



**UNIVERSITÉ  
DE GENÈVE**

**GENERESO**

Université de Genève

Certificat complémentaire en géomatique

## **Dossier de stage**

# **Cartographie du riz en Inde et harmonisation de bases minières chez Genereso Sàrl**

Stage effectué chez Genereso Sàrl (juillet – octobre 2025)

Étudiant : Loris Bach

Superviseur : Dr Diego Villagómez

Enseignant responsable : Dr Gregory Giuliani

Membre du comité scientifique : Pr Hy Dao

Genève, novembre 2025

## Table des matières

1.	Liste des abréviations.....	3
2.	Résumé.....	4
3.	Introduction.....	4
4.	Présentation de l'organisation hôte.....	5
5.	Objectifs et contexte technique du stage.....	6
6.	Partie I – Traitement et export des mosaïques Sentinel-2.....	7
6.1.	Objectif .....	8
6.2.	Méthodologie .....	8
6.3.	Résultats.....	12
6.4.	Discussion critique.....	14
7.	Partie II – Harmonisation et structuration des bases minières.....	16
7.1.	Objectif .....	16
7.2.	Méthodologie .....	19
7.3.	Résultats.....	21
7.4.	Discussion critique.....	21
8.	Autres activités et contributions.....	23
8.1.	Application de veille et de surveillance des chaînes d'approvisionnement.....	23
8.2.	Formation et perfectionnement ArcGIS.....	24
8.3.	Réunions et échanges externes.....	24
9.	Réflexion personnelle .....	26
10.	Conclusion et perspectives.....	27
11.	Annexes.....	28
12.	Bibliographie et références de données .....	28
12.1.	Références de données utilisées.....	28
12.2.	Références institutionnelles et techniques .....	28

## **1. Liste des abréviations**

AI – Artificial Intelligence (intelligence artificielle)

API – Application Programming Interface (interface de programmation)

ESG – Environmental, Social and Governance (critères environnementaux, sociaux et de gouvernance)

FONGIT – Fondation Genevoise pour l’Innovation Technologique

GeoTIFF – GeoTagged Image File Format (format raster géoréférencé basé sur TIFF)

GPS – Global Positioning System (système de positionnement par satellites)

ICMM – International Council on Mining and Metals

ISO – International Organization for Standardization (organisation internationale de normalisation)

ISO3 – Code pays à trois lettres défini par la norme ISO 3166-1 alpha-3 (ex. CHE, FRA)

LLM – Large Language Model (modèle de langage)

MNDWI – Modified Normalized Difference Water Index (indice d’eau “modifié”)

NDMI – Normalized Difference Moisture Index (indice d’humidité)

NDRE – Normalized Difference Red Edge Index (indice “red edge”, bord rouge)

NDVI – Normalized Difference Vegetation Index (indice de végétation)

RMI – Responsible Minerals Initiative

SaaS – Software as a Service (logiciel fourni sous forme de service)

Sàrl – Société à responsabilité limitée

SDK – Software Development Kit (kit de développement logiciel)

SIG – Système d’information géographique

WAPHA – Water and Planetary Health Analytics (organisation active dans l’analyse et la visualisation de données pour l’environnement et la durabilité)

WGS84 – World Geodetic System 1984 (système géodésique de référence)

## **2. Résumé**

Réalisé entre juillet et octobre 2025 au sein de Genereso Sàrl à Genève, ce stage du Certificat complémentaire en géomatique de l'Université de Genève avait pour objectif de mettre en pratique des compétences en géomatique, en télédétection et en gestion de données dans un contexte professionnel.

Le travail s'est organisé autour de deux axes principaux. Le premier a porté sur la préparation et l'analyse d'images Sentinel-2 dans le cadre d'un mandat pour un client privé, visant à produire des jeux de données exploitables pour l'étude des surfaces rizicoles dans deux États agricoles du nord de l'Inde, le Punjab et l'Haryana. Le second axe a concerné l'harmonisation de bases de données minières internationales afin de définir un modèle commun, faciliter les correspondances entre sources et permettre leur intégration dans la plateforme géospatiale de Genereso.

Ces travaux ont conduit à la mise en place de procédures reproductibles pour le traitement d'images satellitaires et à la structuration d'un schéma de données minier unifié et documenté. Le stage a également permis de mieux comprendre les contraintes d'un environnement de startup, où l'autonomie et la rapidité d'itération doivent être conciliées avec la rigueur méthodologique et la qualité des livrables.

## **3. Introduction**

Le Certificat complémentaire en géomatique de l'Université de Genève vise à relier la formation académique à la pratique professionnelle dans le domaine des systèmes d'information géographique et de la télédétection. Le stage effectué chez Genereso Sàrl s'inscrit dans cette perspective et avait pour finalité d'appliquer les savoirs acquis à des problématiques concrètes de traitement, d'intégration et de valorisation de données géospatiales. Cette mise en situation professionnelle a permis de travailler sur des projets appliqués où les contraintes opérationnelles, la reproductibilité des méthodes et la qualité des livrables jouent un rôle central.

Genereso Sàrl est une startup genevoise active dans le domaine de l'intelligence géospatiale et du suivi des chaînes d'approvisionnement en métaux. L'entreprise développe des solutions combinant géomatique, télédétection et analyse de données, dans un environnement de recherche et développement où la polyvalence et l'adaptation rapide aux besoins des projets sont essentielles.

Les travaux menés se sont articulés autour de deux axes principaux. Le premier a porté sur la préparation et l'analyse d'images Sentinel-2 dans le cadre d'un mandat pour un client privé, avec pour objectif de produire des mosaïques exploitables et des premières sorties cartographiques relatives aux surfaces rizicoles dans deux États agricoles du nord de l'Inde, le Punjab et l'Haryana. Le second a consisté à harmoniser plusieurs bases de données minières internationales selon un modèle commun, afin de faciliter les correspondances entre sources et d'assurer leur intégration dans la plateforme géospatiale de Genereso. Ces deux volets s'inscrivent dans une logique de standardisation et de mise à disposition de données structurées pour l'analyse et la visualisation.

En parallèle de ces activités principales, le stage a permis de contribuer à d'autres tâches ponctuelles de l'entreprise, notamment la rédaction de documentation technique, le perfectionnement sur les outils ArcGIS et la participation à des échanges avec des acteurs externes. L'ensemble de ces expériences a renforcé des compétences techniques et organisationnelles, tout en offrant une compréhension plus concrète du travail en startup, entre autonomie, itérations rapides et exigences de qualité.

#### **4. Présentation de l'organisation hôte**

Genereso Sàrl est une entreprise genevoise spécialisée dans l'intelligence géospatiale appliquée au secteur des métaux et des matières premières critiques. Elle développe une plateforme SaaS (Software as a Service) intégrée qui centralise des données géographiques, minières et environnementales afin de renforcer la transparence et la durabilité des chaînes d'approvisionnement. Ses applications web combinent imagerie satellite, données cadastrales et indicateurs sectoriels à différentes échelles régionales, offrant aux utilisateurs des outils d'analyse et de suivi en temps réel des activités minières à travers le monde. Genereso s'appuie sur les infrastructures cloud de Google, Amazon et Esri avec qui la startup a des partenariats pour déployer ses solutions interactives, tout en promouvant une exploitation responsable et documentée des ressources naturelles.

L'entreprise, fondée et dirigée par Diego Villagómez, est issue de l'écosystème de l'innovation genevoise et bénéficie du soutien de la FONGIT (Fondation genevoise pour l'innovation technologique). Elle collabore avec plusieurs institutions publiques et privées, ainsi qu'avec des acteurs du domaine minier et de la géomatique, afin de développer des outils de suivi et d'aide à

la décision fondée sur la donnée géospatiale. Son positionnement à l'intersection entre la recherche appliquée, la durabilité et la technologie en fait un acteur émergent du domaine de la traçabilité des chaînes d'approvisionnement.

Genereso fonctionne selon une organisation souple, propre aux startups technologiques, où les membres de l'équipe participent conjointement à la conception des outils, à la gestion des bases de données et à la communication scientifique. Les travaux sont menés dans un environnement collaboratif mêlant compétences en télédétection, en intelligence artificielle et en systèmes d'information géographique. Cette structure permet de valoriser la polyvalence, la prise d'initiative et la réactivité, qualités particulièrement sollicitées au cours du stage.

