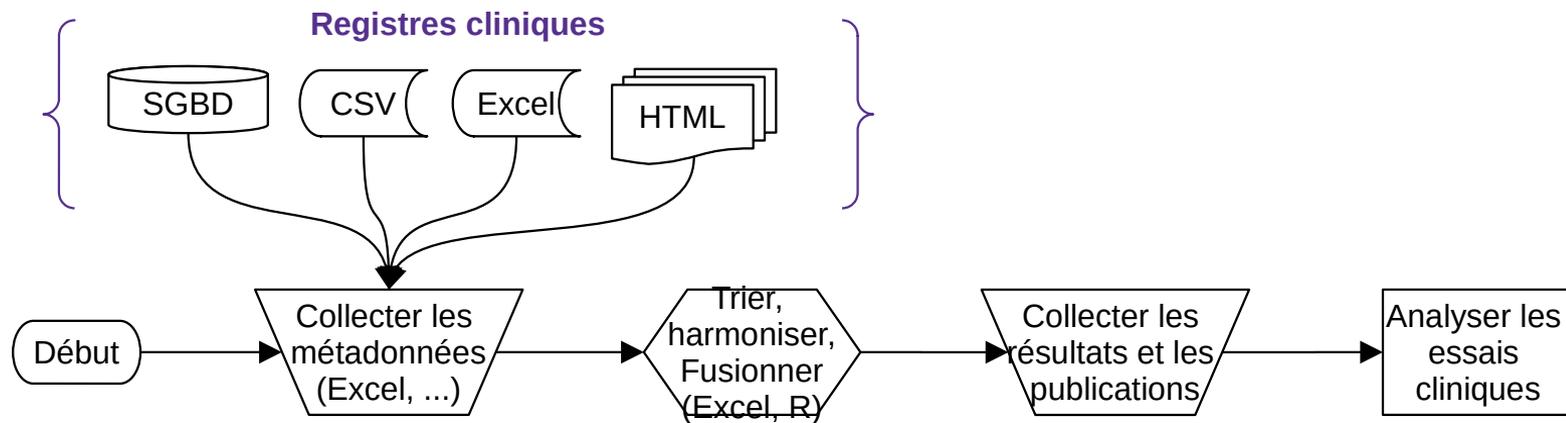


# Extraction des métadonnées des essais cliniques sur le COVID-19

Françoise CONIL, AG APRIL, 16 mars 2024

# Le contexte : analyse des essais cliniques

- Équipe METHODS <sup>[1]</sup> du CRESS (Centre of Research in Epidemiology and Statistics)
- Méta-analyse des essais cliniques menées dans le monde
- Objectif : **Analyser la qualité des études et l'efficacité des traitements testés**
- Beaucoup d'étapes manuelles



[1] <https://cress-umr1153.fr/index.php/methods/>

# Déclaration des essais cliniques : obligation

- La déclaration de tous les essais « interventionnel » est considérée comme une **obligation scientifique, éthique et morale** <sup>[1]</sup>
- La **déclaration d'Helsinki** <sup>[2]</sup> stipule que chaque essai devrait être enregistré dans une **base de données publiquement accessible** avant le recrutement du premier patient
- Les essais sont obligatoirement déposés dans un registre
- Les essais sont dispersés dans plusieurs registres avec des doublons entre les registres
- L'OMS collecte les données de registres (data providers <sup>[4]</sup>) qui respectent un ensemble de critères <sup>[3]</sup>

[1] OMS : <https://www.who.int/news-room/questions-and-answers/item/clinical-trials>

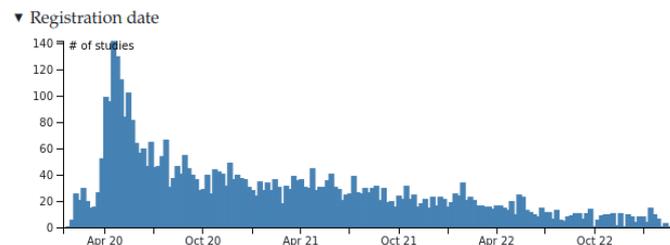
[2] <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>  
(points 34 et 35)

[3] <https://www.who.int/clinical-trials-registry-platform/network/who-data-set/data-providers>

[4] <https://www.who.int/clinical-trials-registry-platform/network/registry-criteria>

