



LA GESTION DE DONNÉES SENSIBLES : L'ANONYMISATION MULTILINGUE POUR LES ADMINISTRATIONS PUBLIQUES PROJET MAPA

Victoria Arranz, ELDA
arranz@elda.org
20 octobre 2021



CONTEXTE POUR LE PARTAGE DE DONNÉES PUBLIQUES

- Plusieurs directives pour le partage de données :
 - Directive PSI (2003 / 2013) -> encourage le partage et la réutilisation des données provenant du secteur public
 - RGPD (2016) (Règlement général sur la protection de données personnelles) -> exige que les informations personnelles soient éliminées
- Complexité pour les organismes publics :
 - la suppression des informations sensibles dans les textes est une tâche complexe
- Besoin :
 - Des outils d'anonymisation / "désidentification"

TECHNIQUES POUR LA GESTION DES DONNÉES PERSONNELLES

- *Anonymisation:*
 - Des informations sensibles sont traitées (par ex., remplacement par vides ou XXX) et aucune identification de la personne n'est possible
 - Les données anonymisées : leur source ne peut pas être repérée, mais leur valeur et réutilisation sont très limitées
- *Pseudo-anonymisation:*
 - Technique de "désidentification" qui remplace les identifiants par des pseudonymes (des catégories, des éléments de la même catégorie, etc.)
 - Les données "désidentifiées" sont conformes au RGPD et réutilisables



MAPA

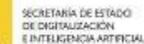


ANONYMISATION MULTILINGUE POUR LES ADMINISTRATIONS PUBLIQUES

MAPA

ANONYMISATION MULTILANGUE POUR LES ADMINISTRATIONS PUBLIQUES

- Action CEF – Secteur Télécommunications
- Durée : 01/01/2020 - 31/12/2021
- Partenaires :



MAPA

- Objectifs :
 - Développement d'une boîte à outils pour la « désidentification » multilingue de données personnelles
 - Docker indépendant – connexion API
 - Domaines : légal et médical
 - Langues : 24 langues de l'UE
 - Soutien pour les administrations publiques européennes et eTranslation
 - Conformité au RGPD
 - Création de corpus bruts (1M de phrases par langue) et de corpus annotées (une fraction de données brutes)
- Approche :
 - Basé sur la reconnaissance d'entités nommées
 - Apprentissage sur des données annotées
 - Modèles mono- et multilingues

MAPA

- Etapes :
 - Définition des entités (informations personnelles) à “désidentifier” : spécifications
 - Entités définies en relation avec les domaines concernés : médical et légal
 - Annotation manuelle de données identifiées (pour l’entraînement)
 - Obstacles :
 - Difficulté d’obtention de données sensibles sur lesquelles travailler :
 - les rapports cliniques sont confidentiels
 - les données du tribunal sont déjà anonymisées dans certains pays (France)
 - Développement pour le français sur :
 - EUR-LEX
 - Cour de Cassation : données “reconstituées” automatiquement avec des listes d’entités
 - Cas cliniques traités automatiquement pour enrichir leur contenu en entités

Cassation

Mme MOUILLARD, président

Arrêt n° 681 FS-P+B
Pourvoi n° W 16-28.281

R É P U B L I Q U E F R A N Ç A I S E

AU NOM DU PEUPLE FRANÇAIS

ARRÊT DE LA COUR DE CASSATION, CHAMBRE COMMERCIALE, FINANCIÈRE ET ÉCONOMIQUE, DU 4 NOVEMBRE 2020
M. O... B..., domicilié [...], a formé le pourvoi n° W 16-28.281 contre l'arrêt rendu le 13 novembre 2019
1°/ à la société Cooper International Spirits, dont le siège est [...],
2°/ à la société Etablissements Gabriel Boudier, société anonyme, dont le siège est [...],
3°/ à la société St Dalfour, société par actions simplifiée, dont le siège est [...],
défenderesses à la cassation.