Le stage s'est déroulé sous la supervision directe de Diego Villagómez, cofondateur et directeur de Genereso, et sous le suivi académique du Dr Gregory Giuliani, enseignant-chercheur à l'Université de Genève. Le travail s'est réparti entre le bureau de la FONGIT à Plan-les-Ouates et des sessions de télétravail. Cet encadrement mixte, à la fois professionnel et académique, a permis de maintenir un équilibre entre les exigences de l'entreprise et les objectifs pédagogiques du Certificat complémentaire en géomatique.

## **5. Objectifs et contexte technique du stage**

Le stage s'est déroulé dans un contexte de développement et de prototypage au sein de Genereso Sàrl, où des données géospatiales issues de sources hétérogènes sont intégrées et exploitées dans un environnement majoritairement cloud. Les projets menés s'inscrivaient dans la logique de l'entreprise consistant à structurer des données et des traitements reproductibles afin de produire des sorties directement exploitables pour l'analyse et la cartographie.

L'objectif principal du stage était d'appliquer des compétences en géomatique, en télédétection et en gestion de données à la conception et à la mise en œuvre de traitements opérationnels. Le travail s'est articulé autour de deux volets complémentaires. Le premier concernait la préparation et l'analyse d'images Sentinel-2 dans le cadre d'un mandat portant sur les surfaces rizicoles au Punjab et à l'Haryana, avec pour finalité de produire des mosaïques exploitables et des premières sorties cartographiques. Le second volet portait sur l'harmonisation de plusieurs bases de données minières internationales (ICMM, RMI, Jasansky, Hudson-Edwards) selon un modèle de données

commun, afin de faciliter les correspondances entre sources et leur intégration dans la base centrale de Genereso.

Ces volets répondaient à des besoins distincts. D'une part, disposer d'un flux reproductible permettant de générer des produits Sentinel-2 et des indicateurs spatialisés sur de grandes zones d'étude. D'autre part, construire une base unifiée et documentée permettant d'analyser les liens entre sites miniers, matières premières et acteurs économiques à partir de sources multiples.

Sur le plan technique, les traitements ont été réalisés à l'aide d'un environnement Jupyter hébergé sur Vertex AI, en s'appuyant sur Google Earth Engine pour l'accès et le traitement des images satellitaires, et sur Google Cloud Storage pour l'export et l'échange des produits. Python a été utilisé pour structurer et automatiser les étapes de traitement et de préparation des données, tandis qu'ArcGIS Pro a servi à produire certaines sorties cartographiques et à mettre en œuvre des traitements complémentaires côté SIG. L'ensemble du travail a été documenté de manière à permettre la reprise des traitements et la vérification des paramètres utilisés.

## **6. Partie I – Traitement et export des mosaïques Sentinel-2**

Le premier volet du stage a porté sur la mise en place d'une chaîne de traitement et d'analyse d'images Sentinel-2 appliquée à deux États agricoles du nord de l'Inde, le Punjab et l'Haryana. Ce travail s'inscrivait dans un mandat réalisé pour un client souhaitant disposer d'indicateurs sur l'étendue des surfaces rizicoles dans ces régions, où la culture du riz constitue un secteur majeur de l'économie agricole. L'objectif était d'obtenir des mosaïques homogènes malgré la forte couverture nuageuse liée à la mousson, puis d'exploiter ces images pour établir une première caractérisation du couvert rizicole à l'aide d'indices spectraux.

Dans ce cadre, le traitement des scènes Sentinel-2 a été conçu de manière à fournir des mosaïques directement utilisables pour l'analyse, tout en assurant une reproductibilité suffisante pour être appliqué à de larges régions d'étude. Ces mosaïques ont ensuite servi de base au calcul d'indices permettant de distinguer les zones rizicoles à différentes étapes du cycle cultural, et d'élaborer une classification simple riz / non-riz destinée à fournir une première estimation des surfaces concernées.

## **6.1. Objectif**

Le travail mené dans ce premier volet avait pour objectif de préparer et d'analyser des mosaïques Sentinel-2 couvrant les États du Punjab et de l'Haryana, dans le cadre d'un mandat destiné à estimer l'étendue des surfaces rizicoles. La démarche visait d'abord à produire des images homogènes malgré les contraintes atmosphériques liées à la mousson. Les périodes d'intérêt ont été définies juste avant et juste après la mousson, lorsque des images partiellement dégagées restent disponibles, tout en correspondant à des phases clés du cycle rizicole (semis et transplantation, puis récolte). L'enjeu consistait à combiner plusieurs scènes afin d'obtenir un rendu continu et représentatif des conditions au sol, en limitant l'influence des nuages et de leurs ombres.

Le traitement devait être suffisamment flexible pour s'adapter à différentes zones d'étude et permettre la génération rapide de mosaïques correspondant à des périodes précises du cycle agricole. Ces mosaïques avaient pour vocation de servir de base au calcul d'indices spectraux utiles à la caractérisation du couvert rizicole, puis à la production d'une première sortie cartographique riz/non-riz fondée sur des règles simples dérivées de l'analyse des pixels situés à l'intérieur des parcelles de riz de référence fournies par le client (relevées sur place au GPS). L'objectif final était de disposer d'un flux reproductible et de produits prêts à l'emploi pour une première estimation spatialisée, tout en préparant d'éventuelles analyses plus détaillées si des données de référence complémentaires deviennent disponibles.

## **6.2. Méthodologie**

Le traitement des images Sentinel-2 a été réalisé dans un environnement Jupyter hébergé sur Vertex AI, en s'appuyant sur Google Earth Engine pour l'accès aux données et les opérations de filtrage, de masquage et de mosaïquage. Cette configuration permet de travailler directement avec de grands volumes d'images satellitaires et de mettre en place un flux de traitement automatisé et reproductible. Les traitements ont porté sur deux États du nord de l'Inde, le Punjab et l'Haryana, qui constituent des régions agricoles majeures pour la culture du riz.

Les images utilisées sont des produits Sentinel-2 de niveau L2A, déjà corrigés des effets atmosphériques. La sélection s'est concentrée sur deux périodes clés du cycle rizicole, correspondant à la phase de semis et transplantation en début d'été, et à la phase de récolte en début d'automne. Ces fenêtres temporelles ont été retenues car elles correspondent à des moments



où la signature spectrale du riz diffère fortement de celle des autres cultures présentes dans la région. Elles se situent également juste avant et juste après la mousson, période durant laquelle la couverture nuageuse est généralement trop importante pour permettre une exploitation fiable des images optiques.

Pour chaque période d'intérêt, une fenêtre temporelle élargie a été définie afin d'inclure des scènes acquises plusieurs jours avant et après la date centrale. Cette approche permet d'augmenter le nombre d'images candidates et de compenser la présence résiduelle de nuages en combinant plusieurs scènes partielles. Les images acquises pendant la mousson proprement dite ont été largement exclues, la couverture nuageuse y étant quasi permanente sur l'ensemble de la zone d'étude.

La qualité des scènes sélectionnées a été contrôlée à l'aide de la couche de classification de scène, utilisée pour masquer les pixels correspondant aux nuages, aux cirrus et aux ombres. Les scènes présentant une proportion de nuages trop élevée ont été écartées. Les images retenues ont ensuite été combinées par composition médiane afin de produire, pour chaque région et chaque période, une mosaïque représentative des conditions au sol. Ce choix méthodologique permet de limiter l'influence des valeurs extrêmes et d'obtenir un rendu radiométriquement homogène sur l'ensemble de la zone d'étude. Les bandes disponibles à 10 m de résolution ont été conservées telles quelles, tandis que les bandes à 20 m ont été rééchantillonnées à 10 m afin de produire des mosaïques multispectrales cohérentes.

Les mosaïques finales ont été exportées au format GeoTIFF vers Google Cloud Storage, puis intégrées dans les environnements Python et ArcGIS Pro utilisés par Genereso pour les étapes d'analyse ultérieures. À partir de ces mosaïques, plusieurs indices spectraux ont été calculés afin de caractériser le couvert végétal et les conditions d'humidité des parcelles.

Quatre indices ont été retenus pour l'analyse. Le NDVI (Normalized Difference Vegetation Index) renseigne sur la vigueur du couvert végétal et augmente fortement au cours de la croissance du riz. Le NDMI (Normalized Difference Moisture Index) est sensible à l'humidité du couvert et permet d'identifier les parcelles saturées en eau, en particulier lors de la transplantation. Le MNDWI (Modified Normalized Difference Water Index) a été calculé à l'aide des bandes B3 et B8A, plutôt qu'avec la formulation classique utilisant le proche infrarouge à ondes courtes, cette variante offrant un contraste plus marqué pour les rizières en début de cycle, lorsque l'eau libre et la

végétation émergente coexistent. Enfin, le NDRE (Normalized Difference Red Edge) exploite une bande red-edge, située à la transition entre le rouge et le proche infrarouge, et permet de suivre la densification progressive du couvert végétal grâce à sa sensibilité à l'activité chlorophyllienne.