# La pandémie COVID-19

- **Explosion des essais cliniques** sur le COVID-19
- 20 annotateurs à temps plein complètent, analysent et synthétisent ces données
- Le CRESS fait part de son besoin d'automatisation de tâches
  
- Avril 2020 : le CNRS lance un appel à volontaires
- Une équipe hétérogène et dispersée se constitue



- [1] CRESS : <https://cress-umr1153.fr/>  
[2] LRI => LISN : <https://www.lisn.upsaclay.fr/>  
[3] LIMOS : <https://limos.fr/>  
[4] LIRIS : <https://liris.cnrs.fr/>

# Registres d'essais cliniques : pb de qualité

- Standards sur l'échange de données provenant d'études cliniques <sup>[1]</sup> mais pas sur les métadonnées concernant ces études <sup>[2]</sup>
- Registres plus ou moins détaillés, **données non homogènes selon les registres**, accès variables et registres parfois inaccessibles (se produit encore)
- Nombreuses publications sur les **problèmes de qualité de données** dans le registre OMS et dans les registres primaires et depuis longtemps
  - Obstacles to the reuse of study metadata in ClinicalTrials.gov <sup>[2]</sup>
  - The Quality of Registration of Clinical Trials: Still a Problem <sup>[3]</sup>
  - Exploring Data Quality Management within Clinical Trials <sup>[4]</sup>

[1] <https://www.cdisc.org/standards/data-exchange/define-xml>

[2] <https://www.nature.com/articles/s41597-020-00780-z>

[3] <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0084727>

[4] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5801732/>

# Hétérogénéité des sources / registres

- **OMS** : fichier CSV
- **USA** : accès à une base PostgreSQL
- **Europe** : scraping
- **Iranien** : XML et scraping
- Indien : scraping
- Australien : XML et Excel
- Anglais : API qui délivre du XML

*En gras : les sources qui sont actuellement utilisée en production*

**OMS** (ko) : <https://www.who.int/clinical-trials-registry-platform>

**USA** : <https://clinicaltrials.gov>

**Europe** (va bientôt changer) : <https://www.clinicaltrialsregister.eu/>

**Iranien** (changé) : <https://en.irct.ir>

Indien : <http://ctri.nic.in/Clinicaltrials/advancesearchmain.php>

Australien : <https://www.anzctr.org.au/TrialSearch.aspx>

Anglais : <https://www.isrctn.com/>

# Pb qualité des données : source OMS

Quelques exemples de valeurs, toutes les variations ne sont pas présentes !

champ « <b>Inclusion age</b> »		champ « <b>Inclusion gender</b> »		champ « <b>Date registration</b> »	
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				

# Pb qualité des données : registre Européen

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

Un export texte plein, sans séparateur, difficilement exploitable

[1] <https://www.clinicaltrialsregister.eu/ctr-search/rest/download/trial/2021-004016-26/FI>

# Scrapping du registre Européen

## Summary

EudraCT Number:	2021-004016-26
Sponsor's Protocol Code Number:	Clin_COVID-19_Corok
National Competent Authority:	Finland - Fimea
Clinical Trial Type:	EEA CTA
Trial Status:	Ongoing
Date on which this record was first entered in the EudraCT database:	2021-09-14

## A. Protocol Information

A.1	Member State Concerned	Finland - Fimea
A.2	EudraCT number	2021-004016-26
A.3	Full title of the trial	Substudy "Responses to Covid-19 vaccines" in research "Clinical picture, immunology, genetics and pathogenesis of COVID-19 infection" "COVID-19 infektion taudinkuva, immuunivaste, genetiikka ja patogeneesi"- tutkimuksen alatutkimus "Koronarokotteen aikaan saama vaste"

# Objectif : créer un entrepôt de données

- Automatiser la récupération, l'extraction et la transformation des données depuis les sources
- Historiser les données et indiquer les nouveaux essais, les essais modifiées, les données modifiées
- Permettre aux annotateurs de travailler directement dans l'entrepôt

Current available trials in the data warehouse : **4964** --> [Explore/Export trials](#)

Sources are currently :

