible, have the power of revoking them).

4 Technical Specifications for the “Connect” Scenario

V1 – February, 2018



Requirements

The connect scenario needs a solution able to:

- register and authenticate D-DCs
- allow the D-DC to ask a specific scope of data to the S-DC
- present this scope of data to the data-subject
- restrict D-DC access to this scope only
- present the finality of the D-DC’s request and ask the data-subject the consent about this
- identify and authenticate the data-subject.

OAuth 2.0 protocol implementation

The OAuth 2.0 protocol is a well-known and widely used protocol on the web, by billions of users several times a day.

It can manage almost all requirements:

- register and authenticate D-DCs
- allow the D-DC to ask a specific scope of data to the S-DC
- present this scope of data to the data-subject
- restrict D-DC access to this scope only
- present the finality of the D-DC's request and ask the data-subject the consent about this
- identify and authenticate the data-subject.

That is why we recommend the use of OAuth 2 to implement the Dataaccess "Connect" scenario, and describe how to add the missing feature.

Added feature to the typical OAuth scenario

- The D-DC must present data scopes, purpose and duration before starting the OAuth process.
- The consent screen should handle optional scopes. Technical suggestion: display all scopes as optional, and flag the required scopes, suffixing them with a "--required" flag.
- In case of delayed access to data: provide accurate availability delay. Prepare a media (email, through the D-DC, ...) to advise the data subject of expected/effective data availability.
- Send confirmation email on user consent.

Access duration: token lifetime

Limit the access duration by default sounds safe. But it should not damage the user experience. Some balanced examples:

- Ask the user to renew his/her consent at list every 1 or 2 year.
- Access_token: the lifetime should be sufficiently long to get the answer of several calls done consecutively. The lifetime can be between 5 and 15 minutes.
- Refresh_token: in the case of an asynchronous or recurrent portability, the refresh_token must be sufficiently long to enable a refresh of the access_token prior to calling the portability APIs after a portability initialization has been done. The lifetime of the refresh_token can be between 1 month to 1 year.

5 Back-office and Consent Management

What S-DCs should develop beyond authentication and consent

V1 - February, 2018



Back-office features (S-DC)

- BO1. Activity Monitoring
- BO2. D-DC contract management
- BO3. Consent reception, storage and back-office management (eg, managing end dates)
- BO4. Portability receipt: the S-DC sends its customer an email receipt of the consent he/she just gave. The receipt should contain his identity, the data shared, the timestamp of the consent, the duration of the consent (if applicable), the identity of the D-DC, the way the consumer can access the shared data, and the way he can revoke his consent
- BO5. Consent access and management by users: A space where customers can access the consents they have given and revoke them
- BO6. Troubleshooting and ticket management.

6 Appendix 1: How to Select Which Data are Portable

The following contains recommendations. The Dataaccess specifications do not mandate a specific scope for portable data.

V1 – February, 2018

WG29's 5 questions (5 "yes" means the data is portable)

Source: http://ec.europa.eu/information_society/newsroom/image/document/2016-51/wp242_en_40852.pdf

1. Is it personal data concerning the data subject?

- Yes, but: the data-subject is not the contract holder
- Yes, but: the service is used by many data-subjects at a time

2. Is the data processing carried out by automated means?

- No: I handle my list of client in my paper notebook (example)
- Yes: all other cases

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

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

4. Is it data "provided" by the data subject?

Explicitly provided data, transaction data, observed data= Yes Inferred data, e.g. profiling, recommendation... = No
*"Given the policy objectives of the right to data portability, the term "provided by the data subject" must be interpreted broadly, and only to exclude "inferred data" and "derived data". A data controller can exclude those inferred data but **should include all other personal data provided by the data subject through technical means provided by the controller**. Thus, the terms "provided by" includes personal data that relate to the data subject activity or result from the observation of an individual's behaviour but not subsequent analysis of that behaviour."*

5. Would the portability NOT adversely affect the rights and freedoms of others?

Data model and format

The Data must be provided in a "structured, commonly used and machine-readable format".

- If there is a standard widely used for this kind of data: use it!
 - ▶ vCard for contact, iCalendar, mbox for emails, ofx for bank accounts ...
 - ▶ Industry-specific common formats, etc.
- There isn't (yet) a standard: make things as easy as possible for users!
 - ▶ use commonly used structured text formats (as json)
 - ▶ datatypes should follow standards (iso8601 for dates, iso3166 for countries, ...).
 - ▶ document the data and make the documentation easily accessible online
- A data controller can provide different formats for the same data.