Les indices spectraux ont été calculés à partir des mosaïques, puis les valeurs ont été extraites pour l'ensemble des pixels inclus dans les polygones de parcelles de riz disponibles pour le Punjab et l'Haryana. Ces polygones, fournis par le client, correspondent à des parcelles relevées sur le terrain au GPS. L'analyse ne porte donc pas sur des statistiques calculées parcelle par parcelle, mais sur une distribution globale construite à partir de tous les pixels appartenant aux zones de riz de référence. Les distributions ont été analysées au moyen de boîtes à moustaches afin de caractériser des plages de valeurs typiques du riz aux périodes de semis et de récolte, et d'observer les différences entre ces deux phases du cycle cultural. La figure 1 met en évidence un décalage des distributions entre les deux fenêtres temporelles pour plusieurs indices, ainsi que la présence de nombreuses valeurs aberrantes et d'un recouvrement partiel entre les distributions. Ces distributions ont ensuite servi de base à la définition de seuils simples utilisés pour la suite de l'analyse.

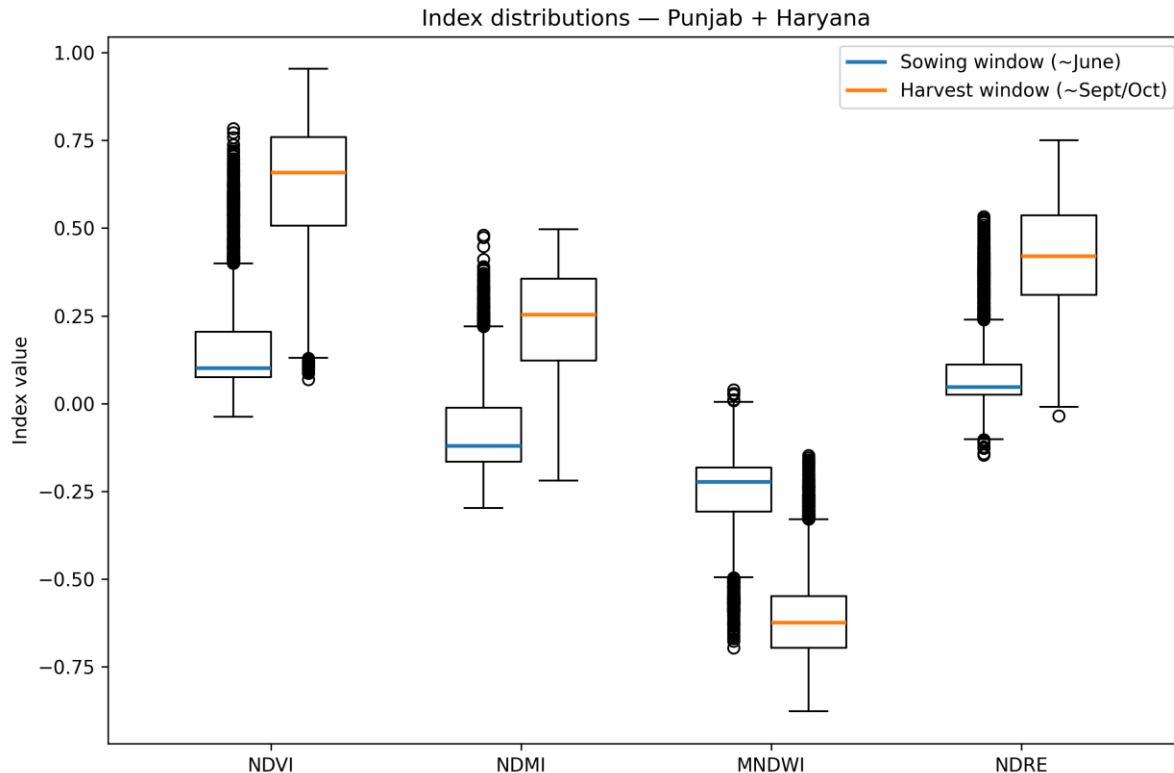


Figure 1. Boîtes à moustaches des indices NDVI, NDMI, MNDWI et NDRE pour le riz au Punjab et à l'Haryana, en juin et en septembre–octobre.

La classification riz/non-riz n'a pas été réalisée directement dans le notebook, mais dans ArcGIS Pro. Les seuils dérivés de l'analyse des distributions d'indices ont été intégrés dans un modèle construit à l'aide de ModelBuilder, complété par des scripts ArcPy pour automatiser certaines opérations. Comme l'illustre la figure 2, le flux de traitement calcule les indices (NDVI, NDMI, NDRE et NDWI) séparément pour les mosaïques de juin et de septembre, applique pour chaque période un jeu de bornes dédié, puis combine les deux classifications intermédiaires afin de produire une sortie binaire finale. Ce schéma de traitement explicite permet de relier l'analyse exploratoire conduite dans l'environnement Python à une mise en œuvre opérationnelle dans ArcGIS Pro, aboutissant à des cartes riz/non-riz exploitables et reproductibles.

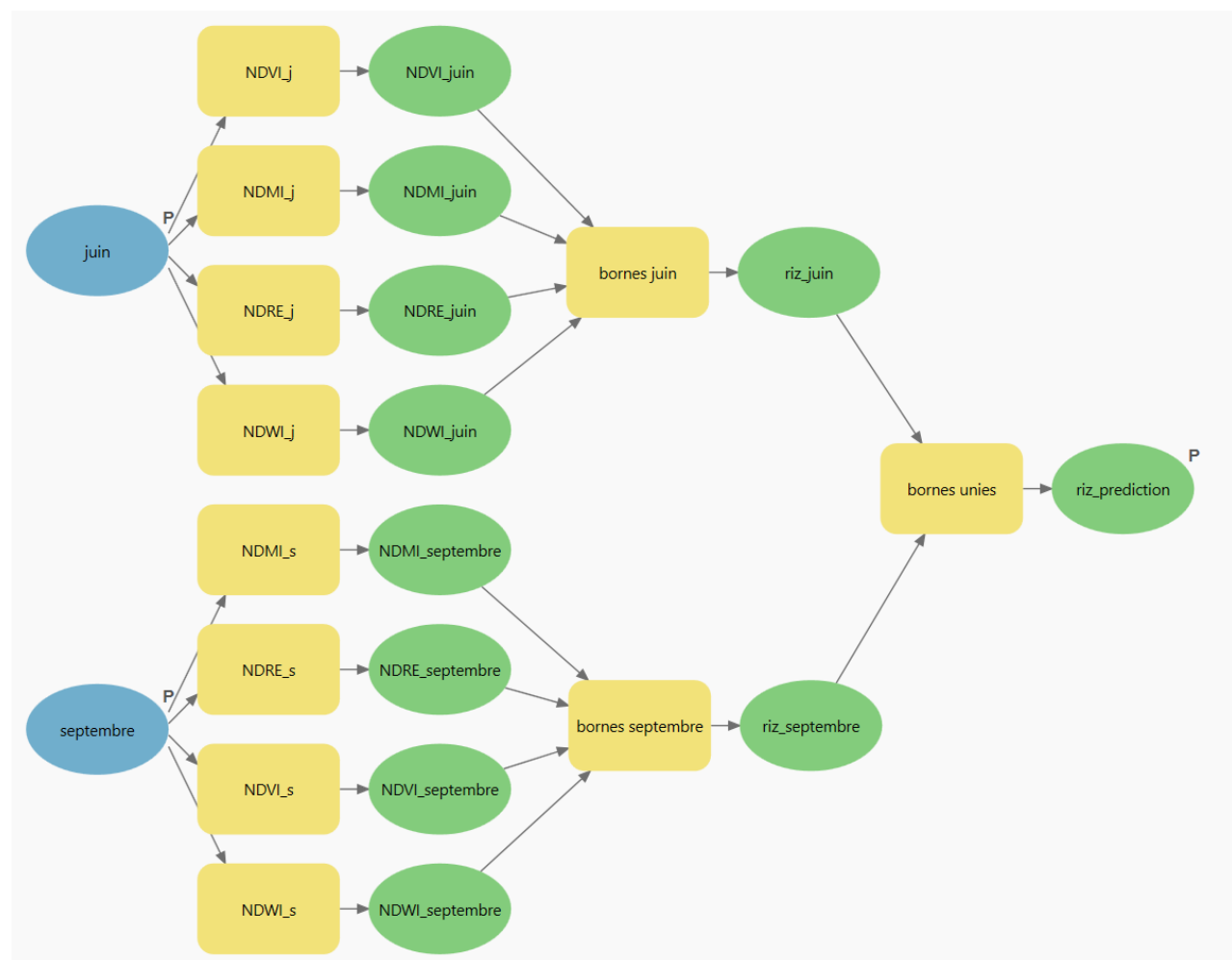


Figure 2. Flux ModelBuilder (ArcGIS Pro) appliquant les seuils d'indices pour produire la classification riz/non-riz

### 6.3. Résultats

La chaîne de traitement mise en place a permis de produire des mosaïques Sentinel-2 homogènes pour les deux régions d'étude, le Punjab et l'Haryana, à deux périodes correspondant à des phases clés du cycle rizicole, en début et en fin de saison. Les mosaïques obtenues présentent une continuité spatiale satisfaisante sur l'ensemble des zones couvertes et constituent une base exploitable pour le calcul d'indices spectraux et la production de sorties cartographiques.