- **EU CLINICAL TRIALS REGISTER** : <https://www.clinicaltrialsregister.eu> - **422** entries (last refreshed on **2024-03-07 08:00:05**)  
(last successfull import date : 2024-03-07 08:00:05 --> 1503 entries)
- **INTERNATIONAL CLINICAL TRIALS REGISTRY PLATFORM (ICTRP)** : <https://www.who.int/clinical-trials-registry-platform> - **958** entries (last refreshed on **2022-07-07 13:30:02**)  
(last successfull import date : 2024-03-06 16:14:56 --> 7293 entries)
- **IRANIAN REGISTRY OF CLINICAL TRIALS (IRCT)** : <https://en.irct.ir> - **513** entries (last refreshed on **2023-12-28 04:00:02**)  
(last successfull import date : 2024-01-12 12:00:02 --> 1062 entries)
- **CLINICALTRIALS.GOV** : <https://aact.ctti-clinicaltrials.org> - **3071** entries (last refreshed on **2024-03-06 16:00:03**)  
(last successfull import date : 2024-03-06 16:00:03 --> 9649 entries)

# Entrepôt : historisation 1/2

- 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.	<a href="https://www.clinicaltrialsregister.eu/ctr-search/trial/2021-005059-35/IT">https://www.clinicaltrialsregister.eu/ctr-search/trial/2021-005059-35/IT</a>	Terminated	
EUCTR2020-002214-40-IT	March 1, 2022, 8:58 p.m.	March 7, 2024, 8 a.m.	March 7, 2024, 8 a.m.	<a href="https://www.clinicaltrialsregister.eu/ctr-search/trial/2020-002214-40/IT">https://www.clinicaltrialsregister.eu/ctr-search/trial/2020-002214-40/IT</a>	Terminated	

# Entrepôt : historisation 2/2

Center ?

single-center

Last imported at : March 15, 2023,  
4 a.m.

Source : [ClinicalTrials.gov](#)

Study aim ?

Long covid

Last imported at : Oct. 26, 2020,  
11:31 p.m.

Source : [ClinicalTrials.gov](#)

Inclusion criteria ?

Last imported at : March 2, 2024,  
8 a.m.

Source : [ClinicalTrials.gov](#)

inclusion criteria: screening procedures will be performed as part of this study. in order to be eligible to participate in this study, an individual must meet all of the following criteria: male or female, aged 18 to 80 years previous infection with sars-cov2 at least 4 weeks prior to enrollment, confirmed by laboratory or a healthcare provider. participants with a documented positive home antigen test will need confirmation of positive covid-19 antibodies (positive anti-nucleocapsid antibody test, or positive anti-spike antibody test if unvaccinated at the time of the antibody test) to be considered. presence of physical limitations or significant fatigue since covid-19 as demonstrated by: total score  $\leq 19$  on the promis short form for physical function or total score  $\geq 9$  on the promis short form for fatigue, and score  $\geq 1$  on the patient global rating of flu severity and patient global assessment of interference with daily activities absence of post-exertional malaise (pem) as

# Entrepôt : technologies utilisées

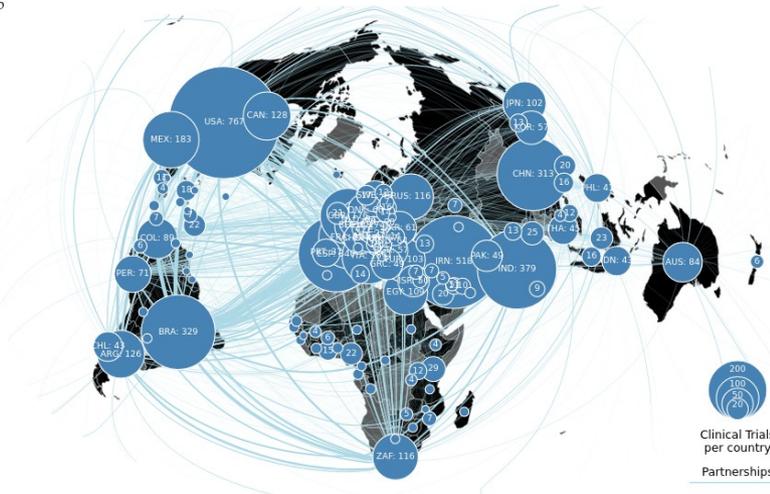
- Framework Django et langage Python
- Différentes librairies Python : Requests, Beautiful Soup, Pandas, Jupyter Notebooks, lxml, Selenium
- Base de données PostgreSQL
- Déploiement sur VM Proxmox avec Ansible, Docker pour le développement

Démarche d'ouverture du code en cours.

