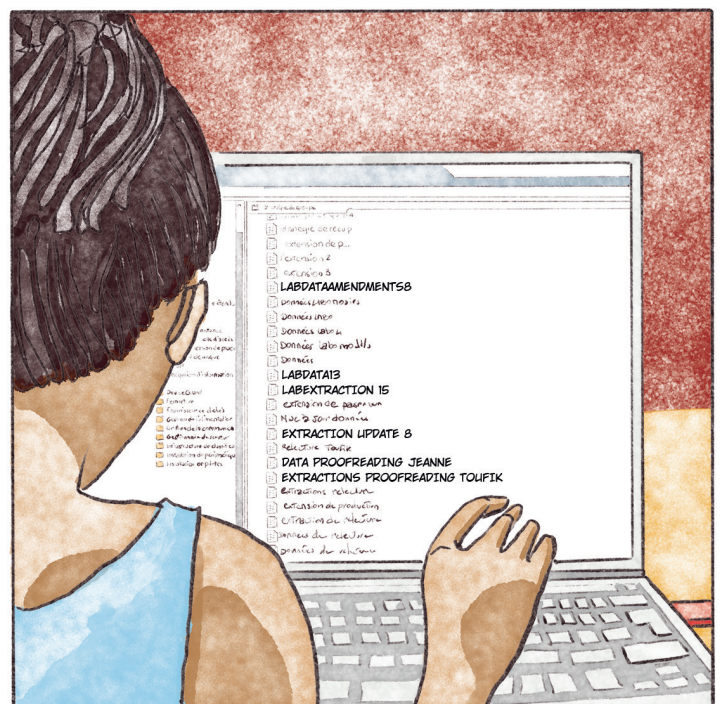


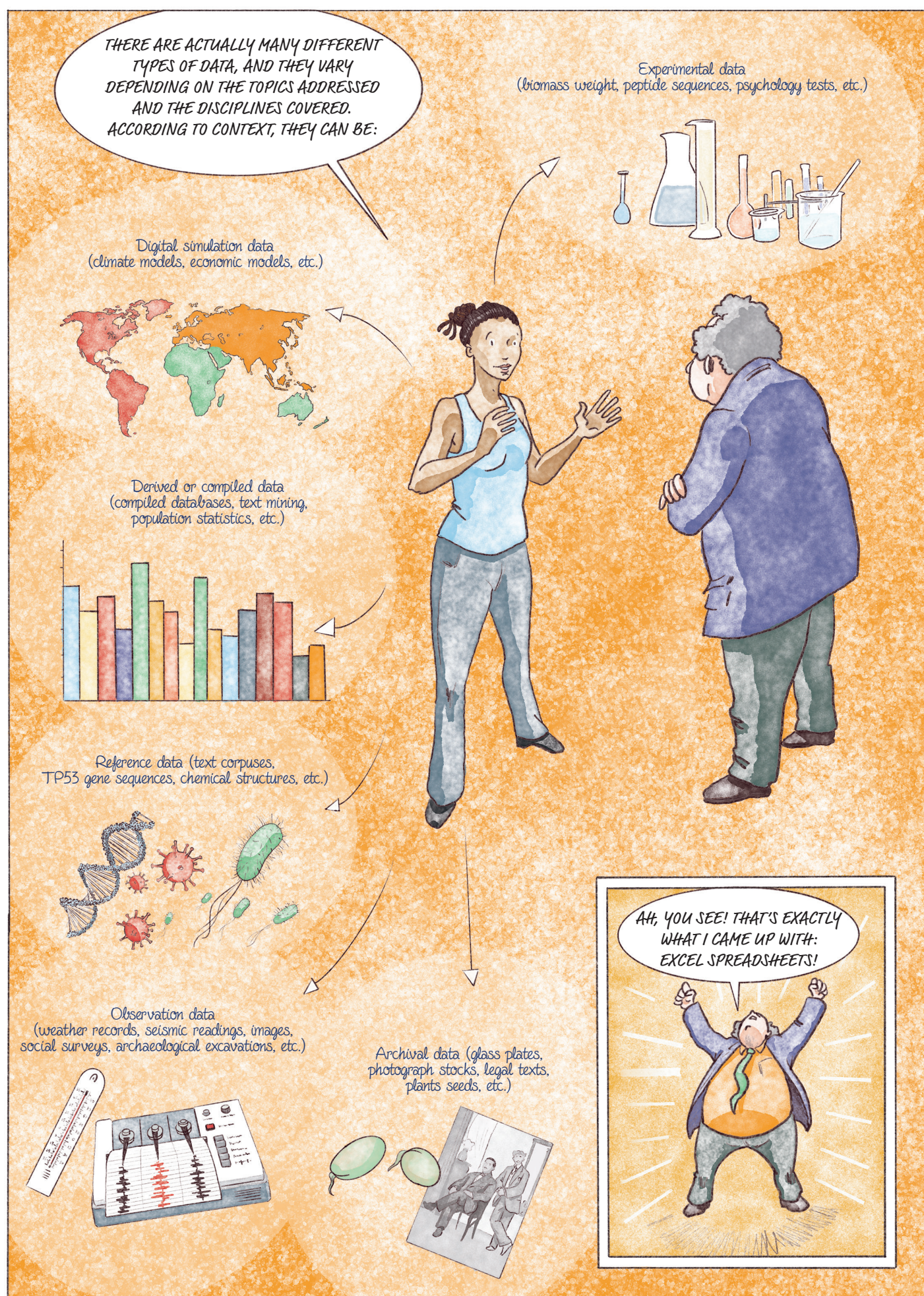
LET'S TAKE STOCK of **research data** with **Sorella!**

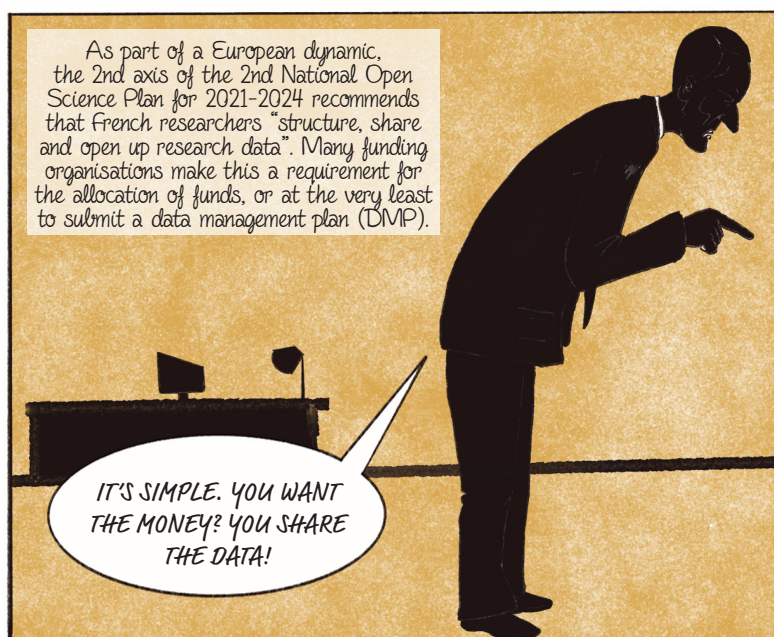
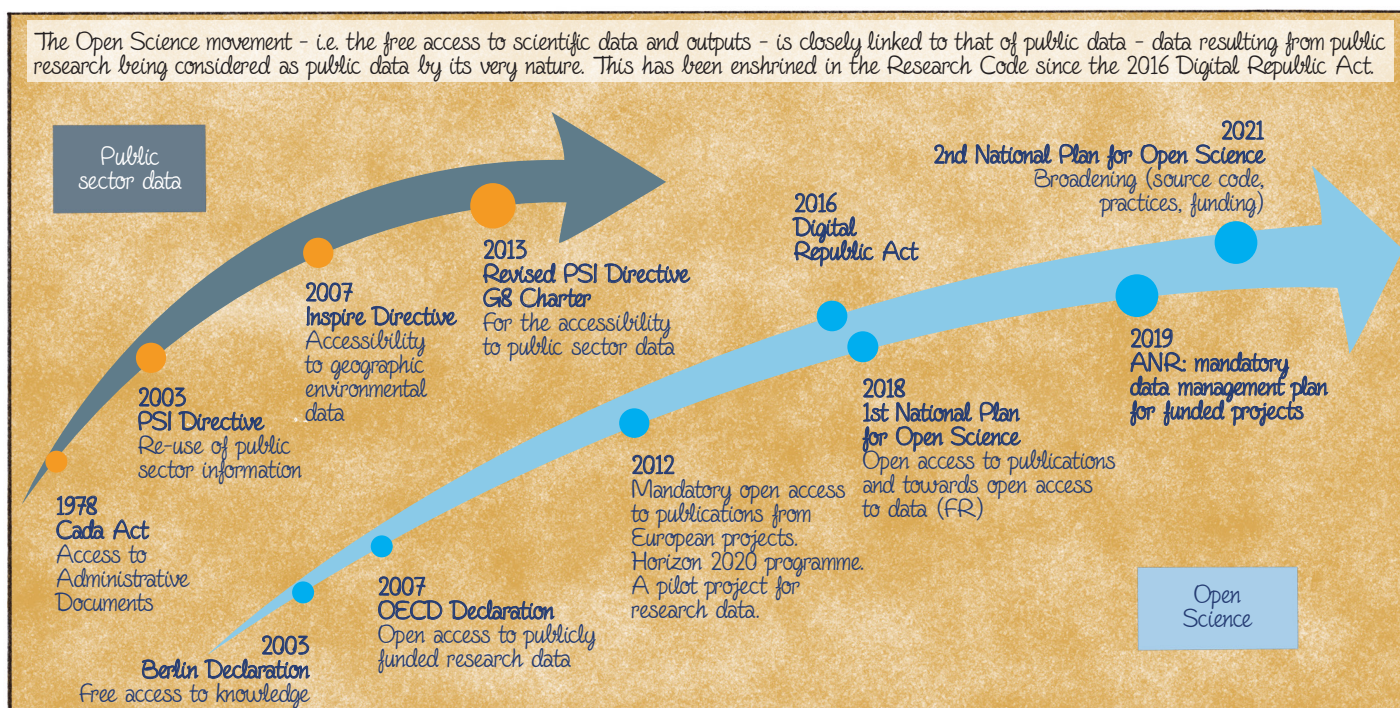
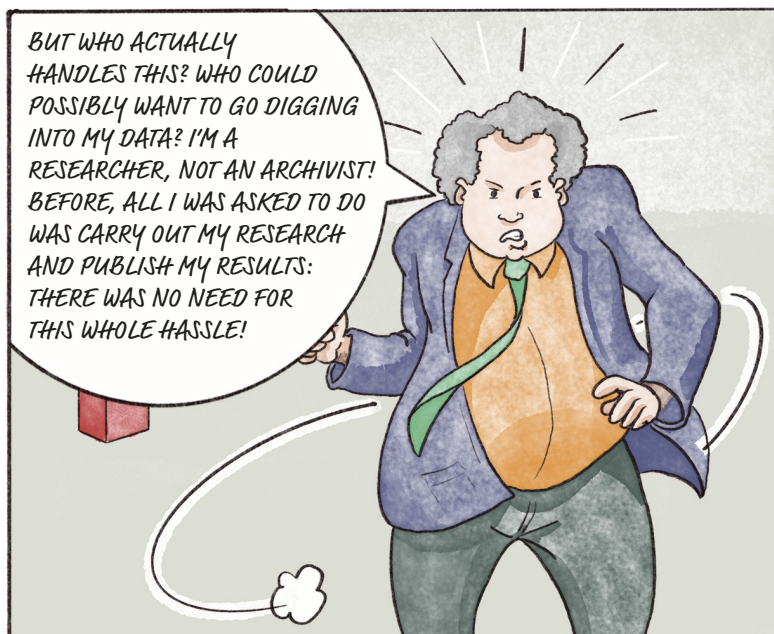