Les produits générés sont des mosaïques multi-bandes à 10 m de résolution, couvrant environ 100 × 50 kilomètres par région. Les fichiers ont été exportés au format GeoTIFF et intégrés dans ArcGIS Pro pour les étapes suivantes du traitement. Cette organisation permet une exploitation directe des données dans un environnement SIG, tant pour la visualisation que pour l'application de règles de classification.

À partir des mosaïques, les indices spectraux NDVI, NDMI, MNDWI et NDRE ont été calculés pour les deux périodes considérées. Ces indices mettent en évidence des contrastes spatiaux entre zones cultivées et autres types de couverture, ainsi qu'entre le début et la fin du cycle cultural. À titre illustratif, la figure 3 présente la carte du NDVI sur la fenêtre de récolte au Punjab, calculée à partir des mosaïques Sentinel-2, et permet de visualiser la distribution spatiale des valeurs sur la zone d'étude.

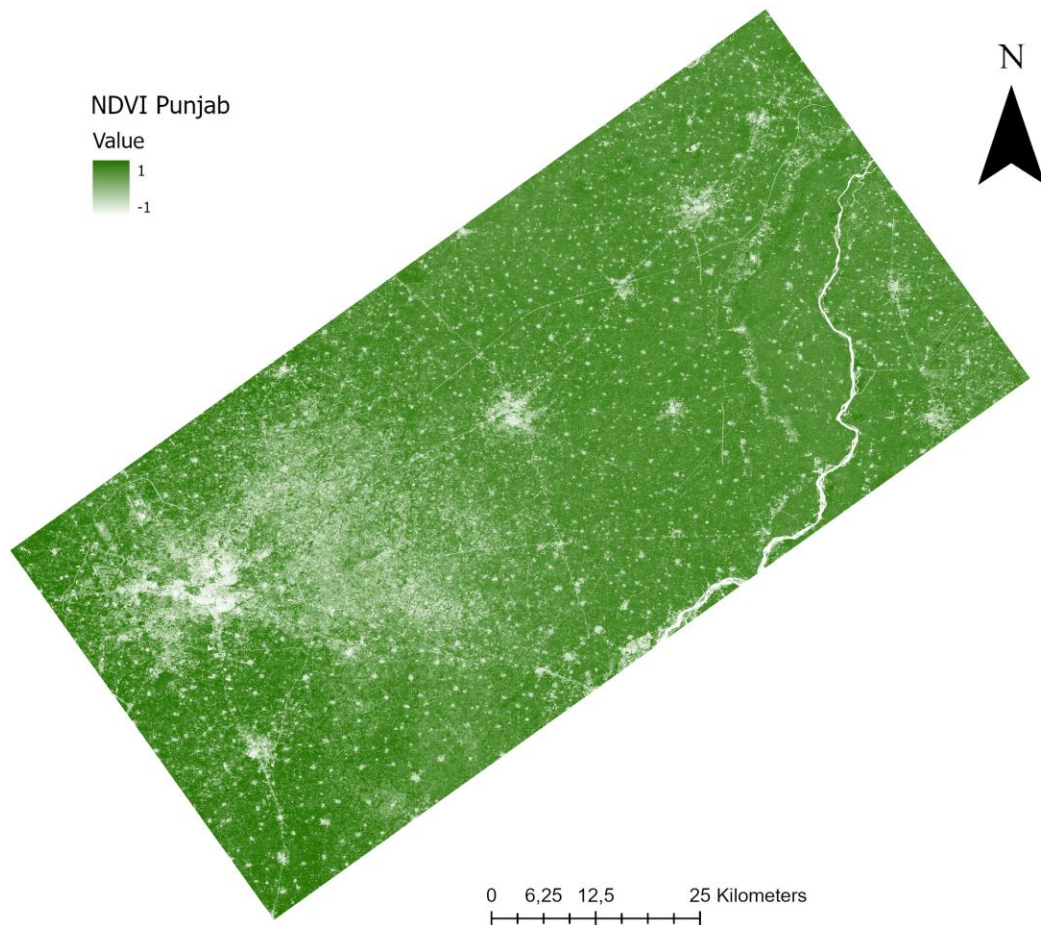


Figure 3. Carte de l'indice NDVI du Punjab pour la période de récolte.

Sur la base des seuils définis à partir de l'analyse des distributions d'indices, une classification binaire riz/non-riz pixel par pixel a été produite dans ArcGIS Pro à l'aide d'un modèle ModelBuilder. La figure 4 présente le résultat de cette étape pour le Punjab. La carte fournit une estimation spatialisée des pixels classés comme rizières à une résolution de 10 m sur la zone couverte par les mosaïques, et peut servir d'entrée à des analyses ou agrégations ultérieures (par exemple par unité administrative ou par zone d'intérêt).

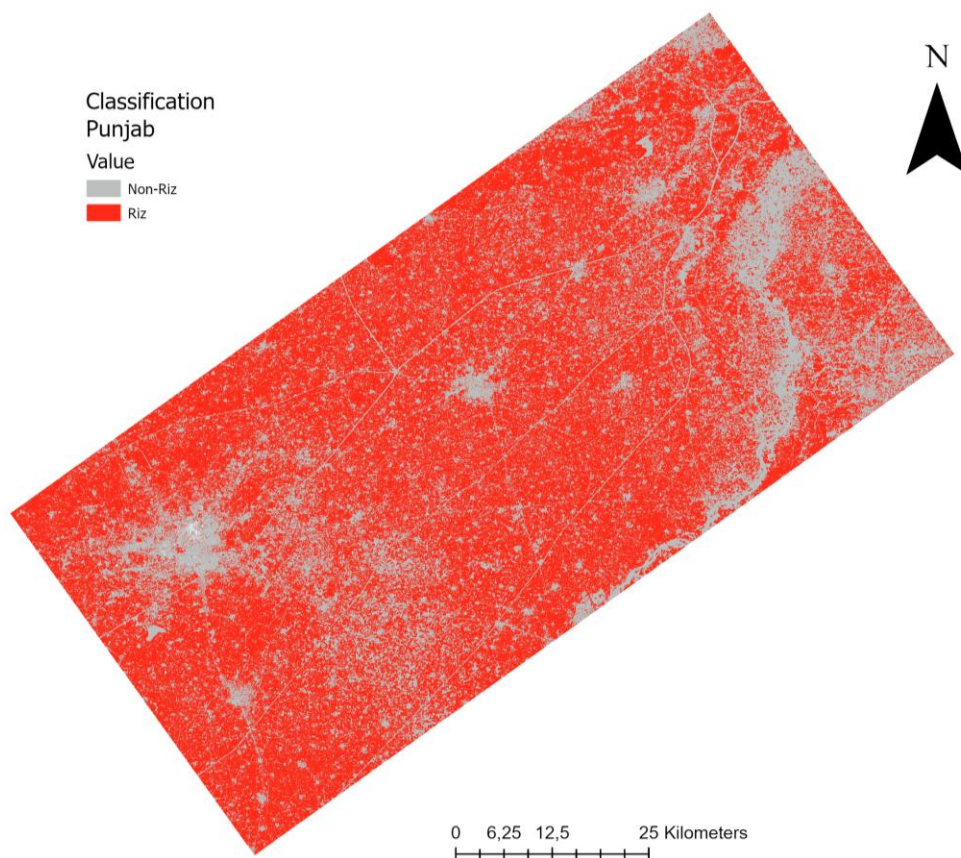


Figure 4. Prédications riz/non-riz pour le Punjab.

