Antoine Lacassagne
CENTRE DE LUTTE CONTRE LE CANCER DE NICE

OncoPaca-Corse
DISPOSITIF SPÉCIFIQUE RÉGIONAL DU CANCER

Utilisations de l'intelligence artificielle dans la recherche clinique en oncologie

Ateliers Recherche Clinique DSRC OncoPaca-Corse - 9 Avril 2026

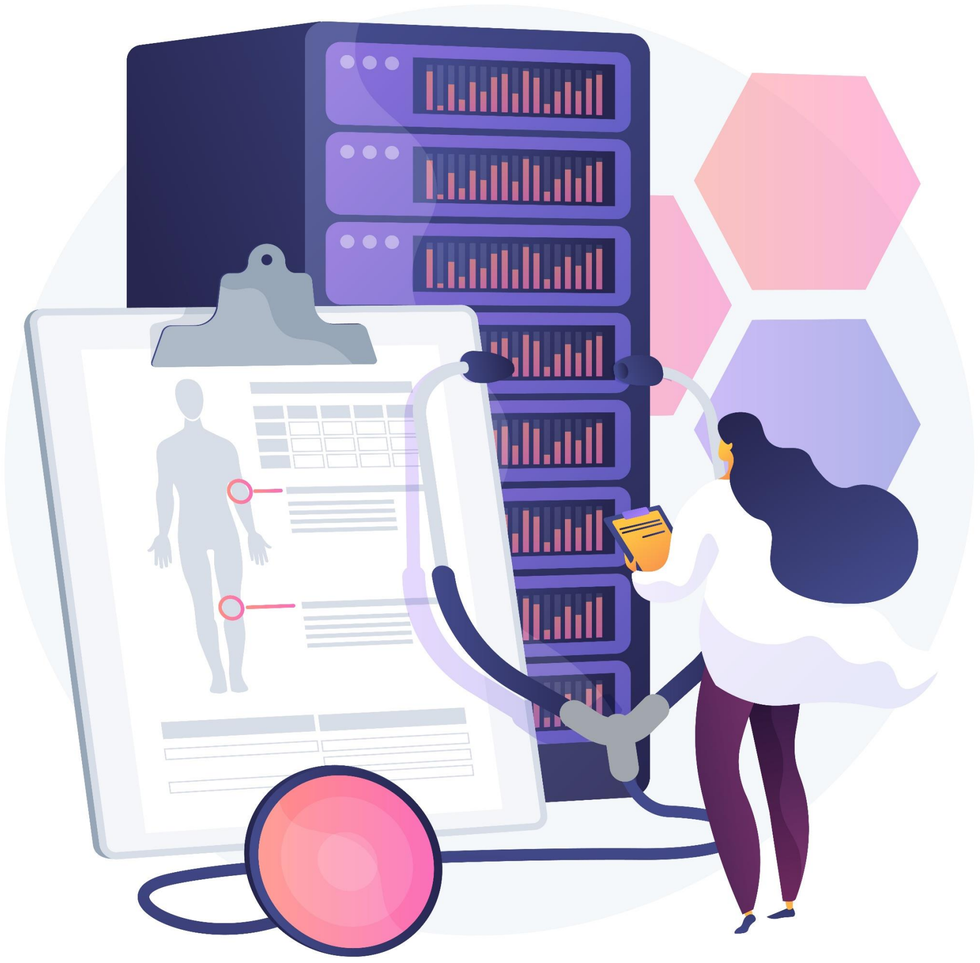
Renaud Schiappa – Responsable Adjoint
Département d'Epidemiologie, Biostatistique et Données de Santé – Pr Chamorey

Antoine Lacassagne
CENTRE DE LUTTE CONTRE LE CANCER
unicancer NICE

Utilisation de l'intelligence artificielle en recherche clinique

OncoPaca-Corse
DISPOSITIF SPÉCIFIQUE RÉGIONAL DU CANCER

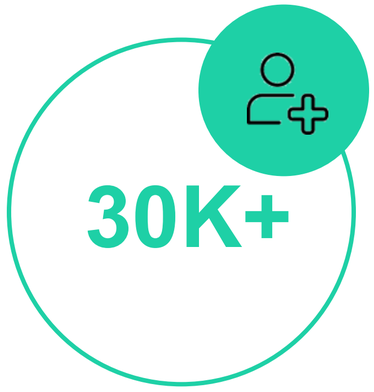
Comment sélectionner les patients ?



Applications pour la recherche clinique

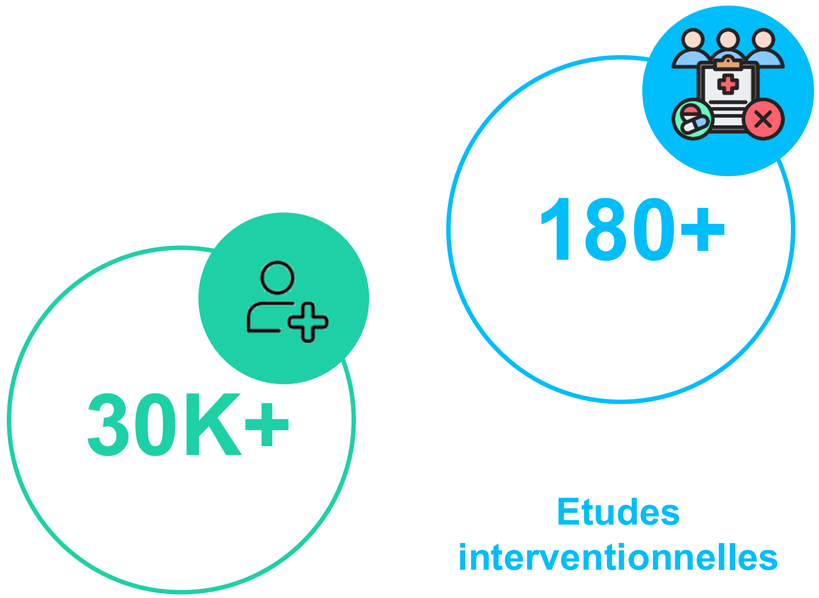
Comment sélectionner les patients ?

Comment sélectionner les patients ?



Patients par année

Comment sélectionner les patients ?



Comment sélectionner les patients ?

180+

Etudes interventionnelles

30K+

Patients par année

2000+

Critères d'inclusions

Comment sélectionner les patients ?

30K+

Patients par année

180+

Etudes interventionnelles

2000+

Critères d’inclusions

2800+

Critères de non inclusion

Comment sélectionner les patients ?

30K+

Patients par année

180+

Etudes interventionnelles

2000+

Critères d’inclusions

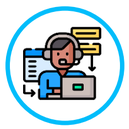
2800+

Critères de non inclusion

20%

Données structurées

Comment sélectionner les patients ?



Manuellement

RegEx

NLP

LLM

Comment sélectionner les patients ?

Manuellement

Comment sélectionner les patients ?

RegEx

« une expression régulière est une chaîne de caractères qui décrit, selon une syntaxe précise, un ensemble de chaînes de caractères possibles. »

Comment sélectionner les patients ?

RegEx

« une expression régulière est une chaîne de caractères qui décrit, selon une syntaxe précise, un ensemble de chaînes de caractères possibles. »

Valeur brute dans le texte
ECOG : 0
ECOG 2
IK : 90%
OMS = 0
OMS 0
OMS :0
Performance status 2
PS = 2
PS à 0

Comment sélectionner les patients ?

RegEx

« une expression régulière est une chaîne de caractères qui décrit, selon une syntaxe précise, un ensemble de chaînes de caractères possibles. »

Valeur brute dans le texte	
ECOG : 0	"ecog[ ]*[:=àa]?[ ]*[0-5]"
ECOG 2	
IK : 90%	"oms[ ]*[:=àa]?[ ]*[0-5]"
OMS = 0	"i?n?d?i?c?e?[ ]*d?e?[ ]*karnofsky[ ]*[:=àa]?[ ]*[0-9][0-9]?%"
OMS 0	"IKA?[ ]*[:=aà]?[ ]*[0-9][0-9]?%"
OMS :0	"KPS[ ]*[:=àa]?[ ]*[0-9][0-9]?%"
Performance status 2	
PS = 2	
PS à 0	

"PS[ ]*[:=àa]?[ ]*[0-5]"

"perform[ae]nce[ ]*status[ ]*[:=àa]?[ ]*[0-5]"

Comment sélectionner les patients ?

RegEx

« une expression régulière est une chaîne de caractères qui décrit, selon une syntaxe précise, un ensemble de chaînes de caractères possibles. »

Valeur brute dans le texte		Valeur Base de données
ECOG : 0	"ecog[ ]*[:=àa]?[ ]*[0-5]"	0
ECOG 2		2
IK : 90%	"oms[ ]*[:=àa]?[ ]*[0-5]"	0
OMS = 0	"i?n?d?i?c?e?[ ]*d?e?[ ]*karnofsky[ ]*[:=àa]?[ ]*[0-9][0-9]?%"	0
OMS 0	"IKA?[ ]*[:=aà]?[ ]*[0-9][0-9]?%"	0
OMS :0	"KPS[ ]*[:=àa]?[ ]*[0-9][0-9]?%"	0
Performance status 2		2
PS = 2		2
PS à 0		0

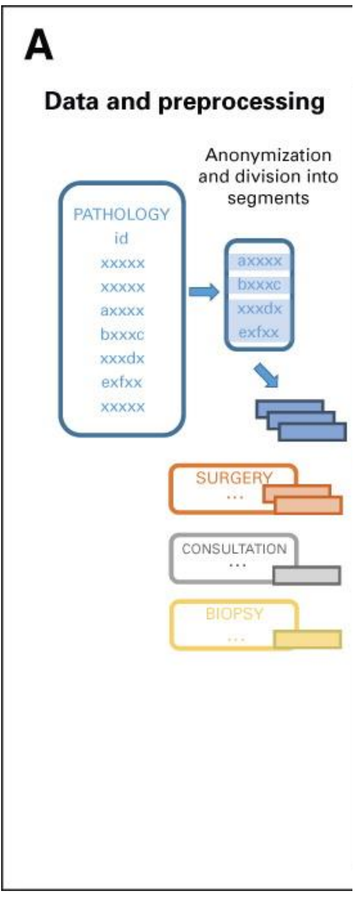
"PS[ ]*[:=àa]?[ ]*[0-5]"

"perform[ae]nce[ ]*status[ ]*[:=àa]?[ ]*[0-5]"

Comment sélectionner les patients ?

Capacité à traiter le langage naturel

NLP



Schiappa et al., RUBY: Natural Language Processing of French Electronic Medical Records for Breast Cancer Research. *JCO Clin Cancer Inform, 2022*

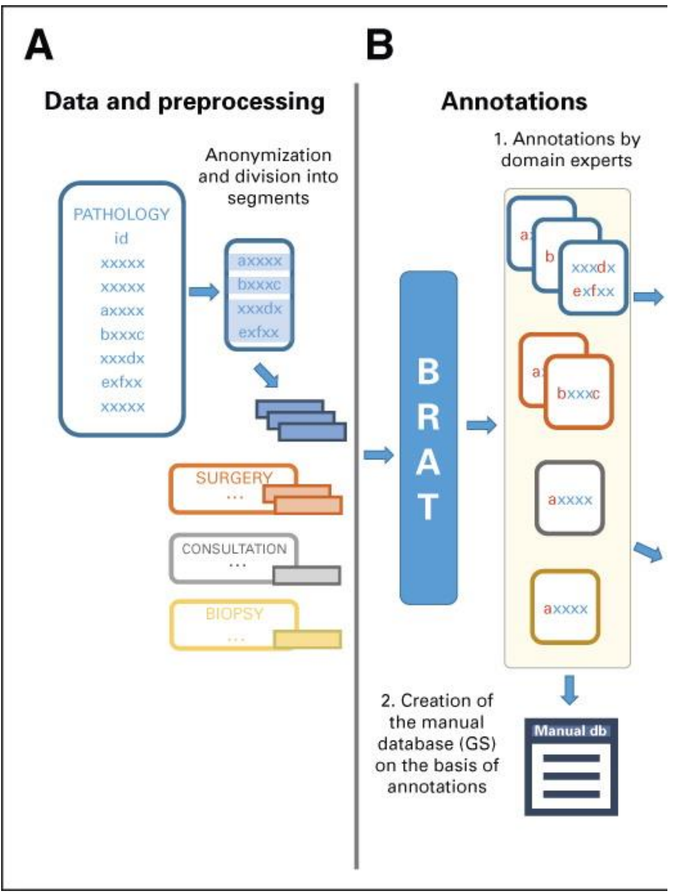
Schiappa et al., Validation of RUBY for Breast Cancer Knowledge Extraction From a Large French Electronic Medical Record System. *JCO Clin Cancer Inform, 2023*

Culie et al, Enhancing Thyroid Pathology With Artificial Intelligence: Automated Data Extraction From Electronic Health Reports Using RUBY. *JCO Clin Cancer Inform, 2024*

Comment sélectionner les patients ?