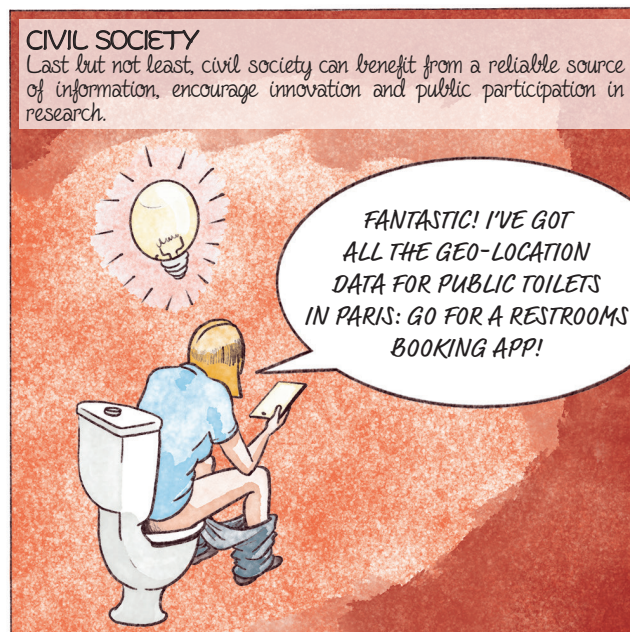
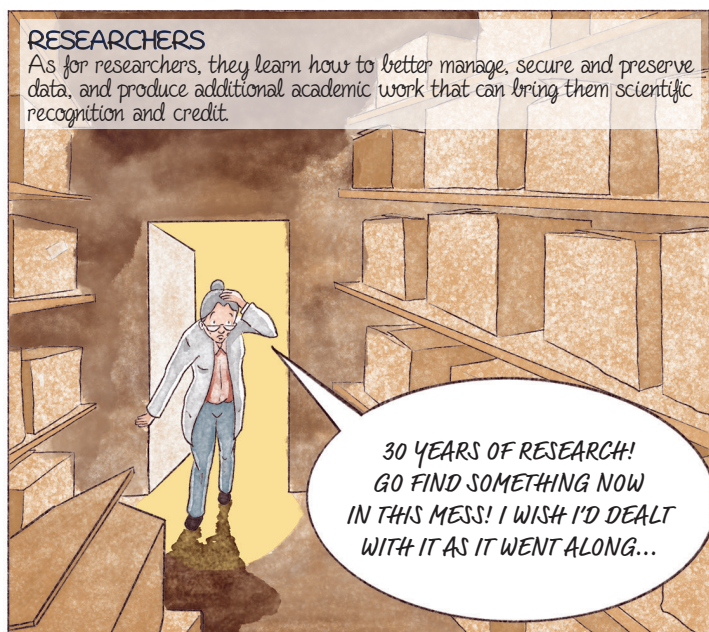
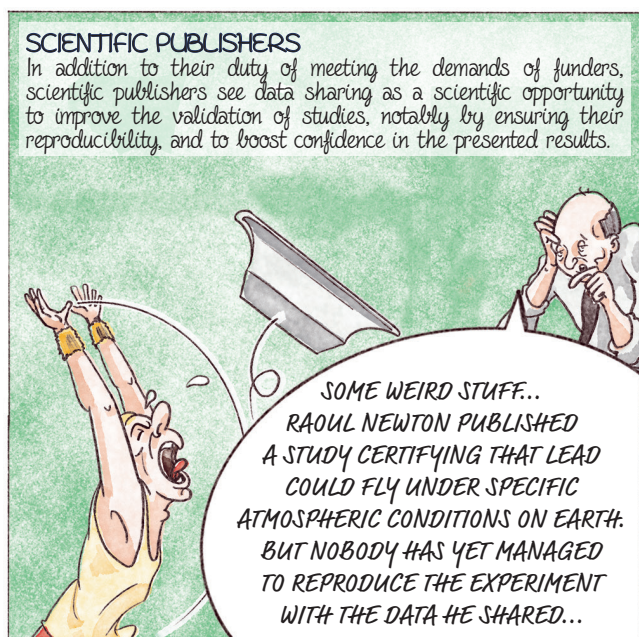
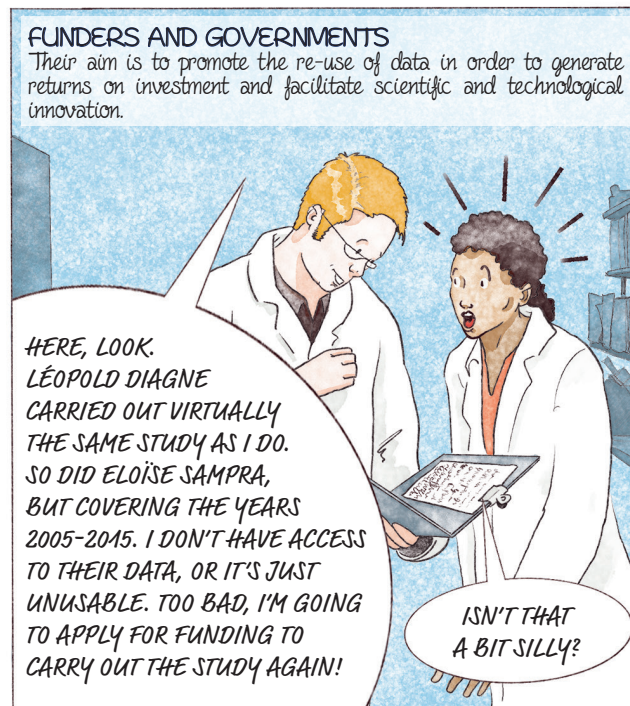
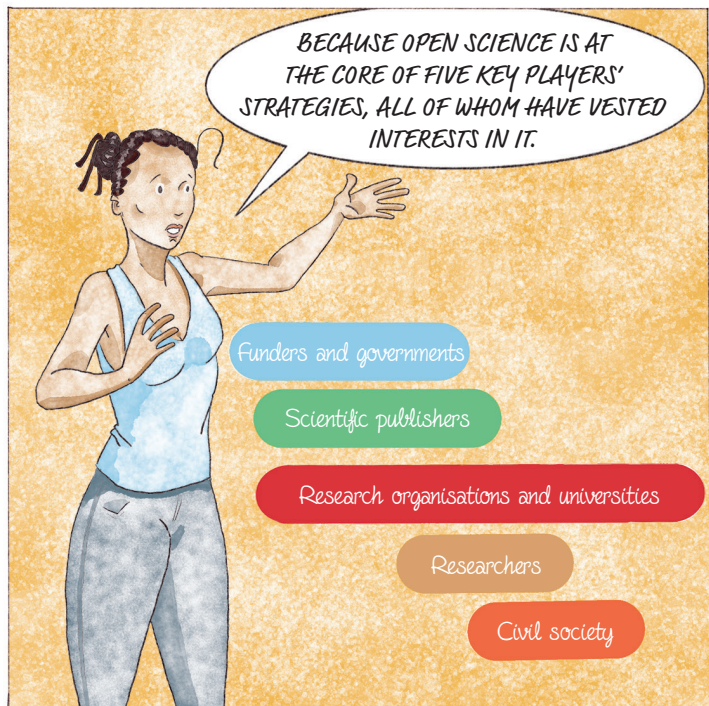


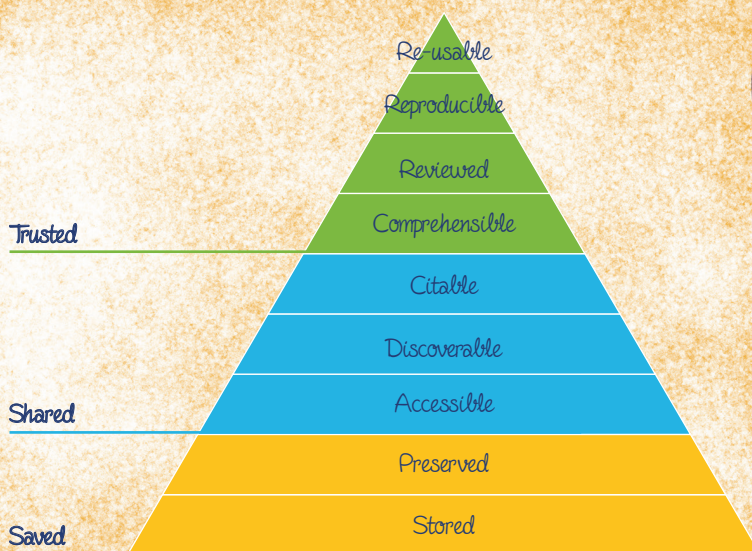
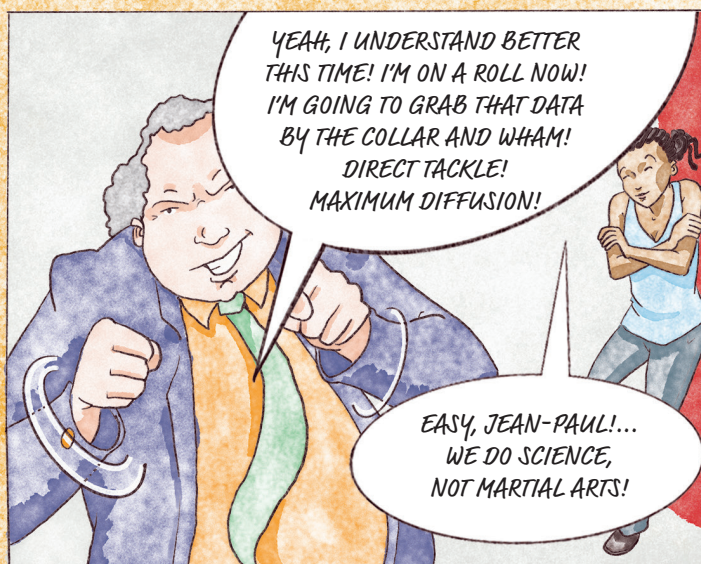


1. What are research data?





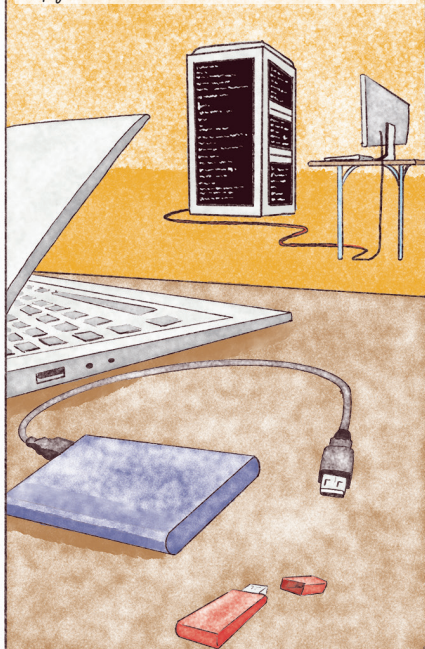




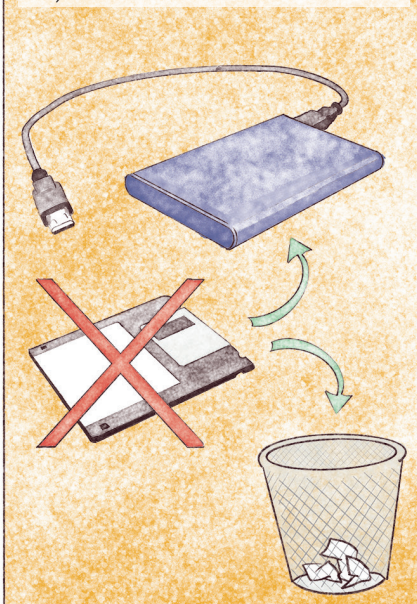
From A. de Waard. The Mendeley Data management platform: Research data management from a publisher's perspective. (2017) in Danielle Descoteaux, Chiara Farinelli, Marina Soares e Silva, Anita de Waard: Playing Well on the Data FAIRground: Initiatives and Infrastructure in Research Data Management. Data Intelligence 2019; 1 (4): 350-367. doi: https://doi.org/10.1162/dint_a_00020

Step 1: Saved data

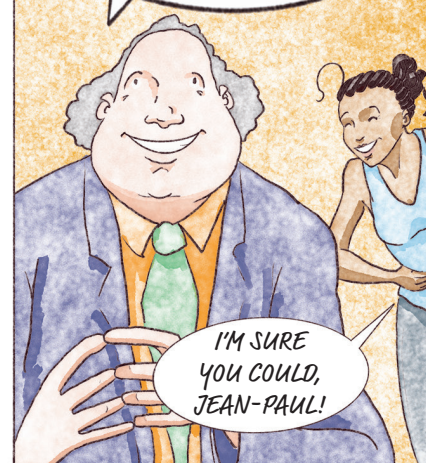
Data must first be **stored**, ideally according to the 3-2-1 rule: 3 identical copies, stored on 2 different media + 1 off-site copy.