#### 6.4. Discussion critique

La chaîne de traitement développée a rempli son objectif principal en permettant la production de mosaïques Sentinel-2 exploitables sur le Punjab et l'Haryana à deux périodes clés du cycle rizicole. L'automatisation du flux sur Google Earth Engine a apporté un gain opérationnel important, en standardisant les étapes de sélection, de masquage, de composition et d'export des images, tout en assurant une bonne reproductibilité des résultats. Cette organisation facilite l'intégration des produits dans les environnements SIG et cloud utilisés par Genereso.

Certaines limites subsistent néanmoins, en particulier en lien avec les conditions atmosphériques de la région étudiée. Même en travaillant en marge de la mousson, la disponibilité d'images réellement exploitables reste variable et des résidus de nuages fins ou d'ombres peuvent persister localement dans les mosaïques. Par ailleurs, le recours à une composition temporelle médiane,

bien adapté pour obtenir des images homogènes, tend à lisser certaines variations spectrales locales et peut réduire la sensibilité à des contrastes plus ponctuels.

La principale contrainte méthodologique du travail concerne toutefois la disponibilité des données de référence. Les données terrain fournies ne comprenaient que des polygones de rizières, sans information équivalente pour les zones non-riz. En l'absence d'échantillons négatifs, il n'était pas possible de mettre en œuvre une classification supervisée classique ou d'entraîner un modèle de machine learning de manière rigoureuse. Dans ce contexte, l'approche fondée sur l'analyse des distributions d'indices spectraux et sur la définition de seuils simples à partir des boîtes à moustaches constituait une solution pragmatique et cohérente avec les données disponibles. Elle permet d'exploiter l'information contenue dans les parcelles de riz connues sans introduire artificiellement des exemples non-riz. En revanche, l'absence de données de référence pour la classe non-riz empêche d'évaluer quantitativement la classification, en particulier d'estimer la proportion de faux positifs. Les résultats produits doivent donc être interprétés comme une première estimation spatialisée, utile pour l'exploration et la production d'un aperçu régional, mais insuffisante pour une validation robuste sans données complémentaires.

Une amélioration majeure des résultats serait envisageable si des données complémentaires devenaient disponibles. La présence d'une carte des parcelles agricoles couvrant la région d'étude, même sans attributs thématiques, permettrait d'agréger les indices à l'échelle des champs et d'obtenir des sorties plus cohérentes spatialement. Surtout, la disponibilité de polygones décrivant explicitement des zones non-riz (par exemple des champs de blé ou de canne à sucre) ouvrirait la voie à une évaluation quantitative (matrice de confusion, estimation des faux positifs et faux négatifs) et à des approches supervisées plus avancées.

Enfin, des améliorations futures pourraient porter sur le renforcement du contrôle qualité des mosaïques et sur la validation systématique des sorties avant diffusion. À plus long terme, l'intégration de données complémentaires, en particulier radar, offrirait une alternative pertinente pour atténuer les limitations liées à la couverture nuageuse persistante pendant la mousson et permettre un suivi plus continu des surfaces cultivées.

## **7. Partie II – Harmonisation et structuration des bases minières**

Le second volet du stage a porté sur la conception d'une base de données minière harmonisée rassemblant des sources internationales hétérogènes. Ce travail s'inscrit dans la volonté de Genereso de disposer d'un socle de données géospatiales unifié, cohérent et documenté, destiné à l'analyse et à la visualisation des activités extractives à l'échelle mondiale.

L'harmonisation des données constitue un préalable essentiel à toute exploitation analytique ou cartographique. Elle vise à rendre comparables des jeux de données issus de contextes différents, à corriger les incohérences de structure et de terminologie, et à assurer la traçabilité des transformations opérées. Dans le cadre du stage, l'objectif était de développer une approche reproductible et transparente pour intégrer ces informations dans le modèle de données interne de Genereso.

Le travail a également préparé la mise en place d'un processus de dédoubleage des enregistrements, afin d'éviter la redondance entre projets décrits par plusieurs sources. Ce second volet complète le précédent en s'attachant non plus à la production d'images, mais à la structuration des données descriptives qui constituent la base de l'analyse géospatiale au sein de l'entreprise.

### **7.1. Objectif**

L'objectif de cette seconde partie était de concevoir une base de données minière harmonisée à partir de plusieurs sources internationales traitant des sites d'extraction et de transformation des matières premières. Ce travail visait à établir un modèle commun permettant de rassembler des informations dispersées dans des formats et des vocabulaires hétérogènes, afin de créer un ensemble cohérent et directement exploitable dans les environnements géospatiaux de Genereso. La figure 5 illustre l'un des schémas sources utilisés (Jasansky et al., 2023), structuré en tables thématiques reliées à une table centrale, et met en évidence les champs clés permettant de lier les informations entre modules tout en conservant la traçabilité des attributs.

Le projet reposait sur quatre ensembles de données principaux : ceux de l'International Council on Mining and Metals (ICMM), de la Responsible Minerals Initiative (RMI), de la compilation scientifique de Jasansky et collaborateurs, et de la base de données de Hudson-Edwards et collaborateurs. Ces sources couvrent une grande diversité de thématiques : localisation des sites, opérateurs et propriétaires, métaux extraits, volumes de production et contexte environnemental.



L'harmonisation devait permettre de relier ces informations selon une structure interne commune tout en conservant la traçabilité des champs d'origine.

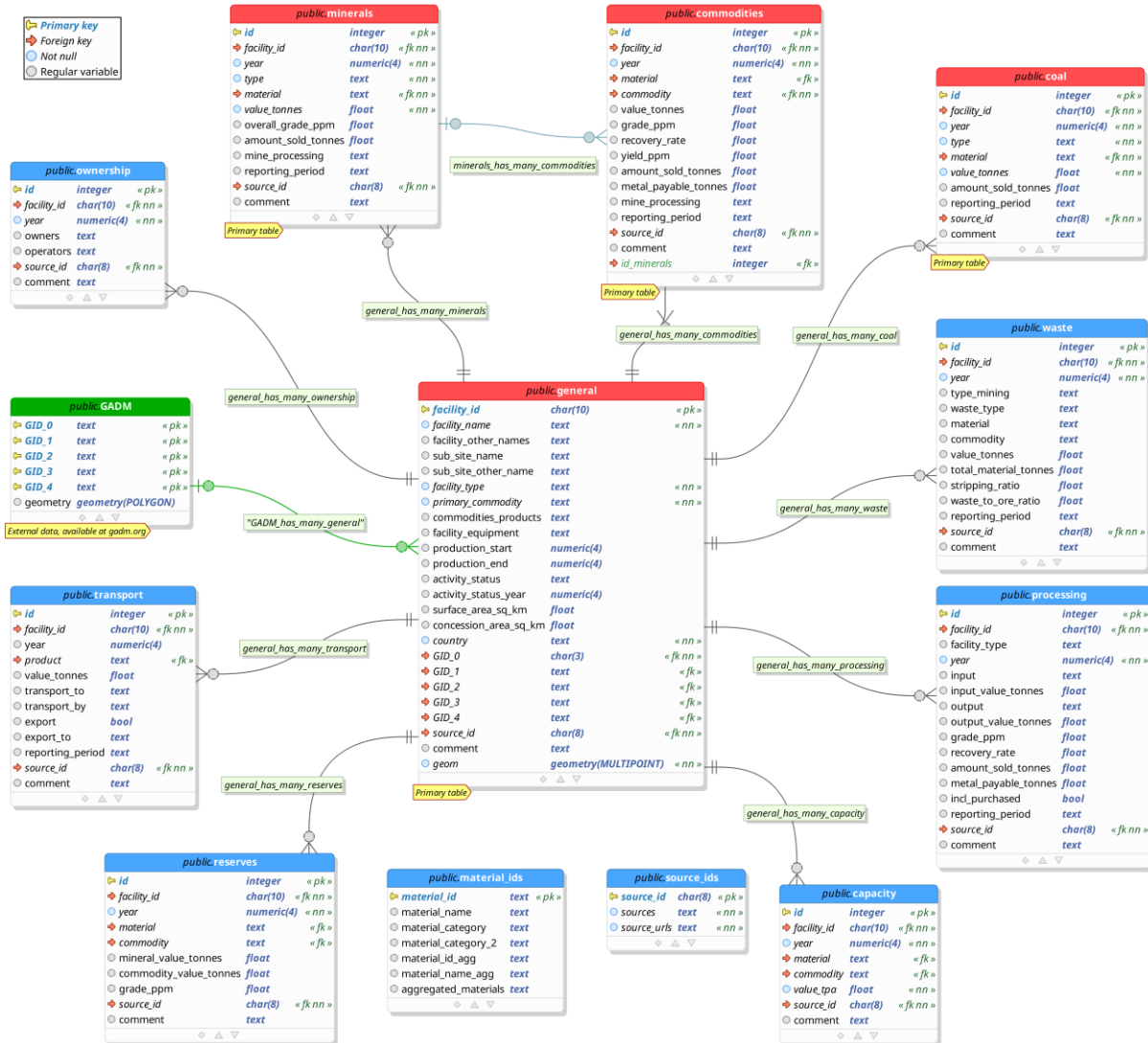


Figure 5. Structure de la base de données Jasansky. Source : Jasansky et al., 2023.

