
DMP du projet "Traitement de données hétérogènes image / signal pour l'analyse des trajectoires neurodéveloppementales des nouveau-nés prématurés"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "Science Europe : modèle structuré" fourni par Science Europe.

Renseignements sur le plan

Titre du plan	DMP du projet "Traitement de données hétérogènes image / signal pour l'analyse des trajectoires neurodéveloppementales des nouveau-nés prématurés"	
Version	Version initiale	
Domaines de recherche (selon classification de l'OCDE)	Computer and information sciences	
Langue	fra	
Date de création	2023-04-01	
Date de dernière modification	2024-07-06	
Licence	Nom	Creative Commons Attribution Share Alike 4.0 International
	URL	http://spdx.org/licenses/CC-BY-SA-4.0.json

Renseignements sur le projet

Titre du projet Traitement de données hétérogènes image / signal pour l'analyse des trajectoires neurodéveloppementales des nouveau-nés prématurés

Acronyme HINT

Résumé L'imagerie médicale fournit un large spectre de modalités donnant accès à des données de nature physique et de dimensions variées. Chaque modalité d'imagerie fournit des informations précieuses pour le diagnostic et le suivi de patients. Néanmoins, la mise en relation de ces informations est rendue complexe par l'hétérogénéité entre ces données, encore accrue par les différences de matériels et de protocoles d'acquisitions, notamment dans le cas d'études multicentriques. En néonatalogie, et plus particulièrement dans les services prenant en charge les nouveau-nés prématurés, les technologies d'imagerie sont indispensables, tant pour le traitement quotidien des patients que pour les recherches visant à appréhender les mécanismes de maturation fonctionnelle du cerveau, et ainsi développer de nouveaux protocoles de soin. Néanmoins, il n'existe que peu d'outils appropriés pour la mise à profit des données sources de telles études (IRM, échographiques, EEG). Notre but est de développer des méthodes numériques et outils logiciels permettant aux cliniciens d'étudier les trajectoires de neurodéveloppement en néo- et périnatologie. Dans ce cadre, les images disponibles sont toutefois beaucoup moins nombreuses que chez l'adulte, ajoutant aux problématiques d'hétérogénéité des difficultés liées à la faible volumétrie de jeux de données, susceptibles de mettre en échec les paradigmes usuels d'extraction d'information. Nous adressons ces problématiques sous plusieurs angles : la mise en correspondance de données multiphysiques, multidimensionnelles hétérogènes (IRM 3D vs. échographie 2D, multicentriques) ; l'aide à l'analyse d'information en signal (EEG) et imagerie par des approches d'IA dans un contexte hétérogène "small data" ; la mise à profit du couplage image / signal, par le développement d'outils de visualisation multiphysique. Ces outils seront mis à profit pour le traitement de différentes cohortes de données néonatalogiques (CALIN, Epirmex, Lytonepal).

Sources de financement

- Agence Nationale de la Recherche : ANR-22-CE45-0034

Date de début 2023-03-01

Date de fin 2028-02-29

Partenaires

- GROUPE DE RECHERCHE SUR L'ANALYSE MULTIMODALE DE LA FONCTION CÉRÉBRALE - UMR INSERM_S 1105
- KITWARE
- LABORATOIRE DE TRAITEMENT DE L'INFORMATION MÉDICALE
- Centre de REcherche en STIC <https://ror.org/016wdna72>

Produits de recherche :