Appendix 2: Putting Data Portability into Practice

7

A step-by-step approach

V1 - February, 2018

You can learn more with the booklet "Data Portability in Practice" at http://infolabs.io/dp17_en (published in April 2018)



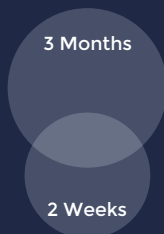


By the participants of the portability working group, from *Parlez-vous data?* and *MesInfos*, bringing together *Fing*, *Maif*, *Orange*, *la CNIL*, *Carrefour*, *GRDF*, under the direction of *Guillaume Jacquart*, *Charles Nepote* and *Armelle Gilliard*.

Reaching out to actors outside the organization



Duration



Actors

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

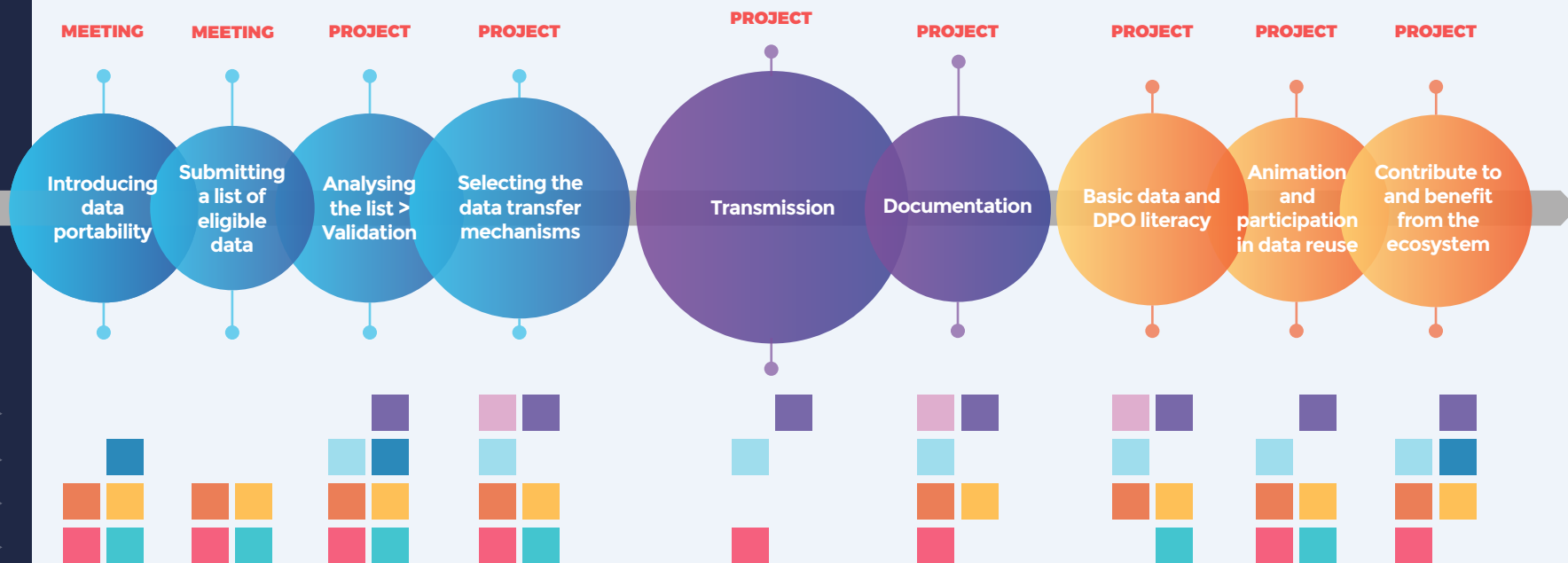
Licence



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

DATA PORTABILITY IN PRACTICE

Reaching out to actors outside the organization



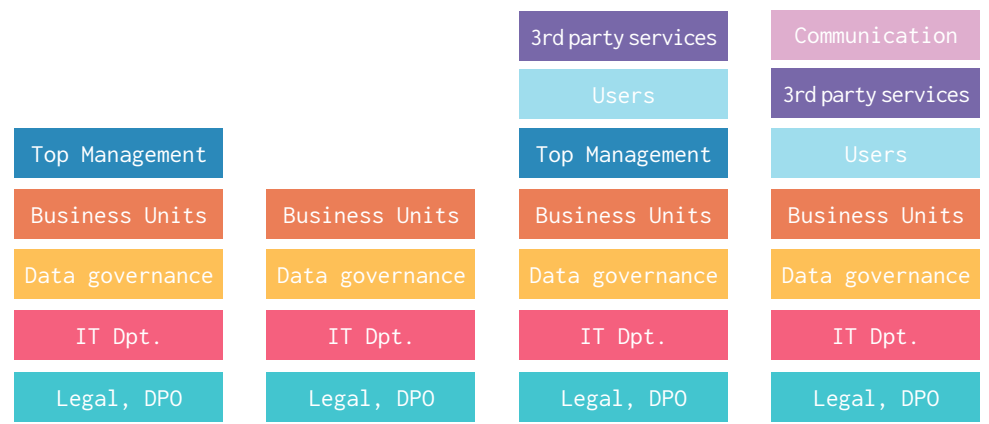
IN-HOUSE
PRESENTATION
OF THE OPPORTUNITY
opportunity & risks for each party - schedule

