



LIRIS



UMR 5205 CNRS

Comment transformer les données en connaissances ?

Les décodeuses du numérique, 7 janvier 2025, CNRS et Rectorat de Paris

Françoise CONIL, Ingénieure d'études CNRS, laboratoire LIRIS Lyon



Laboratoire d'InfoRmatique en Image et Systèmes d'information



cnrs



INSA



— université
— LUMIÈRE
— LYON 2



Attrait pour les sciences

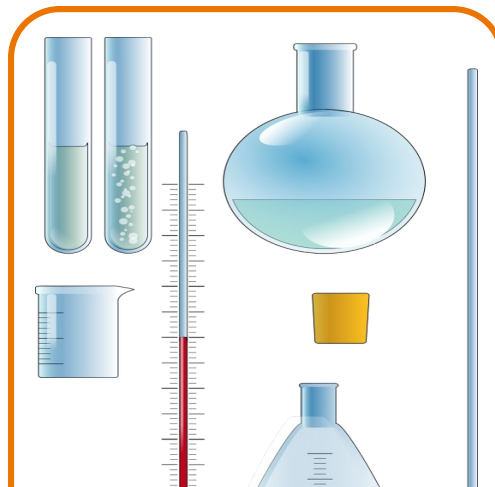


- Portrait « décodeuse » : <https://www.ins2i.cnrs.fr/fr/francoise-conil-coder-pour-un-monde-meilleur>
- « Amazing Grace » de Pablo Stanley : <http://compscigail.blogspot.com/2015/01/a-comic-about-grace-hopper.html>

Ingénieure en développement logiciel



Bac
Scientifique



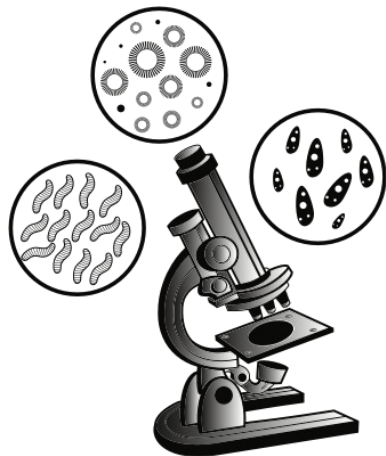
École
d'ingénieur



Télécom
puis CNRS

- Illustration « Bac scientifique » CC0 1.0 : <https://openclipart.org/detail/248081/reading>
- Illustration « École d'ingénieur » CC0 1.0 : <https://openclipart.org/detail/1739/chemistry-set>
- Illustration « Télécom puis CNRS » : © Léa Castor / CNRS Sciences informatique, <https://www.ins2i.cnrs.fr/fr/les-decodeuses-du-numerique>