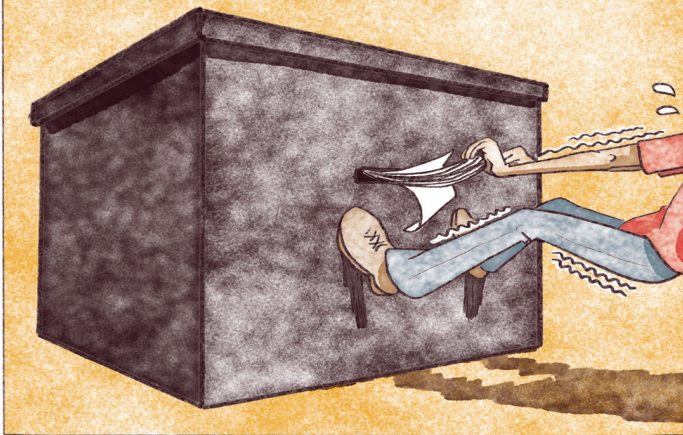
This is however not sufficient. To prevent IT technologies obsolescence, data must be **preserved**, i.e. selected on the basis of various criteria and varying lifespans (storage for 6 months, 3 years, 10 years, etc.).



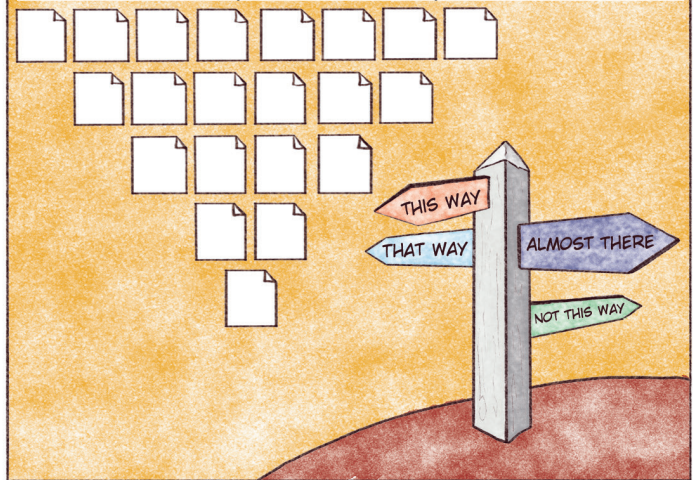
AH, NOW THAT BRINGS ME BACK! WHEN I WAS STILL YOUNG AND HANDSOME, I HAD MY FIRST STUDY STORED ON 52 FLOPPY DISKS. DO YOU THINK I COULD ASK A TRAINEE TO RESTORE THAT DATA FOR ME?



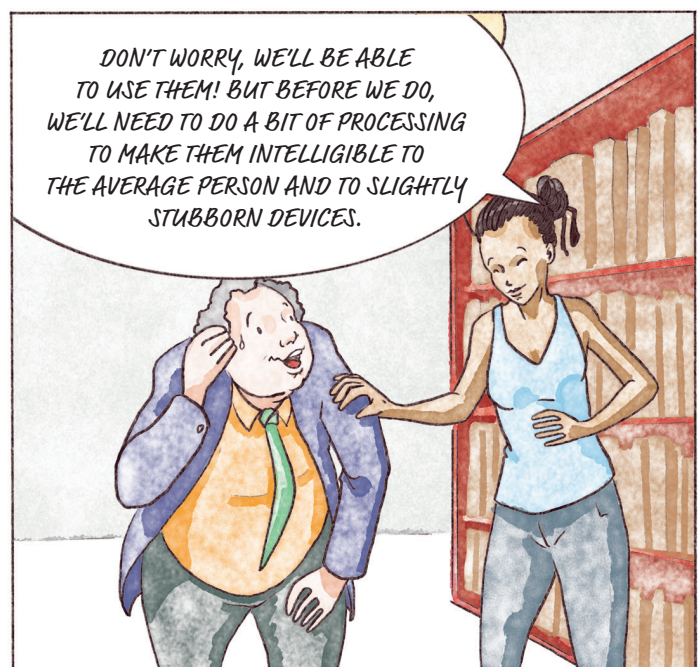
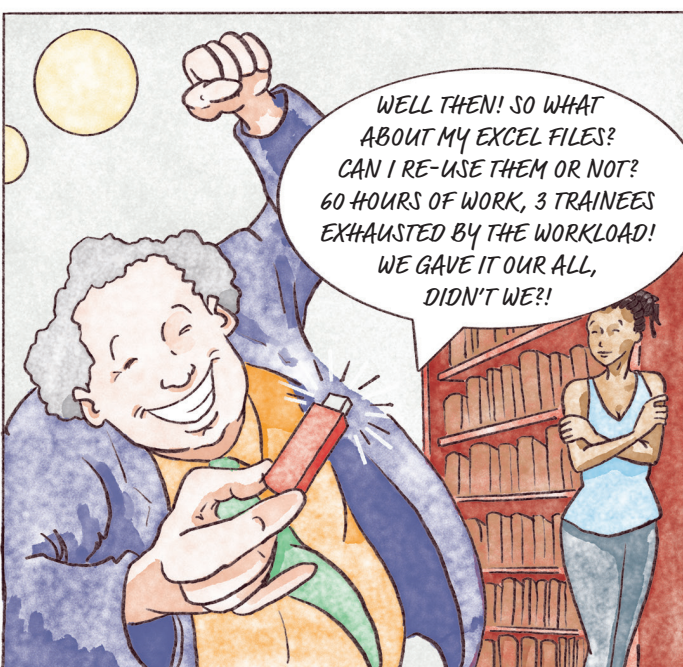
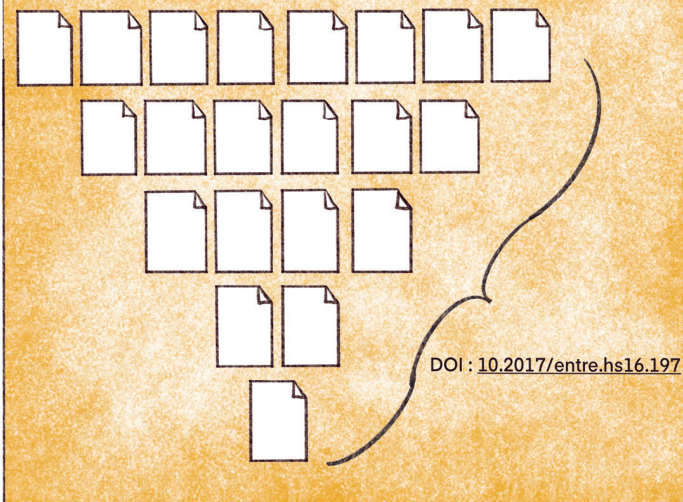
Even when data is stored and preserved, it is not necessarily available to researchers and other devices seeking to interrogate it. At the very least, it should be made **accessible** online.



However, even online, research data is not always easy to find by other **researchers**. It is therefore important to make them more widely known - for example by improving the quality of their description.



Furthermore, in order to monitor the re-use of this data and ensure that researchers who produced it receive the scientific credit they deserve, it is advisable to make it **citeable**, for example by assigning it a DOI (Digital Object Identifier) or linking it to a data-paper.



Step 3: Trusted data

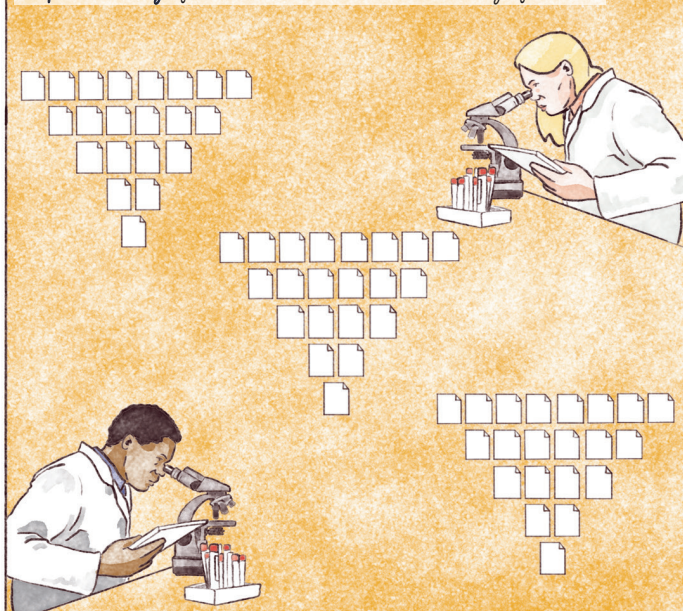
Data that has been collected for internal use is not necessarily **comprehensible** to a third party. It is therefore important to document the data collection: what units of measure were used? What is the context? What abbreviations and parameters were employed? It is also necessary to describe them as accurately as possible - particularly by means of exhaustive and precise metadata.



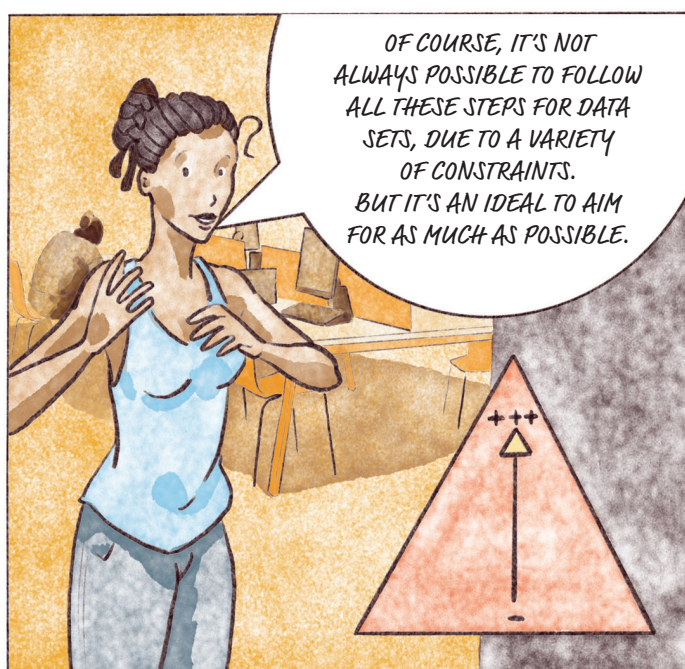
In order to scientifically **validate** this data, it can also be useful to have it reviewed by peers. Reviewing systems also exist for research data (such as data-papers).