Reaching out to actors outside the organization

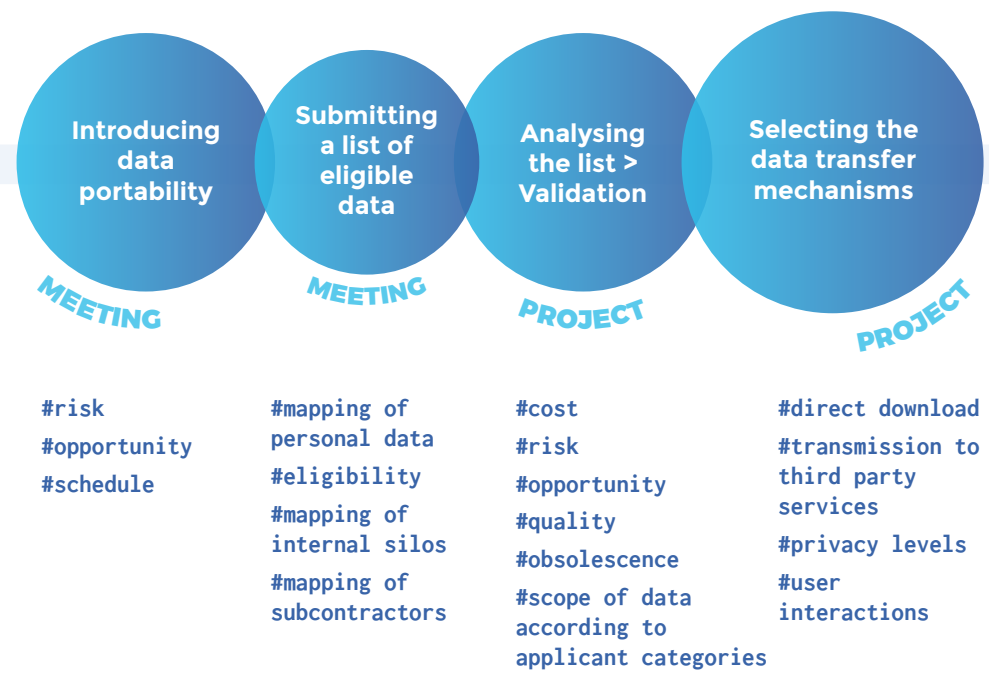
1. Planning

In detail

Actors



The program



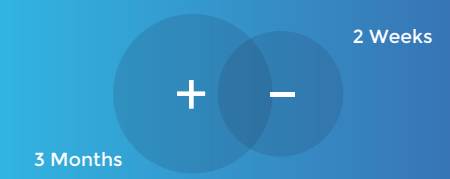
IN-HOUSE PRESENTATION OF THE OPPORTUNITY

Opportunity & risks for each party - schedule

Steps



Duration



Reaching out to actors outside the organization

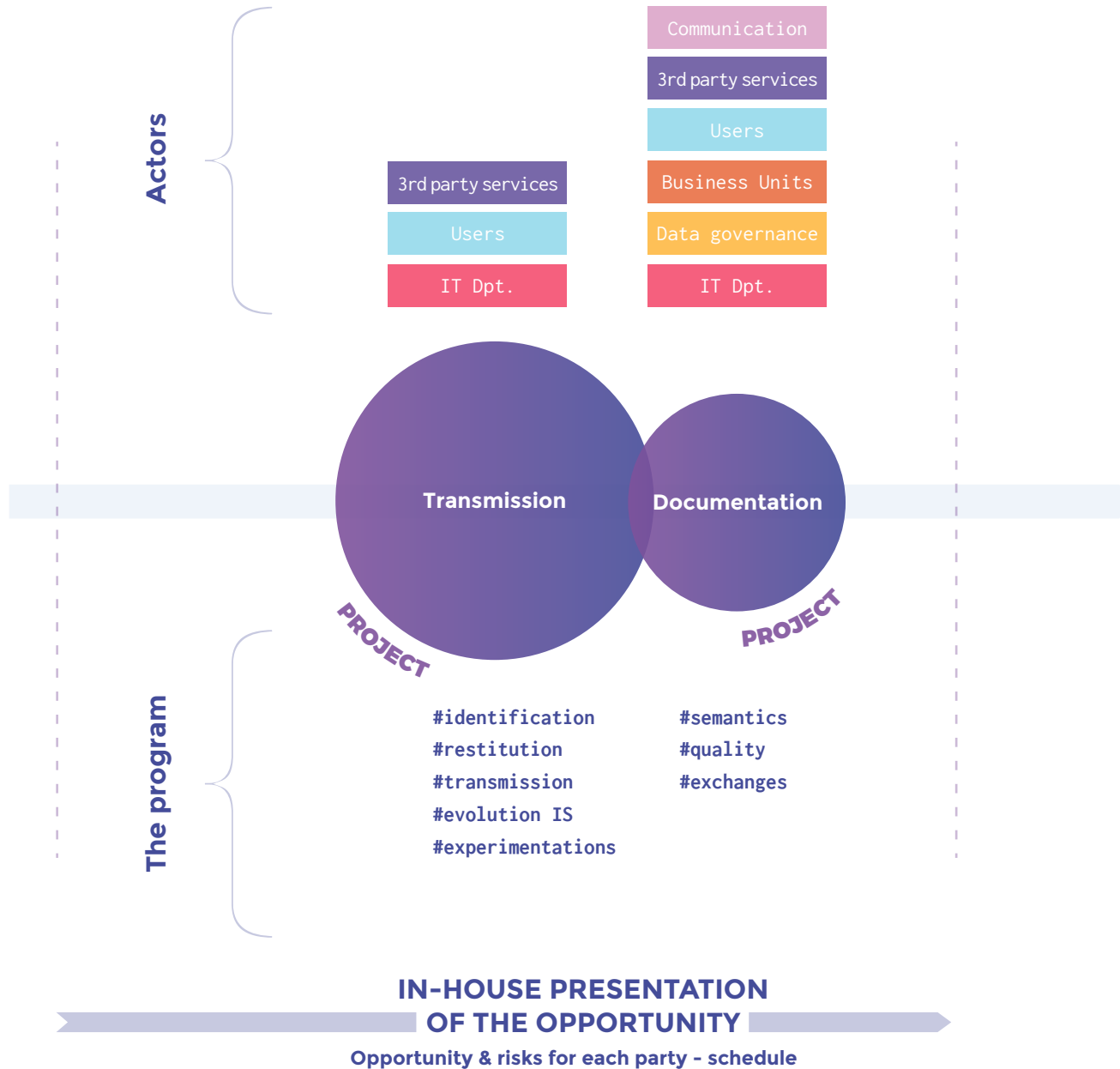


Licence



2. Implementation