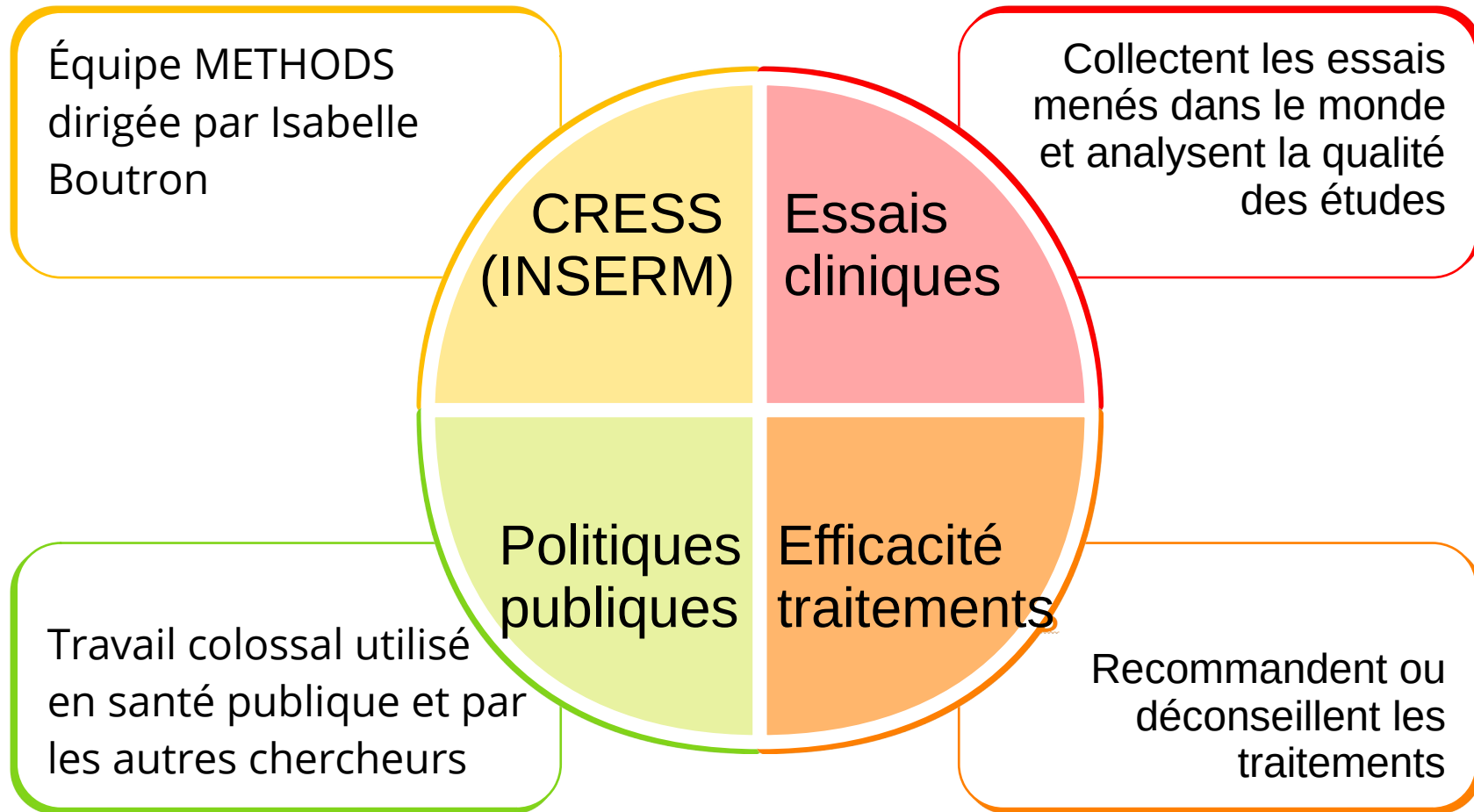
Intégration des données COVID-19



« L'objectif d'un **essai clinique** portant sur un **produit de santé** (médicament, dispositif ou thérapie cellulaire et génique) est d'évaluer la **sécurité** et l'**efficacité** de ce dernier chez des volontaires sains ou malades. »

- <https://www.inserm.fr/nos-recherches/recherche-clinique/essais-cliniques-recherches-interventionnelles-portant-sur-produit-sante/>
- https://fr.wikipedia.org/wiki/Essai_clinique
- Illustration CC0 1.0 : <https://openclipart.org/detail/297875/microscope-and-virus>

Évaluation des traitements médicaux



Dispersion : OMS data providers



- OMS : <https://www.who.int/clinical-trials-registry-platform/network/data-providers>

Systemes d'écriture

Le fichier CSV de l'OMS nous permet de récupérer les métadonnées pour des registres qui nous seraient plus compliquées à traiter.

お知らせ

重要なお知らせ

利用者の皆様 【重要】データ利用に関するお願い

↑ jRCTでは、公開されているデータの適正な利用をお願いしています。個人利用の範囲を超えた大量データ収集はお控えください。プログラムを利用した自動操作等による意図的な大量データ収集は個人利用の範囲を超えた利用とみなされま
す。皆様に快適にご利用いただくために、ご理解をよろしくお願いいたします。

利用者の皆様

↑ jRCT ほか国内の臨床試験情報登録センター（UMIN-CTR）に登録された臨床試験の情報を横断的に検索する場合は以下の URL をご利用ください。
rctportal.niph.go.jp

Registre japonais

1

بررسی اثربخشی و ایمنی تجویز داروی هیدروکسی کلروکین در پیشگیری از ابتلا به COVID-19 بعد از مواجهه با بیماران مبتلا به COVID-19

از ابتلا به COVID-19 بعد از مواجهه با بیماران مبتلا به COVID-19 طراحی: کارآزمایی بالینی تصادفی (...): در این مطالعه افراد با سن بیشتر از 18 سال که تماس نزدیک با بیمار مبتلا به COVID-19 در حداقل 4 روز اخیر (...) به COVID-19 مبتلا شدند، افرادی که علائم شبه آنفولانزا دارند و افرادی که سابقه مسافرت در دو هفته اخیر