1. Base de donnée CALIN (IRM, US, EEG) (Jeu de données)
2. Base de données GRAMFC (Jeu de données)

Contributeurs

Nom	Affiliation	Rôles
Dollé Guillaume - 0000-0002- 5844-2536	Laboratoire de Mathématiques de Reims - 201019023N	
Finet Julien - 0000-0002- 6462-9049	Kitware	
Loron Gauthier - 0000-0001- 9536-9510	CReSTIC - 200415064L	<ul style="list-style-type: none"> • Collecteur de données (CALIN) • Personne contact pour les données (CALIN) • Responsable de la conservation à long terme des données (CALIN) • Responsable de la documentation des données (CALIN) • Responsable de la production ou de la collecte des données (CALIN) • Responsable de la protection des données (CALIN) • Responsable du dépôt et de la diffusion des données (CALIN) • Responsable du stockage des données (CALIN) • Responsable du traitement et de l'analyse des données (CALIN)
Passat Nicolas		<ul style="list-style-type: none"> • Responsable du plan de gestion de données
PASSAT Nicolas	Centre de REcherche en STIC - 200415064L	<ul style="list-style-type: none"> • Coordinateur du projet
Rousseau François - 0000-0001- 9837-7487	LATIM - 201220084H	
Wallois Fabrice - 0000-0003- 2928-5428	GRAMFC - 201220180M	<ul style="list-style-type: none"> • Personne contact (GRAMFC, GRAMFC, GRAMFC, GRAMFC, GRAMFC, GRAMFC, GRAMFC, GRAMFC, GRAMFC, GRAMFC)

Budget

Type de coût (Titre)	Montant	Etape du cycle de vie
Stockage	0 EUR	Coûts liés à la collecte/production des données - CALIN
Stockage	0 EUR	Coûts liés à la réutilisation de données existantes - CALIN
Matériel informatique (Serveurs)	0 EUR	Coûts liés à la réutilisation de données existantes - CALIN
Matériel informatique (Serveurs)	0 EUR	Coûts liés à la collecte/production des données - CALIN
Stockage	0 EUR	Coûts liés à la collecte/production des données - GRAMFC
Matériel informatique (Serveurs)	0 EUR	Coûts liés à la collecte/production des données - GRAMFC
Stockage	0 EUR	Coûts liés à la réutilisation de données existantes - GRAMFC
Matériel informatique	0 EUR	Coûts liés à la réutilisation de données existantes - GRAMFC

Droits d'auteur :

Le(s) créateur(s) de ce plan accepte(nt) que tout ou partie de texte de ce plan soit réutilisé et personnalisé si nécessaire pour un autre plan. Vous n'avez pas besoin de citer le(s) créateur(s) en tant que source. L'utilisation de toute partie de texte de ce plan n'implique pas que le(s) créateur(s) soutien(nen)t ou aient une quelconque relation avec votre projet ou votre soumission.

DMP du projet "Traitement de données hétérogènes image / signal pour l'analyse des trajectoires neurodéveloppementales des nouveau-nés prématurés"

Description des données et collecte ou réutilisation de données existantes

Base de donnée CALIN (IRM, US, EEG)

Description générale du produit de recherche

Nom Base de donnée CALIN (IRM, US, EEG)

Description Un jeu de données anonymes est créé à partir d'une base de données existantes issues du PHRCI-18-029 sur la cohorte CALIN. Cette base contiendra des données d'acquisitions brutes décrites dans la table ci-dessous.

Type de données d'acquisition	Détails	Format de stockage	Quantité
• IRM - Image par résonance magnétique		NiftI-2	N/A
• US - Échographie		NiftI-2	N/A
• EEG - Électroencéphalogramme		EDF+	N/A
• aEEG - Électroencéphalogramme d'amplitude		EDF+	N/A

Les données seront structurées au format **BIDS** (Brain Imaging Data Structure) initié originellement par le dépôt [Openneuro](#). Elles seront préparées pour une ouverture éventuelle vers ce dépôt ou assimilé. Ces données anonymes seront hébergées chez le partenaire producteur (CHU de Reims) et identifiées par un PID non précisé à ce stade.

Type Jeu de données

Workpackage WP1

Mots clés

- IRM (Image par résonance magnétique)
- US (Échographie à ultrason)
- EEG (Électroencéphalogramme)

Mots clés (texte libre)

Langue fra

Type d'identifiant Identifiant local

Contient des données personnelles ? Non

Contient des données sensibles ? Oui

Prend en compte des aspects éthiques ? Oui

Est-ce que des données existantes seront réutilisées ?

Justification

Les données collectées (cf. 1.1) sont des données extraites du PHRCI, anonymisées, puis réutilisées dans le cadre du projet ANR pour développer des outils de traitement et d'analyse à destination du milieu clinique.

Les données d'origine sont premièrement qualifiées pour l'étude, anonymisées intégralement (métadonnées et données) avant d'être intégrées à la nouvelle base.

Chaque partenaire du projet dispose d'un accès contrôlé à la base.