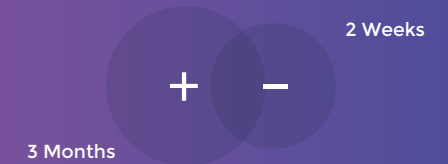
In detail



Steps

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

Duration



Reaching out to actors outside the organization

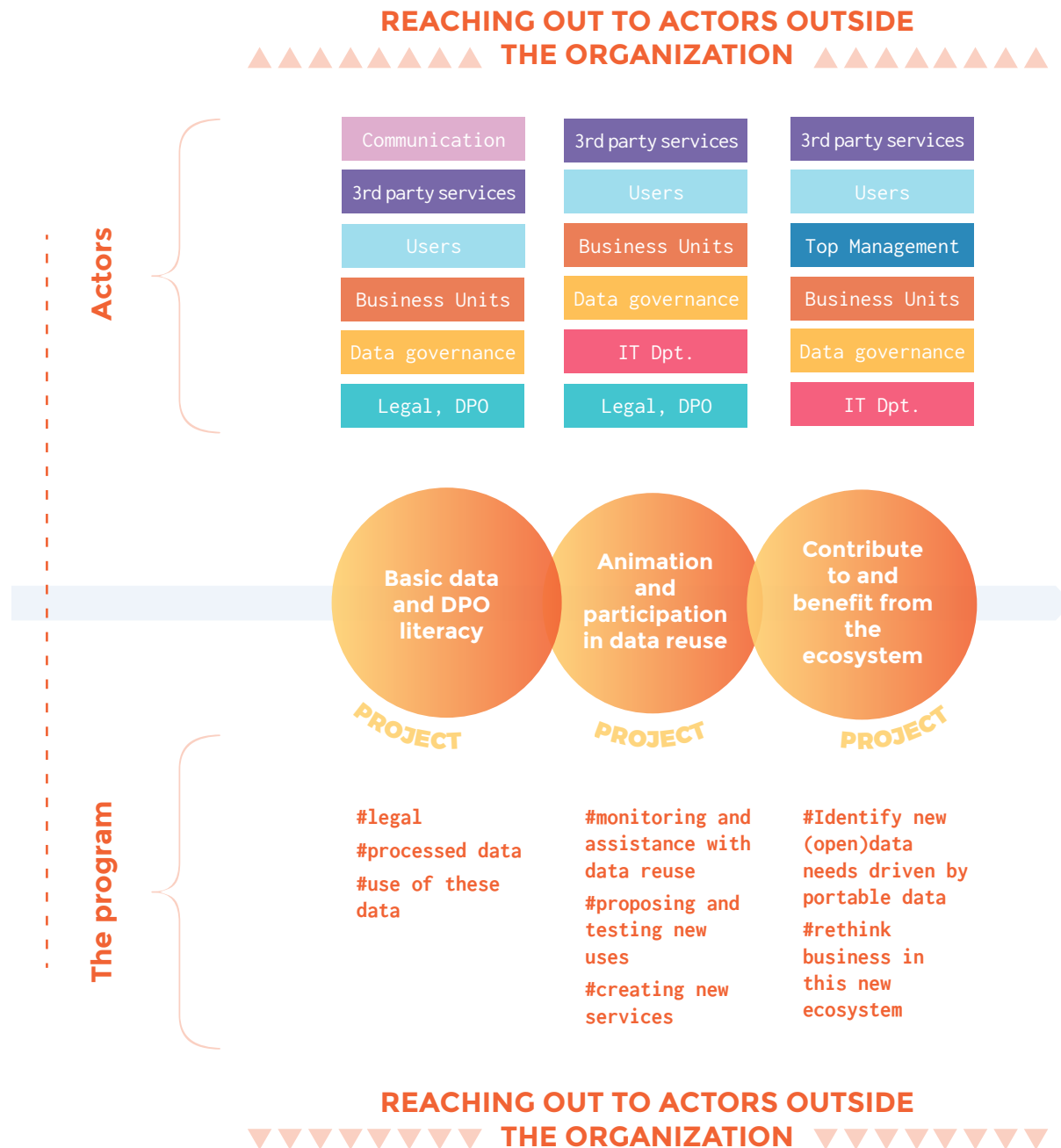


Licence



3. Training, coaching assistance

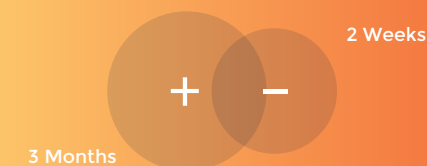
In detail



Steps

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

Duration



Reaching out to actors outside the organization



Licence



8 Appendix 3: “Backlog”

Issues that should be dealt with in future versions of the Dataaccess specs

V1 – February, 2018

BL1. Identification and authentication

BL1.1- Federated ID usage

- Use case: User has a certified ID with one or several federated ID system and wished to use it in order to authenticate him/herself with the S-DC
- Example: France Connect (English-language description), Estonia's ID Card...