IRCTID: IRCT20130917014693N10

کارآزمایی‌های مشابه

Registre iranien

Problèmes de qualité : métadonnées OMS

Quelques exemples de problèmes dans les métadonnées :

champ « Inclusion age min »		champ « Inclusion gender »		champ « Date registration »	
18 Years	5819	All	6813	20/04/2020	75
18	510	Both	2296	2020-03-15	33
18 years	487	 Female: yes Male: yes 	573	31-08-2020	15
no limit	93	Both, male and female	420		
18Y	93	Female	232		
20years-old	82	Male and Female	191		
>= 20age old	68	Both males and females	177		
no minimum age	57	-	147		
19 Year(s)	52	Male	98		
Not applicable	51	Male/Female	65		
1 Month	16	Females	7		
18?(Year)	8	 Female: yes Male: no 	4		
N/A (No limit)	8	BOTH	4		
0 Day(s)	2	F	3		
1months-old	1	B	2		
18 años	1	Not Specified	2		
2??	1	M	1		
0.1	1				

Problèmes de qualité : registre Européen

Un **export texte** plein, sans séparateur, **difficilement exploitable** :

Summary^M

EudraCT Number: 2021-004016-26^M

Sponsor's Protocol Code Number: Clin_COVID-19_Corok^M

National Competent Authority: Finland - Fimea ^M

Clinical Trial Type: EEA CTA^M

Trial Status: Ongoing^M

Date on which this record was first entered in the EudraCT database: 2021-09-14^M

Link: <https://www.clinicaltrialsregister.eu/ctr-search/trial/2021-004016-26/FI/>^M

^M

F.3 Group of trial subjects^M

F.3.1 Healthy volunteers: Yes^M

F.3.2 Patients: No^M

F.3.3 Specific vulnerable populations: No^M

F.3.3.1 Women of childbearing potential not using contraception ^M

(For clinical trials recorded in the database before the 10th March 2011 this question read: "Women of childbearing potential" and^M

did not include the words "not using contraception". An answer of yes could have included women of child bearing potential whether^M

or not they would be using contraception. The answer should therefore be understood in that context. This trial was recorded in the^M

database on 2021-09-14) : Yes^M

F.3.3.2 Women of child-bearing potential using contraception: Yes^M

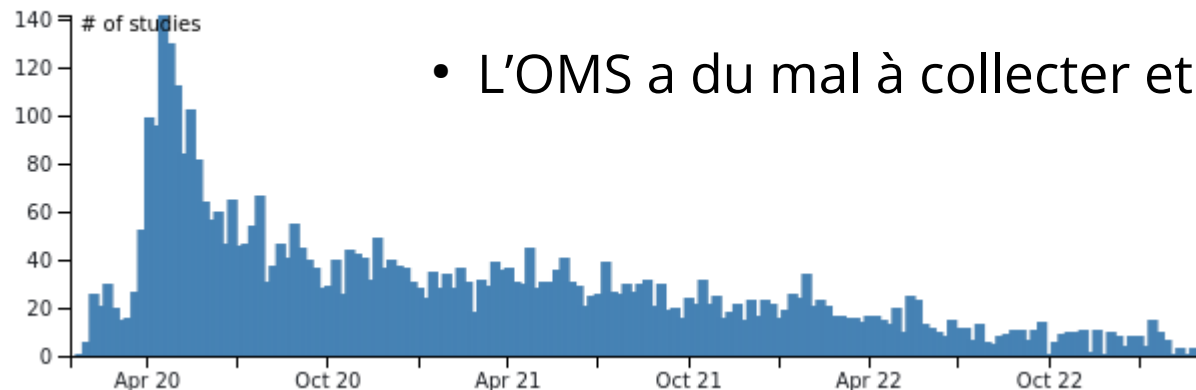
Traitements manuels et pandémie



En 2020, le **nombre d'essais cliniques explose**

- Impossible de traiter les essais manuellement ^[1]
- Les serveurs sont régulièrement inaccessibles
- L'OMS a du mal à collecter et à fournir les données

▼ Registration date



(1) 20 annotateurs mobilisés

- Graphique issu du site COVID-NMA : <https://covid-nma.com/>

- Illustration : <https://www.vecteezy.com/vector-art/14320081-cute-bacteria-germ-and-virus-characters> par Yuliya Pauliukevich sur Vecteezy

Automatisation informatique



Avril 2020

Appel à volontaires

Équipe pilotée par

Sarah Cohen-Boulakia



© Léa Castror / INRS21



Automatisation de la collecte

Au bout de 4 mois : la collecte et l'extraction sont automatisées

Il faudra presque **2 ans** pour améliorer la récupération et la transformation des données et ainsi que pour créer l'**entrepôt de données**.