Données réutilisées

- PHRCI : PHRCI-18-029

Coûts

- Stockage : 0 EUR
- Matériel informatique : 0 EUR

Comment seront produites/collectées les nouvelles données ?

Nom de la méthode

Protocole CALIN

Description

De nouvelles données pourront être introduites suivant les modalités d'acquisition décrites en 1.1 en suivant le protocole CALIN détaillé dans le PHRCI (en cours de collecte).

Nature des données

Données expérimentales

Coûts

- Stockage : 0 EUR
- Matériel informatique : 0 EUR

Base de données GRAMFC

Description générale du produit de recherche

Nom	Base de données GRAMFC												
Description	<p>Production d'un jeu de données issue d'une base générée dans le cadre de l'ANR MAIA est introduit dans le projet. Cette base contiendra des données d'acquisitions brut précisées dans la table ci-dessous.</p> <table border="1"> <thead> <tr> <th>Type de donnée d'acquisition</th> <th>Détails</th> <th>Format de stockage</th> <th>Quantité</th> </tr> </thead> <tbody> <tr> <td>• NIRS - Spectroscopie proche infrarouge</td> <td></td> <td>• Matlab</td> <td>• N/A</td> </tr> <tr> <td>• HR EEG - Électroencéphalogramme haute résolution</td> <td>• 64/128 canaux</td> <td>• EDF+</td> <td>• N/A</td> </tr> </tbody> </table> <p>Les données seront structurées au format BIDS (Brain Imaging Data Structure initié par le dépôt Openneuro). Ces données anonymes seront hébergés chez le partenaire producteur de données.</p>	Type de donnée d'acquisition	Détails	Format de stockage	Quantité	• NIRS - Spectroscopie proche infrarouge		• Matlab	• N/A	• HR EEG - Électroencéphalogramme haute résolution	• 64/128 canaux	• EDF+	• N/A
Type de donnée d'acquisition	Détails	Format de stockage	Quantité										
• NIRS - Spectroscopie proche infrarouge		• Matlab	• N/A										
• HR EEG - Électroencéphalogramme haute résolution	• 64/128 canaux	• EDF+	• N/A										
Type	Jeu de données												
Workpackage	WP1												
Mots clés (texte libre)	HR EEG Electroencéphalographie Haute résolution, HD NIRS, High Density Near Infrared spectroscopy												
Langue	anglais												
Date de publication	2024-04-11												
Identifiant pérenne													
Type d'identifiant	Identifiant local												
Contient des données personnelles ?	Non												
Contient des données sensibles ?	Oui												
Prend en compte des aspects éthiques ?	Oui												

Est-ce que des données existantes seront réutilisées ?

Justification	<p>Les données collectées sont des données extraites de l'ANR MAIA et des données obtenus dans le cadre clinique, anonymisées, puis réutilisées dans le cadre du projet ANR pour développer des outils de traitement et d'analyse à destination du milieu clinique.</p> <p>Les données d'origine sont premièrement qualifiées pour l'étude, anonymisées intégralement (métadonnées et données) avant d'être intégrées à la nouvelle base.</p> <p>Chaque partenaire du projet dispose d'un accès contrôlé à la base.</p>
Données réutilisées	<ul style="list-style-type: none"> Données ANR MAIA :
Coûts	<ul style="list-style-type: none"> Stockage : 0 EUR Matériel informatique : 0 EUR

Comment seront produites/collectées les nouvelles données ?

Nom de la méthode	Analyse multimodale de la fonction cérébrale
Description	Acquisition multimodale simultanée ou pas de l'électroencéphalographie Haute Résolution et de la spectroscopie dans le proche infrarouge haute définition
Nature des données	Données expérimentales
Références associées	<ul style="list-style-type: none"> • EEG HR : • NIRS HD :
Coûts	<ul style="list-style-type: none"> • Stockage : 0 EUR • Matériel informatique : 0 EUR

Documentation et qualité des données

Base de donnée CALIN (IRM, US, EEG)	
Quelles métadonnées et quelle documentation (par exemple mode d'organisation des données) accompagneront les données ?	
Description	Les données seront structurées au format standard BIDS (https://bids.neuroimaging.io/) qui définit la convention de nommage des fichiers.
Standards de métadonnées/données	<ul style="list-style-type: none"> • JSON-LD (JavaScript Object Notation for Linked Data) : https://rdamsc.bath.ac.uk/msc/m95 • Brain Imaging Data Structure (BIDS) : https://rdamsc.bath.ac.uk/api2/m113
Code langue des métadonnées	eng
Logiciel de documentation	Dataverse
Quelles seront les méthodes utilisées pour assurer la qualité scientifique des données ?	
Description	Les données sont acquises dans le cadre d'une PHRCi (CALIN : enregistrement ansm 2020-A00382-37) qui définit notamment les protocoles d'acquisition.

