

1. Informations sur le Plan de Gestion de Données		
1.1.	Auteur(s) du PGD	ENNINGA Jost ( <a href="mailto:jost.eninga@pasteur.fr">jost.eninga@pasteur.fr</a> )
1.2.	Date de la première version du PGD	jeudi 24 mars 2022
1.3.	Version actuelle du PGD et date	V1, jeudi 24 mars 2022
1.4.	Lieu de conservation du PGD	Les versions initiales et intermédiaires sont stockées sur un serveur interne atlas/dihp et mises en ligne sur le site du projet. La version finale est déposée sur un entrepôt ( Zenodo ).
2. Informations sur le projet		
2.1.	Nom du financeur du projet et programme de financement	Le projet est financé par the European Commission
2.2.	Acronyme du projet	EndoSubvert
2.3.	Titre du projet	Common mechanisms of host membrane trafficking subversion by intracellular pathogens to rupture bacterial containing vacuoles
2.4.	Identifiant du projet	Consolidator Grant (CoG), LS6, ERC-2015-CoG
2.5.	Coordinateur du projet	ENNINGA Jost ( <a href="mailto:jost.eninga@pasteur.fr">jost.eninga@pasteur.fr</a> )
2.6.	Institution et unité de rattachement du coordinateur du projet	Institut Pasteur, Département biologie cellulaire et infection, Dynamique des interactions hôte pathogène ( Jost Enninga chef d'unité )
2.7.	Date de début du projet	1-1-17
2.8.	Date de fin du projet	31-12-22
3. Présentation générale des données		
3.1.	Quel est l'objectif de la collecte/génération des données ?	Ce projet comprend des acquisitions microscopiques (électronique, spinning, et widefield ) de Shigelles infectant différentes lignées de cellules

3.2.	Combien de jeu(x) de données allez-vous générer durant ce projet ?	1 jeu de données
3.3.	Quelle est la nature et le format des données générées/collectées ?	Les données sont générées par microscopie électronique et confocale et analysées par les logiciels Fiji, imod, serialEM et LasX. GraphPad sera utilisé pour l'analyse statistique. Excel sera utilisé pour la gestion des données, les représentations graphiques et l'analyse des données.
3.4.	Indiquez le volume prévisionnel des données générées pour ce projet	50 TB stockées sur les serveurs internes et moins de 50 Go publiables sur zenodo
3.5.	Des données existantes seront-elles réutilisées ? Si oui, précisez leur origine.	Aucune donnée ancienne ne sera utilisée dans ce projet. Tous les livrables du projet proviendront des données originales acquises au cours de cette période.
<b>4.</b>	<b>Ressources nécessaires à la gestion des données</b>	
4.1.	De quelles ressources matérielles avez-vous besoin pour la gestion de vos données ?	Nous avons assez d'espace de stockage et de matériels, et nous gérons au plus précis nos datas
4.2.	Qui assure la gestion opérationnelle des données au cours du projet de recherche ?	Le chercheur est responsable : - de la collecte, du traitement et de l'analyse des données - de la génération des métadonnées et de la documentation liée aux données - du stockage des données - de l'archivage et du partage des données
4.3.	Quelles formations ou quel accompagnement vous semblent nécessaires pour vous aider à la gestion de vos données ?	Une formation particulière a été suivie par le correspondant informatique nécessaire pour conserver et gérer les données générées dans le cadre de ce projet. L'équipe ( 10 personnes ) peut avoir accès à des réponses techniques : métadonnées, standards de métadonnées et archivage lors d'opendesk organisé à l'institut Pasteur