Import Data USA

Result of last imports :



Date	Duration	Number of trials	Import status	Extracted, transformed and load	ETL duration
2024-03-27 20:00:03	00:04:54	9677	Success	Yes	00:09:22
2024-03-26 20:00:02	00:04:50	9674	Success	Yes	00:09:21
2024-03-25 16:00:02	00:04:46	9674	Success	Yes	00:09:29
2024-03-25 12:00:03	None	0	Fail	Yes	00:00:00

Entrepôt de données

Historiser les données, indiquer les **nouveaux essais**, les essais modifiées, les **données modifiées**

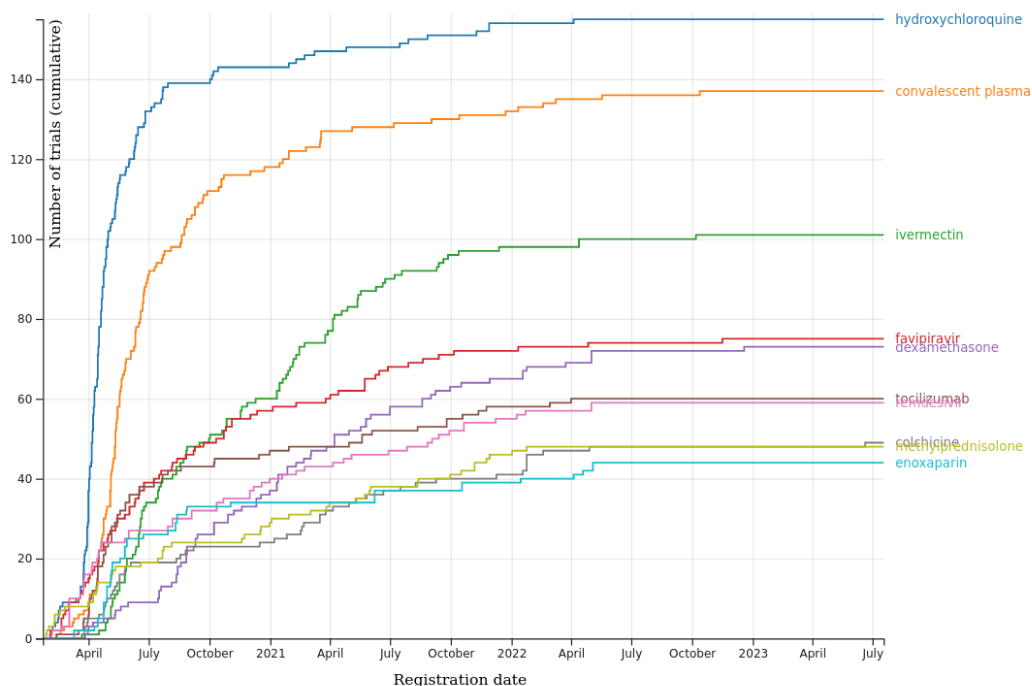
Les **annotateurs** travaillent directement dans l'entrepôt

- Highlighted colors informs about last update timing:
 - **red** mean updated in the last 7 days
 - **orange** mean updated in the last 30 days
 - **yellow** mean updated since more than 30 days
- Newly imported trials have **green** background color
- To identify which trial characteristics have been updated: click on the correspondent trial registration number to access the dedicated trial page. Then identify any column highlighted in yellow, orange or red
- To access history of changes: click on the red highlighted hyperlink last modified at

Trial Registration Number	First imported	Last refreshed	Last update	Full text link	Recruitment status	Actions
EUCTR2021-005059-35-IT	June 7, 2022, noon	March 7, 2024, 8 a.m.	March 7, 2024, 8 a.m.	https://www.clinicaltrialsregister.eu/ctr-search/trial/2021-005059-35/IT	Terminated	
EUCTR2020-002214-40-IT	March 1, 2022, 8:58 p.m.	March 7, 2024, 8 a.m.	March 7, 2024, 8 a.m.	https://www.clinicaltrialsregister.eu/ctr-search/trial/2020-002214-40/IT	Terminated	

Visualisations COVID-NMA

- L'automatisation, l'homogénéisation et la normalisation des données ont permis d'**analyse en continu des données**
- L'analyse au niveau mondial met en évidence les **études redondantes** (hydroxychloroquine)
- **Forte pression OMS qui utilise ces visualisations**
- La visualisation est faite par une autre équipe LIRIS