Base de données GRAMFC	
Quelles métadonnées et quelle documentation (par exemple mode d'organisation des données) accompagneront les données ?	

Description	Les données seront structurées au format standard BIDS qui définit la convention de nommage des fichiers.
Standards de métadonnées/données	<ul style="list-style-type: none"> • JSON-LD (JavaScript Object Notation for Linked Data) : https://rdamsc.bath.ac.uk/msc/m95 • Brain Imaging Data Structure (BIDS) : https://rdamsc.bath.ac.uk/api2/m113
Code langue des métadonnées	anglais
<hr/>	
Quelles seront les méthodes utilisées pour assurer la qualité scientifique des données ?	
Description	Les données sont acquises de manière monocentrique, et bénéficient donc d'une homogénéité en termes de matériel et de protocole d'acquisition.

Exigences légales et éthiques, code de conduite

Base de donnée CALIN (IRM, US, EEG)	
Quelles sont les contraintes juridiques (sensibilité des données autres qu'à caractère personnel, confidentialité, ...) à prendre en compte pour le partage et le stockage des données ?	
Description	Cas particulier de l'identification morphométrique dans la donnée IRM : la résolution spatiale des images IRM pouvant permettre une éventuelle reconnaissance des traits faciaux, un traitement spécifique est mis en oeuvre afin de limiter l'identification par recoupement (cropping de la zone intracrânienne) et garantir l'anonymat.
<hr/>	
Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?	
Description	La collecte des données est réalisée à partir des données issues dans le cadre du PHRCi (cf. DMP PHRCi, référence ci-avant dans le document) avec l'aval du comité d'éthique. Les données recueillies à partir de cette base sont, pour leur part, anonymes.
Références associées	<ul style="list-style-type: none"> • PHRCi (CALIN) : 2020-A00382-37
<hr/>	
Base de données GRAMFC	
Quelles sont les contraintes juridiques (sensibilité des données autres qu'à caractère personnel, confidentialité, ...) à prendre en compte pour le partage et le stockage des données ?	
Description	
<hr/>	
Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?	

Description

La collecte des données est réalisée à partir des données obtenues dans le cadre de différents projets de recherches avec l'aval du comité d'éthique. Les données recueillies sont, pour leur part, anonymes.

Références associées

- CPP : ID-RCB : 2021-A02556-35
- CPP : ID-RCB : 2024-A00758-33
- CPP : ID-RCB : 2019-A01534-53
- CPP : ID-RCB : 2008-A00728-47
- CPP : CCPPRN : 04H23 DGS : 2005/0015
- CPP : ID-RCB : 2011-A01452-39
- CPP : ID-RCB : 2011-A01295-44

Traitement et analyse des données

Base de donnée CALIN (IRM, US, EEG)**Comment et avec quels moyens seront traitées les données ?****Description**

Les données sont analysées à l'aide de différents outils de visualisation et de traitement : Deltamed Coherence, SAS, 3DSlicer sur les postes opérationnels et sur les serveurs du CHU.

Les données seront également traitées sur les plateformes des partenaires pour les différents traitements, notamment sur le mésocentre régional ROMEO.

Equipements, plateformes techniques

- ROMEO : <https://cat.opidor.fr/index.php/ROMEO>

Base de données GRAMFC**Comment et avec quels moyens seront traitées les données ?****Description**

Les données sont acquises à partir de dispositifs médicaux pour l'EEG et de dispositifs de NIRS ayant l'accord de l'ANSM. Les données sont ensuite traitées via des codes développés sous Matlab ou Python ou à l'aide de logiciels tels que Curry.