# COVID-NMA: cartographe et analyser des essais cliniques 1/3

► User Guide

▼ Map



Download the data

▼ 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	NCT05595369
(1) Imidazolethyl ethanamide pentanoic acid vs (2) Placebo	233	Mild	Valentia Pharm JSC	NCT05216497
(1) Isotretinoin vs (2) Isotretinoin vs (3) Standard of care	100000	Critical	Katereisheikh University	NCT04353180
(1) Interferon gamma vs (2) Standard of care	630	Healthy volunteers	SPP Pharmacion 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

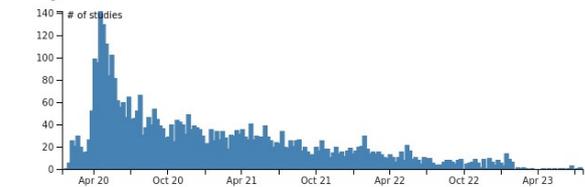
## Filters

All trials selected (4256) | Reset all

Search...

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

▼ Registration date

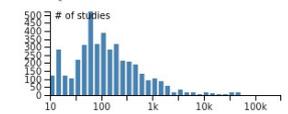


To filter by Registration dates, click and drag to create a range.

▼ 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)

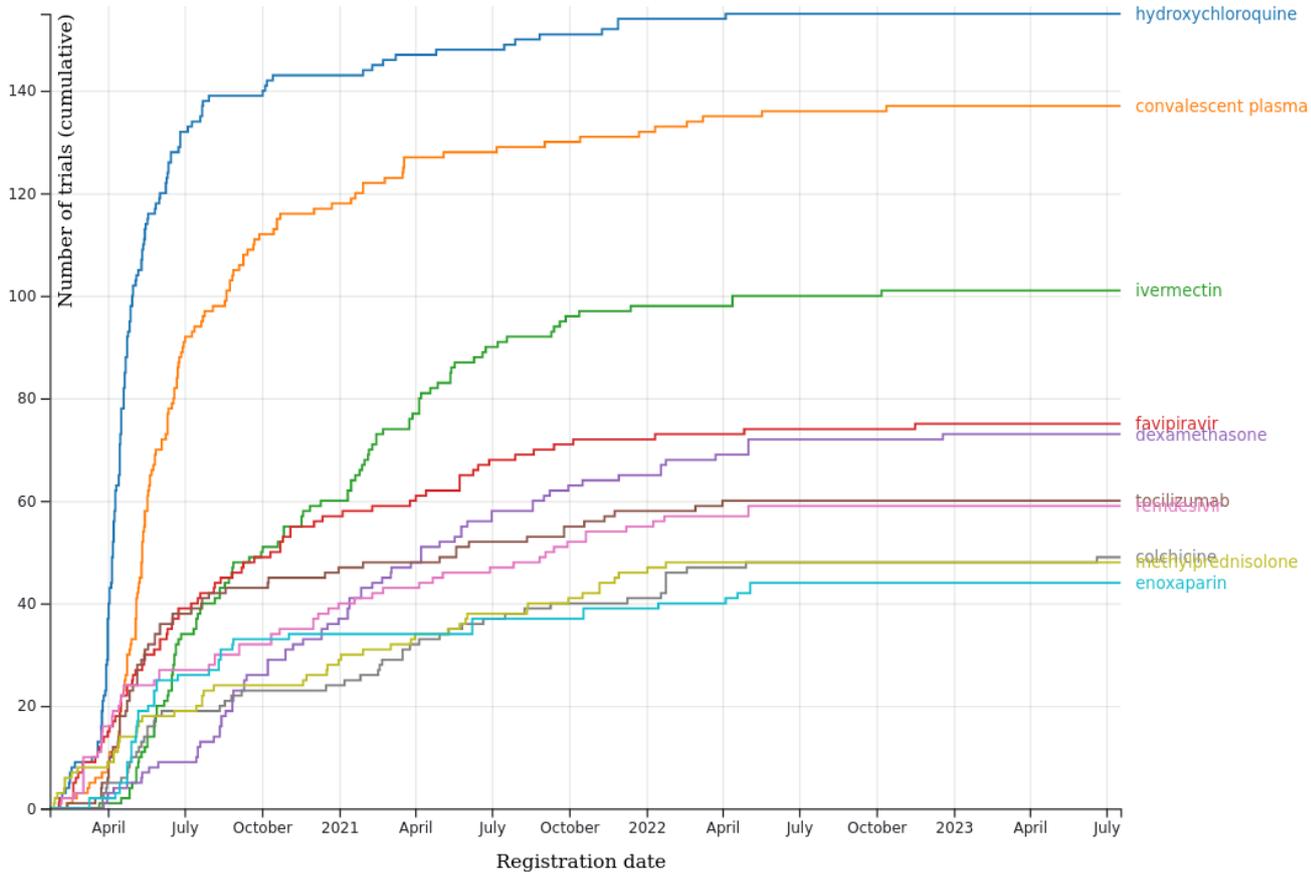
▼ Publication status

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

▼ Registry status

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

# COVID-NMA: cartographier et analyser des essais cliniques 2/3



# COVID-NMA: cartographier et analyser des essais cliniques 3/3

<p>NCT04359810</p> <p>O'Donnell M, J Clin Invest, 2021</p> <p><a href="#">Full text</a></p> <p><a href="#">Commentary</a></p> <p><a href="#">Commentary</a></p>	Public/non profit	Convalescent plasma	Standard plasma	RCT	<p>Patients with confirmed COVID-19 (mild-critical) admitted to 5 centers in Brazil and USA</p>	N=223	<p> Low</p> <p><a href="#">Details</a></p>	<a href="#">Full description</a>
<p>IRCT20151227025726N24</p> <p>Tolouian R, J Investig Med, 2021</p> <p><a href="#">Full text</a></p> <p><a href="#">Commentary</a></p> <p><a href="#">Commentary</a></p>	No specific funding	Bromhexine hydrochloride	Standard care	RCT	<p>Patients with confirmed COVID-19 pneumonia (symptomatic, unclear severity) admitted to a single center in Iran</p>	N=111	<p> High</p> <p><a href="#">Details</a></p>	<a href="#">Full description</a>