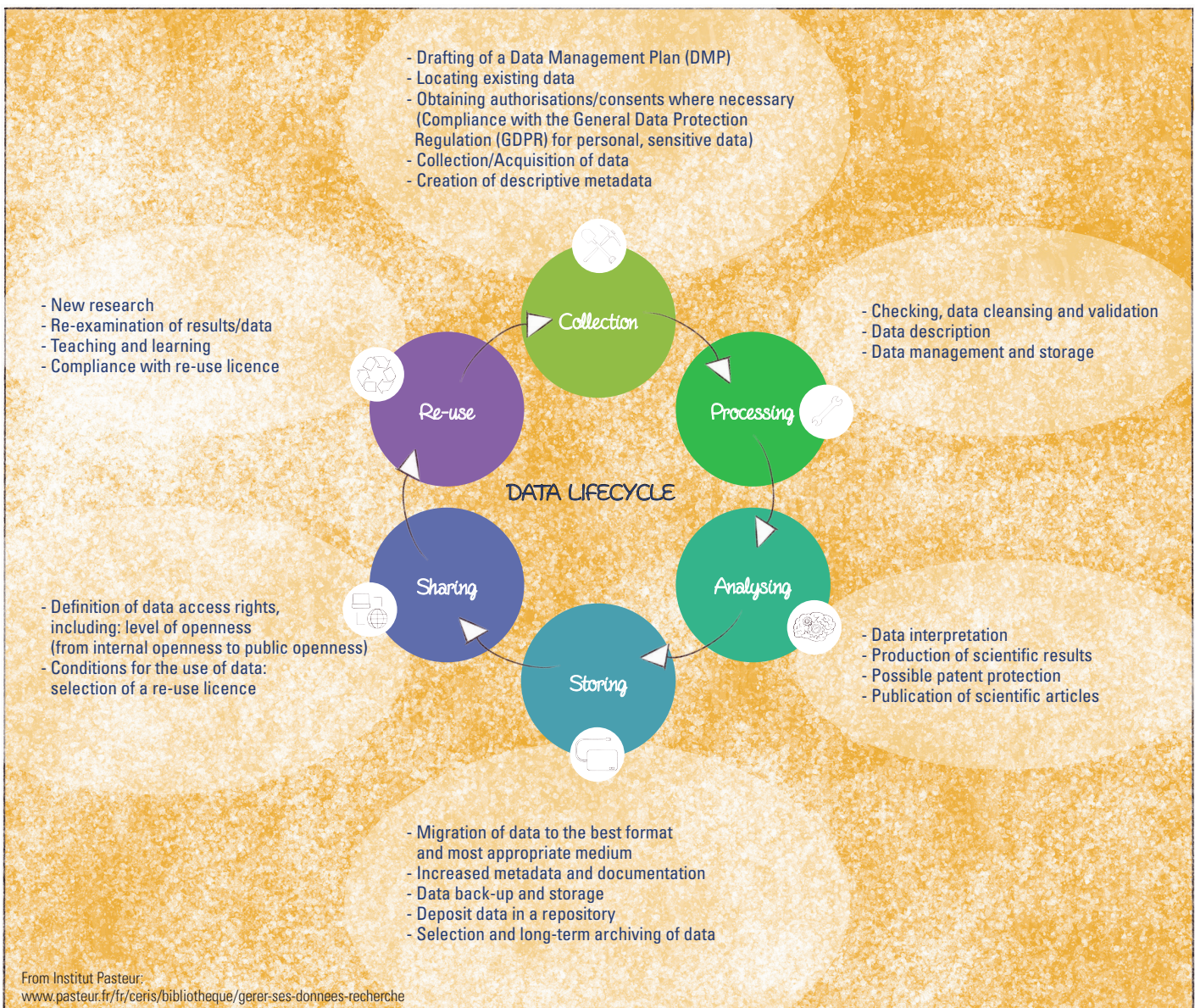
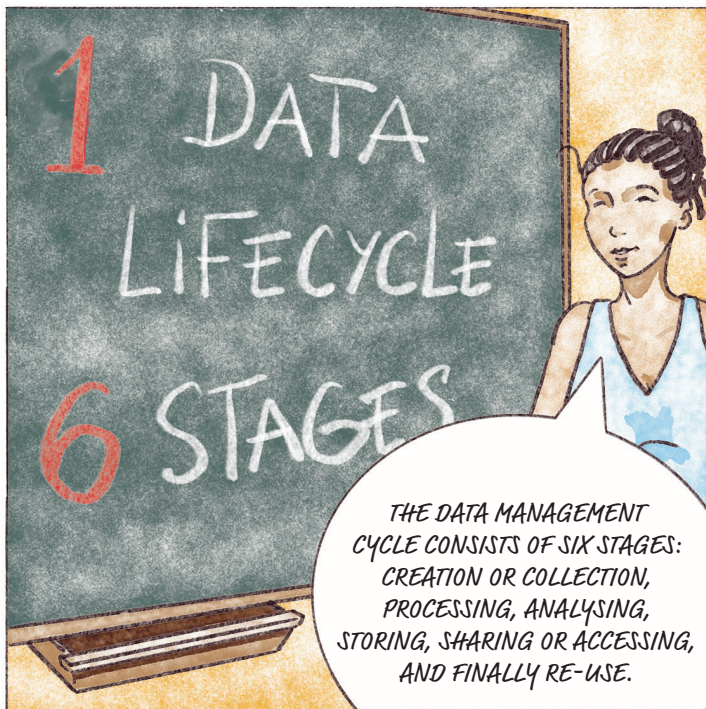
Reproducibility of research increases the credibility of results.



Research data that is comprehensible, reliable and reproducible is more likely to be re-used and cited by other researchers. A user licence should therefore be applied to the data, providing a strict framework for **re-use** by other researchers.



2. How to manage and disseminate research data?





LEGAL ISSUES: OPEN OR NON-OPEN DATA?

Data that can be distributed

- Free communication if (cf. Digital Republic Act, Oct. 2016):
- data resulting from a research activity at least semi-financed by public funds
 - not protected by a specific law
 - made public by the researcher or the institution (the institution decides which data will be open, where and under what conditions it will be deposited).
 - Compulsory disclosure of some geographical and environmental data (cf. Inspire convention and Arrhus convention)

Data that can be distributed under specific conditions

- Data presenting risks for the protection of the nation's scientific and technical potential (cf. "protected unit" laboratory)
- Restricted areas: physical and digital access subject to authorisation
- Data protected by copyright and other intellectual property laws
- Personal data (see General Data Protection Regulation (GDPR))

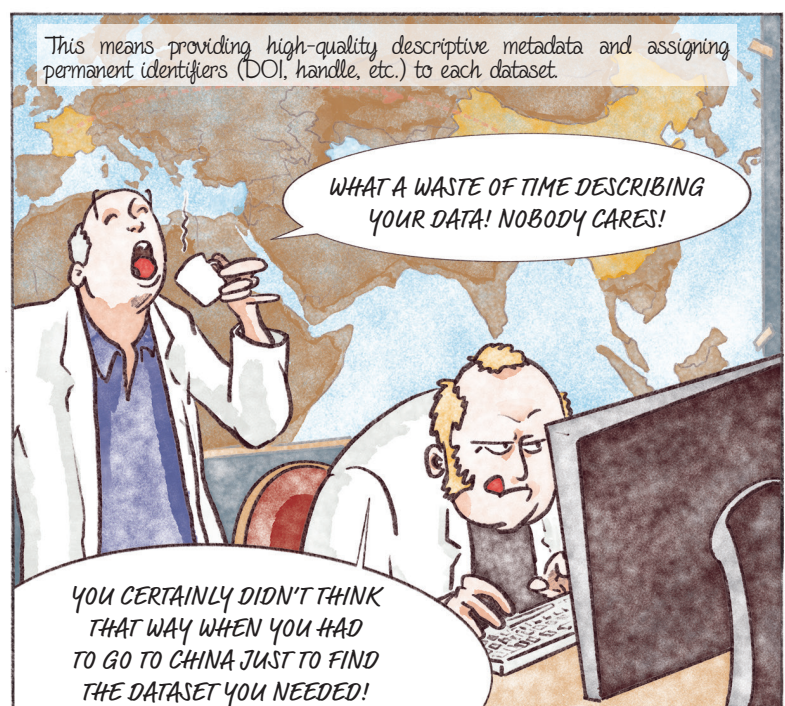
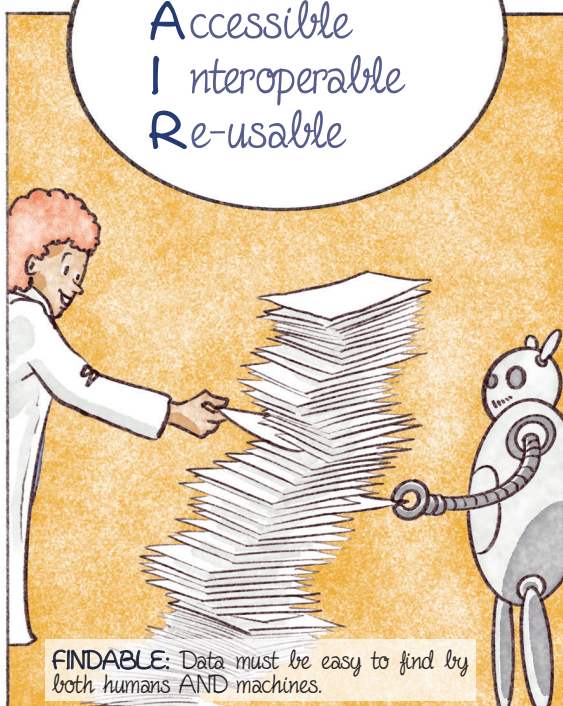
Data prohibited for distribution