Capacité à traiter le langage naturel

NLP



Schiappa et al., RUBY: Natural Language Processing of French Electronic Medical Records for Breast Cancer Research. *JCO Clin Cancer Inform, 2022*

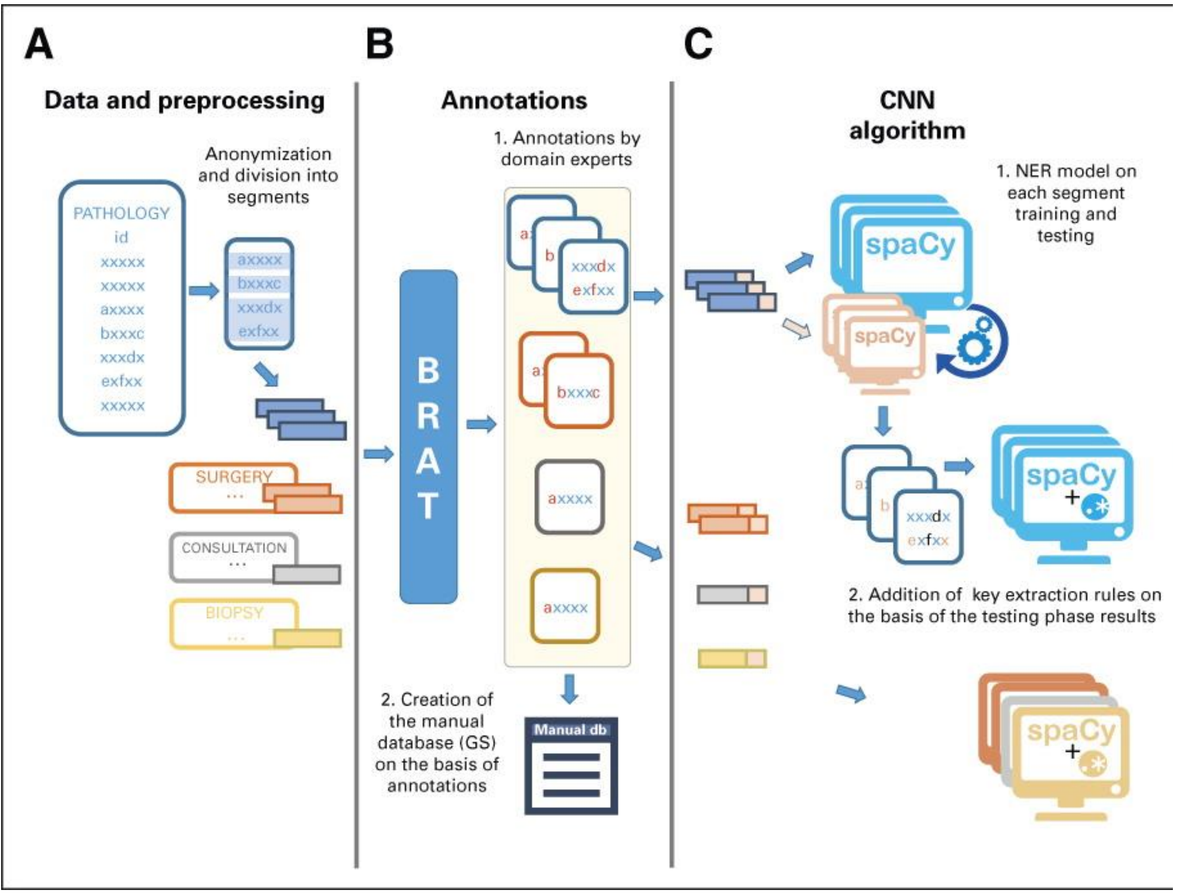
Schiappa et al., Validation of RUBY for Breast Cancer Knowledge Extraction From a Large French Electronic Medical Record System. *JCO Clin Cancer Inform, 2023*

Culie et al, Enhancing Thyroid Pathology With Artificial Intelligence: Automated Data Extraction From Electronic Health Reports Using RUBY. *JCO Clin Cancer Inform, 2024*

Comment sélectionner les patients ?

Capacité à traiter le langage naturel

NLP



Schiappa et al., RUBY: Natural Language Processing of French Electronic Medical Records for Breast Cancer Research. *JCO Clin Cancer Inform, 2022*

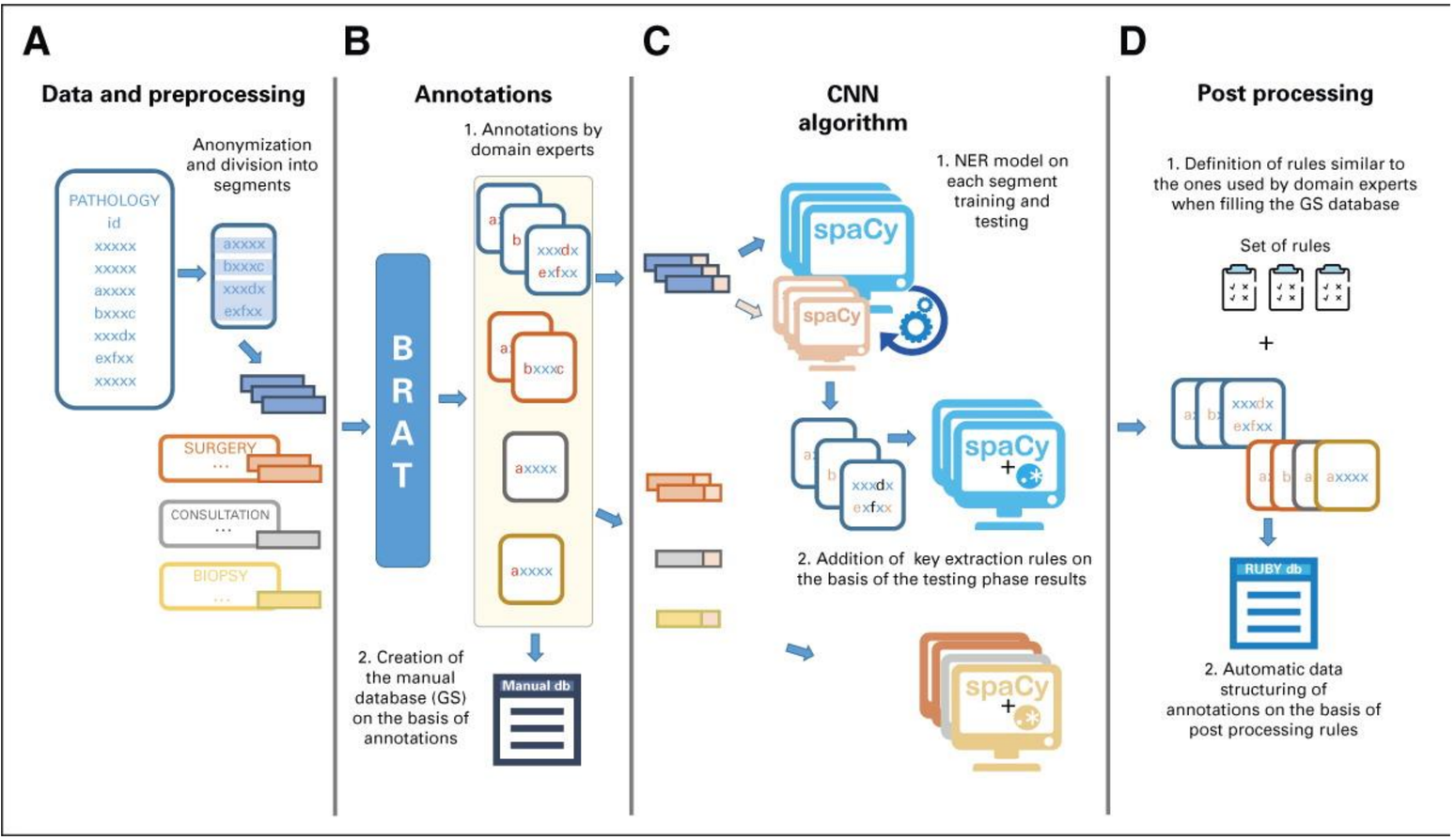
Schiappa et al., Validation of RUBY for Breast Cancer Knowledge Extraction From a Large French Electronic Medical Record System. *JCO Clin Cancer Inform, 2023*

Culie et al, Enhancing Thyroid Pathology With Artificial Intelligence: Automated Data Extraction From Electronic Health Reports Using RUBY. *JCO Clin Cancer Inform, 2024*

Comment sélectionner les patients ?

Capacité à traiter le langage naturel

NLP



Schiappa et al., RUBY: Natural Language Processing of French Electronic Medical Records for Breast Cancer Research. *JCO Clin Cancer Inform, 2022*

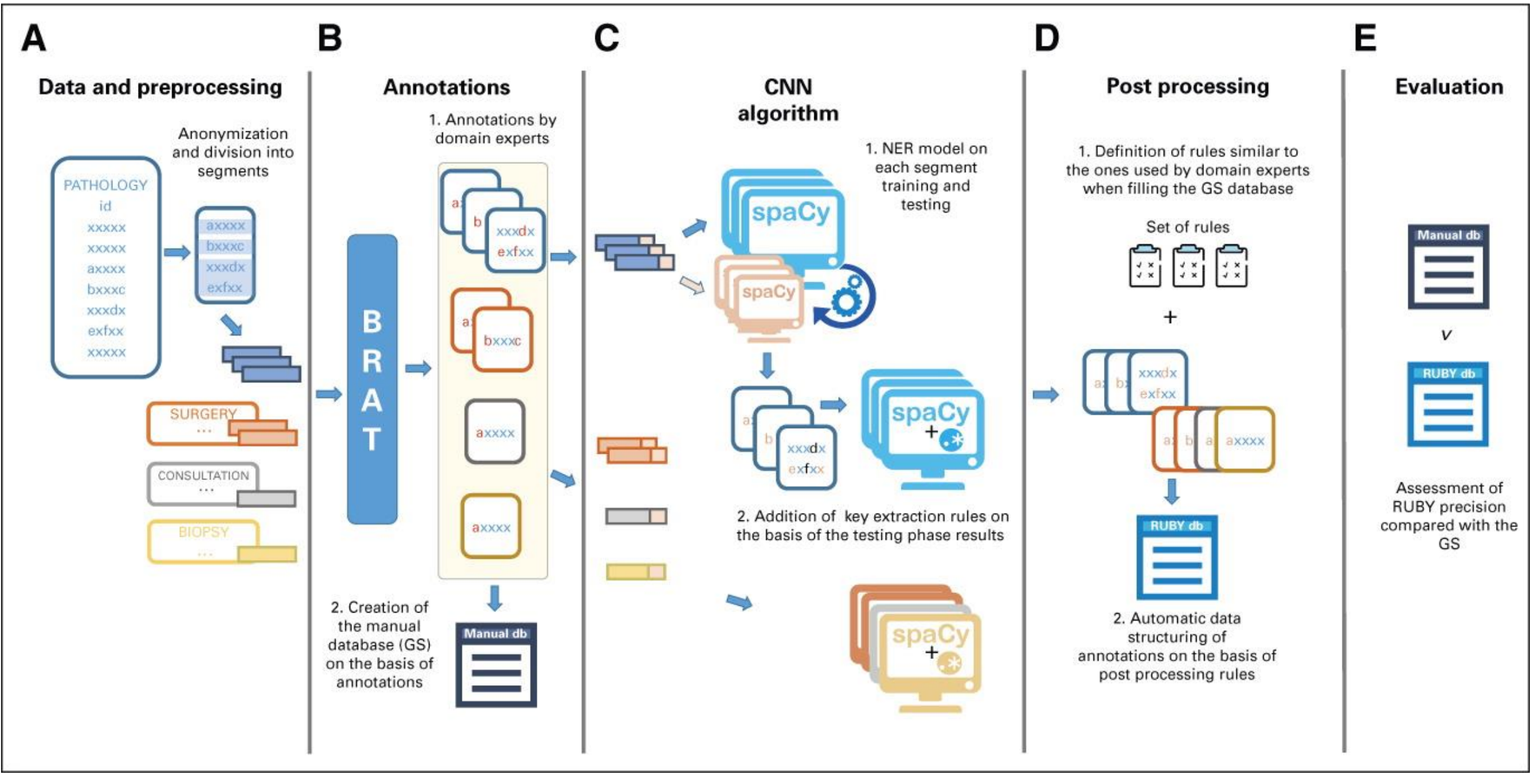
Schiappa et al., Validation of RUBY for Breast Cancer Knowledge Extraction From a Large French Electronic Medical Record System. *JCO Clin Cancer Inform, 2023*

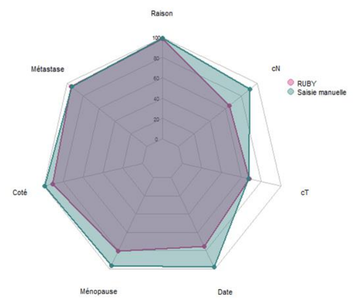
Culie et al, Enhancing Thyroid Pathology With Artificial Intelligence: Automated Data Extraction From Electronic Health Reports Using RUBY. *JCO Clin Cancer Inform, 2024*

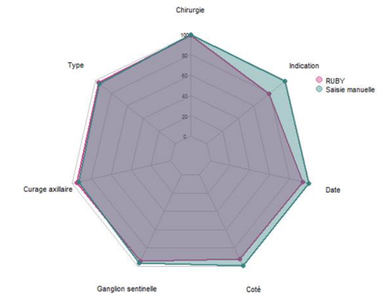
Comment sélectionner les patients ?

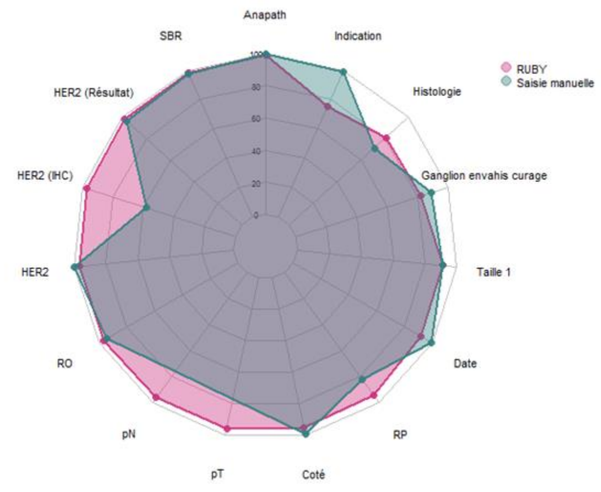
Capacité à traiter le langage naturel

NLP









Schiappa et al., RUBY: Natural Language Processing of French Electronic Medical Records for Breast Cancer Research. *JCO Clin Cancer Inform, 2022*

Schiappa et al., Validation of RUBY for Breast Cancer Knowledge Extraction From a Large French Electronic Medical Record System. *JCO Clin Cancer Inform, 2023*

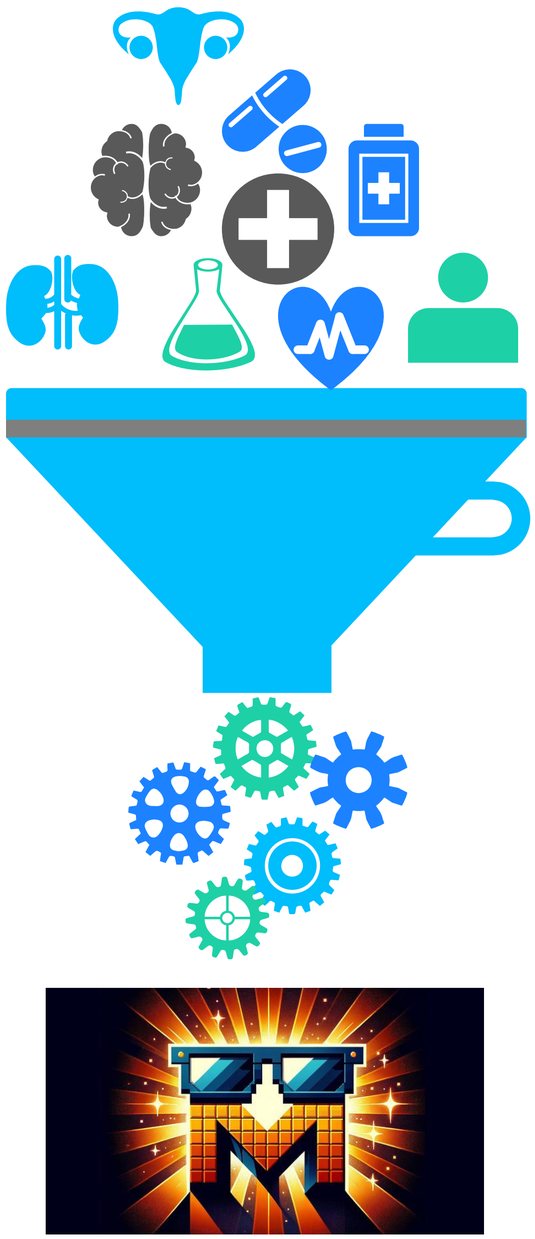
Culie et al, Enhancing Thyroid Pathology With Artificial Intelligence: Automated Data Extraction From Electronic Health Reports Using RUBY. *JCO Clin Cancer Inform, 2024*

Comment sélectionner les patients ?