4.4.	Quel budget avez-vous prévu pour la gestion de vos données ? Comment comptez-vous financer ces coûts ?	L'institut et le laboratoire disposent d'un budget dédié pour assurer le stockage et la maintenance des données en interne.  En outre, le stockage dans un dépôt de données ouvert sera couvert par la Commission européenne si la convention de subvention le permet.
<b>5.</b>	<b>Aspects juridiques et éthiques</b>	
5.1.	Votre projet comporte-t-il des données à caractère personnel ?	Non, le projet ne comporte pas des données à caractère personnel.
5.2.	Votre projet comporte-t-il d'autres données soumises à une obligation contractuelle, réglementaire ou légale ? Si oui, de quel type ?	Non, le projet n'est pas soumis à une obligation contractuelle, réglementaire ou légale.
<b>6.</b>	<b>Gestion des données au cours du projet</b>	
6.1.	Quel est le lieu de stockage de vos données au cours du projet ?	Les données sont stockées sur un serveur de stockage partagé fourni par l'Institut Pasteur. De plus, les données en attente de traitement sont également stockées sur des NAS et disques durs externes au sein de notre unité.
6.2.	Utilisez-vous un plan de classement pour gérer vos fichiers de données ? Indiquez brièvement comment celui-ci est organisé.	Oui. Nous utilisons un système de dossiers hiérarchiques allant du type d'expérience, des données brutes aux dossiers de données traitées. Les données brutes ne sont jamais manipulées et sont enregistrées dans un dossier différent de celui des données traitées, analysées et compilées.
	Quelles règles de nommage utilisez-vous pour vos	Chaque fichier est nommé avec la date, le sujet et l'utilisateur qui l'a acquis. Les fichiers traités et analysés

6.3.	données ? Quelles règles utilisez-vous pour un versionnage explicite ?	suivent les données d'analyse et la version du document. Les documents textuels suivent la date et la version ainsi que les initiales des auteurs tout au long de la révision. Toutes les informations sont séparées par un trait de soulignement.
6.4.	Quelles sont les mesures mises en place pour garantir la qualité des données ?	Pour garantir la qualité des données, différentes mesures ont été mises en place : - Répétition des expériences de manière indépendante (minimum de trois répétitions sur trois jours différents) - Standardisation de la collecte des données paramètres des microscopes, culture cellulaire etc - Revue régulière des données avec le PI
<b>7.</b>	<b>Sélection des données pour l'archivage à long terme</b>	
7.1.	Existe-t-il des contraintes réglementaires de conservation de vos données ? Si oui, lesquelles ?	Pas de contrainte réglementaire particulière
7.2.	Quels sont les jeux de données ayant une valeur à long-terme et qui devraient être conservés ? Quels sont les jeux de données à détruire ?	Les jeux de données analysés et significatifs doivent être conservés en raison des difficultés de reproductibilité et du temps nécessaire à leur régénération. Leur conservation est essentielle pour assurer la reproductibilité des résultats présentés dans les publications et pour pouvoir les comparer aux données qui seront générées ultérieurement. Les jeux de données non exploitables sont détruits
7.3.	Sur quelle plateforme ou dans quel entrepôt les jeux de données à conserver seront-ils archivés ? Cette plateforme est-elle certifiée pour une conservation et une gestion à long terme ?	A la fin du projet, le jeu de données sera transféré sur l'entrepôt ZENODO qui assure un archivage durable des données de recherche finales. Les données brutes resteront stockées sur les serveurs sécurisés de l'Institut Pasteur.
7.4.	Précisez les formats choisis pour l'archivage.	Les formats PDF, TIFF, PNG et CSV seront utilisés. Ainsi que les formats DV, nd2, graphpad, tiff