DONNÉES
“RECONSTITUÉES”

Cassation

Mme MOUILLARD, président

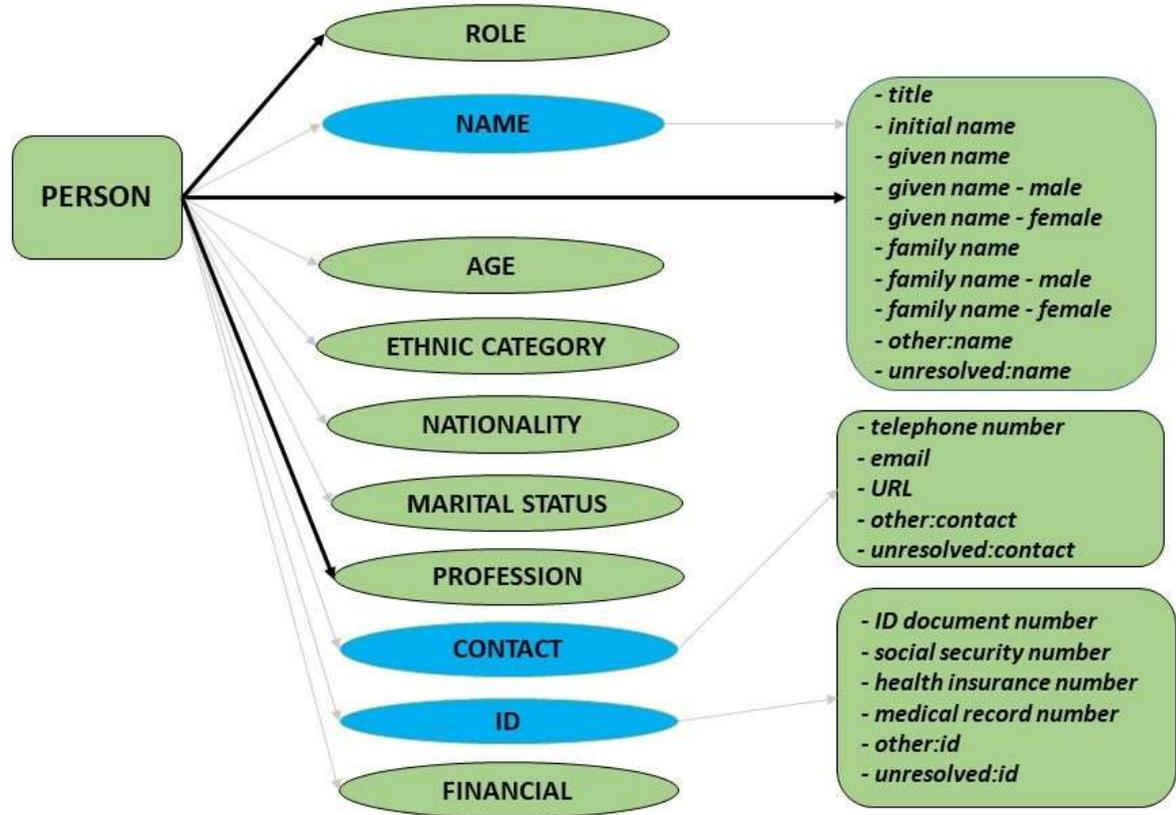
Arrêt n° 681 FS-P+B
Pourvoi n° W 16-28.281

