

Transcription



Hello everyone, my name is Michael Nauge. I am the **Data Management Cluster Coordinator** at the **University of Poitiers**.

First and foremost, I would like to say : the **University of Poitiers** is **actively engaged to Open Science** in its broadest sense. We believe that making research results—from publications to raw data—more transparent and accessible is key to **scientific progress** and, ultimately, to **generating a real impact for our citizens**.

We have a dedicated website on this subject.

Open Science : national strategy



In France, we've had a national policy on **Open Science** for several years now. These goals are **detailed** in a **strategic roadmap**

This strategy is built around few main pillars:

- **First**, we want to **make open access the standard** for all our publications.
- **Second**, we are focusing on **research data**—ensuring it is well-structured, shared, and as open as possible.
- **Third**, we aim to promote and open up the **source code** produced by our researchers.
- **And finally**, our ultimate goal is to **transform practices** so that Open Science becomes the default principle for everyone."

Open Science : local strategy



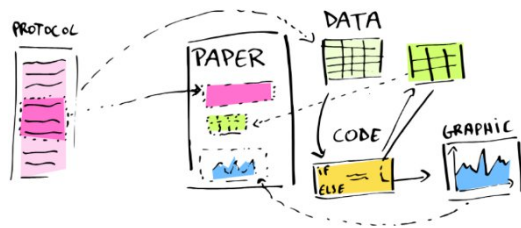
Deeply inspired by the **National roadmap** our university has developed its own local version: the **University of Poitiers Open Science Roadmap**.

As you can see, there are **many similarities**. We wanted to show that we **support** national and European ambitions on this topic at a local

level.

Scientific Outputs

Questions/Scientific hypothesis > Protocol > DATA > Interpretation > Results > Dissemination



To **understand the challenges**, let me give you a quick reminder of what we mean by **research products**

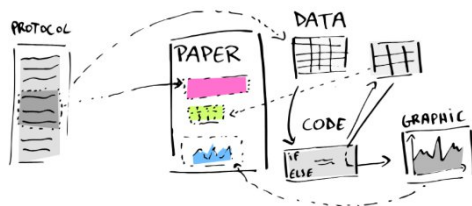
Everything starts with a **question**—or a set of questions—and **scientific hypotheses** to try and answer them. To test these hypotheses, we need **evidence**, verifiable facts, and **credible data**. This requires a formal **protocol** to collect that data.

Usually, this data is **raw** and extensive, so we need to process it to make it **interpretable**. To do that, we rely on **software, scripts, and code**.

Thanks to **refined data** and **visualizations**, we can now interpret the results and **share the new knowledge** we've gained. This is usually done by **summarizing** the work in a **scientific paper**.

Scientific Outputs: Focus on Paper

Questions/Scientific hypothesis > Protocol > DATA > Interpretation > Results > Dissemination



Regarding publications, an area we've **known well for years** now.

I'll keep this brief.

The French Digital Republic law

The Law for a Digital Republic was adopted on 28 September 2016 and was published in the Official Journal on 7 October 2016.
loi pour une République numérique

The law allows the author of "a scientific article financed for the most part by public funds and which has been published in a periodical that appears at least once a year" to deposit his or her author's version (= final version of the manuscript accepted for publication) in an Open Archive, even after having granted exclusive rights to a publisher.

Once the article is published, the author version can be immediately deposited in HAL, with an **embargo period**:

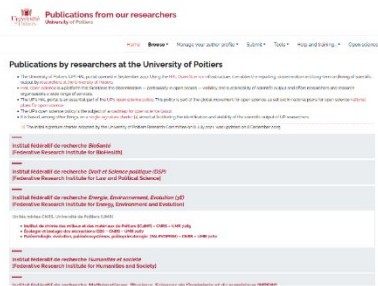
- 6 months for STM articles
- 12 months for SHS articles

STM : Science Technique et Médecine
SHS : Science Humaine et Sociale



Briefly, we're lucky in France to have a specific law. It allows our authors to **freely share** their versions in **open archives**, like our national platform, named in french **HAL**. This is a **Green Open Access**.