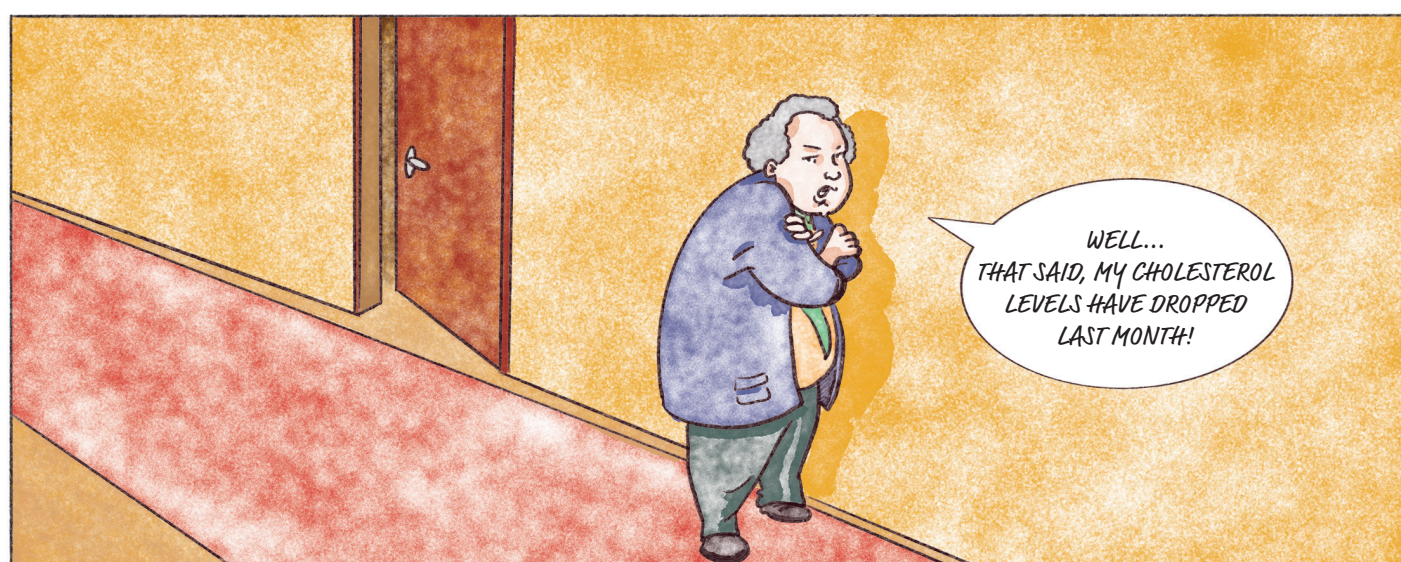
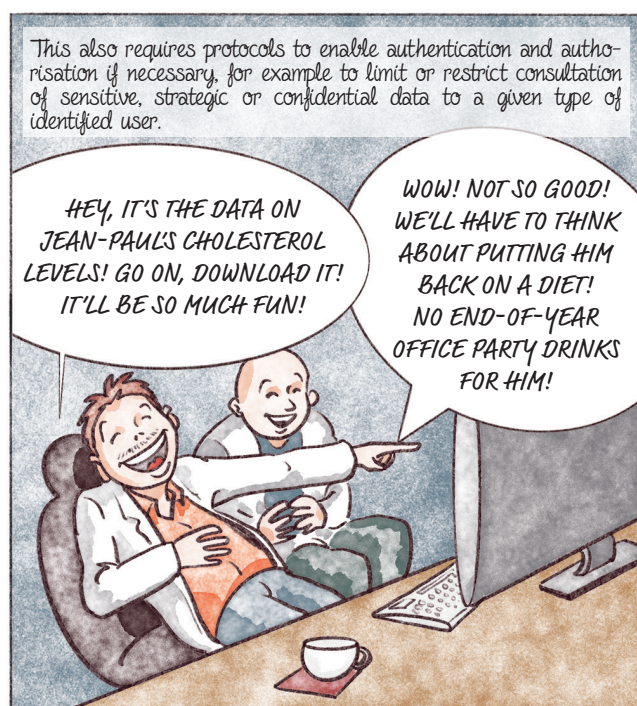
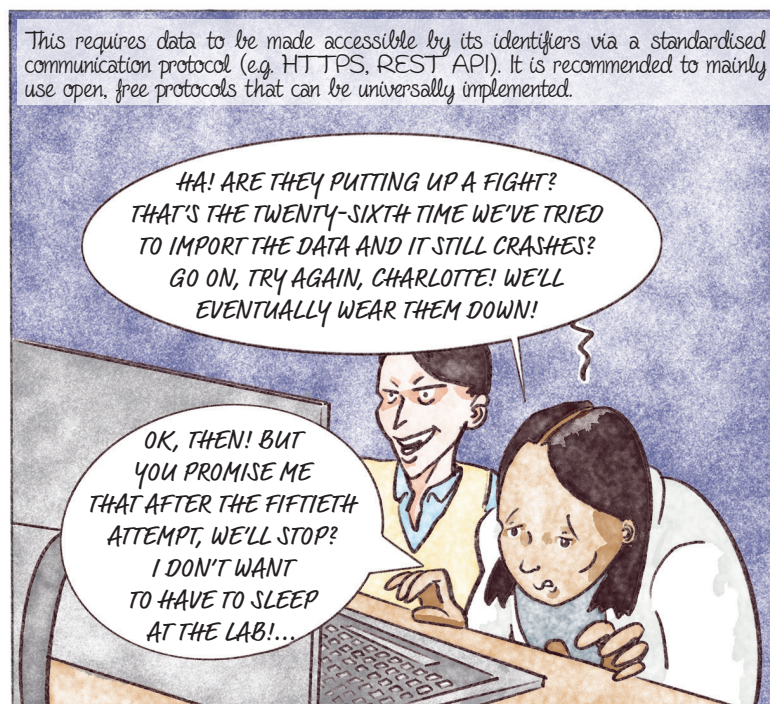
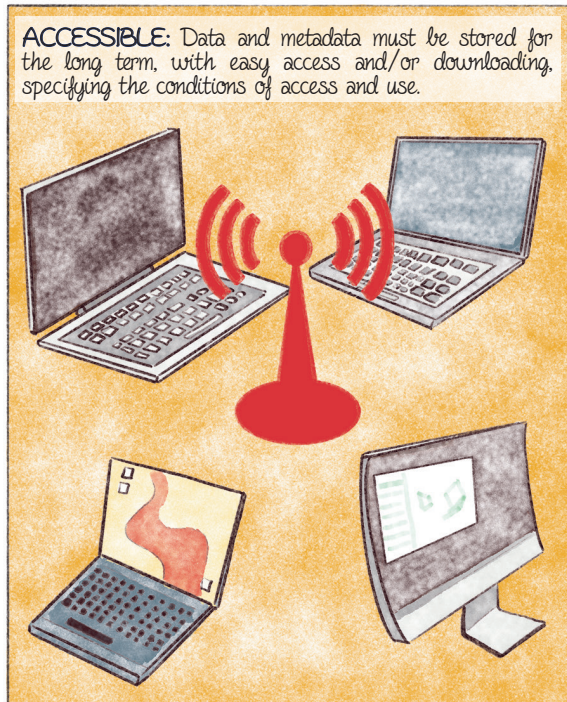
- Data presenting risks for the protection of national defense secrets
- Data presenting risks for the security of the State, public safety or the security of the institution
- Professional secrecy or confidentiality (medical secrecy, investigation secrecy, banking and tax secrecy)

In order to achieve this goal, data sharing must comply with a set of best practices that enable it to be discovered and used by humans - but also by machines. This necessary commitment is summed up by the acronym "FAIR".

From Dominique L'Hostis, From management plan to data-paper, June 2019
<https://gricad-media.univ-grenoble-alpes.fr/video/plan-gestion-donnees-au-data-paper>

F indable
A ccessible
I nteroperable
R e-usable





INTEROPERABLE: Data and metadata must be downloadable, usable, intelligible and combinable with other data, by humans AND by machines.

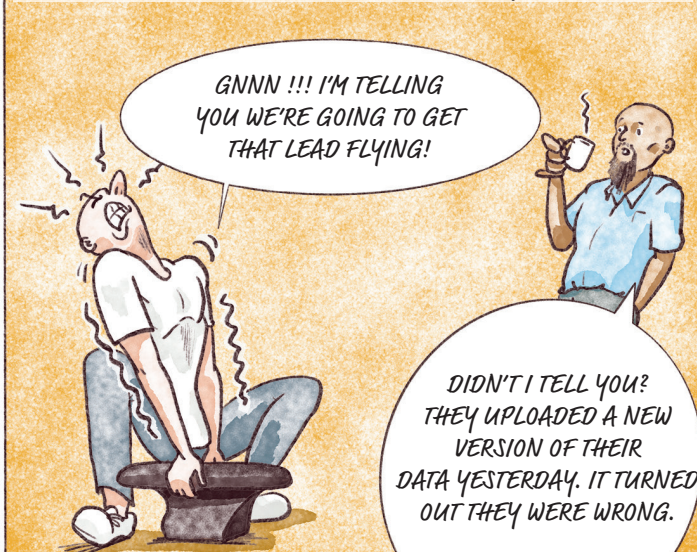


Data and metadata should use a formal, accessible and shared language that is widely applicable to knowledge representation, such as Semantic Web technologies. It is recommended to use standard ontologies and controlled vocabularies.

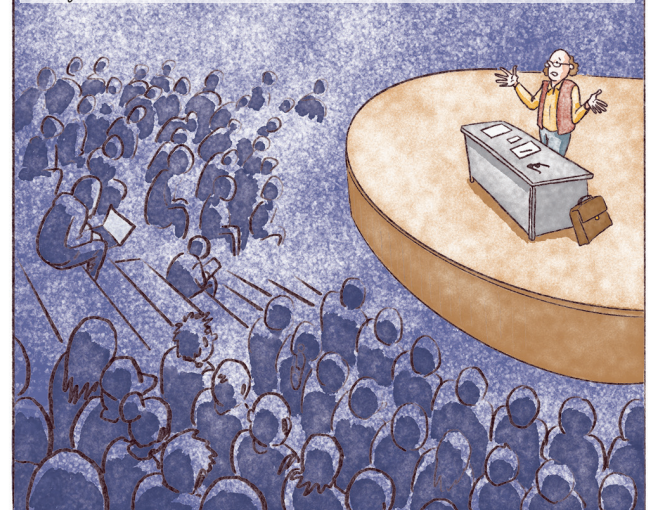
I CAN'T FIND ANY MORE INFORMATION IN THE DATA FROM THE "BEETLE" LABORATORY. TO DESCRIBE THEIR STUDY OF CARS, THEY USED THE WORDS "AUTOMOBILE", "VEHICLE", "RIDE" AND "CHARIOT". AT SOME POINT THEY EVEN REFERRED TO THEM AS "JALOPIES" AND "CLUNKERS"! AND NOW I CAN'T FIND ANYTHING!



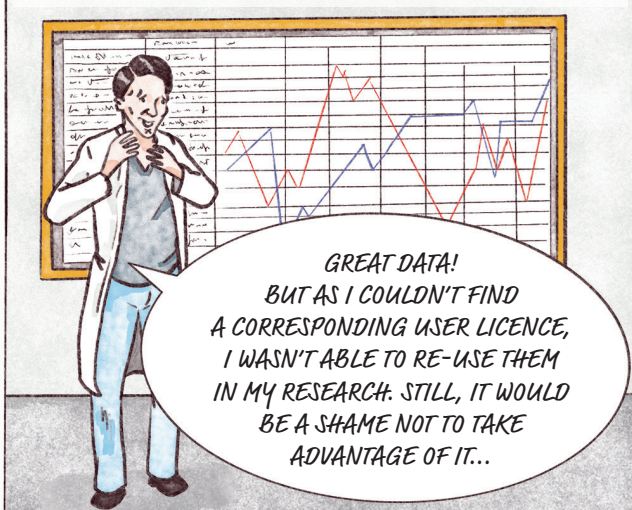
Data and metadata may include links to other (meta)data, previous or more recent versions, additional data or articles citing the data.