Capacité à comprendre, résumer, générer et prédire de nouveaux contenus

LLM

Comment sélectionner les patients ?



Capacité à comprendre, résumer, générer et prédire de nouveaux contenus

LLM

Comment sélectionner les patients ?

Capacité à comprendre, résumer, générer et prédire de nouveaux contenus

LLM

Côté de la tumeur ?
Quel est le stade de la maladie ?
Statut ménopausique ?
Type histologique ?

Comment sélectionner les patients ?

Capacité à comprendre, résumer, générer et prédire de nouveaux contenus

LLM

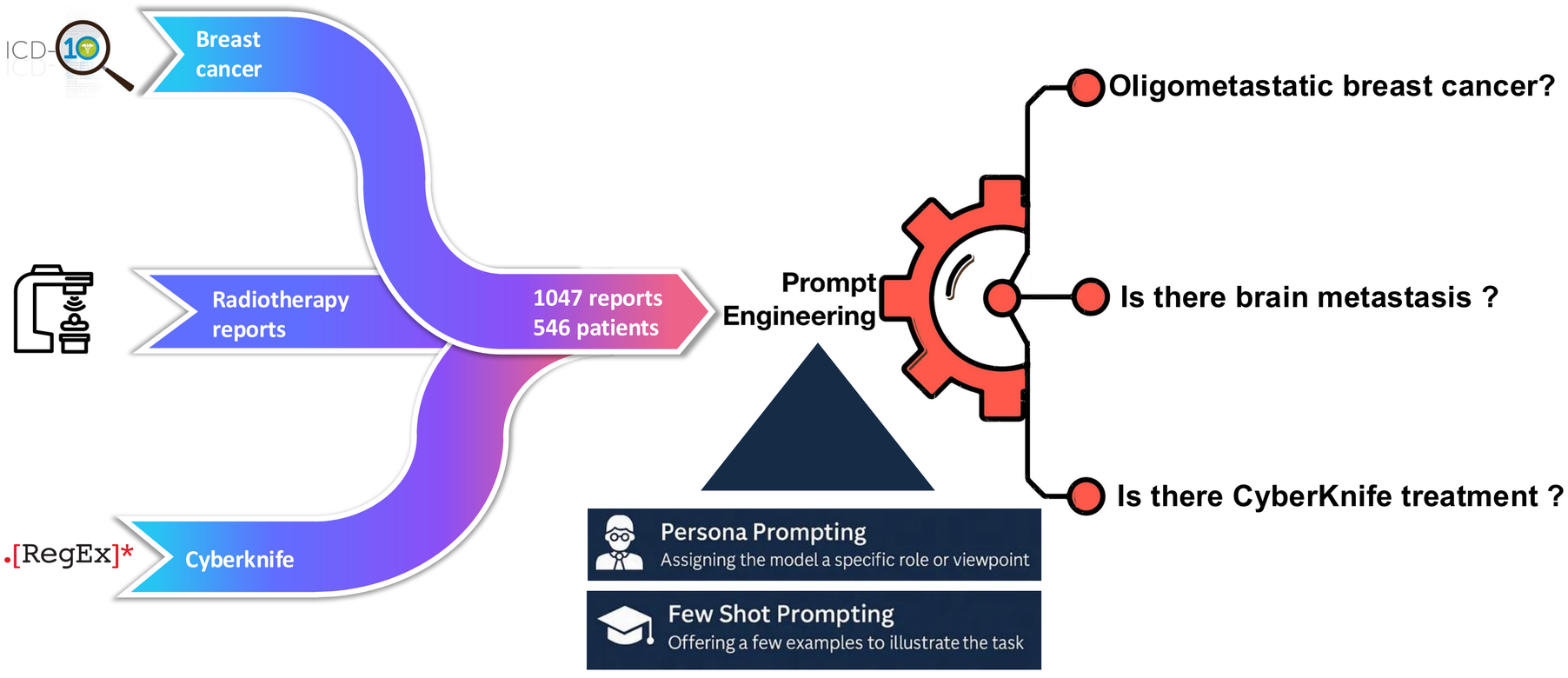
Consultation							Biopsy				
Patient	Date	Side	T	N	M	Menopause	Biopsy	Date	Side	Type	...
1	06/07/2015	LEFT	1	0	0	NO	YES	06/08/2015	LEFT	NST	...
1	04/06/2017	**LEFT**	2	1	1	YES	**YES**	06/07/2017	**LEFT**	LOBULAR	...
2	01/01/2012	RIGHT	3	0	0	**YES**	**YES**	01/02/2012	RIGHT	MIXED	...
2	10/05/2012	**LEFT**	4	1	1	**YES**	**YES**	08/06/2012	**LEFT**	NST	...

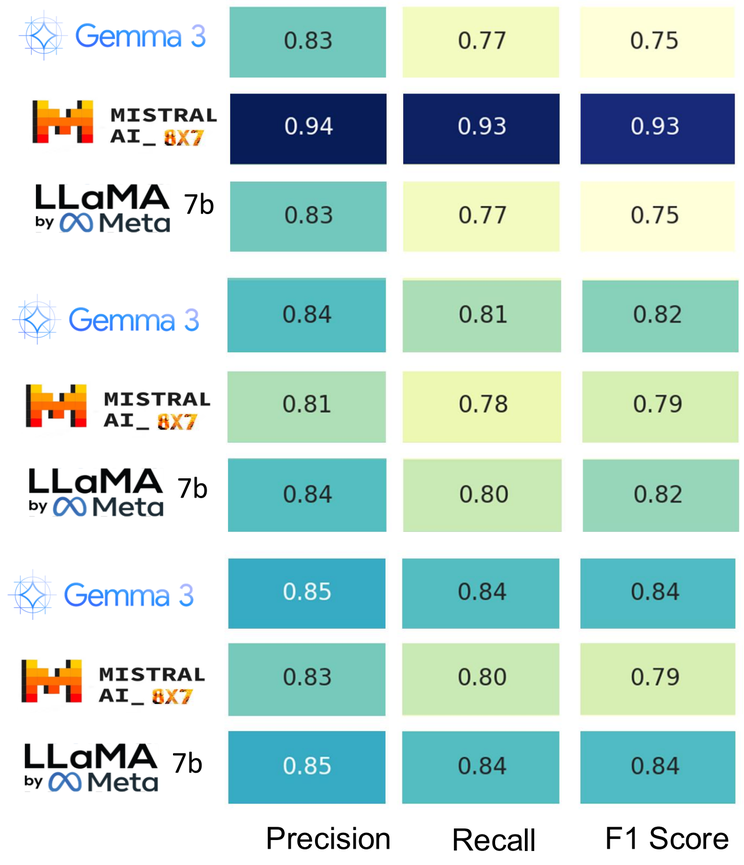
Côté de la tumeur ?
Quel est le stade de la maladie ?
Statut ménopausique ?
Type histologique ?

Leveraging Large Language Models for Identifying CyberKnife Radiotherapy in Oligometastatic Breast Cancer: A Time-Saving Tool for Retrospective Studies

R. Schiappa[1], S. Ben Dhia[2], S. Contu[1], E. Chamorey[1]
[1]Departemenf of Epidemiology, Biostatistics and Health Data - Centre Antoine Lacassagne, Nice, France,
[2]Radiotherapy - Centre Antoine Lacassagne, Nice, France

Large Language Models (LLMs) have the potential to streamline retrospective oncology research by extracting structured insights from free-text clinical reports. At the Centre Antoine Lacassagne (CAL), in Nice, where precise manual selection is time-consuming and resource-intensive, we explored their use for identifying breast cancer patients who received CyberKnife radiotherapy, in an oligometastatis context without brain metastasis.





LLMs, while not flawless, offer a valuable balance between performance and efficiency for retrospective studies. Their use can significantly reduce manual workload in identifying patients, especially when high recall is less critical than filtering relevance.

ESMO AI & DIGITAL ONCOLOGY Annual Congress
BERLIN GERMANY 12-14 NOVEMBER 2025
200P

Renaud.schiappa@nice.unicancer.fr

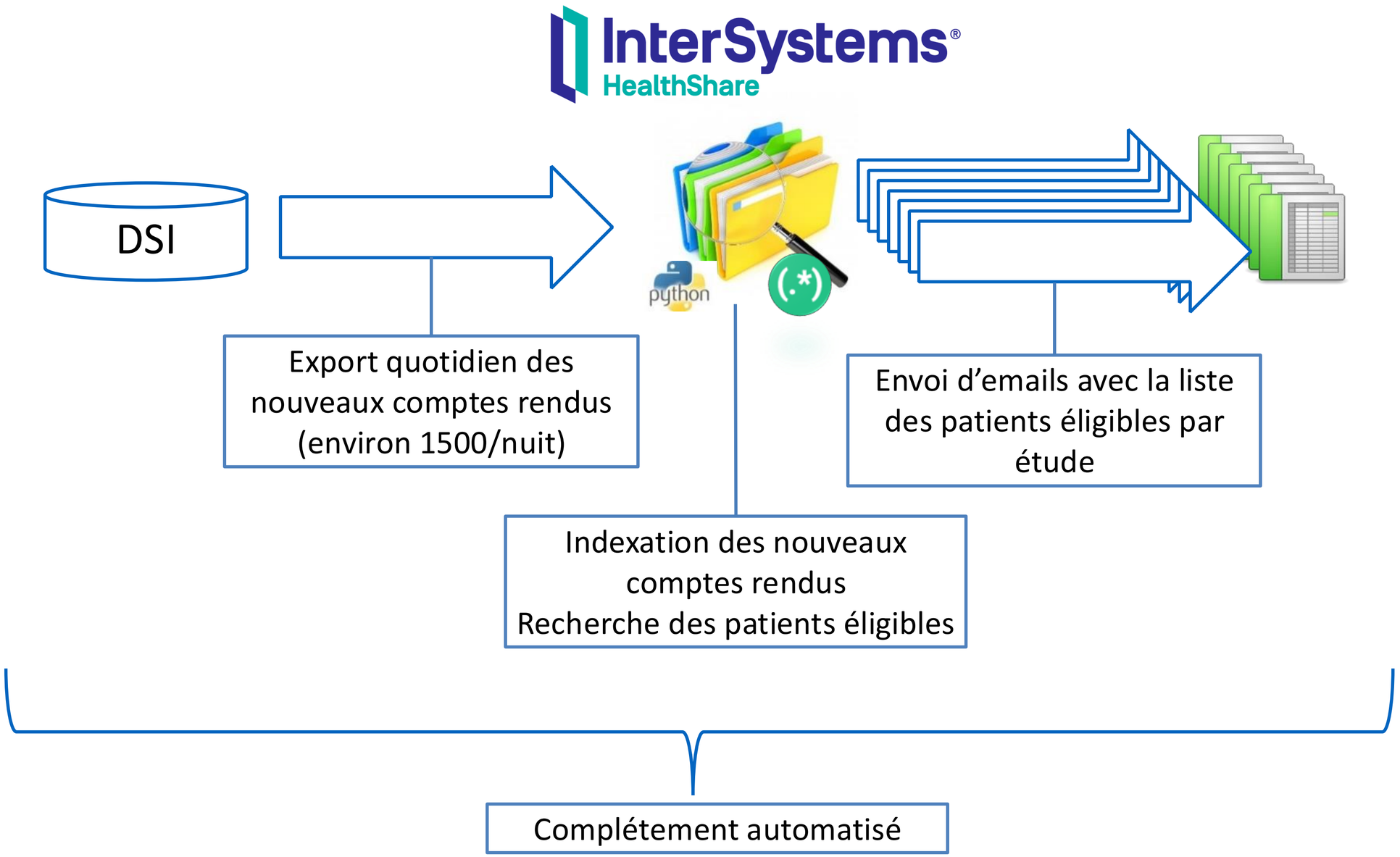
Comment sélectionner les patients ?