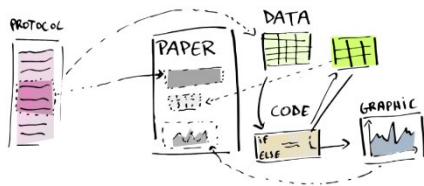
National Open Archive: HAL



Within this national platform, we have a **dedicated space** for our university. There, you can access **many publications** from our labs **without needing a subscription**

Scientific Outputs: Focus on Data

Questions/Scientific hypothesis > Protocol > DATA > Interpretation > Results > Dissemination



Now, I'd like to **focus more** on **research data** which, I must admit, is one of my **favorite topic**.

Research DATA : Precious Assets

Imperative for research

- Testing hypotheses
- Continuing previous work

Data Engineering and Processing:

- Resource-intensive: High financial and time costs
- Highly complex: Technical and organizational challenges

Cost + Complexity => High Risk combination for any project



As previously discussed, research data is key to testing our hypotheses. It's what allows science to be truly cumulative.

But generating data often requires very expensive technical equipment, particularly in imaging. Take, for instance, a confocal microscope or an MRI scanner.

Sometimes, high-cost equipment is unnecessary; for instance, in **SSH**, data collection relies heavily on **labor-intensive** processes, such as archival research or conducting surveys in the field.

Moreover, as research becomes increasingly **interdisciplinary**, excellent organization is required to coordinate various contributors. Bringing together the expertise of researchers, from many country, many support staff, and sometimes multiple **service providers** is never an easy task....

The combination of high costs and technical complexity naturally elevates the project's risk profile.

To get a real return on investment on high-cost and complex step, we can't let data sit in a silo. Making science sharable, cumulative is the only way to justify these financial costs.

Research DATA : Potential Improvements



« The European Commission report Cost of not having FAIR research data (2019) estimates that the cost of poor management of research data amounts to 3 billion euros for France, due to wasted time, failure to optimization of storage costs, licensing fees, research duplication issues, lack of cross fertilization »
<https://www.eurostat.ec.europa.eu/press-room/2019/03/2019-03-20-01>

Accessibility of publication data :

« publications provide access to around 10% of data, with the rest remaining available but unused on computer hard drives »
<https://bibliotek.berlin.de/berlin/berlin/>

As you can see, the European Commission reports a **massive waste of resources** due to poor data management

In reality, **only a fraction of the data** is actually easy to access.

As a result, there is **significant room for improvement**. we have **a great opportunity** to make things much better.

Benefits of Effective Data Management

- Avoid losing them
- Comparing results
- Meta-analysis
- Replicability
- Credibility
- Cumulative Science



Findable - Accessible

If the money talk doesn't work, I focus on the **personal benefits**, the financial argument fails to convince.

For PhD students, I often remind them that sharing high-quality data will **boost their credibility**. And for senior researchers, I enjoy asking them: **'Do you actually know how to access all the data from your last PhD student?'**

Solution

Univ-Poitiers promotes the use of **trusted data repositories** for dissemination :

- France : Recherche Data Gouv Repository
- France (example SHS) : Nakala / TGIR Huma-Num
- SHS-Linguistic : Ortolang et Cocoon
- European : Zenodo
- List of thematic repositories



Personal laptop and usb key are not enough !
 Findable - Accessible

a colleague mentioned that a student claimed his data was gone forever, lost all his data, because his USB drive **went through the washing machine...**

in our university we promotes the use of trusted data repositories for dissemination, such as Recherche Data Gouv ; Nakala

Benefits of good data management

- Increase your chances of being cited (+25%)
- Encouraging collaboration



Reusable

Another argument I really like: high-quality shared data boosts your **visibility**. The most obvious benefit is an **increase in citations**, but more importantly, it leads to **collaboration invites**. Reusing someone else's data for your own project is great, but inviting the original team to collaborate is even better. Because they bring **expertise and experience**.

Reusable: Solution

- Data are released with a clear and accessible usage licence (statalab 2.0 / CC-BY 4.0)
- Data have persistent identifier (DOI) thanks to trusted repository
- (Meta)data meet domain-relevant community standards
- Data are described with rich metadata and context

Do you have :
 - metadata standards ?
 - file-naming convention ?
 - folder structures ?
 - version control
 - standardized units, values ?
 - readme file ?