Les objectifs spécifiques du travail étaient d'analyser la structure et le contenu de chaque source, puis de définir un schéma cible partagé fondé sur six tables thématiques (Key Facts, Assets, Ownership, Operations, Reserves/Resources et Production). La figure 6 présente ce modèle de données, organisé autour d'une clé commune (Project\_ID) qui permet de relier les tables. Le travail a ensuite consisté à normaliser les attributs géographiques et textuels, à générer des identifiants

uniques stables pour chaque projet afin d'assurer la cohérence des jointures, et à préparer la détection puis la fusion de doublons. Le processus d'harmonisation devait aboutir à une base directement exploitable pour des analyses spatiales et statistiques, ainsi qu'à un ensemble de règles de transformation et de contrôle qualité réutilisable lors des mises à jour. L'ensemble pose les bases techniques d'une intégration progressive de sources hétérogènes au sein de l'écosystème Genereso.

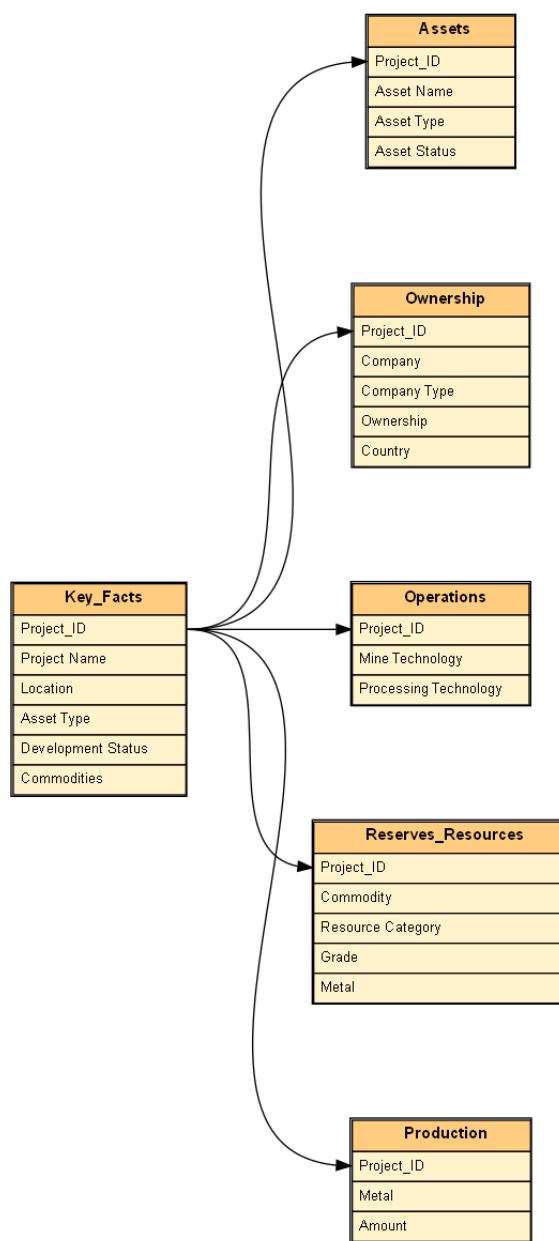


Figure 6. Structure de la base de données Genereso

## 7.2. Méthodologie

Les objectifs présentés ci-dessus ont été traduits en un flux de traitement modulaire, entièrement développé dans un environnement Python exécuté sur Vertex AI. Cette configuration a permis de centraliser les traitements, d'automatiser les vérifications et de garantir la reproductibilité à l'aide de notebooks commentés. Le processus s'est articulé en quatre grandes étapes : analyse des structures d'origine, nettoyage et normalisation, mise en correspondance avec le modèle interne de Genereso, puis export des données harmonisées.

Chaque jeu de données a été importé dans un environnement commun et analysé de façon systématique : vérification du nombre d'enregistrements, inspection des champs, contrôle des coordonnées et détection des valeurs manquantes. Les structures divergentes ont été documentées afin de faciliter la mise en correspondance avec le schéma cible défini par Genereso.

La mise en correspondance a été formalisée sous forme de règles explicites reliant les tables et champs des schémas sources aux six tables cibles (Key Facts, Assets, Ownership, Operations, Reserves\_Resources, Production). La figure 7 illustre cet alignement pour le cas de Jasansky, en montrant comment les informations issues des tables source (par ex. *public.general*, *public.minerals*, *public.reserves*) alimentent les tables cibles de Genereso, ainsi que les dépendances entre modules.

Les valeurs textuelles ont ensuite été normalisées (typographie, noms d'entreprises, produits, unités de mesure) et les coordonnées converties au système de référence WGS84 afin d'assurer la cohérence géographique. Pour uniformiser les champs de localisation, un modèle LLM (Gemini) a été employé afin de standardiser l'information relative au pays. Lorsque la source indiquait un nom de pays non normalisé (orthographe variables, abréviations, dénominations locales), le modèle proposait l'équivalent ISO3. En l'absence d'indication textuelle, il déterminait le pays à partir des coordonnées géographiques.

Les suggestions ont été contrôlées à l'aide de règles automatisées (cohérence géographique, concordance avec l'ISO3 attendu, correspondance entre nom et région), complétées par des vérifications ciblées sur les cas ambigus. Cette approche a permis d'assurer la qualité des métadonnées tout en restant compatible avec un volume de données supérieur à cent mille enregistrements.

Une étape de préfiltrage a été amorcée pour repérer les doublons probables en croisant plusieurs critères tels que la proximité géographique, la similitude de nom et la correspondance du produit principal. Ce module pourra être ultérieurement renforcé en intégrant Gemini comme agent de vérification contextuelle afin d'assister le processus de dédoublement. Les enregistrements suspects sont consignés dans une file de revue en vue d'une validation manuelle ultérieure.

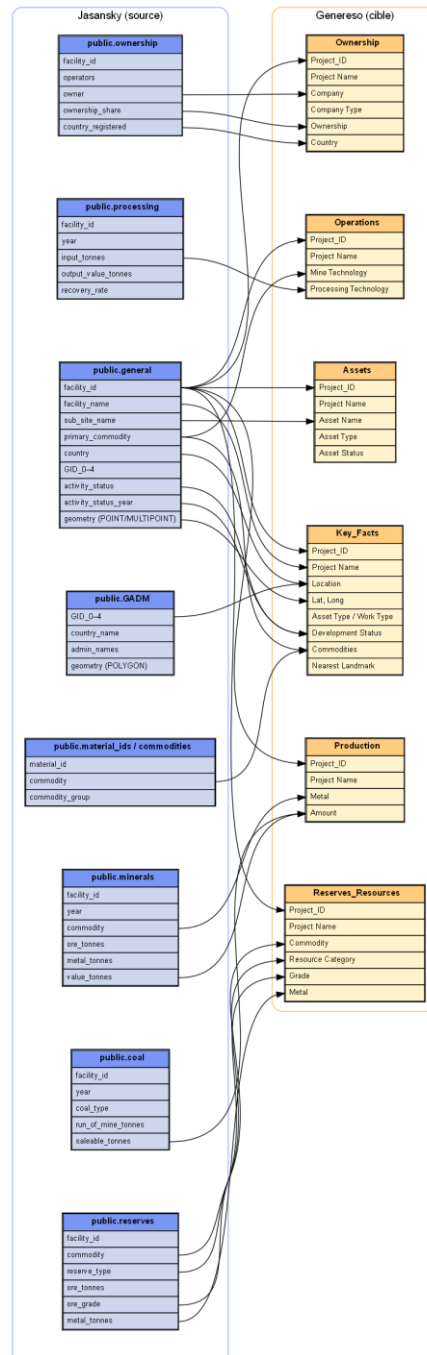


Figure 7. Schéma de correspondance entre Jasansky et Genereso

La base harmonisée est exportée dans un format structuré et versionné, accompagnée d'un dictionnaire de données et d'un journal des transformations. Ces fichiers assurent la traçabilité des modifications et facilitent la réutilisation du processus lors de futures mises à jour. La base harmonisée est destinée à être rendue disponible aux clients payants sur la plateforme ArcGIS Online de Genereso, où elle s'intègre à l'écosystème géospatial de l'entreprise.