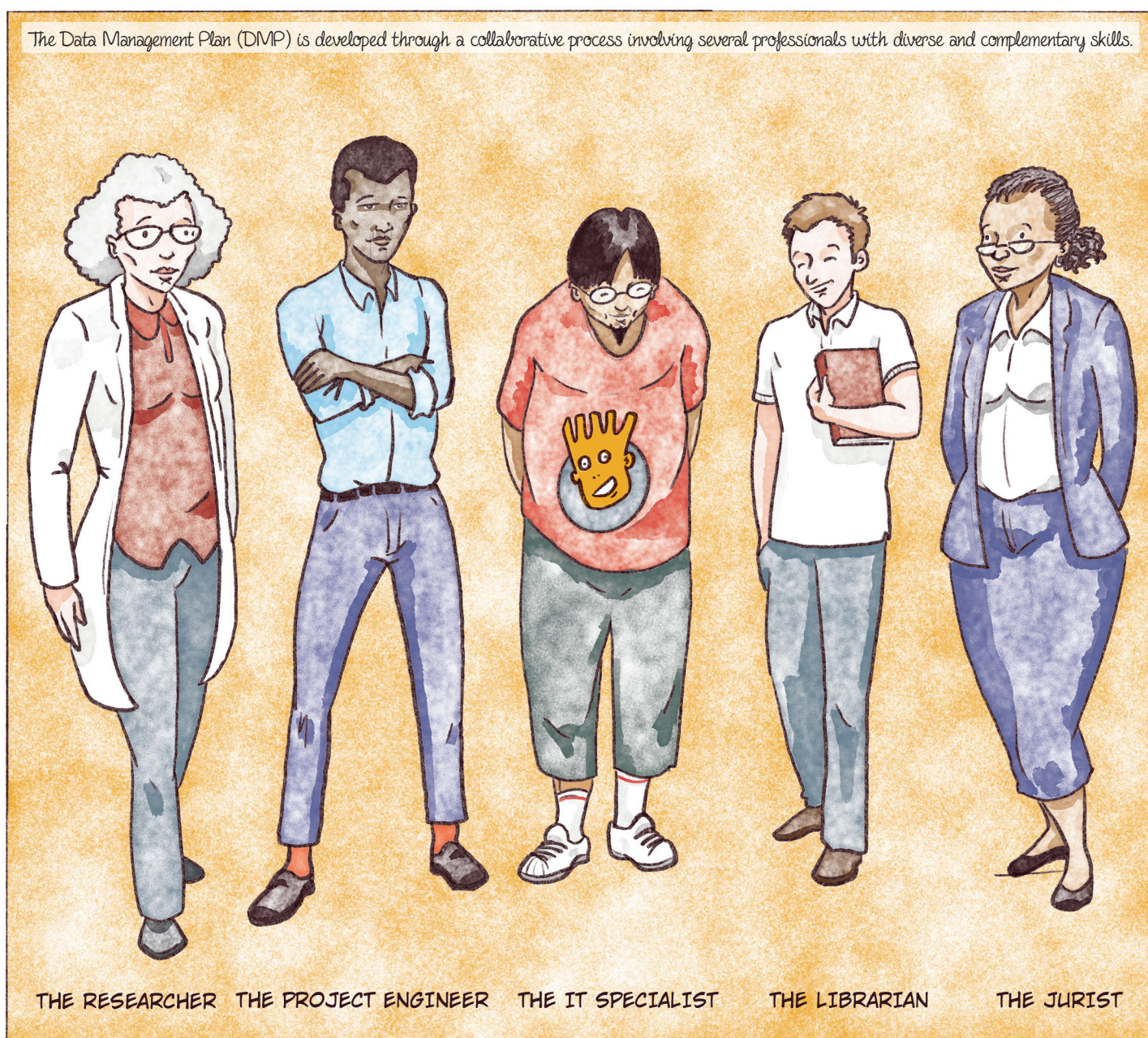
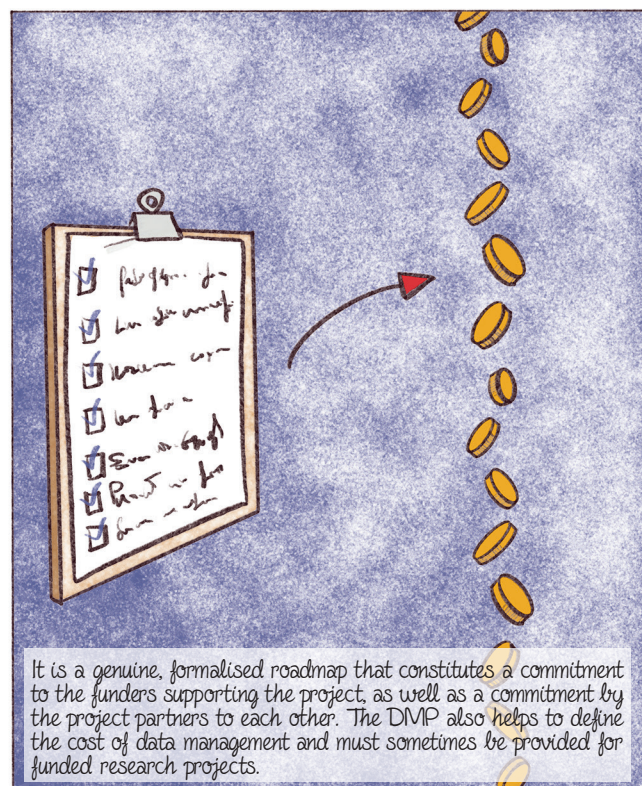
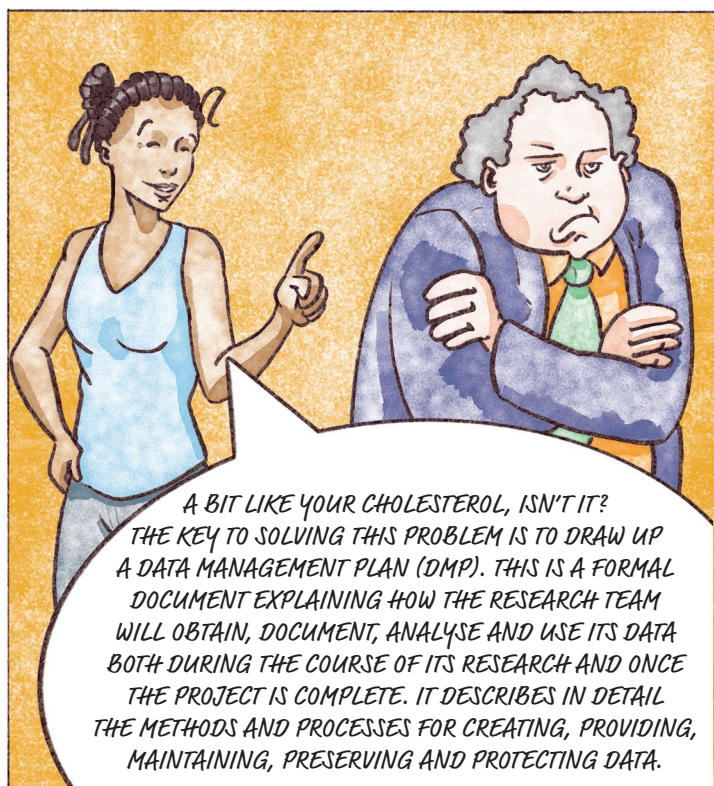


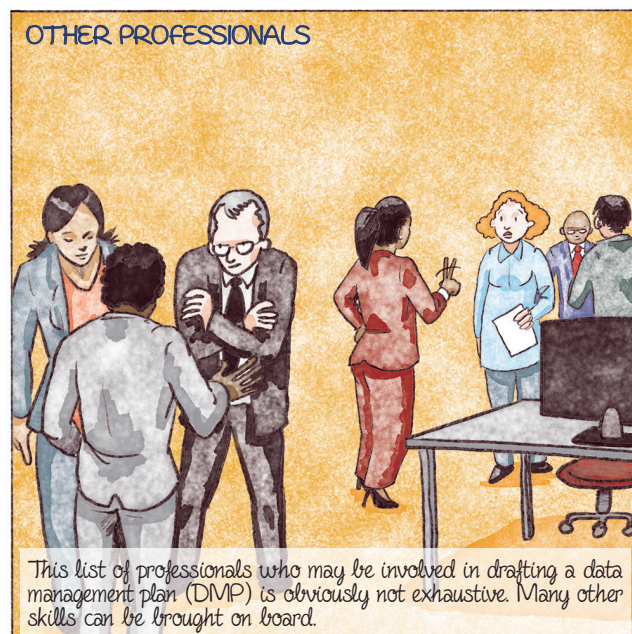
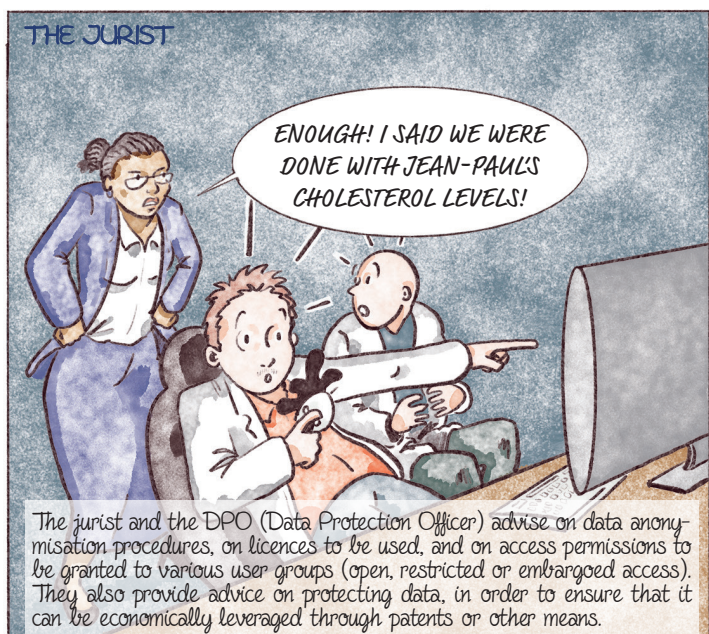
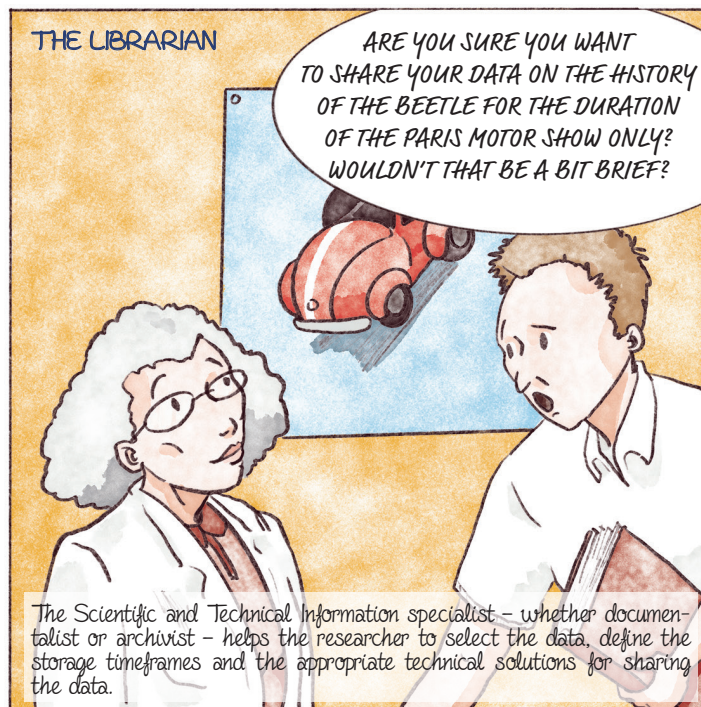
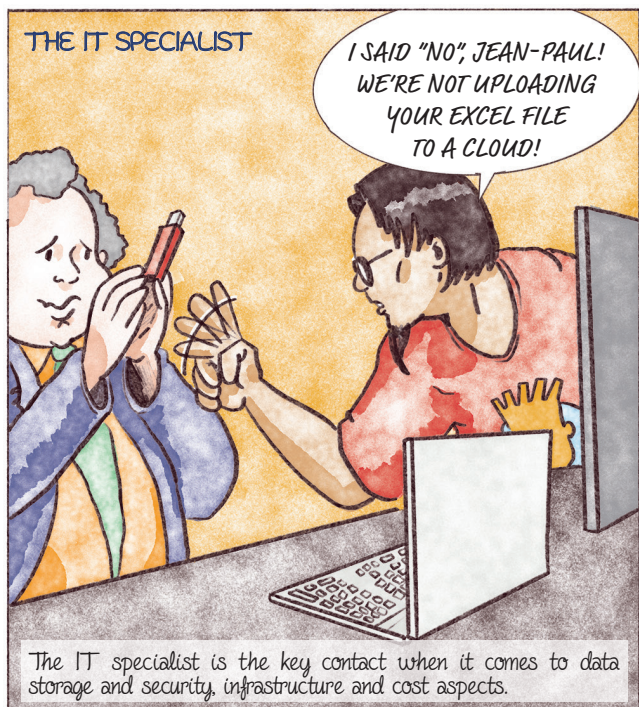
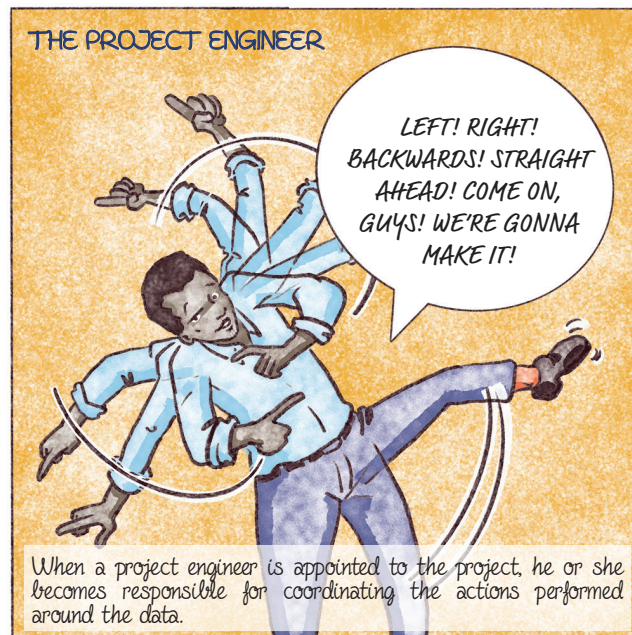
RE-USABLE: Data and metadata must include characteristics that make the data re-usable for future research or for other purposes (teaching, innovation, replication and scientific transparency).