Reusable

In order to provide **clarity for future collaborators**, To avoid **scaring off** potential partners, it's crucial to be very clear about how your data can be reused. This means specifying a license, such as **CC-BY**. This license allows for any kind of transformation or reuse, including **commercial purposes**.

For data to be truly reusable, it has to be easy to understand. That's why we should have **clear structuring** and following **existing standards**.

- Getting started
- Features and Use
- Features**
- Files
- Metadata
- Feedback
- Resources
- Links
- Glossary
- Documentation
- Online form
- Tables and tables
- Model space
- Routing a 2D A-30
- Tutorials
- Accounting a 3D A-30
- Conclusion
- How to
- Translations

[illegible]

Which folder?

Case 1

Ma these v1.doc
MyThesis v1_comments PDF
Ma these v2.doc
Thesis 02/01/2024 v2.doc
Thesis 03-01-2024 v3.doc
Thesis_final.docx
Thesis_final_correction(2).docx
Thesis_final_last_ok.docx
Thesis_final_last_ok_v7.docx
Thesis post-defense correction
Thesis_finaale-final-last_v2.docx
Thesis_finaale-final-last(7)_v1.docx

Case 2

- MyThesis_backup-01.tex
- MyThesis_backup-02.tex
- MyThesis_backup-03.tex
- MyThesis_backup-04.tex
- MyThesis_backup-05.tex
- MyThesis_backup-06.tex
- MyThesis_backup-07.tex
- MyThesis_current.tex
- MyThesis_export.pdf
- Log_backup.txt

1. $\frac{1}{2}$ 2. $\frac{1}{2}$ 3. $\frac{1}{2}$ 4. $\frac{1}{2}$ 5. $\frac{1}{2}$ 6. $\frac{1}{2}$ 7. $\frac{1}{2}$ 8. $\frac{1}{2}$ 9. $\frac{1}{2}$ 10. $\frac{1}{2}$

MyThes
MyThes
ba
My
My
My
My
My
My
My

0 0.5 1

Pr
n

The screenshot shows the official website of the French Ministry of Education (Ministère de l'Éducation Nationale). The header includes the ministry's name and logo, along with navigation links like 'Accueil', 'Actualités', 'Services', 'Régions', 'International', 'Recherche', and 'Contact'. The main banner features the French flag and the text 'L'Éducation Nationale' and 'L'Éducation Nationale, c'est la France de demain.' Below the banner, there are sections for 'Nouveaux arrivés', 'L'Éducation Nationale', and 'Médias'.

sup
lab

- Co
- En
- Co
- Pr

- Coordinated experts network
 - DRInnov, DPD, integrity referent, DSI, SCDS, labo. Members, ...
- From advisory services to concrete assistance
- Courses and events
- Promotes trusted tools & methods



donnees-recherche@univ-poitiers.fr

- donnees-recherche@univ-poitiers.fr

• 1

11. *Journal of the American Medical Association*, 2000; 284: 1039-1044.

Figure 1

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99

Poitiers Data management clusters



Goals

Encourage and Facilitate the writing of Data Management Plans (DMP)

- Individualized advice and proofreading
- Simplify answers to all DMP questions by making recommendations from the university available on the OpidoR DMP platform
- Dedicated events (collaborative writing, DMP, ...)

Supporting data FAIRization

- Help with selecting and depositing data in the appropriate data repository
- Help in selecting a data sharing and distribution license
- Advice and curation on metadata, file formats & naming, data processing workflow

Raising awareness and training in Open Science

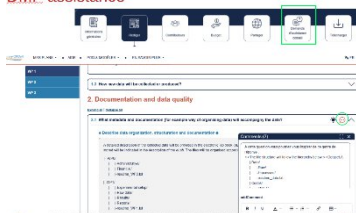
Charter and course

Making data **FAIR** starts with **proper planning**. So we start to encourage and facilitate the DMPs writing.

We provides proofreading, and advices in accordance with the national and local recommendation.

Once a DMP is in place, we assist **our researchers** with depositing their data into **trusted repositories**.