<p>NCT04486508</p> <p>INSPIRATION</p> <p>Bikdeli B, Thromb Haemost, 2021</p> <p><a href="#">Full text</a></p> <p><a href="#">Full text</a></p> <p><a href="#">Commentary</a></p>	Mixed	Intermediate-Dose prophylactic anticoagulation	Standard-Dose prophylactic anticoagulation	RCT	<p>Patients with confirmed COVID-19 (critical) admitted to 10 centers in Iran</p>	N=598	<p> Some concerns</p> <p><a href="#">Details</a></p>	<a href="#">Full description</a>

# Un réel travail collaboratif

- On ne se connaît pas, on est dispersés, on est confinés
- On ne connaît pas le domaine

Il nous faut des outils de communication et de partage

- Chat : Rocket chat avec fédération d'identité
- Visio : Jitsi, Webex, BBB,
- GitLab, NextCloud

Définition de règles communes

- Formater : Black
- Linter : Flake8
- Documentation développeur sur le Wiki du GitLab
- Utilisation des issues et des merge request

# Première livraison au bout de 4 mois

- Déclenchement des collectes par CRON
- Données brutes enregistrées sur disque jusqu'à la prochaine collecte
- Stockage des données collectées dans la base PostgreSQL
- Export CSV / Excel pour chaque source via une interface Web
- On doit reprendre nos projets / enseignements
- Recrutement d'un développeur qui va améliorer ce prototype pendant 2 ans

# Aujourd'hui

- Toutes les données brutes sont conservées en évitant de stocker des données redondantes
- Les données ont été harmonisées grâce à des tables que le CRESS peut renseigner
- Un historique détaillé des modifications des essais existe
- Le déploiement est automatisé (Ansible, Docker)
- Toutes les annotations sont faites dans l'entrepôt directement
- Plus personne pour le développement : ouvrir le code en open source permettrait d'éviter qu'il disparaisse grâce à Software Heritage <sup>[1]</sup>

[1] <https://www.softwareheritage.org/>

# Projet récompensé par un cristal collectif CNRS



# Tous ceux qui ont participé à ce 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)