COVID-NMA : cartographie

► User Guide 📖

▼ Map



▼ Table

Show full table

Treatment (per arm)	Sample size	Severity at enrollment	Sponsor/Funder	Reg. number
(1) Nirmatrelvir + ritonavir vs (2) Nirmatrelvir + ritonavir vs (3) Placebo	900	Patients recovered from covid	Kanecia Obie Zimmerman	NCT05595389
(1) Imidazolyl ethanamide pentandioic acid vs (2) Placebo	233	Mild	Valenta Pharm JSC	NCT05216497
(1) Isotretinoin vs (2) Isotretinoin vs (3) Standard of care	100000	Critical	Kafrelsheikh University	NCT04353180
(1) Interferon gamma vs (2) Standard of care	630	Healthy volunteers	SPP Pharmacoin Ltd.	NCT05054114
(1) Fluticasone vs (2) Standard of care	500	Mild	University of Medicine and Pharmacy at Ho Chi Minh City	NCT05054322
(1) Baricitinib vs (2) Placebo	480	Moderate/severe	Incepta Pharmaceuticals Ltd	NCT05056558
(1) Platelet rich plasma vs (2) Placebo	30	Patients recovered from covid	Stanford University	NCT04406584
(1) Masitinib vs (2) Masitinib vs (3) Masitinib vs (4) Placebo	78	Mild/moderate	AB Science	NCT05047783

Download the data

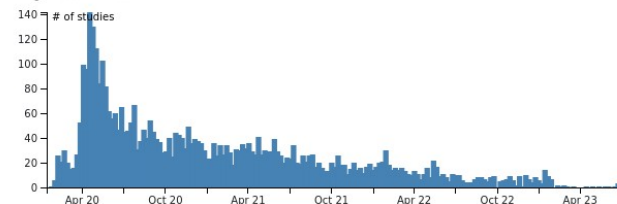
Filters

All trials selected (4256) | Reset all

Search...

Ex: Interferon, antiviral, Spain, Assistance Publique, EUCTR2020...

▼ Registration date

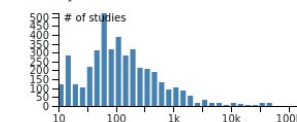


To filter by Registration dates, click and drag to create a range.

▼ Inclusion age

- minimum 18 yo (3,922 studies)
- less than 18 yo (330 studies)
- N/A (4 studies)

▼ Sample size



▼ Recruitment status

- Completed (1,574 studies) ⓘ
- Not recruiting (853 studies) ⓘ
- Recruiting (815 studies) ⓘ
- Terminated (403 studies) ⓘ
- Unknown (362 studies)
- Withdrawn (191 studies) ⓘ
- Not reported (35 studies)
- Suspended (23 studies) ⓘ

▼ Publication status

- Not published (3,427 studies)
- Published (829 studies) ⓘ

▼ Registry status

- No results posted (3,838 studies)
- Results posted in the registry (418 studies)

Logiciels libres et Science Ouverte

Les développements ont utilisé des **logiciels libres**.

Science Ouverte : La pandémie a mis en évidence l'importance d'avoir une **visibilité sur les essais cliniques** menés dans le monde, de **disposer des résultats**.

Projet récompensé par un cristal collectif



Les personnes qui ont participé au projet

Développeur·se·s

- Marinette BOUET (LIMOS – Clermont-Ferrand)
- Nicolas CHAMAND (stagiaire LIRIS – Lyon)
- Sarah COHEN-BOULAKIA (LISN – Paris)
- Françoise CONIL (LIRIS – Lyon)
- Emmanuel COQUERY (LIRIS – Lyon)
- Bastien DOREAU (LIMOS - Clermont-Ferrand)
- Merwan EL ASRI (stagiaire LRI - Paris)
- Benoît GROZ (LRI - Paris)
- Simon KLOPFENSTEIN (stagiaire LIRIS - Lyon)
- Ruben MARTINEZ (LIMOS - Clermont-Ferrand)
- Lamine MBOUP (stagiaire LIRIS - Lyon)
- Brice MEYER
- (stagiaire LIMOS - Clermont Ferrand)
- Virginie NOËL (LISN - Paris)
- Marie SAUVANT (stagiaire LIMOS - Clermont Ferrand)
- Farouk TOUMANI (LIMOS - Clermont-Ferrand)

CRESS

- Hillary BONNET (CRESS – Paris)
- Isabelle BOUTRON (CRESS – Paris)
- Gabriel FERRAND
- Thu Van NGUYEN
- Carolina RIVEROS (CRESS - Paris)