Facilitate the writing of Data Management Plans (DMP) DMP assistance



Let me show you how this works in practice. We use the national platform **DMP OPIDoR**, which allows for **collaborative drafting** and support management.

While writing, researchers can simply click a button to invite us to **review and comment** on their draft. In this example, the researcher hadn't written anything yet regarding **data organization and traceability**. We were able to provide several

comments to guide them. As a result, they used our recommendations to describe their use of **lab notebooks** and their **folder structures**.

Facilitate the writing of Data Management Plans (DMP) DMP: local recommendations



With **another click**, researchers can also access **examples and generic recommendations** that we provide for every single question.

Sorry, these tips are **French-only for now**. But the main point is that we encourage the use of **standards**. In this example, we suggest the **Chemical Entities Ontology**.

Facilitate the writing of Data Management Plans (DMP) DMP: local recommendations



our tips, guidelines are also **compiled into a freely downloadable PDF**

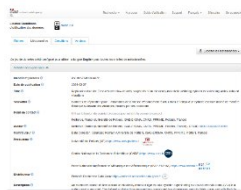
Feel free to take a look

Supporting Data FAIRization Repository: metadata quality



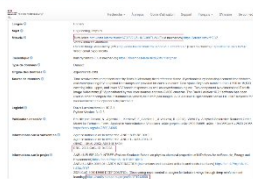
For dissimination, we promote and use **trusted repositories** for sharing our data. But they come with **strict requirements**. You need clear file naming and high quality metadata. Each metadata should follow **existing rules, norms, standard, registries, and ontologies**.

Supporting Data FAIRization
Repository: metadata quality



clear licence, Orcid for researcher/authors

Supporting Data FAIRization
Repository: metadata quality



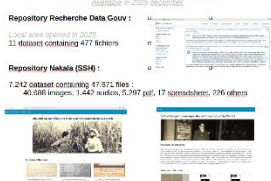
It's not just **plain text** or simple keywords. These are **ontology concepts** with labels in multiple languages, **linked** to a vast network of other concepts.

Repository: local metadata quality improvement



However, for now, researchers don't usually describe their data this way from the start. Things are improving thanks to the work of colleagues like **Guillaume (who I believe is in the audience today)**. For every new dataset, he writes a **curation report** to ensure the quality is constantly getting better.

Data shared by Poitiers

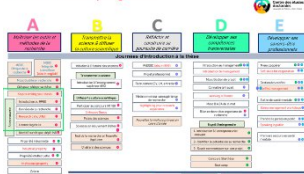


Our researchers use several repositories. The main ones are **Recherche Data Gouv, Nakala, and Zenodo**.

Recherche Data Gouv is quite new, and we've recently set up a dedicated space for our university's labs. In the past, some researchers got into the habit of depositing their data directly with their publishers.

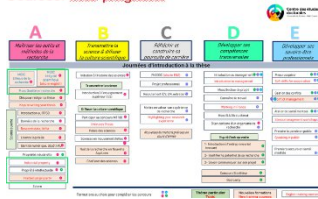
On the other hand, the **Humanities and Social Sciences** community has really embraced **Nakala**. We've even developed a **website generator** to help them showcase and promote their deposits.

Raising awareness and training in Open Science
Courses: PhD programs



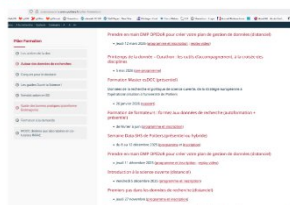
To ensure these Open Science practices stick, we created **dedicated courses** for our PhD students. We focus on key areas such as **data mangement, personal data, digital identity, and open-source/ software licensing**.

Raising awareness and training in Open Science
Courses: PhD programs



These modules **add to** our existing courses on **scientific integrity**, **industrial property**, and **copyright**.

Raising awareness and training in Open Science
Events



For the **rest of our research community**, we also organize various **events**, **workshops**, and **training sessions**.

Raising awareness and training in Open Science
Events and training



All of this is possible because **we work in a strong network**. Regarding training, we have a **close partnership** with the regional **URFIST** unit. Its director, **Raphaëlle**, is here and she's going to tell you more in a second.

Open Science: Access point



To conclude, you can find all the **ways to get in touch with us** right here. Thanks for listening!