Manuellement | RegEx | NLP | LLM

	Manual	RegEx	NLP	LLM
Temps	👎	👍	👎 👍	👍
Précision	👎	👍	👍	👍
Sensibilité	👍	👍	👍	👍

Applications

Applications – Pré-screening automatique



Applications - Cohorting

EMBER-4

A Randomized, Open-Label, Phase 3 Study of Adjuvant Imlunestrant vs Standard Adjuvant Endocrine Therapy in Patients who have Previously Received 2 to 5 years of Adjuvant Endocrine Therapy for ER+,HER2- Early Breast Cancer with an Increased Risk of Recurrence

TRAK-ER

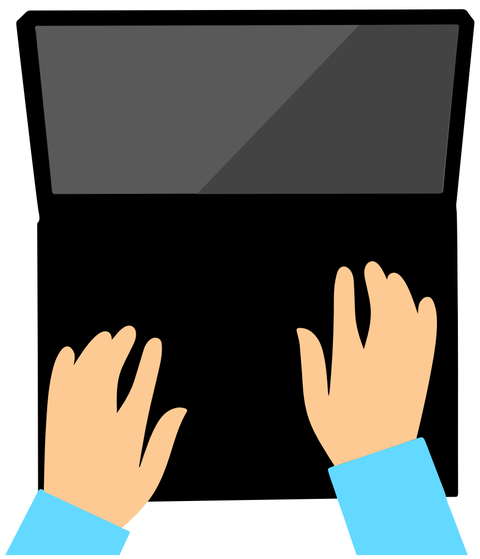
A randomised trial of early detection of molecular relapse with circulating tumour DNA tracking and treatment with palbociclib plus fulvestrant versus standard endocrine therapy in patients with ER positive HER2 negative breast cancer

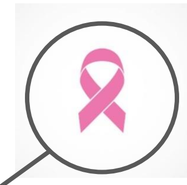
Applications - Cohorting

EMBER-4

TRAK-ER

Breast cancer patients





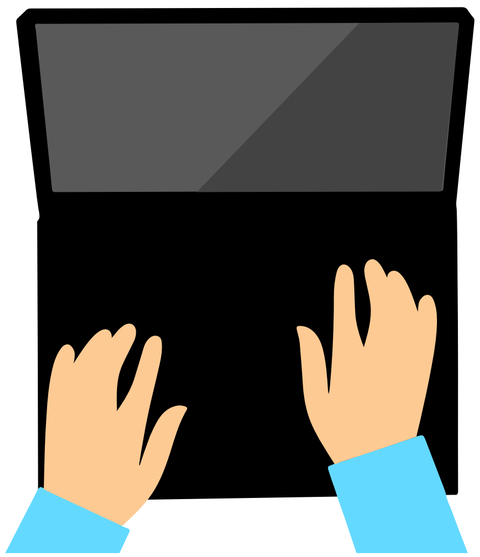
Breast cancer patients

Applications - Cohorting

EMBER-4

Breast cancer patients

Surgery validated by bMDT between 2017 and 2022



TRAK-ER

Breast cancer patients

Surgery validated by bMDT between 2017 and 2022

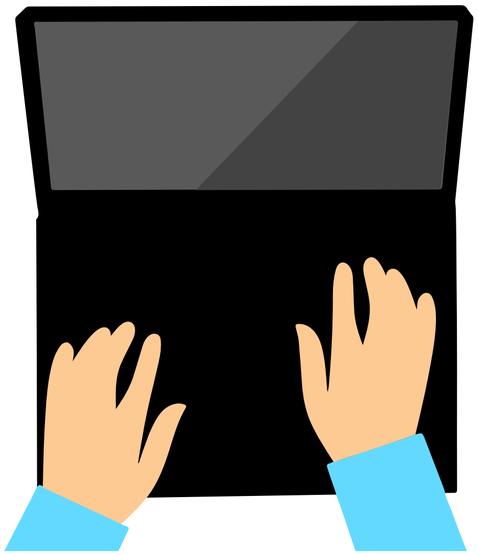
Applications - Cohorting

EMBER-4

Breast cancer patients

Surgery validated by bMDT between 2017 and 2022

ER+, HER2-, early-stage



TRAK-ER

Breast cancer patients

Surgery validated by bMDT between 2017 and 2022

ER+, HER2-, early-stage

Applications - Cohorting

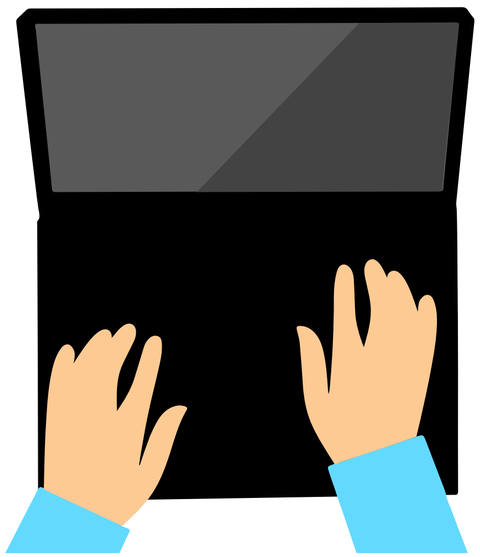
EMBER-4

Breast cancer patients

Surgery validated by bMDT between 2017 and 2022

ER+, HER2-, early-stage

Standard endocrine therapy



TRAK-ER

Breast cancer patients

Surgery validated by bMDT between 2017 and 2022

ER+, HER2-, early-stage

Standard endocrine therapy and no CDK4/6 inhibitor

Applications - Cohorting

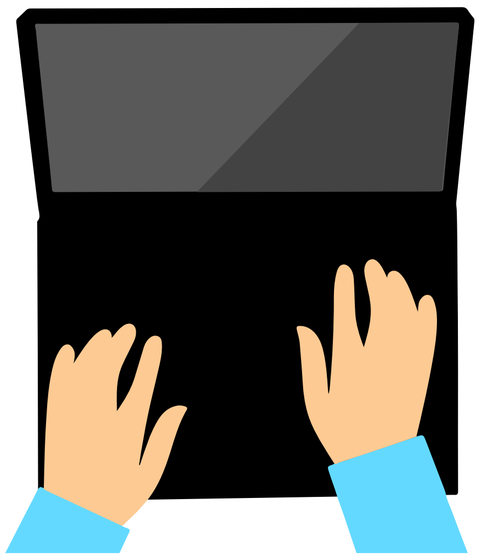
EMBER-4

Breast cancer patients

Surgery validated by bMDT between 2017 and 2022

ER+, HER2-, early-stage

Standard endocrine therapy



TRAK-ER

Breast cancer patients

Surgery validated by bMDT between 2017 and 2022

ER+, HER2-, early-stage

Standard endocrine therapy and no CDK4/6 inhibitor

Applications - Cohorting

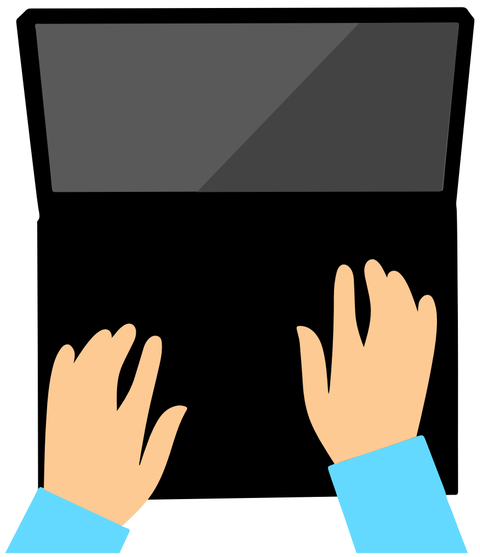
EMBER-4

Breast cancer patients

Surgery validated by bMDT between 2017 and 2022

ER+, HER2-, early-stage

Standard endocrine therapy



TRAK-ER

Breast cancer patients

Surgery validated by bMDT between 2017 and 2022

ER+, HER2-, early-stage

Standard endocrine therapy and no CDK4/6 inhibitor

Applications - Cohorting

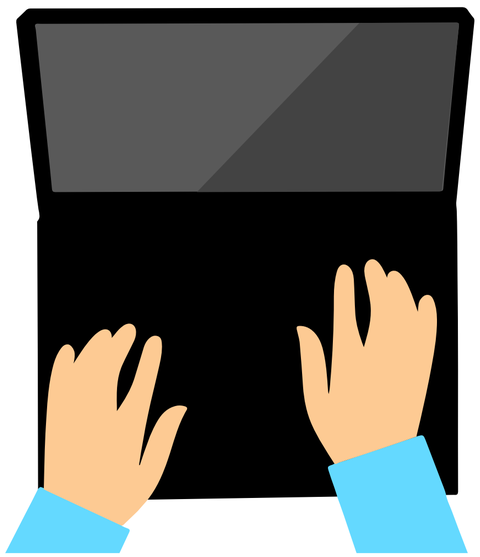
EMBER-4

Breast cancer patients

Surgery validated by bMDT between 2017 and 2022

ER+, HER2-, early-stage

Standard endocrine therapy



TRAK-ER

Breast cancer patients

Surgery validated by bMDT between 2017 and 2022

ER+, HER2-, early-stage

Standard endocrine therapy and no CDK4/6 inhibitor

Applications - Cohorting

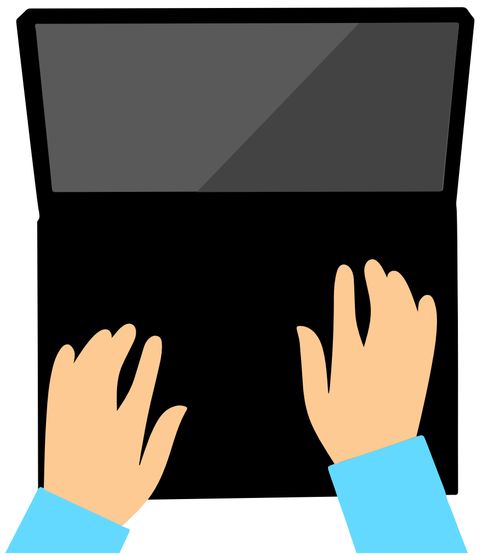
EMBER-4

Breast cancer patients

Surgery validated by bMDT between 2017 and 2022

ER+, HER2-, early-stage

Standard endocrine therapy



TRAK-ER

Breast cancer patients

Surgery validated by bMDT between 2017 and 2022

ER+, HER2-, early-stage

Standard endocrine therapy and no CDK4/6 inhibitor

Applications - Cohorting

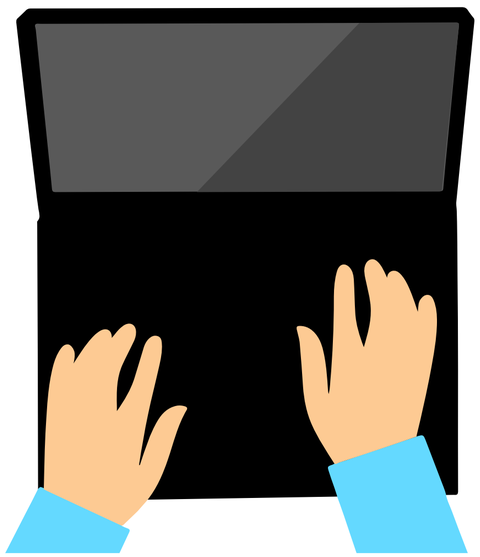
EMBER-4

Breast cancer patients

Surgery validated by bMDT between 2017 and 2022

ER+, HER2-, early-stage

Standard endocrine therapy



TRAK-ER

Breast cancer patients

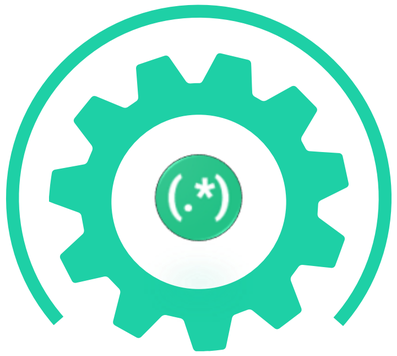
Surgery validated by bMDT between 2017 and 2022