Stockage et sauvegarde des données pendant le processus de recherche

Base de donnée CALIN (IRM, US, EEG)**Comment les données seront-elles stockées et sauvegardées tout au long du projet ?**

Besoins de stockage	<p>Les données seront stockées sur un serveur localisé chez le partenaire producteur de données, à savoir le CHU de Reims. Le stockage s'appuie sur les ressources de stockage propre au CHU qui possède sa politique d'archivage et de sécurité des données. Le stockage initial sera basé sur du stockage NAS avec données stockées à plat selon la hiérarchie proposée par le standard BIDS.</p> <p>Le stockage pourrait évoluer à terme vers un dépôt de données déployé en local, basé sur un outil adapté (par exemple Kitware Girder) permettant l'attribution d'un PID local (URL + unique ID).</p>
Volume estimé des données	300
Unité	Go
Mesures prises pour la sécurité des données	<p>L'outil est déployé en local avec une gestion restreinte des accès et contrôlé derrière un VPN du CHU.</p> <p>Les données sont archivées sur le centre de stockage du CHU qui possède sa propre politique d'archivage sécurisé.</p> <p>En cas de problème technique, les données sont toujours stockées sur des disques chiffrés, structurés mais sans PID.</p>

Base de données GRAMFC

Comment les données seront-elles stockées et sauvegardées tout au long du projet ?

Besoins de stockage	<p>Les données seront stockées sur un serveur localisé chez le partenaire producteur de données, à savoir le CHU d'Amiens. Le stockage s'appuie sur les ressources de stockage propre au CHU qui possède sa politique d'archivage et de sécurité des données. Le stockage initial sera basé sur du stockage NAS avec données stockés à plat selon la hiérarchie proposée par le standard BIDS.</p> <p>Le stockage pourrait évoluer à terme vers dépôt de donnée déployé en local basé sur un outil adapté (par exemple Kitware Girder) permettant l'attribution d'un PID local (URL + unique ID).</p>
Volume estimé des données	1
Unité	To
Mesures prises pour la sécurité des données	<p>Les données sont archivés sur le centre de stockage du CHU qui possède sa propre politique d'archivage sécurisée.</p> <p>En cas de problèmes techniques, les données sont toujours stockées sur des disques chiffrés, structurés mais sans PID.</p>

Partage des données et conservation à long terme

Base de données CALIN (IRM, US, EEG)

Comment les données seront-elles partagées ?

Modalités de partage Les données sont partagées uniquement entre les partenaires du projet ANR. Le jeu de données sera partagé par le biais de disques durs chiffrés. En fonction de l'évolution de la section 5.1, un accès sera fourni aux partenaires du projet concernés.

Entrepôt/Catalogue de données

- Entrepôt Recherche Data Gouv : [https://cat.opidor.fr/index.php/Entrepôt Recherche Data Gouv \(\)](https://cat.opidor.fr/index.php/Entrepôt%20Recherche%20Data%20Gouv)

Comment les données seront-elles conservées à long terme ?

Justification Ce type de jeu de données est difficile à produire, puisqu'il s'agit d'un suivi de patient à long terme de l'ordre de plusieurs années.

Volume estimé des données 300

Unité Go

Base de données GRAMFC

Comment les données seront-elles partagées ?

Modalités de partage Les données sont partagées uniquement entre les partenaires du projet ANR. Le jeu de données sera partagé par le biais de disques durs chiffrés. En fonction de l'évolution de la section 5.1, un accès sera fourni aux partenaires du projet concernés.

Potentiel de réutilisation

Entrepôt/Catalogue de données

- Entrepôt Recherche Data Gouv : : [https://cat.opidor.fr/index.php/Entrepôt Recherche Data Gouv \(\)](https://cat.opidor.fr/index.php/Entrepôt%20Recherche%20Data%20Gouv)

Comment les données seront-elles conservées à long terme ?

Justification Ce type de jeu de données est difficile à produire, puisqu'il s'agit d'un suivi de patient à long terme de l'ordre de plusieurs années.
De plus s'agissant de données pédiatriques celles-ci seront stockées sur du long terme en accord avec la réglementation dans le domaine.

Volume estimé des données 1

Unité To

Date de début 2024-01-10

Date de fin 2040-01-10

Dispositions finales