### **7.3. Résultats**

L'harmonisation des données a permis de définir une structure commune et de tester les principales étapes du processus d'intégration, mais la base de données obtenue n'est pas encore exploitable en l'état. À ce stade, la vérification des enregistrements reste incomplète, notamment en raison de volumes très contrastés selon les sources, allant de quelques centaines d'entrées à plusieurs dizaines de milliers, voire au-delà pour certaines bases.

Les correspondances entre les champs d'origine et le schéma interne ont été établies de manière partielle. Plusieurs ensembles de données ont pu être intégrés à titre expérimental, notamment ceux de Jasansky et collaborateurs pour les localisations, et ceux de l'ICMM et du RMI pour les informations économiques et institutionnelles.

Le processus de nettoyage a mis en évidence des incohérences sur les coordonnées et sur les identifiants de sites. Certaines ont pu être corrigées, tandis que d'autres restent présentes dans les sources intégrées. Un préfiltrage de dédoublement basé sur la proximité géographique et la similarité des noms a également été mis en place, permettant d'identifier des regroupements probables.

La base harmonisée produite durant le stage représente une étape préparatoire encore incomplète. Elle fournit toutefois un cadre structurel clair, des scripts de traitement reproductibles et une documentation du processus, constituant une base de travail pour la poursuite de l'intégration et de la validation.

### **7.4. Discussion critique**

L'harmonisation des bases minières a permis de démontrer la faisabilité d'une intégration systématique de sources hétérogènes dans un modèle commun, tout en assurant la traçabilité et la

transparence des transformations. Le travail réalisé constitue une première étape structurante, mais la consolidation complète des données reste un processus long et complexe.

Plusieurs difficultés ont été rencontrées au cours du traitement. Les divergences de structure et de terminologie entre les bases ont nécessité un important travail d'interprétation, notamment pour les champs textuels comme les noms de sites, les opérateurs ou les types de produits. Certaines sources présentent des niveaux de précision très variables, ce qui a conduit à des arbitrages constants entre la richesse de l'information et la cohérence du modèle final. La qualité des coordonnées géographiques s'est également révélée inégale, en particulier pour des enregistrements anciens ou issus de compilations, dont la localisation et les métadonnées associées peuvent être incertaines.

Le dédoubleage représente un autre point critique. S'il est possible d'identifier automatiquement des regroupements probables à partir de critères spatiaux et lexicaux, la validation manuelle reste indispensable pour éviter les fusions erronées, en particulier pour les projets multi-sites ou ceux désignés différemment selon les sources. À ce titre, le développement d'un agent fondé sur des méthodes d'intelligence artificielle pourrait constituer une aide au préfiltrage, en proposant des correspondances probables et en priorisant les cas à vérifier. Toutefois, un tel système ne remplacerait pas la validation humaine, et sa mise en œuvre reste techniquement exigeante, notamment pour gérer l'ambiguïté des noms et le contexte des entités.

Enfin, l'intégration partielle de bases scientifiques souligne le potentiel d'une approche élargie reliant données géochimiques et environnementales aux informations économiques et opérationnelles. À terme, cette complémentarité pourrait soutenir des analyses plus approfondies sur les impacts environnementaux et la durabilité du secteur minier.

Dans l'ensemble, cette seconde partie a permis d'établir une méthodologie claire, reproductible et documentée pour la gestion et la préparation de données minières à grande échelle. Elle montre que la qualité et l'interopérabilité des structures de données sont des conditions essentielles à toute analyse géospatiale robuste, en particulier lorsque les sources sont multiples, hétérogènes et partiellement documentées.

## **8. Autres activités et contributions**

En parallèle des deux projets principaux consacrés aux mosaïques Sentinel-2 et à l'harmonisation des bases minières, le stage a comporté plusieurs activités complémentaires contribuant au fonctionnement et à la visibilité de Genereso. Ces tâches, parfois ponctuelles, ont concerné notamment en la participation au développement d'un outil de veille géospatiale appliqué au suivi des chaînes d'approvisionnement, ainsi que la participation à des formations et événements professionnels. Ces expériences ont permis d'approfondir la compréhension du fonctionnement global de l'entreprise et d'enrichir les compétences acquises dans un contexte interdisciplinaire.

### **8.1. Application de veille et de surveillance des chaînes d'approvisionnement**

Une autre dimension du stage a concerné la participation au développement agile d'une application de veille géospatiale destinée au suivi des chaînes d'approvisionnement en métaux et matières premières. Ce projet, piloté par Genereso pour un client industriel, vise à centraliser, dans une interface cartographique interactive, les informations relatives aux sites d'extraction, aux acteurs économiques et aux événements susceptibles d'affecter la production ou la durabilité des fournisseurs. L'application a pour objectif d'évaluer les risques associés à chaque acteur ou pays en combinant des flux d'actualités, des données financières et des indicateurs environnementaux.

Le développement technique de l'application n'a pas fait partie du périmètre du stage, mais une contribution a été apportée à la définition de sa structure fonctionnelle et des paramètres nécessaires à la conversion d'informations textuelles en indicateurs exploitables. Cette contribution s'est faite en étroite collaboration avec le client, lors de nombreux échanges organisés à la fois en personne et, plus fréquemment, par visioconférence. Ces réunions ont permis d'adapter progressivement la logique de scoring et de hiérarchisation des signaux aux besoins spécifiques du projet.

Dans une phase exploratoire, le travail a porté sur la création, à l'aide d'un modèle d'intelligence artificielle, d'une première base de nouvelles fictives destinée à tester la structure de l'application avant la connexion à une véritable API d'actualités. Chaque article était structuré selon plusieurs attributs : titre, date, source, pays, type d'événement (social, environnemental, économique, géopolitique), métal concerné et entités impliquées. Ces informations ont ensuite été traduites en scores de risque permettant de quantifier la gravité ou la portée d'un événement.

Le moteur de prédiction attribue à chaque événement un score agrégé (de 0 à 100) calculé à partir de plusieurs paramètres pondérés : fréquence des mentions, pertinence du sujet, tonalité des articles, fiabilité des sources, similarité avec des événements antérieurs et exposition environnementale, sociale et de gouvernance (ESG). Les scores sont ensuite regroupés en cinq niveaux de gravité, allant des *signaux faibles* (0–20 %) aux *perturbations confirmées* (81–100 %).

Cette approche, combinant automatisation et validation humaine, permet de suivre en temps quasi réel l'évolution des signaux médiatiques affectant les chaînes d'approvisionnement. La contribution au projet a consisté à structurer les données de test, à participer à la définition du système de catégorisation et à appuyer la réflexion sur les paramètres de scoring, tout en assurant la liaison entre les besoins du client et l'équipe technique de Genereso.

## **8.2. Formation et perfectionnement ArcGIS**

Le stage a également constitué une période d'apprentissage et de perfectionnement dans l'utilisation des outils de la suite ArcGIS. L'environnement de travail de Genereso reposant en grande partie sur ces applications, plusieurs séances ont été consacrées à l'exploration avancée des fonctionnalités d'ArcGIS Pro, d'ArcGIS Online et de leur intégration dans les environnements de développement.

Les formations suivies sur la plateforme Esri ont permis d'approfondir l'utilisation d'ArcGIS Notebooks et de l'API Python, notamment pour l'automatisation de traitements, la gestion de bases spatiales et la création de visualisations reproductibles. D'autres modules ont porté sur le développement d'applications web avec ArcGIS Maps SDK for JavaScript, afin de mieux comprendre la logique des interfaces interactives utilisées dans les produits de Genereso.

Ces apprentissages ont contribué à renforcer l'autonomie dans la préparation et la publication des données, à améliorer la qualité des livrables cartographiques et à développer une meilleure compréhension des passerelles entre ArcGIS, les environnements cloud et les outils de développement utilisés par l'entreprise.