ER+, HER2-, early-stage

Standard endocrine therapy and no CDK4/6 inhibitor

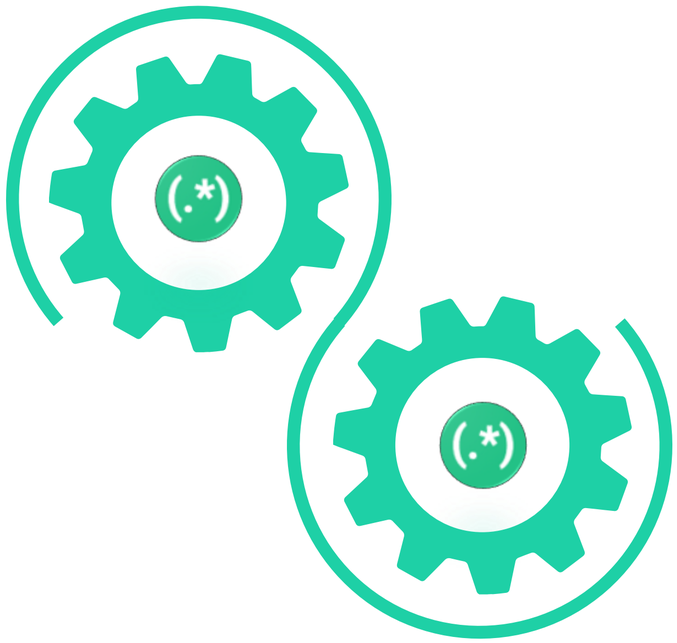
Applications - Cohorting

136 619 Multidisciplinary Team
In total



Applications - Cohorting

136 619 Multidisciplinary Team
In total

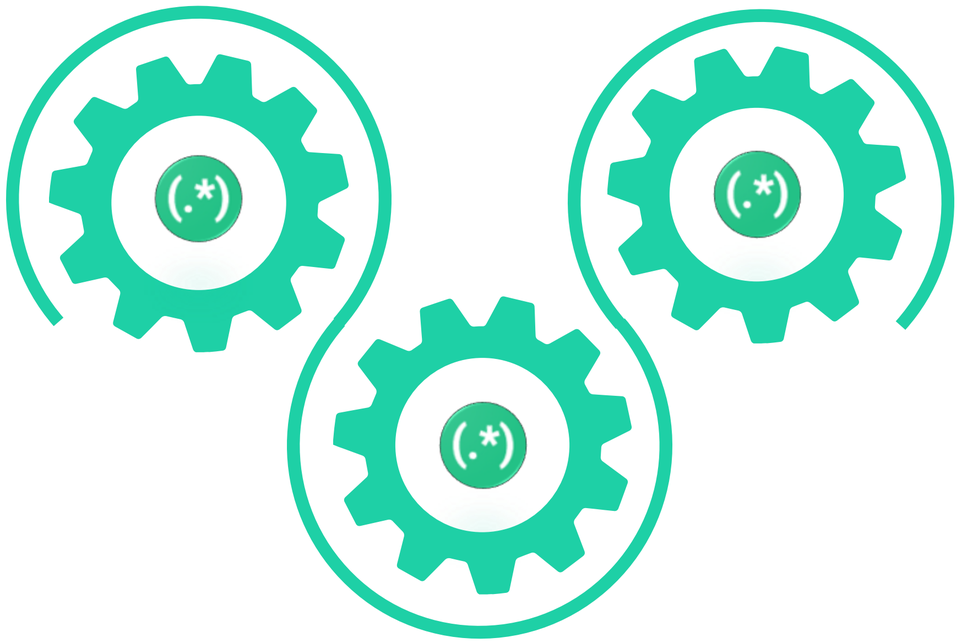


59 679 MDT
2017-2022

Applications - Cohorting

136 619 Multidisciplinary Team
In total

10 787 breast MDT
With notion of surgery and treatment

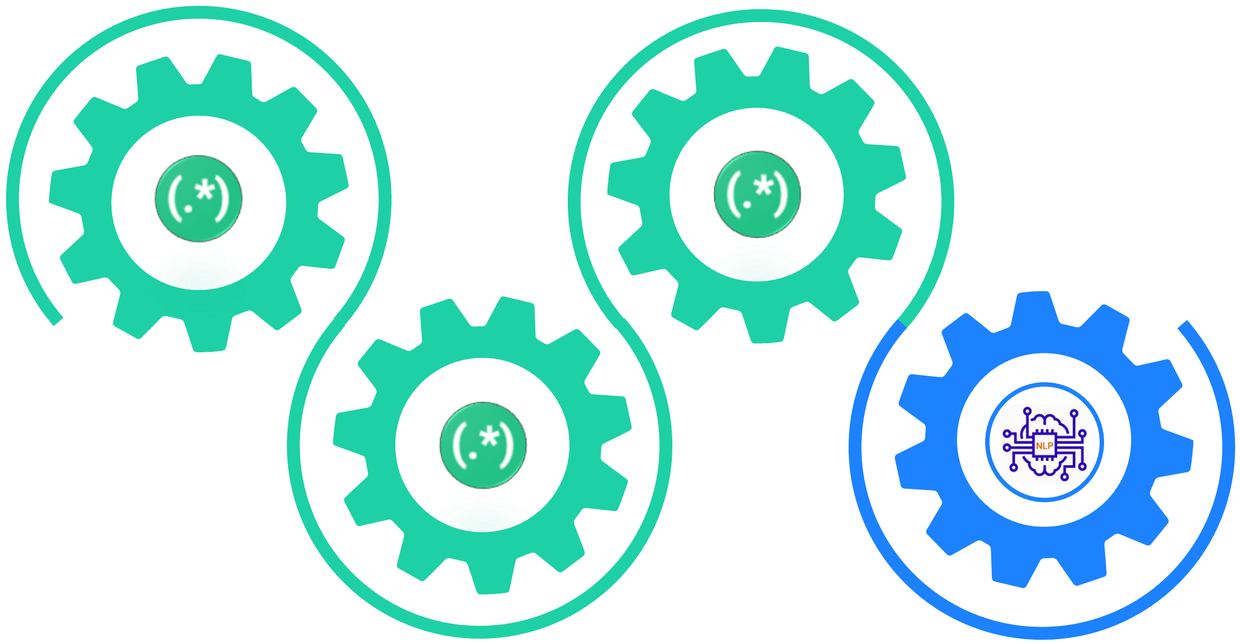


59 679 MDT
2017-2022

Applications - Cohorting

136 619 Multidisciplinary Team
In total

10 787 breast MDT
With notion of surgery and treatment



59 679 MDT
2017-2022

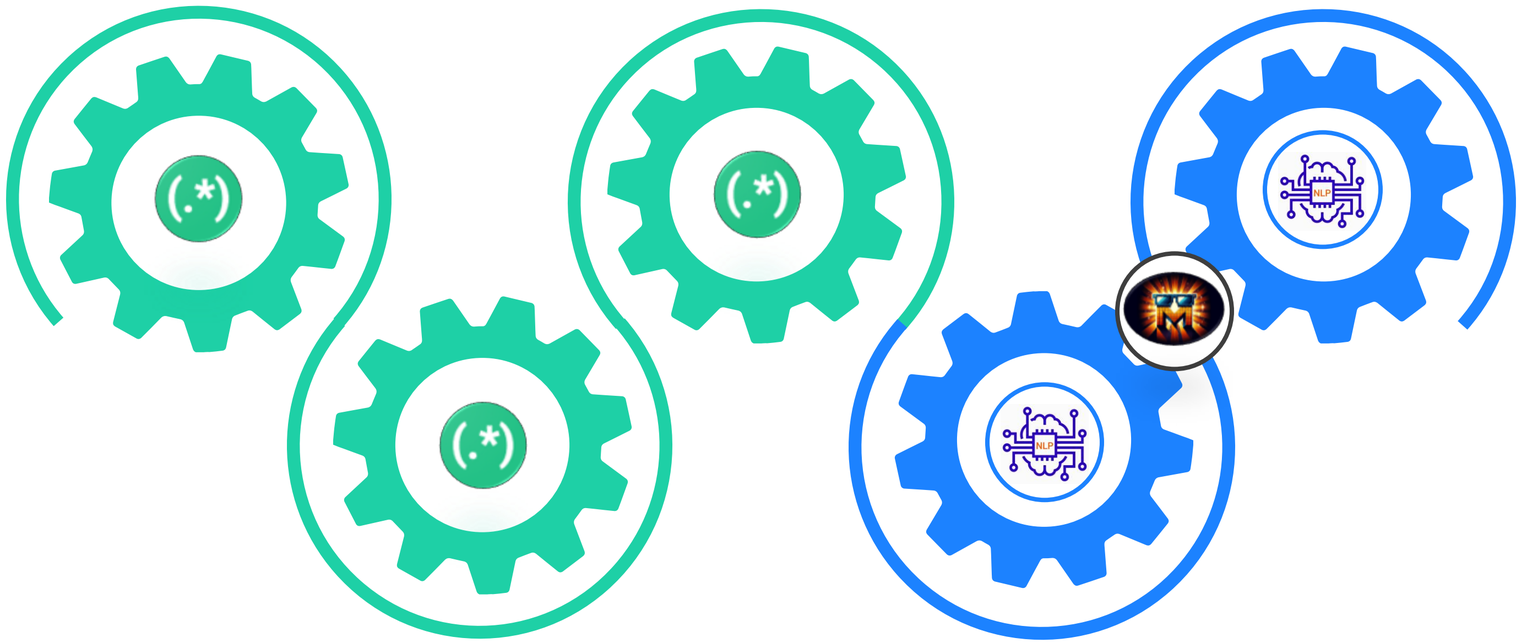
5201 Anapath reports
Automatically structured

Applications - Cohorting

136 619 Multidisciplinary Team
In total

10 787 breast MDT
With notion of surgery and treatment

241 patients prescreened
To be evaluated



59 679 MDT
2017-2022

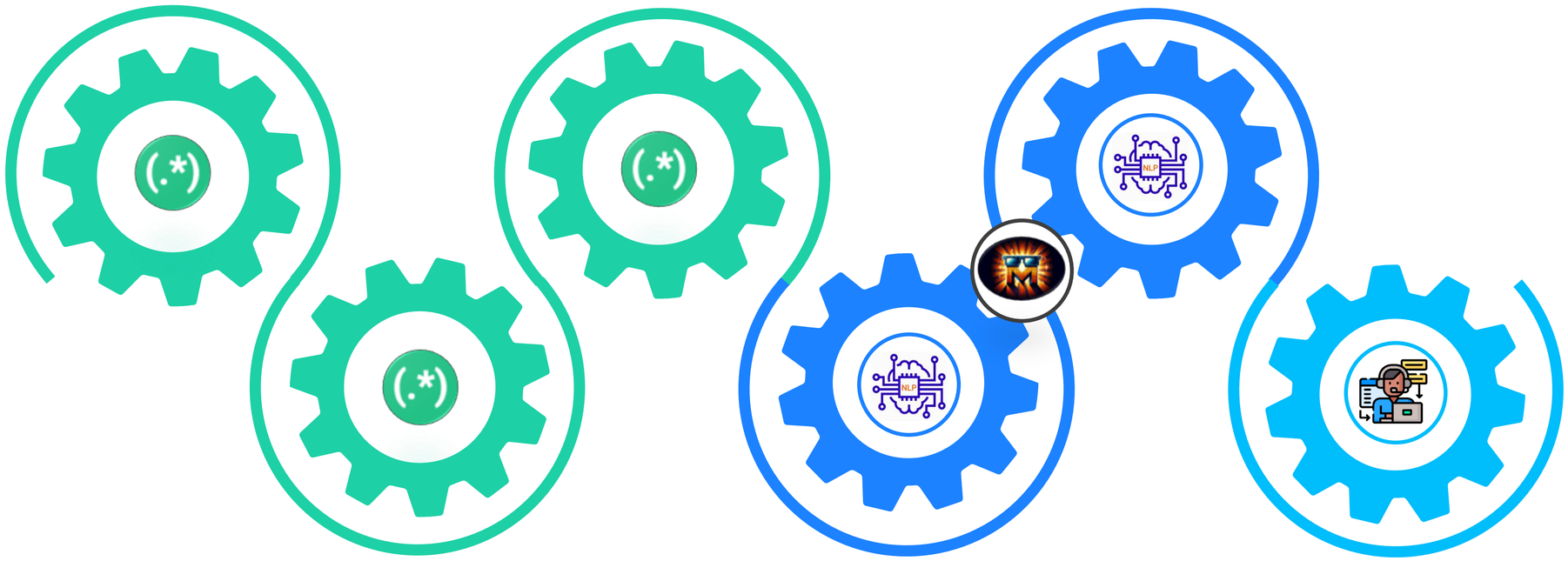
5201 Anapath reports
Automatically structured