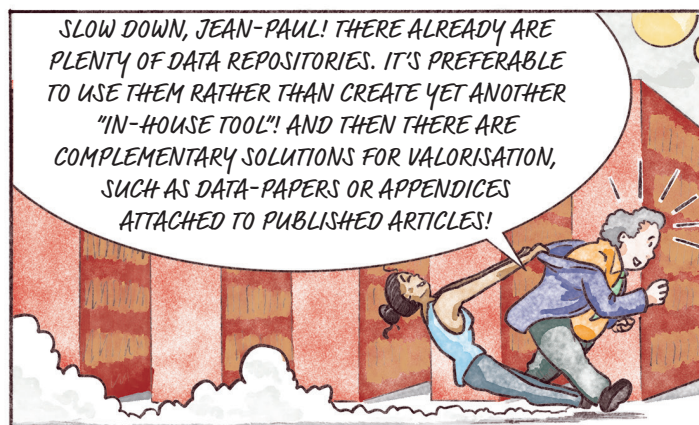
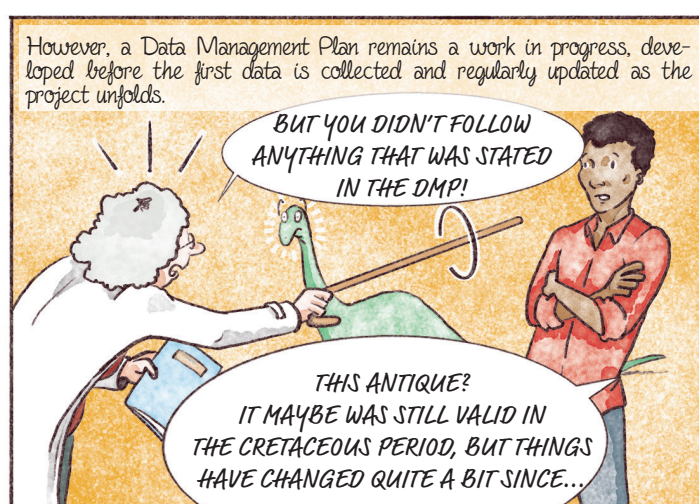
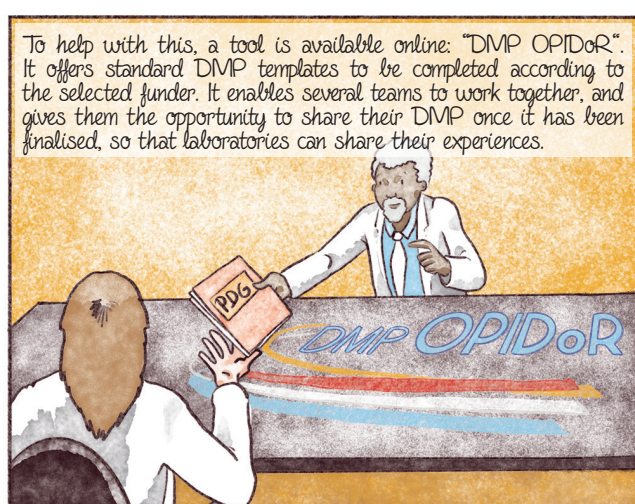
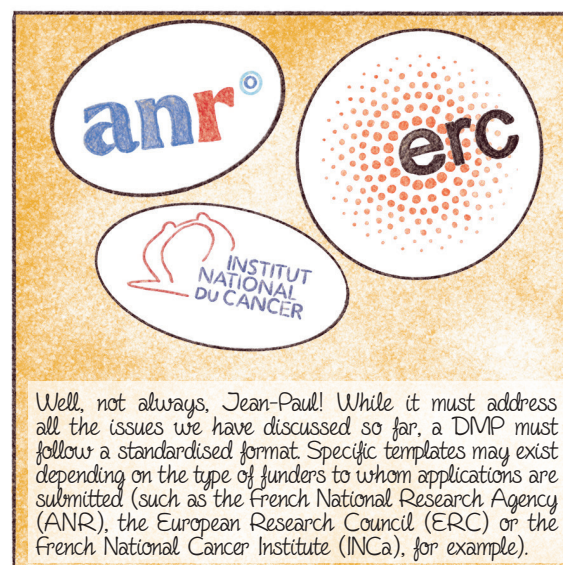


To that end, providing datasets with an explicit and accessible user licence and linking them to their provenance in compliance with the standards of the specified communities will make it easier to re-use them later on.

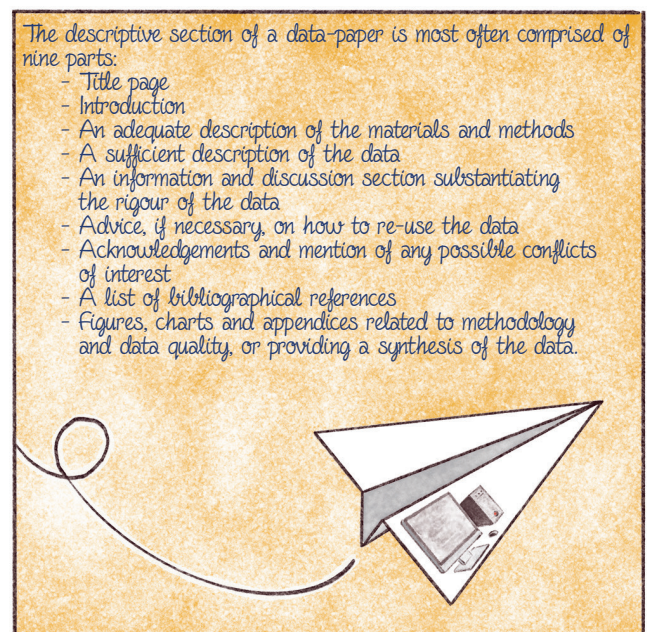
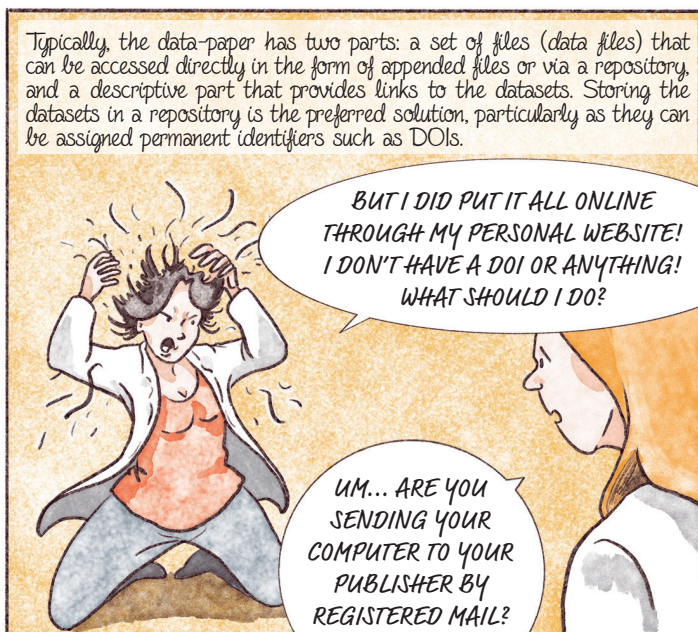
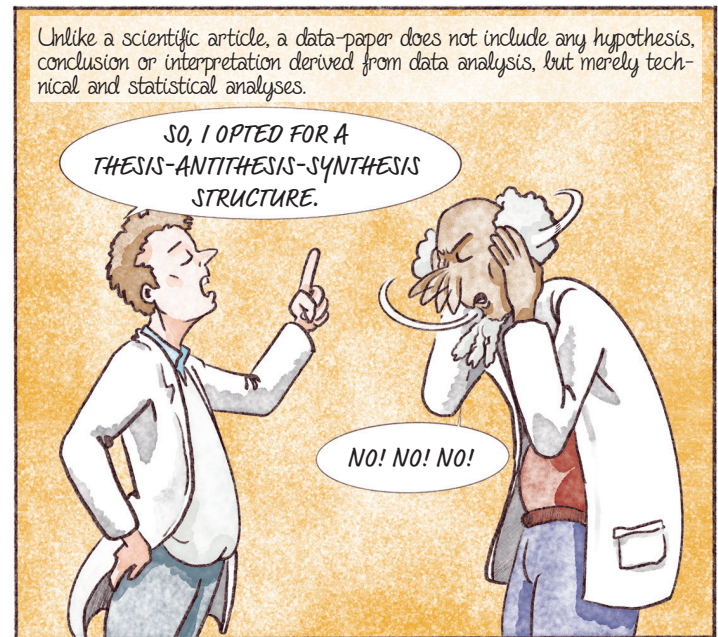
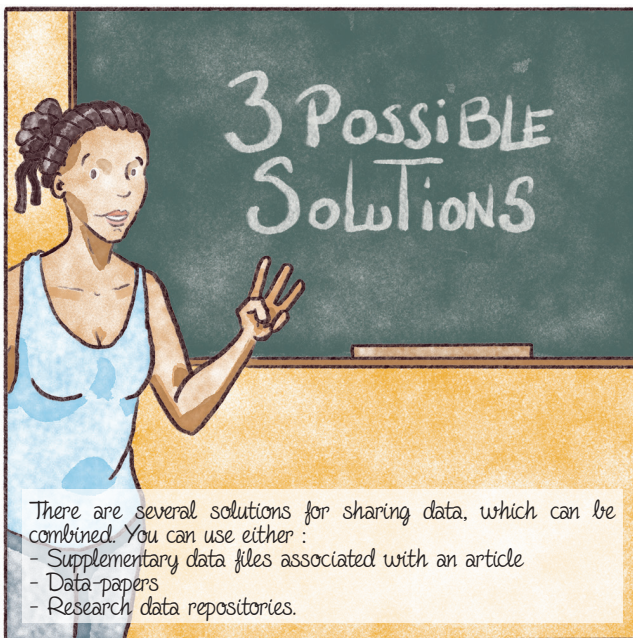