BL1.2- D-DC managed identification

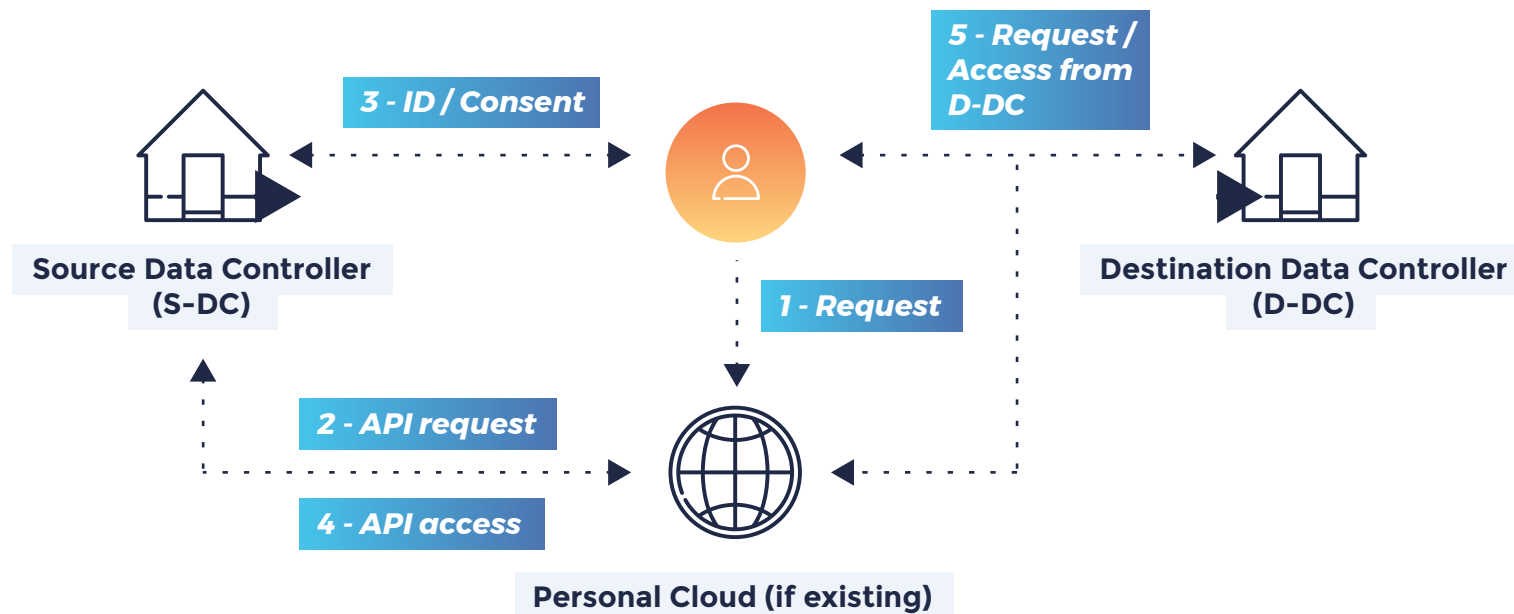
- Use case: User does not have an account (nor, sometimes, a conscious relationship) with the S-DC.
Examples:
 - ▶ Energy smart meters in markets where "distributors" (who transport energy and install the meters) differ from "suppliers" (who sell energy to consumers)
 - ▶ Cars and other data-producing objects that do not automatically come with a contract with a data controller
- The identification/authentication could be at least partially managed by the D-DC, who would handle the responsibility of establishing the link between the object and a person (« this individual is the rightful owner of this object and has rights to use/port the data »)
- Example: GRDF's "Tiers direct" authentication process

BL1.3- Collective vs. individual data

- Example: the contract is with a household, yet usage is individual (telecom...)
- Also for shared objects (e.g., cars)

BL2. The “Connect to cloud” (or “to PIMS”) scenario

- For the S-DC, the user's PIMS/Personal Cloud appears to be a kind of D-DC (and will definitely bear some legal responsibility over handling the data), however, the final added value on data will usually be provided by another party.
- This 4-party scenario has several implications:
 - Legal: How to express and share responsibilities? Is there any kind of relationship between the S-DC and the "final" D-DC?
 - Technical: May not easily fit existing standards such as OAuth, depending on PIMS implementation



BL3. The « Connect » scenario

- Scopes could be standardised
- Scopes could handle data perimeters, but also Read or Write access, and date ranges of access
- S-DC discovery: provide a directory of S-DC, that a user or D-DC can use select possible sources of data. The directory should contain url of privacy policies/terms, technical url of apis, keys, data documentation...



Data-responsible Enterprises

Les entreprises Dataresponsables

User Experience and Technical Specifications

V1 - February, 2018

A project led by



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

Licence

You can use this document as you please.
It is published online under a Creative Commons license,
and is thus freely available for use and reuse (you just need to add the mention
"Based on the document Dataaccess - Data-responsible Enterprises - Fing-CC by 3.0")