7.5.	Pendant combien de temps les données seront-elles conservées ?	Les jeux de données seront conservés pendant la durée maximale permise par l'entrepôt. Pour Zenodo, cela correspond à la durée de vie du laboratoire hôte, le CERN, qui dispose actuellement d'un programme expérimental défini pour les 20 prochaines années au moins. Les données brutes seront conservées pour une durée illimitée tant que l'espace alloué au sein de l'Institut Pasteur est disponible.
7.6.	Quel est le volume final de données archivées ?	50 Tb de données brutes
7.7.	Si une conservation à long terme est requise, comment comptez-vous couvrir ces coûts ?	Les coûts associés à la conservation à long terme seront pris en charge par l'Institut Pasteur.
<b>8.</b>	<b>Jeu de données N°1</b>	
<b>8.1.</b>	<b>Description des données</b>	
8.1.1.	Identifiant et nom du jeu de données	Microscopie électronique, confocale et champs large
8.1.2.	Qui est le fournisseur ou le producteur des données ?	Ce jeu de données est généré par les chercheurs de l'unité Dynamique des interactions hôte pathogène, Institut Pasteur Paris
8.1.3.	Quelle est la nature et le format des données de ce jeu de données ?	Images aux formats PNG et TIFF, lif (leica), nd2 (Nikon ) Données tabulées au format CSV et XLSX.
8.1.4.	Décrivez de façon plus détaillée les données de ce jeu de données	Ce jeu de données comprend des acquisitions microscopiques (électronique, spinning, et widefield ) de Shigelles infectant des cellules
8.1.5.	Décrivez la méthode de collecte et/ou de génération de ces données.	Les données sont générées par microscopie électronique, confocale et champs large et analysées par les logiciels Fiji, imod, serialEM et LasX. GraphPad sera utilisé pour l'analyse statistique. Excel sera utilisé pour la gestion des données, les représentations graphiques et l'analyse des données.
8.1.6.	Décrivez votre jeu de données avec des mots-clés	microscopie électronique, Shigelle, endosome, marquage, infection cellulaire
8.1.7.	Indiquez l'URL ou l'identifiant pérenne permettant d'accéder à votre jeu de données	zenodo. L'identifiant sera disponible au moment du dépôt dans l'entrepôt

8.1.8.	Quel est volume de données dans ce jeu de données ?	50 To
<b>8.2.</b>	<b>Rendre les données librement accessibles</b>	
8.2.1.	Ce jeu de données sera-t-il librement disponible ?	Les données seront disponibles au moment de la prépublication de l'article associé.
8.2.2.	Quel entrepôt de données avez-vous choisi pour stocker et rendre accessible ce jeu de données ?	Les versions initiales et intermédiaires sont stockées sur un serveur interne atlas/dihp et mises en ligne sur le site du projet. La version finale est déposée sur un entrepôt ( Zenodo ).
8.2.3.	Ce jeu de données fera-t-il l'objet d'une demande de brevet ? Si oui, il devra rester confidentiel.	Ce jeu de données ne fera pas l'objet d'une demande de brevet.
8.2.4.	Si ce jeu de données ne peut pas être mis à disposition pour une autre raison, expliquez pourquoi.	Non applicable
8.2.5.	Précisez comment l'accès à ce jeu de données sera fourni au cas où des restrictions s'appliquent	- le jeu de données est déposé sur Zenodo mais en accès restreint. Les conditions d'accès sont précisées sur Zenodo : la personne qui souhaite télécharger le jeu de données doit au préalable expliquer comment elle a l'intention de l'utiliser. Sur la base de cette justification, il sera décidé d'accorder ou de refuser l'accès. - le jeu de données est conservé sur les serveurs de l'Institut Pasteur et accessible sur demande à <a href="mailto:jost.eninga@pasteur.fr">jost.eninga@pasteur.fr</a>
8.2.6.	Quels logiciels sont nécessaires pour visualiser ou accéder aux données ? Fournissez-vous une documentation ou le code open source du logiciel ?	Logiciels utilisés régulièrement : Outils Microsoft office Fiji / ImageJ Excel / GraphPad Imod / SerialEM / LasX
<b>8.3.</b>	<b>Rendre les données trouvables</b>	