R É P U B L I Q U E F R A N Ç A I S E

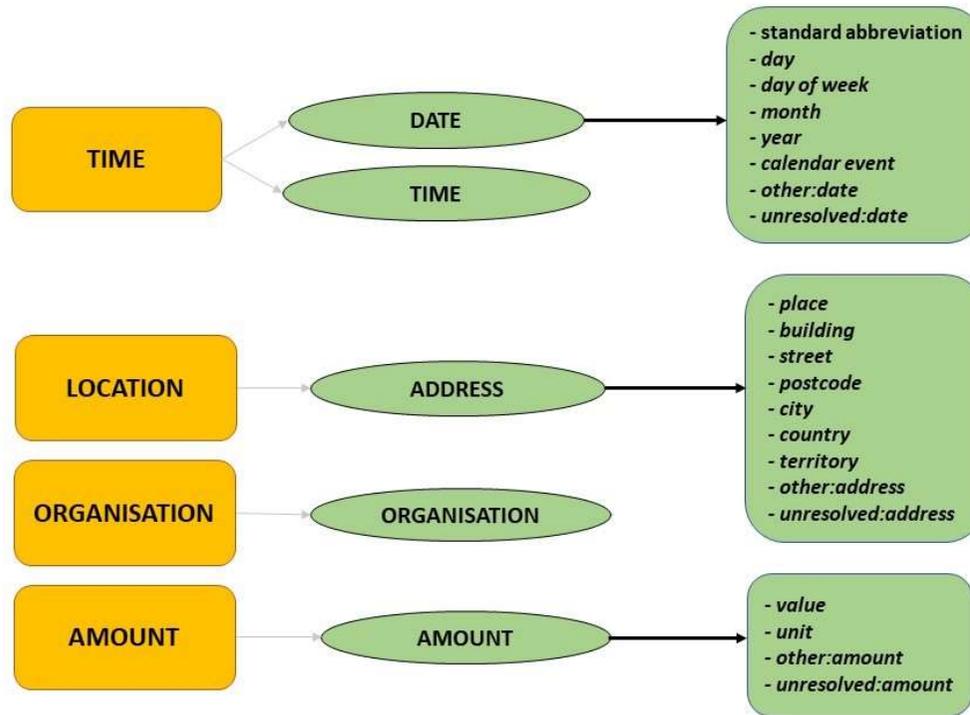
AU NOM DU PEUPLE FRANÇAIS

ARRÊT DE LA COUR DE CASSATION, CHAMBRE COMMERCIALE, FINANCIÈRE ET ÉCONOMIQUE, DU 4 NOVEMBRE 2020
M. Olivier Barbaret, domicilié au 16 Rue du Maréchal Foch, 94480 Ablon-sur-Seine, a formé le pourvoi n° W 16-28.281
1°/ à la société Cooper International Spirits, dont le siège est au 51 avenue du Général Leclerc à Pantin,
2°/ à la société Etablissements Gabriel Boudier, société anonyme, dont le siège est au 23 avenue Jean Jaurès à Pantin,
3°/ à la société St Dalfour, société par actions simplifiée, dont le siège est au 12 avenue du Colonel Fabien à Pantin,
défenderesses à la cassation.

L'HIERARCHIE
D'ENTITÉS
NOMMÉES



L'HIERARCHIE D'ENTITÉS NOMMÉES



L'HIERARCHIE
D'ENTITÉS
NOMMÉES

VEHICLE

VEHICLE

- *licence plate number*
- *colour*
- *type*
- *model*
- *build year*
- *other:vehicle*
- *unresolved:vehicle*

OUTIL D'ANNOTATION

INCEption Projects Dashboard

TEXTE
ANNOTÉ
(DOMAINE LÉGAL)

ELDA: ELDA/JURITEXT00036947082.txt 36-45 / 45 sentences [doc 77 / 274]

36 3°) Alors, d'autre part, que la clause intitulée « article 13 – Clause attributive de juridiction » stipulant, selon les propres constatations du président du tribunal, « En cas de litige ou de contestation, seuls les tribunaux du siège social de l'Opérateur de Transport et/ou de logistique sont compétents, même en cas de pluralité de défendeurs ou d'appels en garantie. Les présentes conditions générales de vente de la Fédération des Entreprises de Transport et Logistique de France (TLF) entrent en vigueur le 3 mai 2010 » est insérée dans un document intitulé « conditions générales de vente de la Fédération des Entreprises de Transport et Logistique de France (TLF) » ; qu'en affirmant néanmoins que cette clause figurait dans les conditions générales de vente du Groupe Bolloré que la société Sopam aurait, selon lui, acceptées, le président du tribunal a dénaturé ces documents en violation de l'article 1134, devenu 1103 du code civil ;

37 4°) Alors, de plus, que l'article « 4.4. Conditions générales de vente » figurant dans le document remis par la société Saga à la société Sopam stipulait « cette cotation est régie par les conditions générales de vente du groupe. Bolloré jointes en annexe » ; qu'il résulte des termes clairs et précis de cette clause que le renvoi aux conditions générales de vente avait pour seul objet la « cotation », c'est-à-dire le prix de la prestation ; qu'en affirmant néanmoins que cette clause renvoyait plus généralement aux conditions générales de vente, y compris à la clause attributive de juridiction qui y serait insérée, le président du tribunal a dénaturé ce document, en violation de l'article 1134, devenu 1103 du code civil ;

38 5°) Alors, par ailleurs, que le juge doit procéder à une analyse même sommaire des éléments qu'il retient au soutien de sa décision ; qu'en retenant, pour dire les conditions générales de vente de la Fédération des Entreprises de Transport et Logistique de France applicables à un contrat conclu en août 2009 nonobstant leur date d'entrée en vigueur le 3 mai 2010, que les parties avaient décidé de les appliquer à leurs contrats, sans procéder à une analyse même sommaire des éléments desquels il retenait une volonté de la société Sopam en ce sens, le président du tribunal de grande instance a méconnu les exigences de l'article 455 du code de procédure civile ;