Applications - Cohorting

136 619 Multidisciplinary Team
In total

10 787 breast MDT
With notion of surgery and treatment

241 patients prescreened
To be evaluated



59 679 MDT
2017-2022

5201 Anapath reports
Automatically structured

58 patients eligible
To be contacted for the study

4 hours

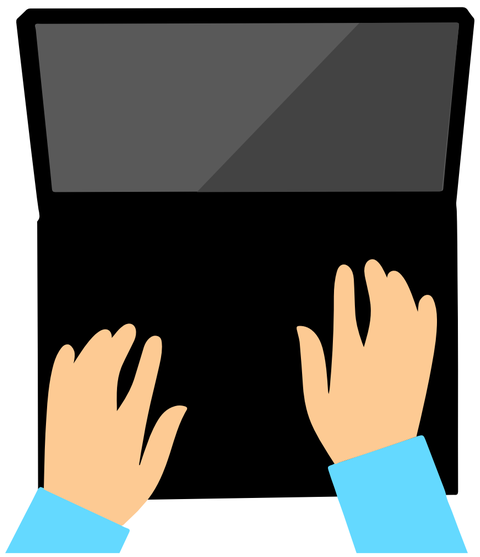
1 hour

12 hours

Applications - Cohorting

EMBER-4

TRAK-ER



Applications - Cohorting

EMBER-4

25 refused to sign

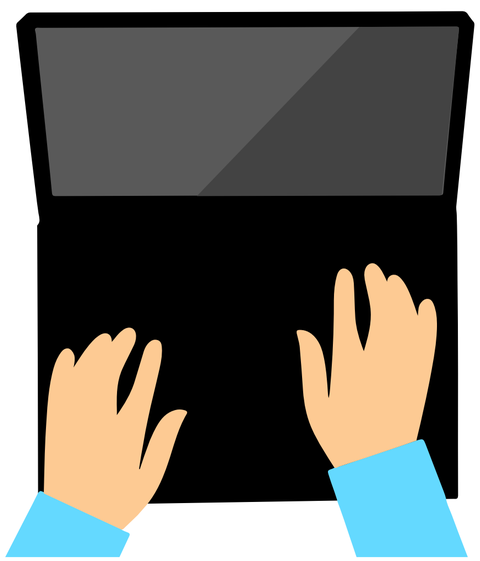
TRAK-ER

Applications - Cohorting

EMBER-4

25 refused to sign

1 inclusion error



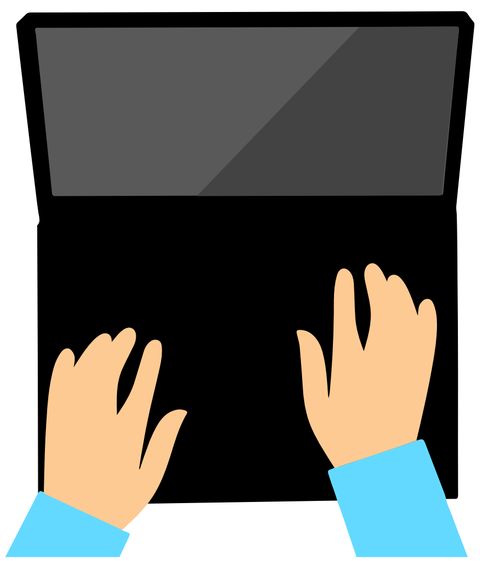
TRAK-ER

Applications - Cohorting

EMBER-4

25 refused to sign

1 inclusion error



TRAK-ER

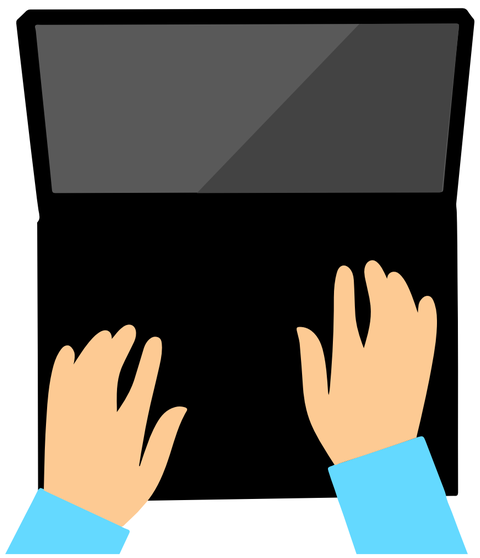
7 screenfailures

Applications - Cohorting

EMBER-4

25 refused to sign

1 inclusion error



TRAK-ER

7 screenfailures

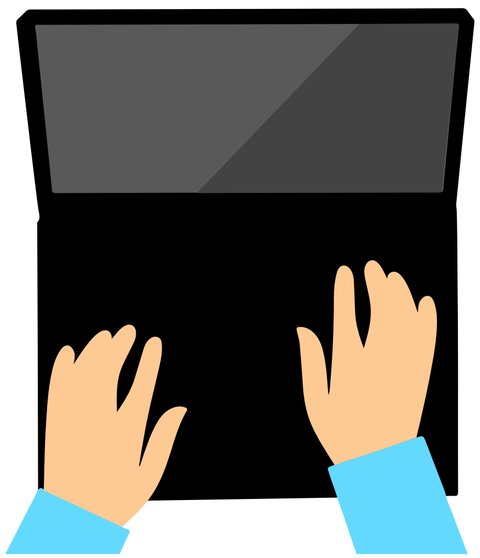
2 are being screened

Applications - Cohorting

EMBER-4

25 refused to sign

1 inclusion error



TRAK-ER

7 screenfailures

2 are being screened

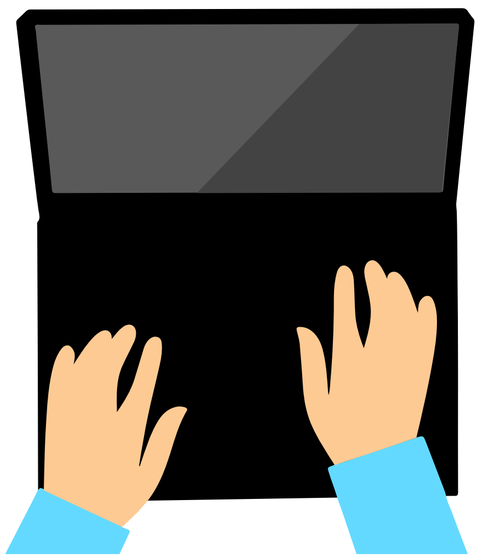
2 are waiting to sign

Applications - Cohorting

EMBER-4

25 refused to sign

1 inclusion error



TRAK-ER

7 screenfailures

2 are being screened

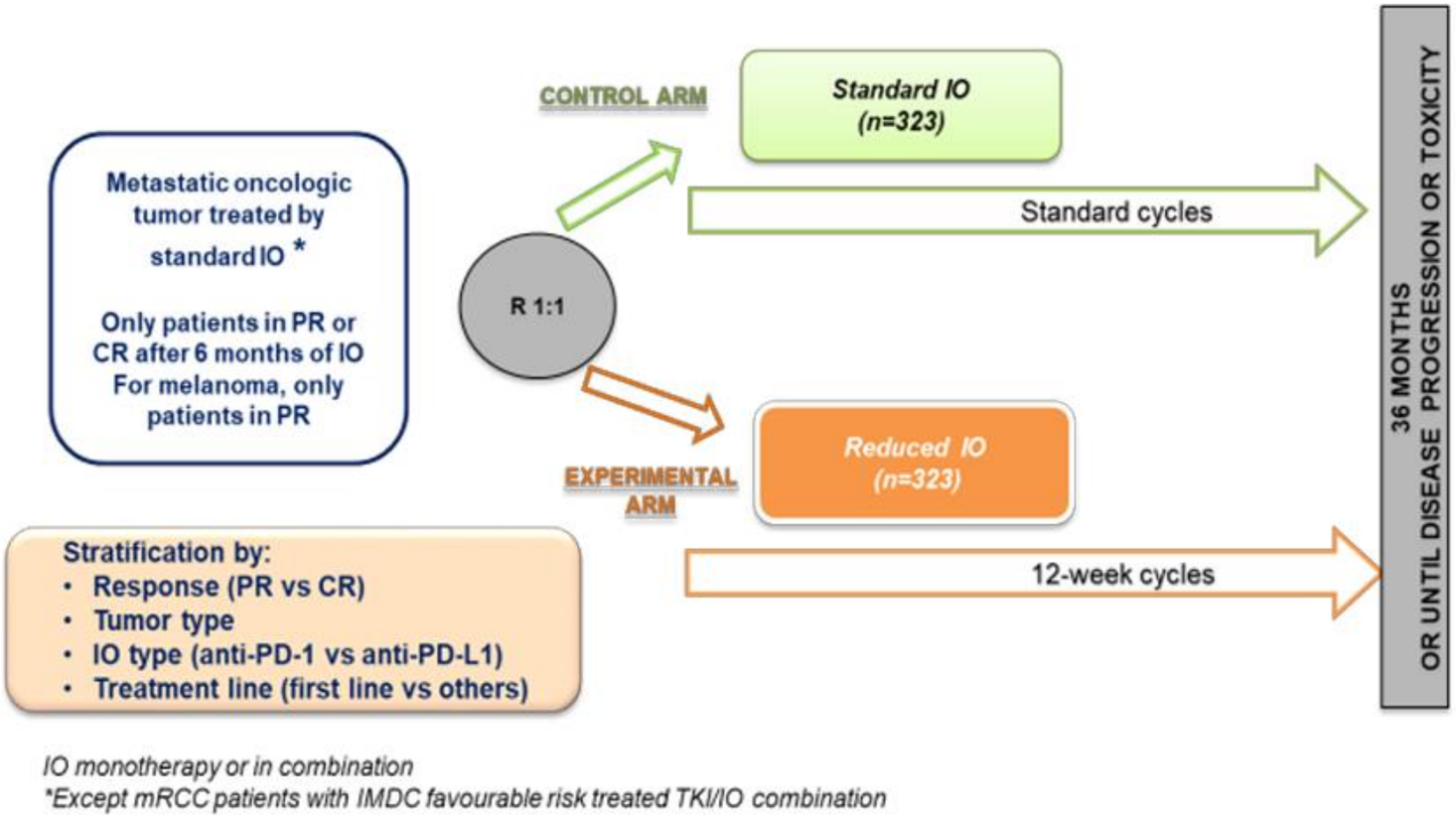
2 are waiting to sign

21 included

Applications – Cohorting automatique MOIO

Essai randomisé de phase III comparant l'immunothérapie (IO) standard, par inhibiteurs des points de contrôle immunitaire, versus une diminution de la dose intensité de l'IO chez les patients ayant un cancer métastatique en réponse après 6 mois d'IO standard.

TRIAL SCHEMA



IO monotherapy or in combination
**Except mRCC patients with IMDC favourable risk treated TKI/IO combination*

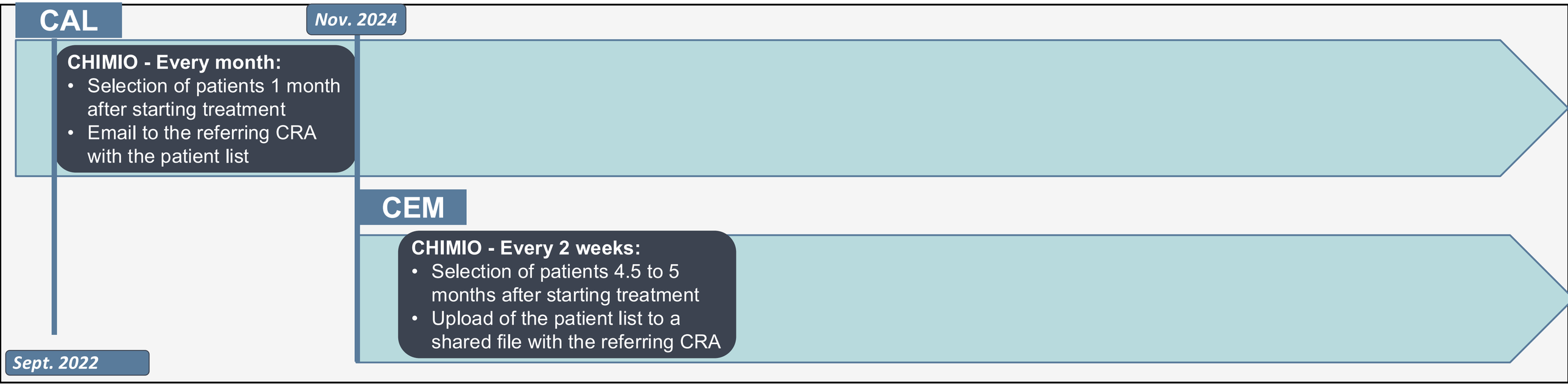
Applications – Cohorting automatique MOIO

CAL

CHIMIO - Every month:
- Selection of patients 1 month after starting treatment
- Email to the referring CRA with the patient list

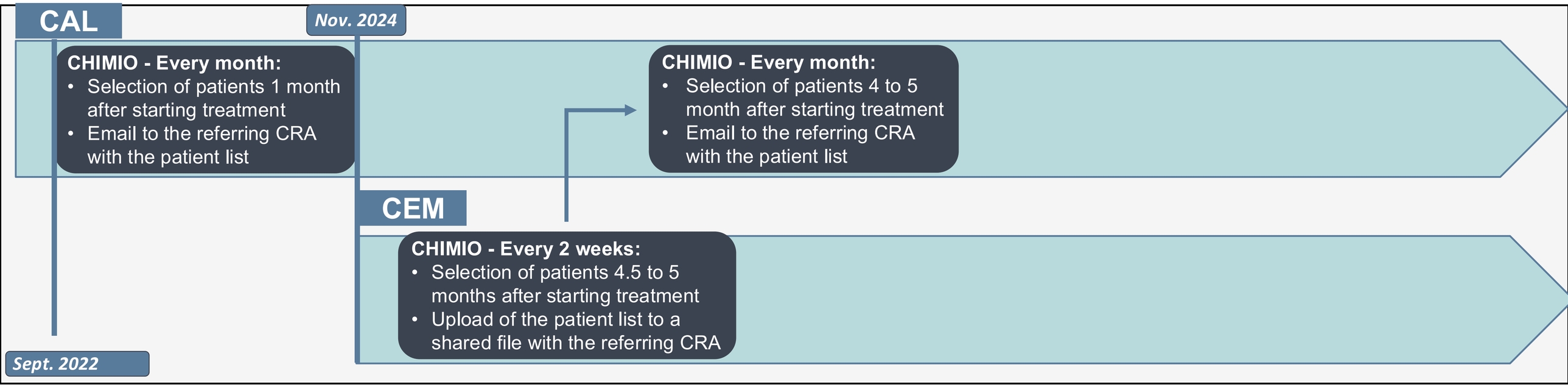
Sept. 2022

Applications – Cohorting automatique MOIO



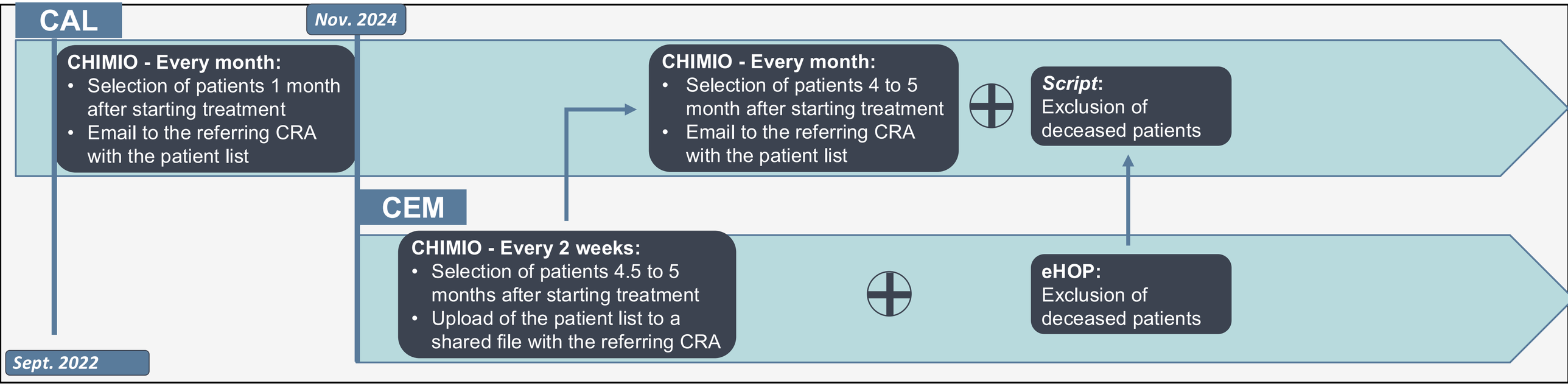
Renaud Schiappa & Lou-Anne Guillotel – ARC DSRC OncoPaca-Corse - 9 Avril 2026

Applications – Cohorting automatique MOIO



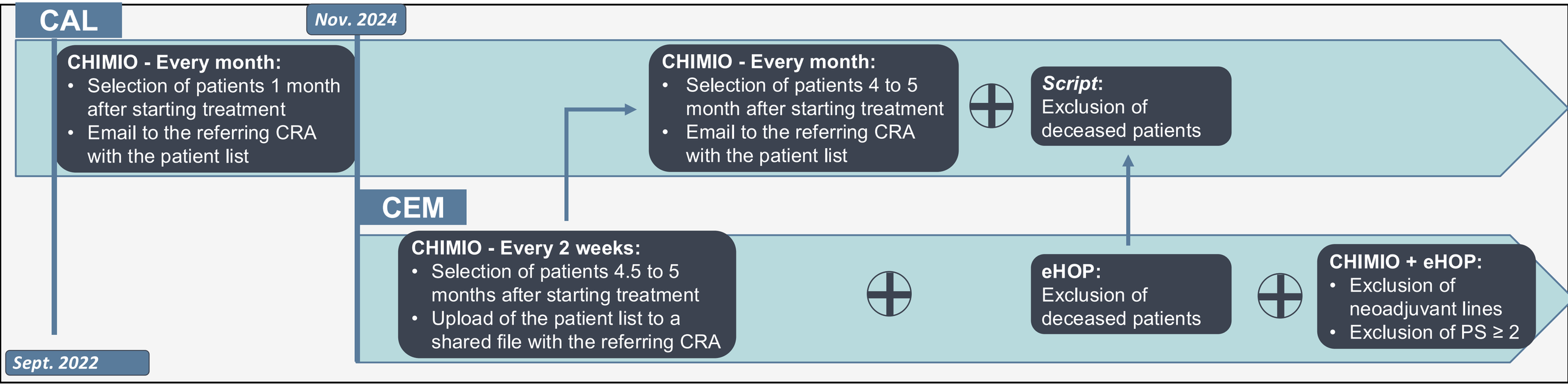
Renaud Schiappa & Lou-Anne Guillotel – ARC DSRC OncoPaca-Corse - 9 Avril 2026

Applications – Cohorting automatique MOIO



Renaud Schiappa & Lou-Anne Guillotel – ARC DSRC OncoPaca-Corse - 9 Avril 2026

Applications – Cohorting automatique MOIO



Renaud Schiappa & Lou-Anne Guillotel – ARC DSRC OncoPaca-Corse - 9 Avril 2026

Applications – Cohorting automatique MOIO

CAL

Nov. 2024

CHIMIO - Every month:
- Selection of patients 1 month after starting treatment
- Email to the referring CRA with the patient list

CHIMIO - Every month:
- Selection of patients 4 to 5 month after starting treatment
- Email to the referring CRA with the patient list