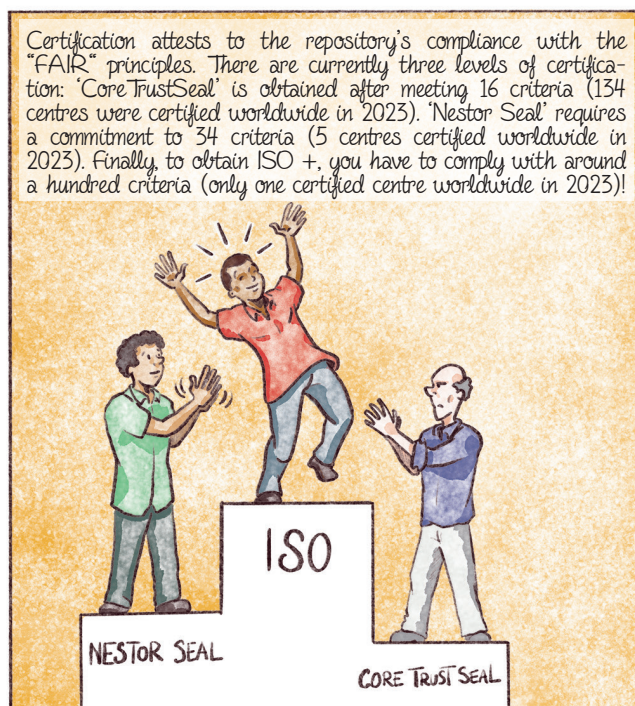
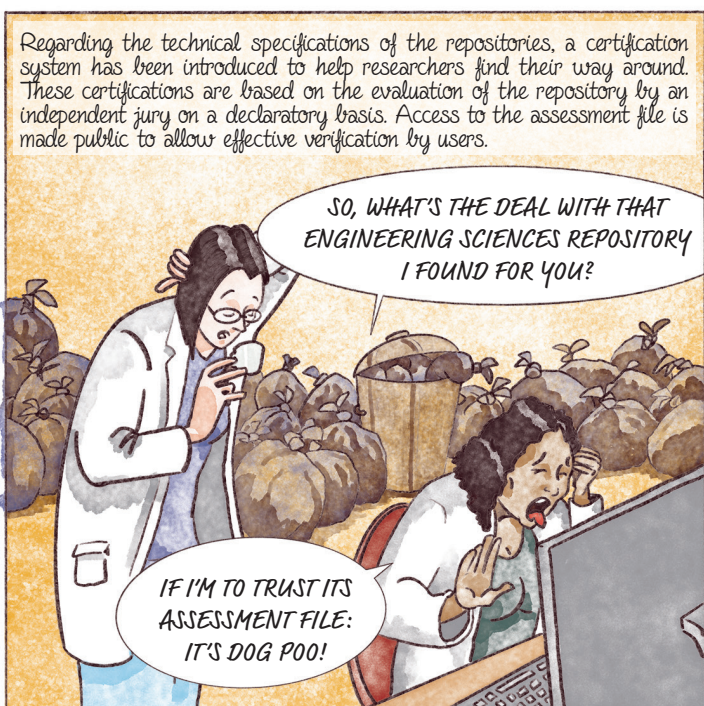
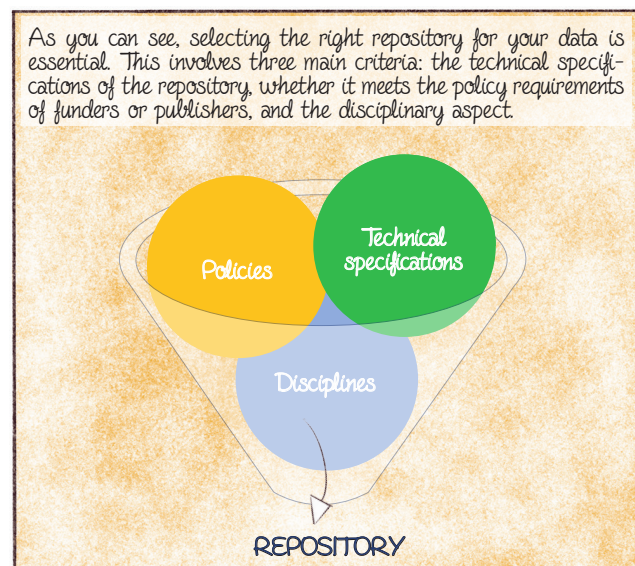
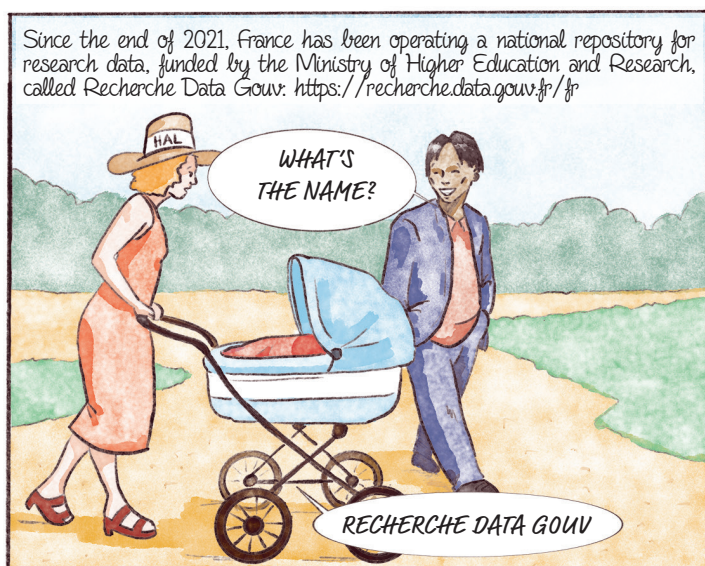
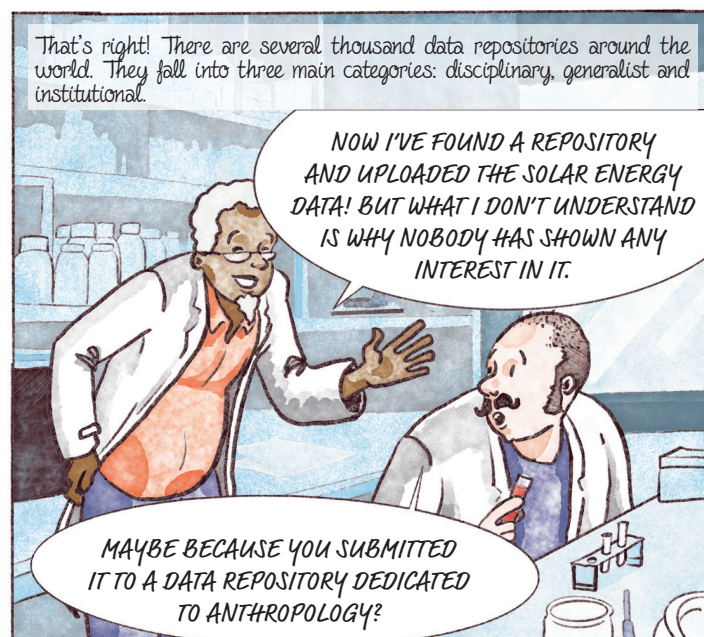
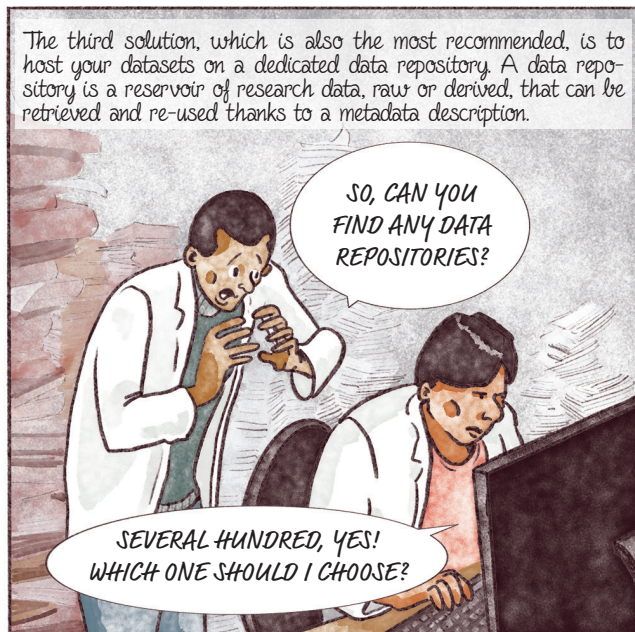


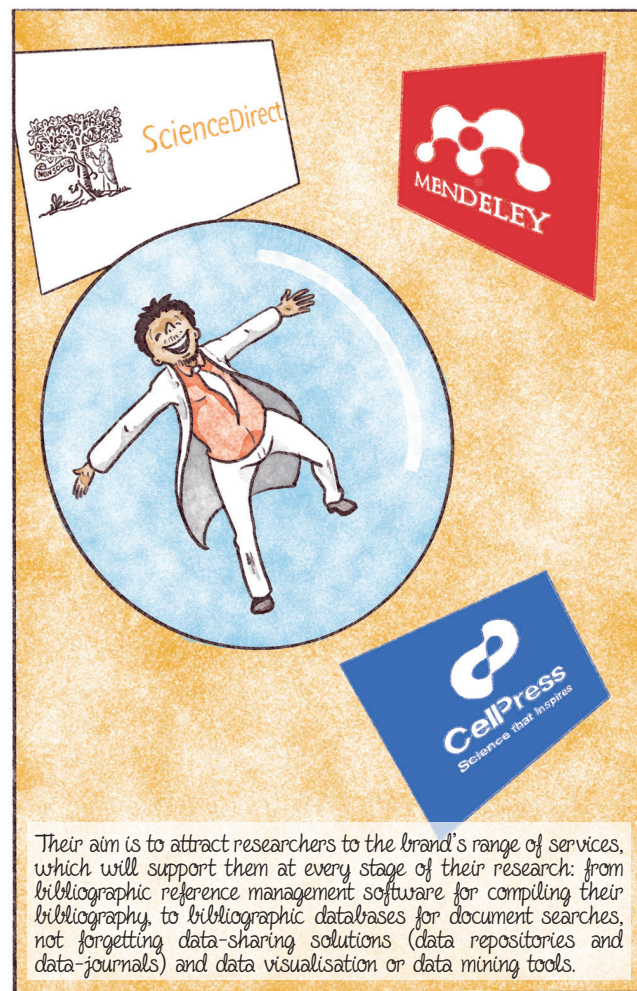
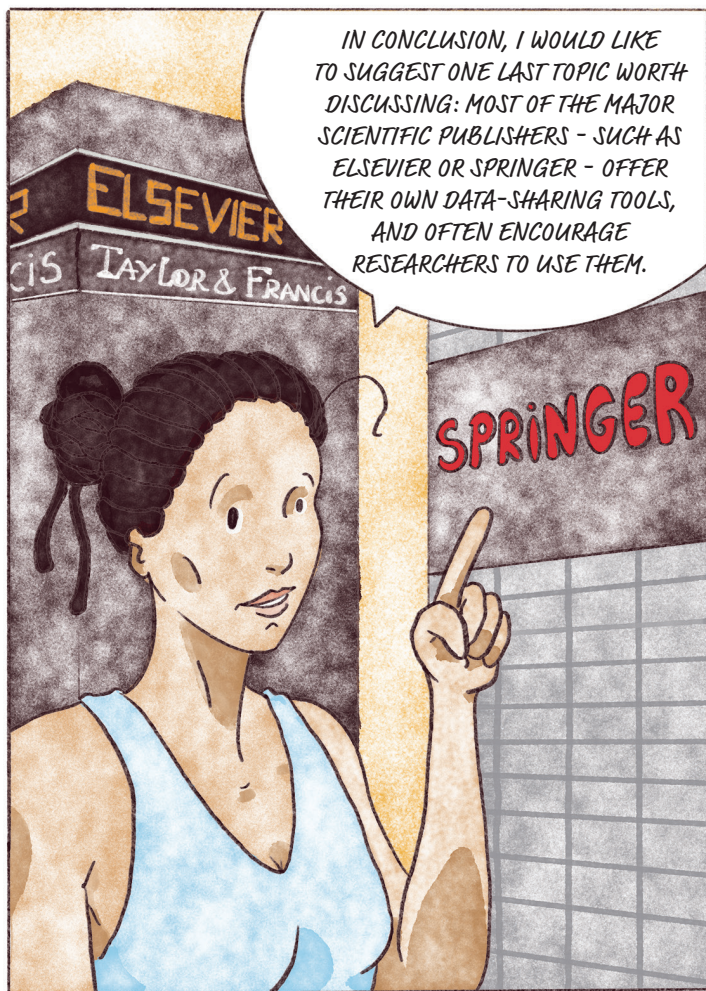




3. Sharing research data: which tools to use?









BIBLIOGRAPHICAL REFERENCES:

Borgman, C. L. (2015). Big Data, Little Data, No Data: Scholarship in the Networked World. Cambridge, MA: MIT Press.

Callisto Formation. Fondation UNIT.
https://callisto-formation.fr/?theme=boostplus_c06&redirect=0

CoopIST : délégation à la formation scientifique et technique, CIRAD.
Gérer les données de la recherche.
<https://doi.org/10.18167/COOPIST/0005>

DoRANum – Données de la recherche : Apprentissage Numérique.
<https://doranum.fr/>

Ouverture des données de recherche – Guide d'analyse
du cadre juridique en France – V2. (2017). Ouvrir la science !
Comité pour la science ouverte.
www.ouvrirlascience.fr/ouverture-des-donnees-de-recherche-guide-danalyse-du-cadrejuridique-en-france-v2

Partager les données liées aux publications scientifiques –
Guide pour les chercheurs. (2022). Ouvrir la Science !
Comité pour la science ouverte.
www.ouvrirlascience.fr/partager-les-donnees-liees-aux-publications-scientifiques-guidepour-les-chercheurs

AUTHOR

Marie Latour, deputy director of the library, University of French Guiana

SCIENTIFIC SUPERVISOR

Annaïg Mahé, lecturer at URFIST Paris

ARTWORK

Olivier Copin

GRAPHIC DESIGN

Bénédicte Sauvage (BCOM)

ENGLISH TRANSLATOR

Stéphane Berland

SCIENTIFIC PROOFREADING

Cyril Heude (data librarian at SciencePo Paris),
Romain Féret (director of Média Normandie),
Amélie Barrio (Co-director of URFIST Occitanie)

Project co-financed by URFIST Paris
(Ecole nationale des Chartes - PSL)