39 6°) Alors, ensuite, que le contrat de commission de transport du 31 mars 2007 par lequel les parties acceptaient de se soumettre aux conditions générales de vente de la Fédération des Entreprises de Transport et Logistique de France entrées en vigueur le 1er octobre 2001 était conclu par la société Saga avec la société

OUTIL D'ANNOTATION



INCEption Projects Dashboard

ELDA: ELDA/filepdf-166-cas_CB.txt 1-6 / 6 sentences [doc 137 / 274]

1 Une jeune femme âgée de 31 ans présentant un diabète immunodépendant (DID) évoluant depuis l'âge de 1 an, en dialyse depuis 9 mois et éligible pour une double greffe rein/pancréas simultanée (SPK) a reçu en Mai 2001 un rein et un pancréas provenant d'un donneur âgé de 19 ans. Le pancréas a été placé dans la fosse iliaque droite avec une dérivation digestive des sécrétions exocrines et une dérivation systémique de l'effluent veineux du pancréas. Le rein gauche a été placé dans la fosse iliaque gauche après ouverture de la gouttière pariéto colique gauche, et implanté sur les vaisseaux iliaques externes par deux anastomoses vasculaires artérielle latéroterminale et veineuse termino-latérale. Le compte rendu opératoire ne fait pas mention d'une repéritonisation de la loge de transplantation rénale. Le traitement immunosuppresseur a associé prednisonne, ciclosporine, MMF et ATG lors de l'induction. Les suites immédiates ont été marquées par une correction rapide du diabète et de insuffisance rénale et Mme Gehant à quittée le service deux semaines après l'intervention avec des glycémies à jeun à 4.8 mmol/l et une créatininémie à 105 µmol/l.

2 Quatre mois après la greffe une élévation de la créatininémie à 212 µmol/l a conduit à une hospitalisation de Mme Gehant qui n'a permis de mettre en évidence qu'une infection urinaire pour laquelle elle a été traitée, ce qui a permis un retour à une valeur normale de créatininémie.

3 Deux mois plus tard Mme Gehant a de nouveau été admise en hospitalisation pour une élévation brutale de la créatinine à 268 µmol/l sans température et sans signe d'accompagnement clinique. Un écho doppler de l'artère rénale réalisé en urgence a constaté une élévation des index de résistance parenchymateux à 0.85 ce qui a conduit à effectuer une biopsie rénale à la recherche d'un rejet aigu, geste qui s'est compliquée d'un hématome intrapéritonéal. A l'anatomopathologie il y avait des lésions tubulaires isolées sans signe de rejet, sans infiltration lymphocytaire ni lésion vasculaire. Aucun traitement spécifique ne fut entrepris et la créatinine redescendit spontanément à 136 µmol/l en moins d'une semaine.

4 Un mois plus tard Mme Gehant fut de nouveau hospitalisée avec un tableau d'asthénie et de douleur du transplant associé à un fébricule et à une élévation de la créatininémie à 568 µmol/l. Par ailleurs l'écho doppler mettait de nouveau en évidence des index de résistance artériels élevés à 0.84. Le tableau clinique fut considéré comme un épisode de rejet aigu et traité d'emblée par bolus de corticoïdes sans qu'un contrôle biopsique puisse être réalisé pour des raisons logistiques. La symptomatologie clinique s'est rapidement amendée et la créatininémie est redescendue en 8 jours à 135 µmol/l.

5 Persistait cependant une petite dilatation du bassinnet à 14 mm pour laquelle une scintigraphie rénale fut prévue pour le mois suivant. Dans les heures qui précédèrent cet examen Mme Gehant se plaignit de douleurs abdominales et de nausées associée à une oligurie, à la constitution d'œdèmes et à une élévation de la créatininémie à 560 µmol/l, le rein étant bien palpé dans la fosse iliaque gauche et légèrement douloureux. Au retour de la scintigraphie un nouveau examen permis de constater une