***Script*:** Exclusion of deceased patients

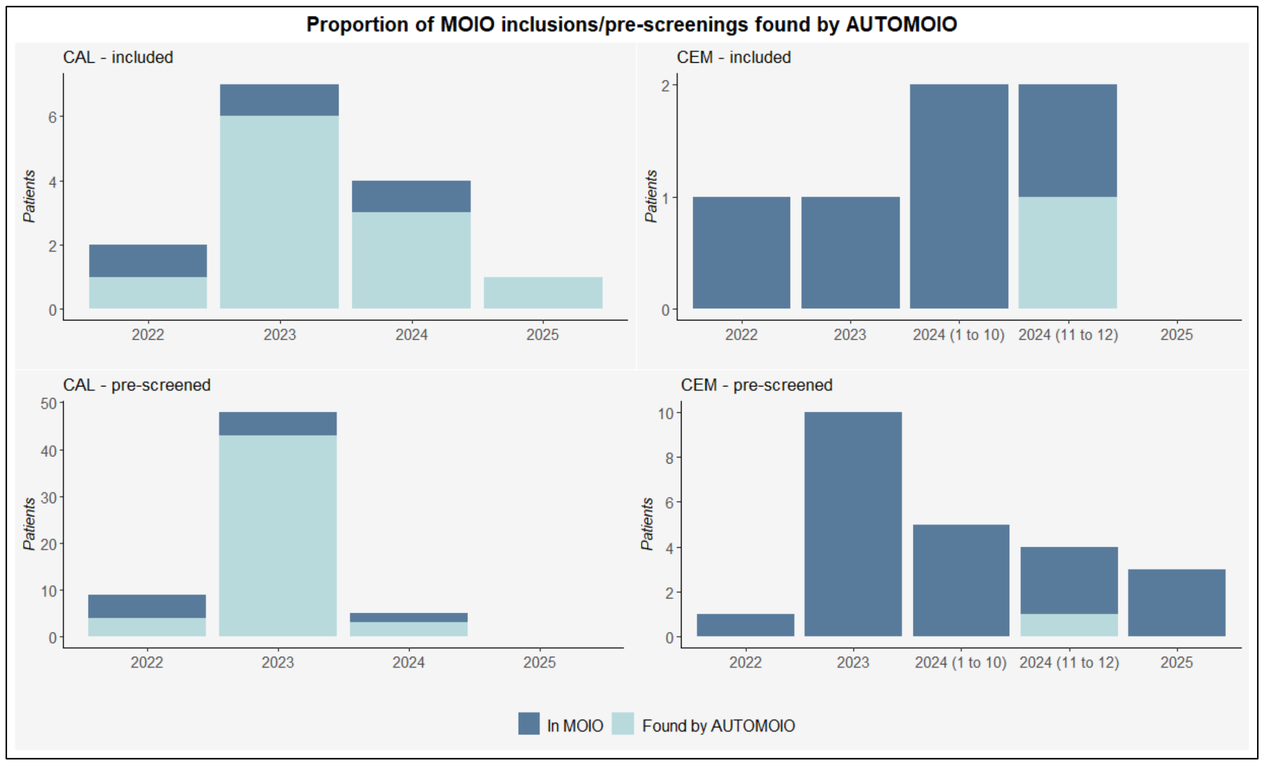
CEM

CHIMIO - Every 2 weeks:
- Selection of patients 4.5 to 5 months after starting treatment
- Upload of the patient list to a shared file with the referring CRA

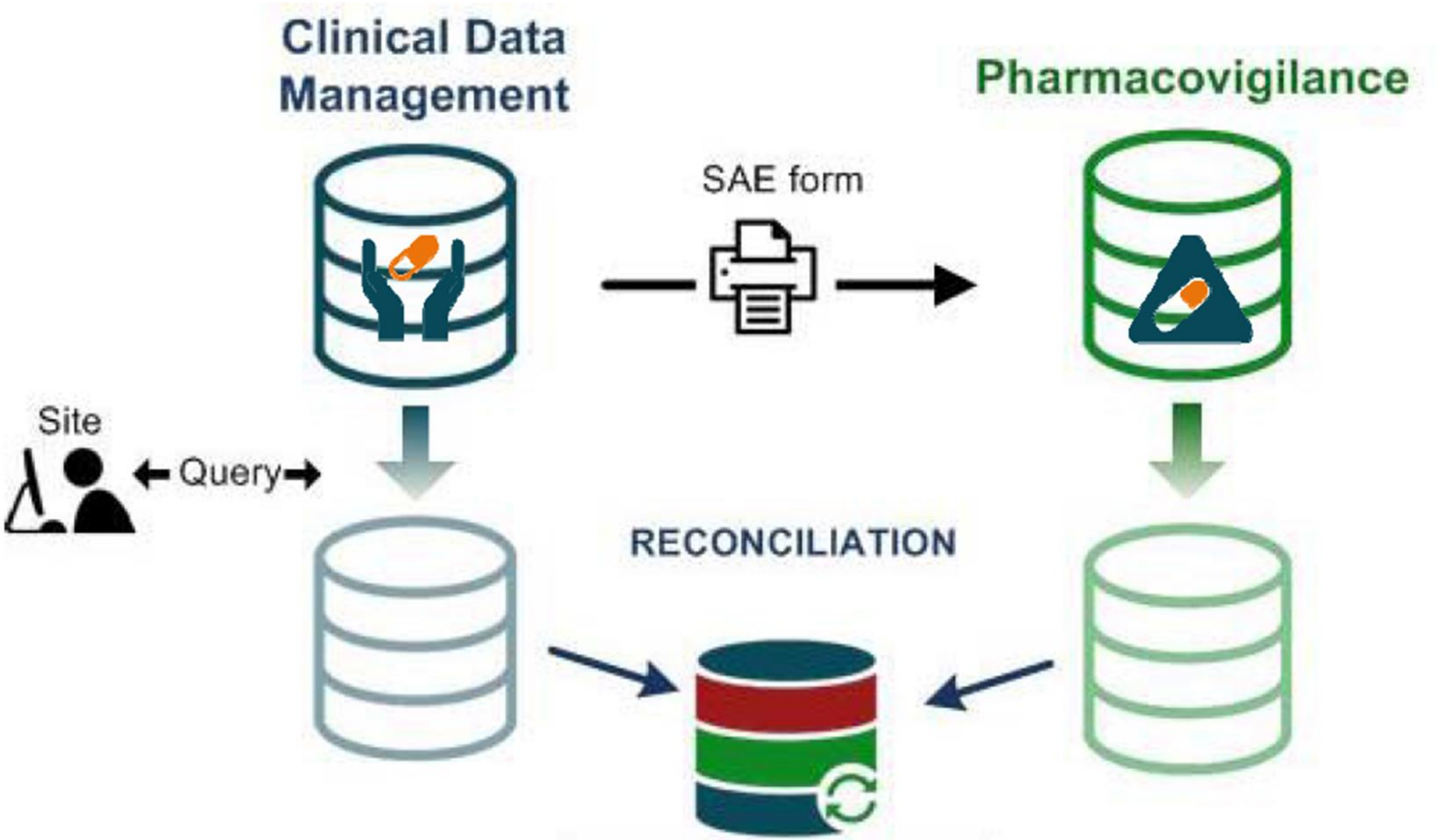
eHOP: Exclusion of deceased patients

CHIMIO + eHOP:
- Exclusion of neoadjuvant lines
- Exclusion of PS ≥ 2

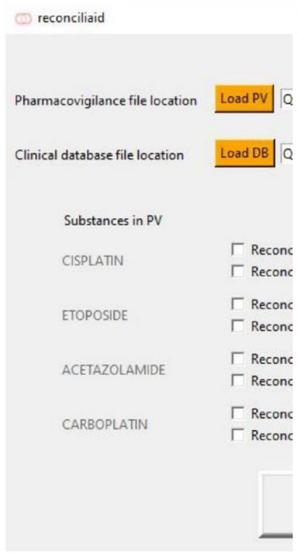
Sept. 2022



Applications – Réconciliation EIG semi-automatique RECONCILIAID



Applications – Réconciliation EIG semi-automatique RECONCILIAID



Study name (Reconciliation number)	NCT number	Total SAEs	Reconciliaid mismatches	Reconciliaid execution time (s)	Reconciliaid total time (min)	Manual total time (min)
Dpdmax (1)	NCT04198727	2	0	0.8	1	6
Fdgimmun (1)	NCT03584334	23	1	0.9	3	50
Fdgimmun (1)*					13	60
Icar (1)	NCT03467360	4	0	0.2	1	10
Iodine (1)	NCT02759133	1	0	0.2	1	3
Opera (1)	NCT02505750	15	1	0.4	3	40
Opera (1)*					11	55
Pazoglio (1)	NCT02331498	8	0	0.3	2	20
Dpdmax (2)	NCT04198727	3	0	0.3	1	5
Dpdmax (2)*					5	19
Fdgimmun (2)	NCT03584334	24	1	0.7	3	50
Halo (1)	NCT02467946	189	5	4.4	15	360
Icar (2)	NCT03467360	4	0	0.2	1	10
Iodine (2)	NCT02759133	1	0	0.2	1	3
Neoscreencovid (2)	NCT04492410	1	0	0.1	1	3
Opera (2)	NCT02505750	15	0	0.5	3	40

Asterisks refer to studies performed by a less experienced user and are execution time is not considered for the evaluation of the average results.
SAE, serious adverse event.

SageJournals

Automatic tool for the reconciliation of serious adverse events for pharmacovigilance: design and implementation of Reconciliaid

Sara Contu [1], Renaud Schiappa [2], Yann Chateau [3], and Emmanuel Chamorey [4]

Applications – Remplissage automatique AUTOMABIO

Terence Viellard & Renaud Schiappa – ARC DSRC OncoPaca-Corse - 9 Avril 2026

Applications – Remplissage automatique AUTOMABIO

Laboratory results

Different lab
Different norms

Retrieve data

Applications – Remplissage automatique AUTOMABIO

Laboratory results

Different lab
Different norms

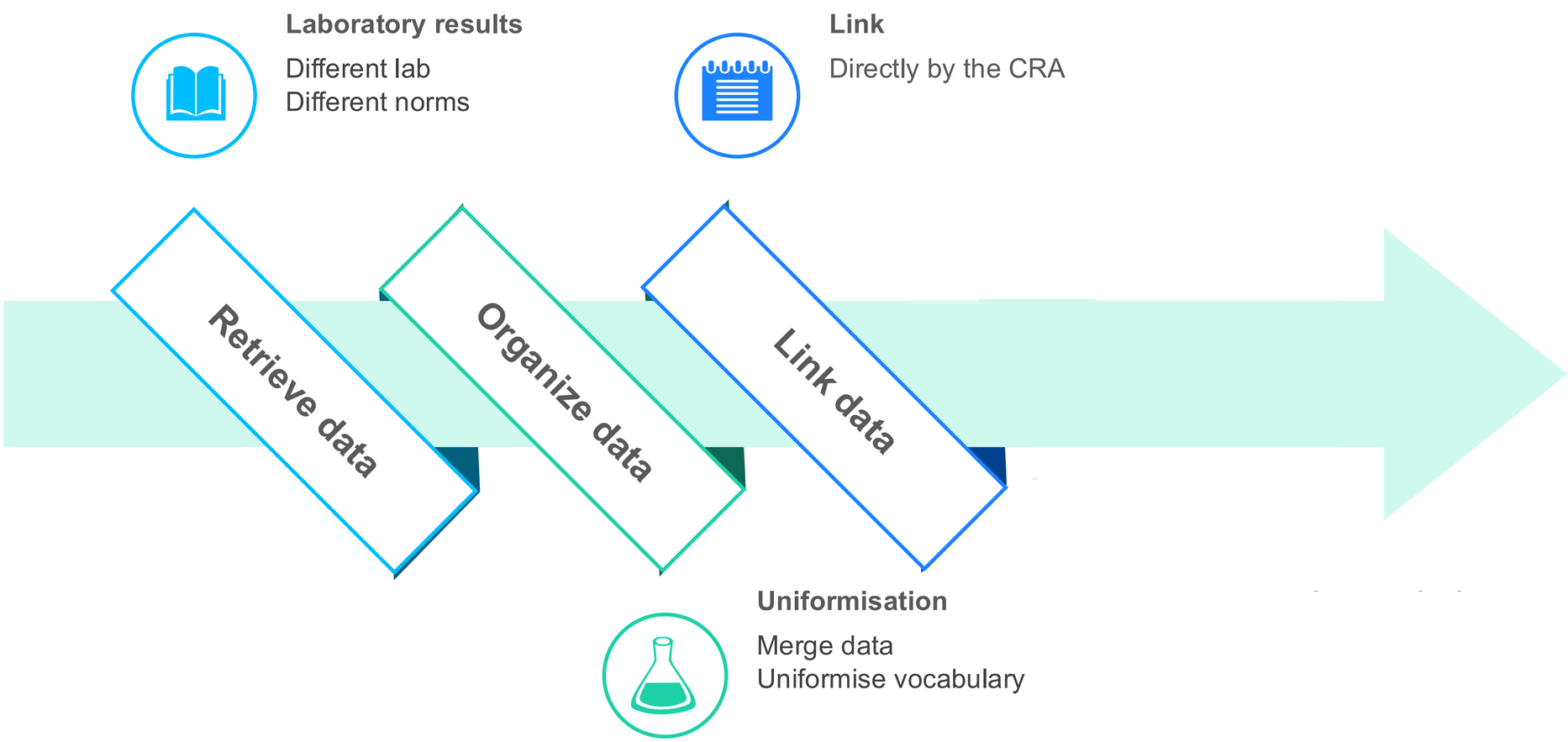
Retrieve data

Organize data

Uniformisation

Merge data
Uniformise vocabulary

Applications – Remplissage automatique AUTOMABIO



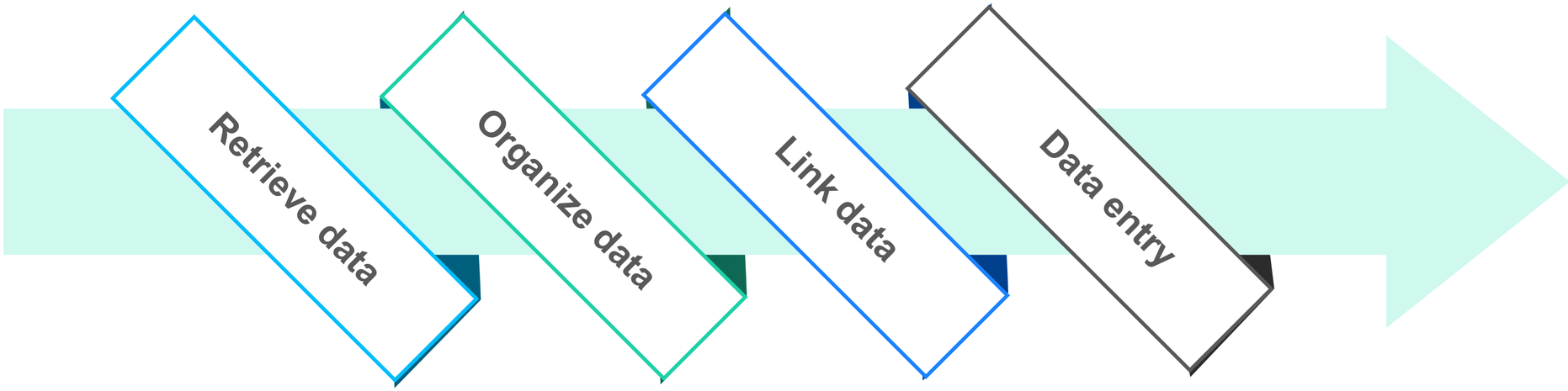
Applications – Remplissage automatique AUTOMABIO