8.3.1.	Ce jeu de données est-il identifié par un identifiant pérenne unique tel que le DOI (Digital Object Identifier) ? Sinon, décrivez comment ce jeu de données et les données sont identifiés.	Les données elles-mêmes ne sont pas identifiées par un DOI mais par un nom clair : topic_doctype_team_date_version ainsi que des dossiers pour chaque expérience. L'ensemble est compilé dans un dossier de jeu de données avec un DOI généré par ZENODO.
8.3.2.	Quels standards de métadonnées utilisez-vous ? Si vous n'utilisez pas de standard de métadonnées, précisez quel(s) type(s) de métadonnées seront créés et comment.	Les métadonnées suivront le standard de métadonnées ou les règles de l'entrepôt choisi pour stocker les données. Pour les acquisitions en microscopie, seront indiquées, les caractéristiques techniques significatives du microscope, ainsi que les cellules infectées, les bactéries ou virus infectant, et l'explication des marquages.
8.3.3.	Ce jeu de données est-il décrit par des mots-clés permettant de le retrouver facilement ?	Oui, ce jeu de données est décrit par des mots-clés
8.3.4.	Fournissez-vous une documentation complémentaire pour décrire plus précisément les données ?	Oui, un fichier est disponible pour chaque donnée pour récapituler les analyses effectuées. ( e-lab read me )
<b>8.4.</b>	<b>Rendre les données interopérables</b>	
8.4.1.	Les données de ce jeu de données sont-elles interopérables du point de vue technique ?	Oui, les images de microscopie sont au format TIFF.
8.4.2.	Sinon, quelles méthodologies appliquerez-vous pour rendre vos données interopérables ?	Nous transformons nos images dans un format ouvert et interopérable

8.4.3.	Précisez si vous utilisez du vocabulaire standard dans votre jeu de données pour permettre une interopérabilité sémantique interdisciplinaire. Dans le cas contraire, fournirez-vous un alignement avec les ontologies les plus fréquemment utilisées ?	Nous n'avons pas utilisé d'ontologie spécifique mais nous utilisons un vocabulaire uniforme dans l'ensemble du jeu de données. Le vocabulaire utilisé est basé sur des termes couramment utilisés dans les sciences du vivant, les termes spécifiques sont précisés dans la documentation associée au jeu de données.
<b>8.5. Accroître la réutilisation des données</b>		
8.5.1.	A la fin du projet, ce jeu de données pourra-t-il être réutilisé par des parties tierces ? Si la réutilisation est restreinte, expliquez pourquoi.	Les données peuvent être réutilisées en interne (par d'autres unités de l'Institut Pasteur) et par les futurs chercheurs qui travailleront sur les futures étapes du développement du projet. Les données seront également disponibles gratuitement sous une licence CC-BY sur le dépôt ZENODO. Elles pourront être réutilisées par tous, et pas seulement en interne.
8.5.2.	Quelle licence sera attribuée à ce jeu de données afin de permettre la réutilisation la plus large possible ?	Licence CC-BY ( <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a> )
8.5.3.	Quand le jeu de données sera-t-il accessible pour la réutilisation ? Le cas échéant, précisez pour quelle raison et pendant combien de temps un embargo est nécessaire.	Les données seront disponibles au moment de la prépublication de l'article associé.
8.5.4.	Précisez la durée pendant laquelle le jeu de données sera réutilisable	Les données seront stockées pendant la durée maximale permise par la capacité de stockage de l'Institut Pasteur et de l'entrepôt choisi pour publier les données. Les données seront également stockées sur des disques durs externes dans le laboratoire.
<b>8.6. Sécurité des données</b>		
8.6.1.	Ce jeu de données doit-il rester confidentiel au cours de votre projet ? Si oui, pouvez-vous préciser auprès de qui il peut être diffusé ?	Les données brutes ne doivent rester accessibles qu'aux chercheurs de l'Institut Pasteur. Les résultats intermédiaires ne seront accessibles qu'aux chercheurs impliqués dans le projet.

8.6.2.	Pendant le projet (avant le dépôt des données dans un entrepôt), ce jeu de données est-il stocké de façon sécurisée ?	Pendant le projet, les données sont stockées sur les serveurs de stockage sécurisés de l'Institut Pasteur. De plus, les données sont également stockées sur un disque dur externe et NAS
8.6.3.	L'entrepôt de données choisi pour conserver ce jeu de données après le projet met-il en œuvre une politique de sécurité pour son système d'information ?	Nous nous sommes assurés que nos données sont bien sécurisées dans l'entrepôt choisi.
8.6.4.	Quelles sont les mesures de sécurité mises en œuvre pour la collecte et l'échange de données ?	Notre projet ne comprend pas de collecte ou d'échange de données. Aucune mesure de sécurité n'est nécessaire.