MAPA : DÉTECTION



- ADDRESS ● AMOUNT ● DATE ● ORGANISATION ● PERSON ● TIME ● VEHICLE ● AGE ● ETHNIC CATEGORY ● FINANCIAL ● ID document number ● MARITAL STATUS ● NATIONALITY ● PROFESSION ● ROLE
- build year ● building ● calendar event ● city ● colour ● country ● day ● day of week ● email ● family name ● family name - female ● family name - male ● given name ● given name - female ● given name - male
- health insurance number ● initial name ● licence plate number ● medical record number ● model ● month ● otheraddress ● otheramount ● othercontact ● otherdate ● otherid ● othersname ● otherrole
- othervehicle ● place ● postcode ● social security number ● standard abbreviation ● street ● telephone number ● territory ● title ● type ● unit ● unresolvedaddress ● unresolvedamount ● unresolvedcontact
- unresolveddate ● unresolvedid ● unresolvedname ● unresolvedvehicle ● url ● value ● year

Enter or paste a text:

que le patient, qui avait fugué de l'établissement le même jour, l'a réintégré le 8 novembre ;
 que, le 14 novembre, le préfet a saisi le juge des libertés et de la détention d'une demande de prolongation de la mesure ;
 Sur la recevabilité du pourvoi en ce qu'il est dirigé contre l'hôpital Sainte-Marie de Privas, relevée d'office après avis donné aux parties conformément aux dispositions de l'article 1015 du code de procédure civile :
 Vu les articles R. 3211-13 et R. 3211-19 du code de la santé publique ;
 Attendu que le pourvoi formé contre l'hôpital Sainte-Marie de Privas, présent à l'audience pour avoir été avisé conformément aux textes précités, mais qui n'était pas partie à l'instance, n'est pas recevable ;
 Sur le moyen unique, pris en sa première branche :
 Vu l'article L. 3211-12-1 du code de la santé publique ;

Available models:

Detect entities Obfuscate entities Replace entities

Run it!

A total of 326 tokens processed in 0.70 seconds using CPU

10 que le patient, qui avait fugué de l'établissement le même jour, l'a réintégré le 8 novembre ;
 11 que, le 14 novembre, le préfet a saisi le juge des libertés et de la détention d'une demande de prolongation de la mesure ;
 12 Sur la recevabilité du pourvoi en ce qu'il est dirigé contre l'hôpital Sainte-Marie de Privas, relevée d'office après avis donné aux parties conformément aux dispositions de l'article 1015 du code de procédure civile :
 13 Vu les articles R. 3211-13 et R. 3211-19 du code de la santé publique ;
 14 Attendu que le pourvoi formé contre l'hôpital Sainte-Marie de Privas, présent à l'audience pour avoir été avisé conformément aux textes précités, mais qui n'était pas partie à l'instance, n'est pas recevable ;

MAPA : OBFUSCATION



- ADDRESS ● AMOUNT ● DATE ● ORGANISATION ● PERSON ● TIME ● VEHICLE ● AGE ● ETHNIC CATEGORY ● FINANCIAL ● ID document number ● MARITAL STATUS ● NATIONALITY ● PROFESSION ● ROLE
- build year ● building ● calendar event ● city ● colour ● country ● day ● day of week ● email ● family name ● family name - female ● family name - male ● given name ● given name - female ● given name - male
- health insurance number ● initial name ● licence plate number ● medical record number ● model ● month ● otheraddress ● otheramount ● othercontact ● otherdate ● otherid ● othername ● otherrole
- othersvehicle ● place ● postcode ● social security number ● standard abbreviation ● street ● telephone number ● territory ● title ● type ● unit ● unresolvedaddress ● unresolvedamount ● unresolvedcontact
- unresolveddate ● unresolvedid ● unresolvedname ● unresolvedvehicle ● url ● value ● year

Enter or paste a text:

Attendu, selon l'ordonnance attaquée, rendue par le premier président d'une cour d'appel, et les pièces de la procédure, que, le 4 novembre 2016, le représentant de l'Etat dans le département a pris, à l'égard de M. Xie, une décision de réadmission en soins sans consentement sous la forme d'une hospitalisation complète, en application des dispositions de l'article L. 3212-3 du code de la santé publique ;
 que le patient, qui avait fugué de l'établissement le même jour, l'a réintégré le 8 novembre ;
 que, le 14 novembre, le préfet a saisi le juge des libertés et de la détention d'une demande de prolongation de la mesure ;
 Sur la recevabilité du pourvoi en ce qu'il est dirigé contre l'hôpital Sainte-Marie de Privas, relevée d'office après avis donné aux parties conformément aux dispositions de l'article 1015 du code de procédure civile ;
 Vu les articles R. 3211-13 et R. 3211-19 du code de la santé publique ;