## **8.3. Réunions et échanges externes**

Le stage a également comporté la participation à plusieurs réunions et échanges externes, en présentiel comme en visioconférence, avec les partenaires technologiques, les investisseurs et les



clients potentiels de Genereso. Ces interactions ont occupé une place importante dans la vie de l'entreprise, où la dimension scientifique et technique s'articule étroitement avec des considérations économiques et stratégiques.

Les échanges avec les partenaires technologiques, en particulier Google et Esri, ont porté sur les services cloud, les environnements de développement et les possibilités d'intégration des outils géospatiaux dans les applications internes. Ces discussions ont permis de mieux comprendre les conditions d'utilisation des infrastructures mises à disposition, ainsi que les perspectives d'évolution offertes par ces écosystèmes.

Les rencontres avec les investisseurs ont porté sur la présentation des projets en cours et les orientations stratégiques de Genereso dans le domaine de la transparence minière et de la géomatique appliquée aux chaînes d'approvisionnement. Ces moments ont offert un aperçu concret du fonctionnement d'une jeune entreprise innovante et des logiques de financement propres au secteur des technologies géospatiales.

Enfin, plusieurs réunions ont été tenues avec des clients et utilisateurs potentiels, intéressés par les services de veille et de cartographie développés par Genereso. Ces échanges ont permis de mieux cerner les besoins opérationnels, d'ajuster la présentation des outils en développement et d'évaluer leur adéquation avec différents cas d'usage.

Ces différentes expériences ont apporté une vision d'ensemble du positionnement de Genereso dans son écosystème, tout en illustrant la manière dont la recherche appliquée et l'innovation technologique s'inscrivent dans des stratégies de partenariat et de développement économique.

La participation au World Resources Forum 2025 a également constitué un moment marquant du stage. Genereso y disposait d'un stand permettant de présenter ses activités et ses projets à un public composé de chercheurs, d'institutions publiques et d'acteurs privés du secteur des ressources. Cette expérience a offert l'occasion d'échanger avec des visiteurs intéressés par les outils de transparence et de veille géospatiale développés par l'entreprise, tout en contribuant à faire connaître ses approches dans le domaine de la durabilité et des chaînes d'approvisionnement.

En parallèle de la présence au stand, le fait d'assister aux conférences et aux tables rondes du Forum a permis d'approfondir la compréhension des enjeux actuels liés au secteur minier, à la gouvernance des ressources et à la transition énergétique. Ces interventions ont donné un aperçu

des débats en cours sur la responsabilité environnementale, les métaux critiques et l'intégration des technologies numériques dans la gestion des ressources naturelles.

Cette expérience a ainsi contribué à élargir la culture professionnelle et à mieux situer le travail mené chez Genereso dans un cadre international, au croisement entre innovation technologique et gouvernance des ressources.

## **9. Réflexion personnelle**

Ce stage a représenté une expérience formatrice à plusieurs niveaux, tant sur le plan technique que professionnel. Il a permis d'appliquer concrètement les compétences acquises au cours du Certificat complémentaire en géomatique et d'en mesurer la portée dans un environnement de recherche et développement. La diversité des tâches effectuées, allant de la télédétection au traitement de bases de données, a offert une vision d'ensemble du cycle complet de la donnée géospatiale, depuis sa production jusqu'à sa valorisation au sein d'applications opérationnelles.

Travailler au sein d'une structure de taille réduite comme Genereso a également constitué une expérience enrichissante. Le contexte de startup impose une grande autonomie et une capacité d'adaptation permanente, tout en valorisant la polyvalence et la créativité. Les échanges fréquents avec le fondateur et les partenaires externes ont permis d'intégrer progressivement les dimensions techniques, stratégiques et économiques propres à un projet d'innovation. Cette proximité avec les processus de décision a renforcé la compréhension des liens entre la production de données et les besoins concrets du marché.

Sur le plan personnel, le stage a consolidé la confiance dans la capacité à conduire de manière autonome des projets de traitement et d'analyse spatiale. Il a aussi mis en évidence l'importance de la rigueur dans la documentation des flux de travail, condition essentielle à la reproductibilité et au transfert de compétences. Enfin, la participation à des réunions professionnelles et à des événements internationaux a permis d'élargir la perspective sur le rôle de la géomatique dans la transition vers des modèles économiques plus durables.

Cette expérience a confirmé l'intérêt pour les approches intégrées de la donnée géospatiale et leur application à des problématiques concrètes de traçabilité et de durabilité. Elle consolide l'orientation professionnelle vers la géomatique appliquée aux ressources naturelles et à

l'environnement, met en évidence des pistes de développement pour la suite du parcours, et a débouché sur une collaboration à plus long terme qui se poursuit au-delà du stage.

## **10. Conclusion et perspectives**

Le stage réalisé au sein de Genereso Sàrl entre juillet et octobre 2025 a permis de relier la formation en géomatique à des applications concrètes mobilisant à la fois la télédétection et la structuration de données. L'expérience a couvert plusieurs étapes du cycle de la donnée géospatiale, depuis la préparation d'images satellitaires et la production de sorties cartographiques jusqu'à l'harmonisation de bases de données et la formalisation de modèles communs.

Les deux volets principaux du stage ont abouti à des résultats réutilisables. Le premier a conduit à la mise en place d'un flux reproductible de traitement Sentinel-2 permettant de produire des mosaïques exploitables sur de grandes zones d'étude, puis d'en dériver des indices spectraux pour une analyse exploratoire. Ce travail, réalisé dans le cadre d'un mandat portant sur les surfaces rizicoles au Punjab et à l'Haryana, a débouché sur une première classification riz/non-riz fondée sur des règles simples et traduite en sorties cartographiques. Le second volet a porté sur l'harmonisation de bases de données minières internationales, avec la définition d'un schéma cible et la documentation des correspondances nécessaires à leur intégration dans la plateforme géospatiale de Genereso.

Ces réalisations ont été complétées par des activités transversales, notamment le perfectionnement sur les outils ArcGIS, et la participation à des échanges liés à l'avancement des projets. Au-delà des aspects techniques, le stage a permis d'expérimenter un environnement de startup où l'autonomie, la rapidité d'itération et la qualité des livrables doivent être conciliées.

À court terme, les perspectives concernent l'amélioration et la validation des sorties liées au projet rizicole, en particulier par l'acquisition de données de référence complémentaires (zones non-riz et/ou découpage parcellaire) permettant d'évaluer les faux positifs et d'explorer des approches plus robustes. À moyen terme, la poursuite du travail sur la base minière harmonisée pourra inclure l'intégration de nouvelles sources, l'automatisation de certaines étapes de nettoyage et le renforcement des mécanismes de dédoublement et de traçabilité.

Ce stage a ainsi constitué une étape importante dans la consolidation de compétences en géomatique appliquée, en montrant comment des méthodes de traitement et d'intégration peuvent être traduites en produits exploitables, qu'il s'agisse d'images prêtes à l'analyse ou de données structurées destinées à alimenter des outils de visualisation et d'aide à la décision.

## 11. Annexes

Les notebooks Jupyter, scripts et outils SIG (notamment les modèles ArcGIS et traitements associés) développés dans le cadre de ce stage ne sont pas joints au présent dossier. Ils peuvent être transmis sur demande, sous réserve des contraintes liées aux données utilisées et des accords avec les partenaires ou clients concernés. Il en va de même pour les rasters produits, notamment des mosaïques Sentinel-2, des cartes d'indices spectraux et des classifications binaires riz/non-riz.

## 12. Bibliographie et références de données

### 12.1. Références de données utilisées

- ICMM. (2025). *Global Mining Dataset 2025*. International Council on Mining and Metals. Consulté le 25 août 2025, depuis <https://www.icmm.com/en-gb/research/data/2025/global-mining-dataset>
- Responsible Minerals Initiative. (2024). *Facility database* [Data set]. Responsible Minerals Initiative. Consulté le 25 août 2025, depuis <https://www.responsiblemineralsinitiative.org/facility-links/facility-database/>
- Jasansky, S., Lieber, M., Giljum, S., & Maus, V. (2023). *Open database on global coal and metal mine production (2000–2021)* [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.7369478>
- Hudson-Edwards, K. A., & Water and Planetary Health Analytics (WAPHA). (2023). *WAPHA global metal mines database* [Data set]. Dryad. <https://doi.org/10.5061/dryad.j3tx95xmg>

### 12.2. Références institutionnelles et techniques

- Esri. *ArcGIS Pro and ArcGIS Online User Documentation*. Redlands, CA, 2024.
- Google Earth Engine. *Developers' Guide*. Mountain View, CA, 2024.