Logiciel : un produit complexe

- Navigateur Web ^[1] : ~10 million de lignes de code
- Logiciel embarqué dans une voiture moderne ^[1] : ~100 million de lignes de code
- Ensemble des logiciels de Google ^[1] : 2 milliard de lignes de code
- Pacemaker ^[2] : ~80 000 lignes de code
- Windows 3.1 (1992) ^[2] : ~2.5 million de lignes de code
- Microsoft Office 2001 ^[2] : ~25 million de lignes de code

Anne-Cécile Orgerie : éco-conception logicielle



© Léa Castor / INS21

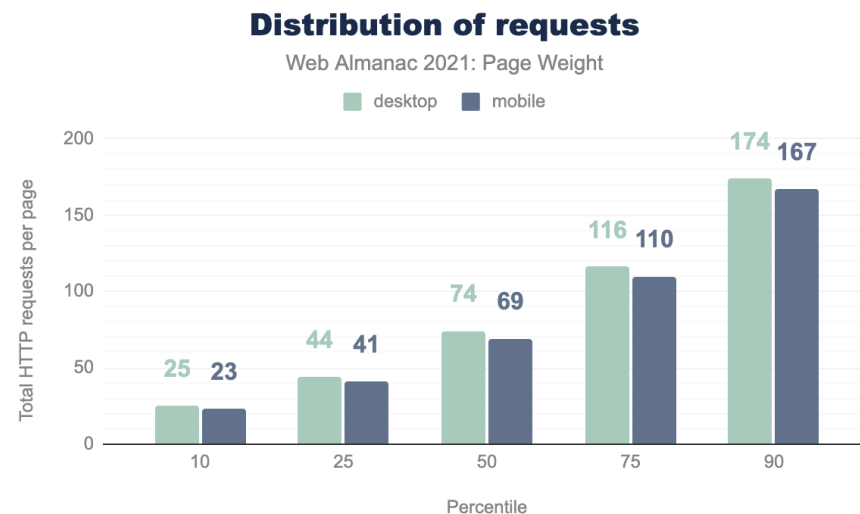
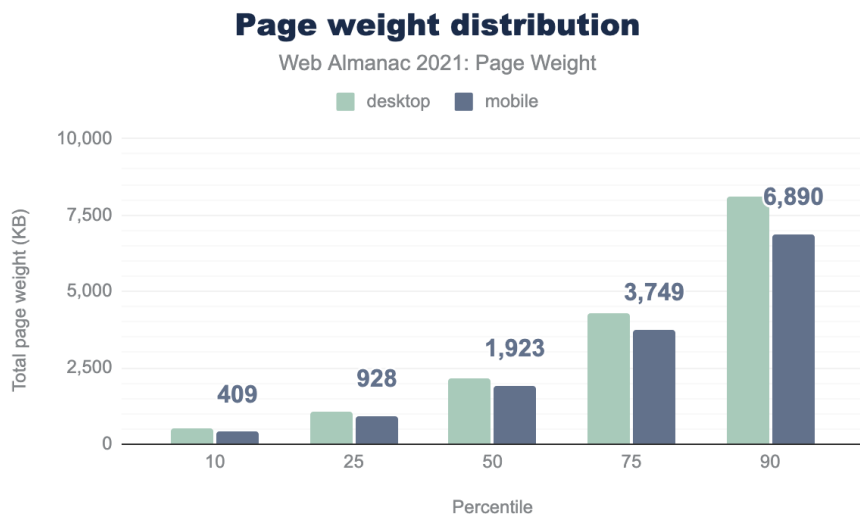
(1) <https://www.college-de-france.fr/agenda/lecon-inaugurale/le-logiciel-entre-esprit-et-la-matiere/le-logiciel-entre-esprit-et-la-matiere>

(2) <https://www.informationisbeautiful.net/visualizations/million-lines-of-code/>

Alourdissement du poids des pages web

En une dizaine d'années, le **poids** des pages web a **augmenté de 356 %**, passant de 484 ko à 2205 ko ^[1].

Impact en terme de **poids** et de **temps de calcul** qui sur la transmission, le traitement et le rendu d'une page.



(1) HTTP archive, web almanach, décembre 2021 : <https://almanac.httparchive.org/en/2021/page-weight>

Respect des données personnelles

Un des avantages des **logiciels libres** est un meilleur respect des données personnelles puisque leur code source peut être étudié.