Available models: Detect entities Obfuscate entities Replace entities

A total of **326 tokens** processed in **0.56 seconds** using **CPU**

9 Attendu, selon l'ordonnance attaquée, rendue par le premier président d'une cour d'appel, et les pièces de la procédure, que, le 4 novembre 2016, le représentant de l'Etat dans le département a pris, à l'égard de M. ***. une décision de réadmission en soins sans consentement sous la forme d'une hospitalisation complète, en application des dispositions de l'article L. 3212-3 du code de la santé publique ;

10 que le patient, qui avait fugué de l'établissement le même jour, l'a réintégré le 8 novembre ;

11 que, le 14 novembre, le préfet a saisi le juge des libertés et de la détention d'une demande de prolongation de la mesure ;

12 Sur la recevabilité du pourvoi en ce qu'il est dirigé contre l'hôpital Sainte***** de ***** , relevée d'office après avis donné aux parties conformément aux dispositions de l'article 1015 du code de procédure civile ;

13 Vu les articles R. 3211-13 et R. 3211-19 du code de la santé publique ;

MAPA : REMPLACEMENT D'ENTITÉS



- ADDRESS
- AMOUNT
- DATE
- ORGANISATION
- PERSON
- TIME
- VEHICLE
- AGE
- ETHNIC CATEGORY
- FINANCIAL
- ID document number
- MARITAL STATUS
- NATIONALITY
- PROFESSION
- ROLE
- build year
- building
- calendar event
- city
- colour
- country
- day
- day of week
- email
- family name
- family name - female
- family name - male
- given name
- given name - female
- given name - male
- health insurance number
- initial name
- licence plate number
- medical record number
- model
- month
- otheraddress
- otheramount
- othercontact
- otherdate
- otherid
- othername
- otherrole
- othervehicle
- place
- postcode
- social security number
- standard abbreviation
- street
- telephone number
- territory
- title
- type
- unit
- unresolvedaddress
- unresolvedamount
- unresolvedcontact
- unresolveddate
- unresolvedid
- unresolvedname
- unresolvedvehicle
- url
- value
- year

Enter or paste a text:

que le patient, qui avait fugué de l'établissement le même jour, l'a réintégré le 8 novembre ;
 que, le 14 novembre, le préfet a saisi le juge des libertés et de la détention d'une demande de prolongation de la mesure ;
 Sur la recevabilité du pourvoi en ce qu'il est dirigé contre l'hôpital Sainte-Marie de Privas, relevée d'office après avis donné aux parties conformément aux dispositions de l'article 1015 du code de procédure civile :
 Vu les articles R. 3211-13 et R. 3211-19 du code de la santé publique ;
 Attendu que le pourvoi formé contre l'hôpital Sainte-Marie de Privas, présent à l'audience pour avoir été avisé conformément aux textes précités, mais qui n'était pas partie à l'instance, n'est pas recevable ;
 Sur le moyen unique, pris en sa première branche :
 Vu l'article L. 3211-12-1 du code de la santé publique ;

Available models: fr_legal+bert_multi
 Detect entities
 Obfuscate entities
 Replace entities

Run it!

A total of **326 tokens** processed in **2.95 seconds** using **CPU**

10 que le patient, qui avait fugué de l'établissement le même jour, l'a réintégré le 15 avril ;
 11 que, le 12 avril, le préfet a saisi le juge des libertés et de la détention d'une demande de prolongation de la mesure ;
 12 Sur la recevabilité du pourvoi en ce qu'il est dirigé contre l'hôpital Sant-Jeanne de Nantes, relevée d'office après avis donné aux parties conformément aux dispositions de l'article 1015 du code de procédure civile :
 13 Vu les articles R. 3211-13 et R. 3211-19 du code de la santé publique ;

MAPA

- Bons résultats préliminaires :
 - Cas cliniques français : 90-92%
 - MEDDOCAN espagnol : jusqu'à 98-99% (données très structurées)
- V2 publique début novembre :
 - Apprentissage avec les dernières données annotées produites
 - Approche de BERT avec optimisation : le manque de données dans certaines langues est compensé par des modèles mono- et multilingues combinés
- Cas d'usage :
 - Européen :
 - Observatoire des plaintes (légal)
 - National
 - Hôpital parisien -> évaluation du système en cours d'organisation (médical)
 - Ministère de la Justice espagnol (légal)



MERCI POUR VOTRE ATTENTION!

Victoria Arranz
arranz@elda.org

Projet MAPA
<https://mapa-project.eu/>