Laboratory results

Different lab
Different norms

Link

Directly by the CRA



Uniformisation

Merge data
Uniformise vocabulary

Automatic data entry

Data is being adapted to the eCRF
Missing data identified
CTCAE calculated

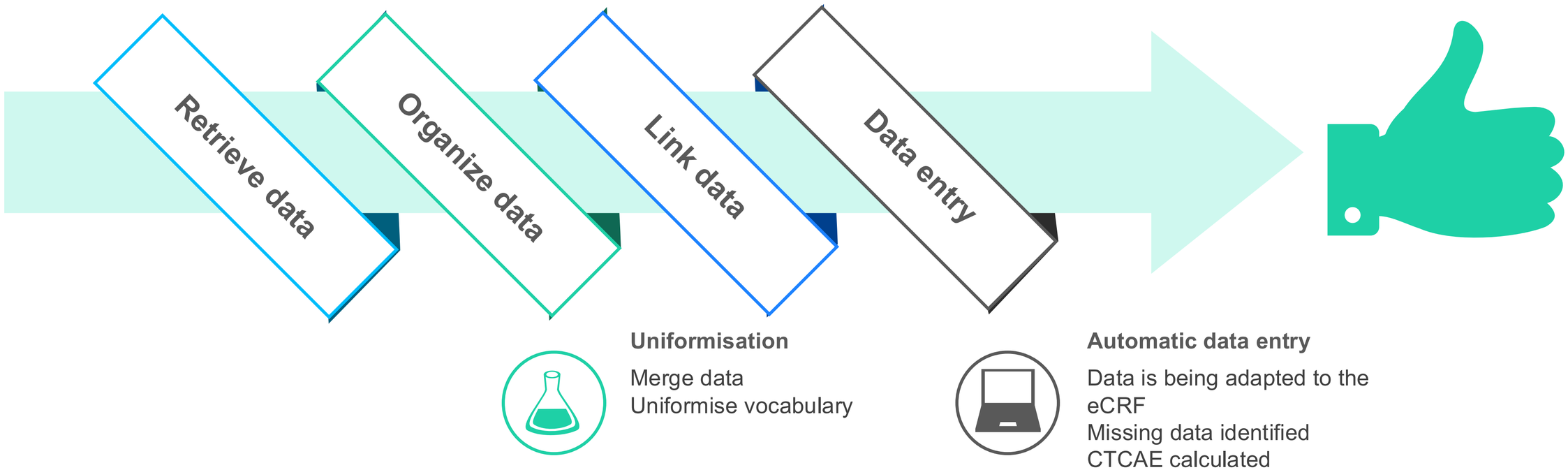
Applications – Remplissage automatique AUTOMABIO

Laboratory results

Different lab
Different norms

Link

Directly by the CRA



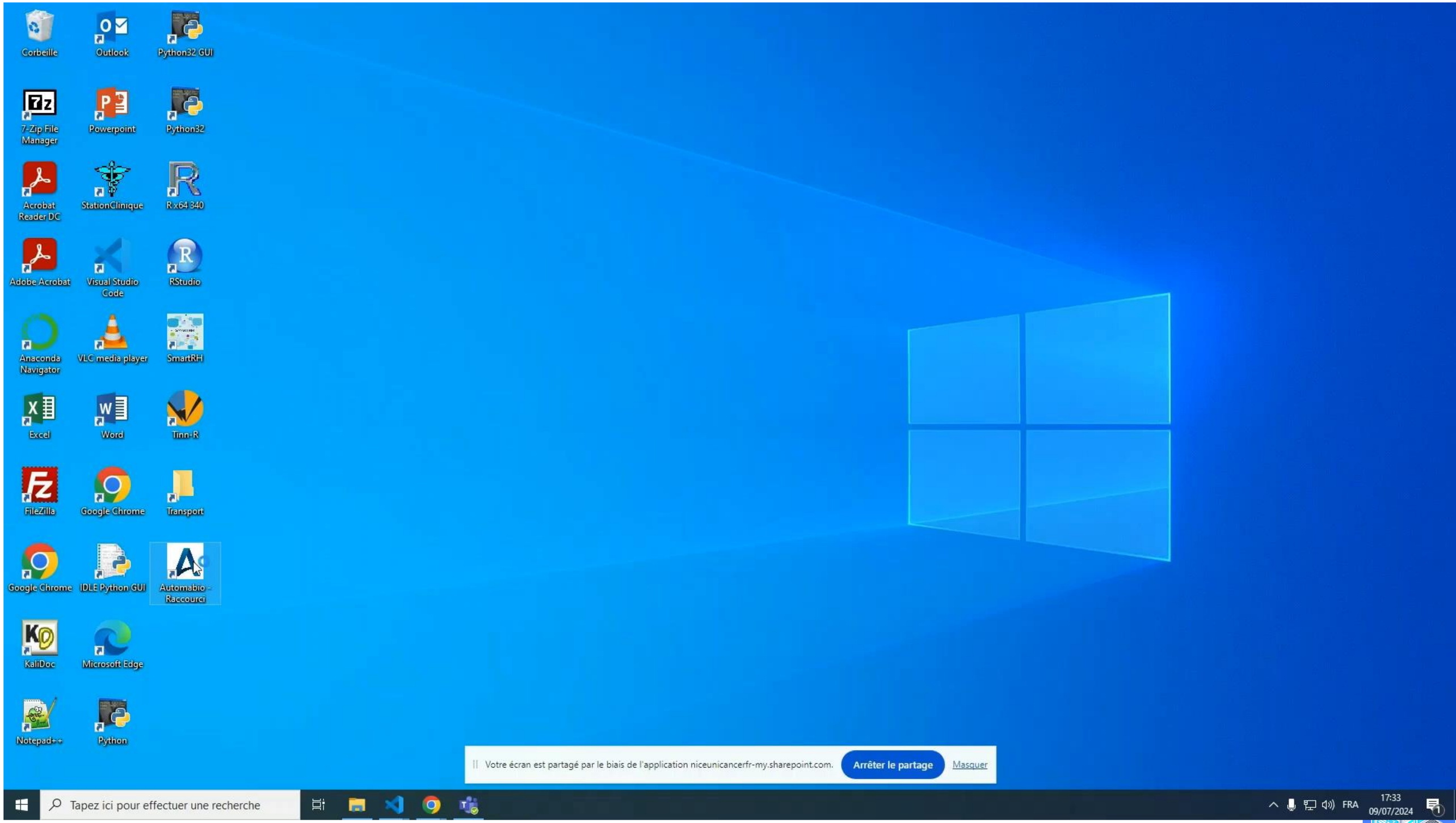
Uniformisation

Merge data
Uniformise vocabulary

Automatic data entry

Data is being adapted to the eCRF
Missing data identified
CTCAE calculated

Applications – Remplissage automatique AUTOMABIO

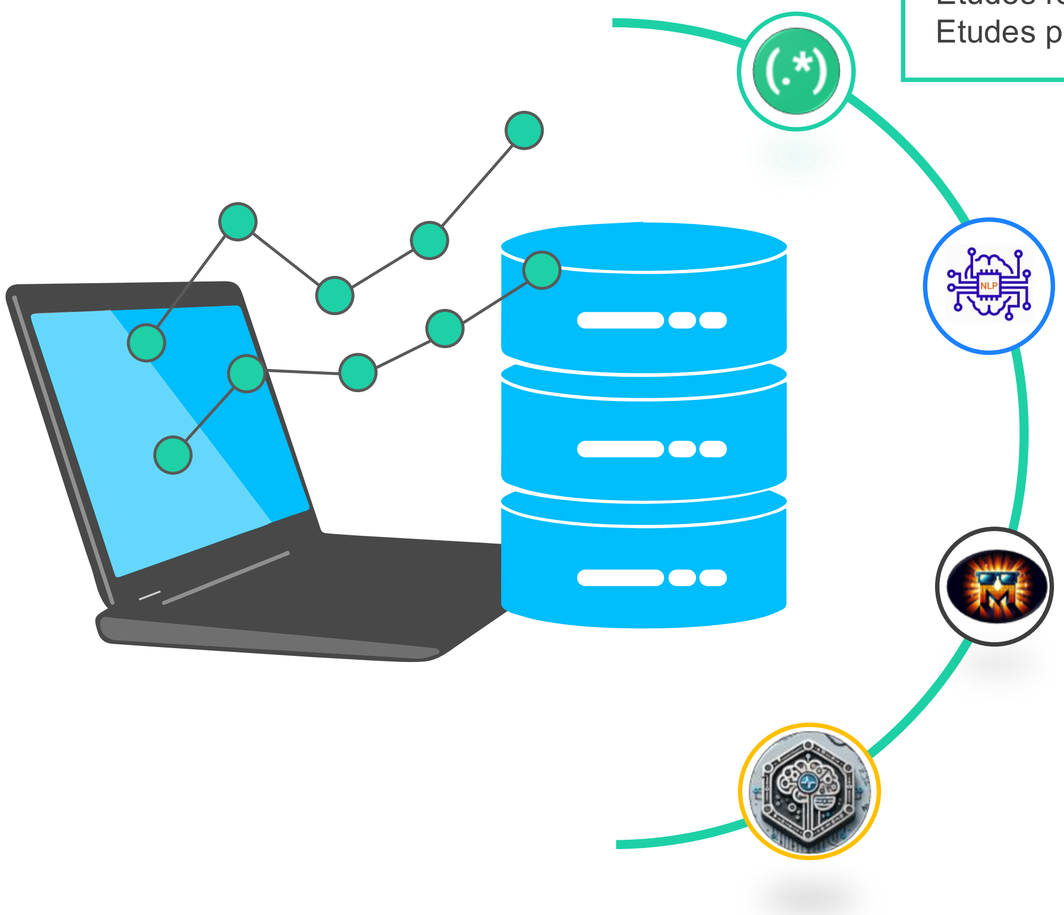


Conclusion

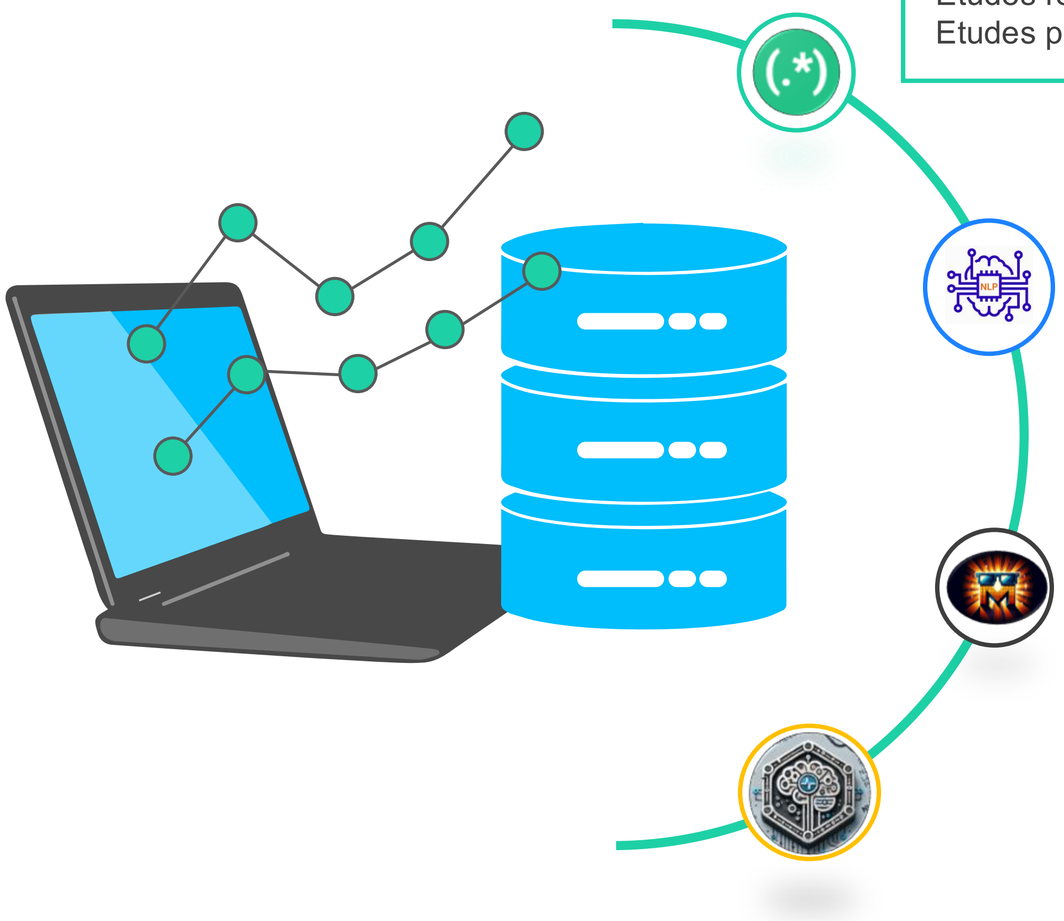
Conclusion

Automatisation

Pre-screening automatisé
Etudes retrospectives : 80-100%
Etudes prospectives : 50% des patients inclus ont été détectés



Conclusion



Automatisation

Pre-screening automatisé
Etudes retrospectives : 80-100%
Etudes prospectives : 50% des patients inclus ont été détectés

Développement

Chronophage au démarrage mais performant et appliqué dans le sein, la thyroïde et en cours sur les traitements oraux

Conclusion

Automatisation

Pre-screening automatisé
Etudes retrospectives : 80-100%
Etudes prospectives : 50% des patients inclus ont été détectés

Développement

Chronophage au démarrage mais performant et appliqué dans le sein, la thyroïde et en cours sur les traitements oraux

Prometteur

Consomme des ressources
Suffisant pour la selection de patients
En étude pour la structuration automatique

Conclusion

Automatisation

Pre-screening automatisé
Etudes retrospectives : 80-100%
Etudes prospectives : 50% des patients inclus ont été détectés

Développement

Chronophage au démarrage mais performant et appliqué dans le sein, la thyroïde et en cours sur les traitements oraux

Prometteur

Consomme des ressources
Suffisant pour la selection de patients
En étude pour la structuration automatique

Le bon outil pour la bonne question

Importance capitale de connaître les données
Pas de “une solution pour les gouverner tous”

Merci !

Renaud SCHIAPPA

Responsable Adjoint

Renaud.schiappa@nice.unicancer.fr