Table des matières

1. Introduction 1
1.1 Présentation de l'employeur 1
1.2 Présentation du projet MAP-X 2
1.3 Principales fonctions et catégories de données de MAP-X 3
1.4 Contexte du stage 6
1.5 Objectifs du stage 7
2. Méthodologie 8
2.1 Identification de services web 8
2.2 Déverrouillage des données de l'Annuaire Statistique 2014 10
2.2.1 Conversion et nettoyage des documents 11
2.2.2 Mise en forme en vue d'une intégration dans la base de données MAP-X 11
2.2.3 Structure des fichiers produits 12
2.3 Intégration de données de conflit dans MAP-X 14
2.4 Test de la plateforme: interface utilisateur et expérience utilisateur 15
2.5 Création du didacticiel vidéo 15
3. Résultats 17
3.1 Identification de services web 17
3.2 Déverrouillage des données de l'Annuaire Statistique 2014 18
3.3 Intégration de données de conflit dans MAP-X 18
3.4 Tests de la plateforme: interface utilisateur et expérience utilisateur 19
3.5 Création du didacticiel vidéo 23
4. Discussion et perspectives 25
5. Conclusion 28
Bibliographie 29
Annexe 1: Integrating great apes and carbon stock corridors data in MAP-X A-1
Annexe 2: Integrating conflict data in MAP-X A-5
Annexe 3: ACLED R script A-10

Liste des figures

Figure 1: Concessions minières par type en RDC..... 3

Figure 2: Chevauchements potentiels entre les limites des concessions minières et les zones protégées (zones surlignées en rouge)..... 4

Figure 3: Volume et localisation des paiements effectués par les entreprises dans le rapport ITIE de 2013 en RDC..... 4

Figure 4: Résultats de la surveillance de l'environnement dans des emplacements sélectionnés pour la région du Katanga. 5

Figure 5: Conflits ayant eu lieu en RDC entre le 11 septembre 2015 et le 10 septembre 2016. 19

Figure 6: Capture d'écran du rapport de Tenke Fungurume mining..... 21

Figure 7: Mise en page alternative pour le descriptif des pays. 22

Liste des tableaux

Tableau 1: Estimation du temps total consacré à chacune des tâches réalisées durant le stage. 8

Tableau 2: Liste des indicateurs de l'Annuaire Statistique 2014 de la RDC sélectionnés afin d'être incorporés dans MAP-X. 10

Tableau 3: Exemple de table transformée dans un format compatible avec la base de données PostgreSQL pour l'indicateur *répartition de l'effectif de la population par province et par sexe* pour l'année 2000. 12

Tableau 4: Liste des services web identifiés au vu d'une intégration dans MAP-X..... 17

Tableau 5: Liste et contenu des vidéos composant le didacticiel de MAP-X. 23

Liste des acronymes

ACLED	Armed Conflict Location & Event Data Project
CIESIN	Center for International Earth Science Information Network
DEWA	Division de l'Alerte Rapide et de l'Evaluation
DRC	Democratic Republic of Congo
EDE	Explorateur de Données Environnementales
GDELT	Global Database of Events, Language, and Tone
GRID	Global Resource Information Database
ICEWS	Integrated Crisis Early Warning System
INS	Institut National de la Statistique
ITIE	Initiative pour la Transparence dans les Industries Extractives
IUCN	International Union for Conservation of Nature
MAP-X	Mapping and Assessing the Performance of Extractive Industries
OFEV	Office fédéral de l'environnement
PNUD	Programme des Nations Unies pour le Développement
PNUE	Programme des Nations Unies pour l'Environnement
PREVIEW	Plate-forme des données sur les risques globaux
RDC	République Démocratique du Congo
RSE	Responsabilité Sociétale des Entreprises
SEDAC	Centre des données socioéconomiques et des applications
SEC	Suitable Environmental Conditions
SIG	Système d'Information Géographique
UCDP GED	Uppsala Conflict Data Program Georeferenced Event Dataset
UNIGE	Université de Genève
WMS	Web Map Service
WWF	World Wide Fund

1. Introduction

Afin de compléter la formation du Certificat complémentaire en géomatique, l'écriture d'un mémoire est nécessaire. Ce dernier peut prendre la forme d'un travail de recherche académique ou être réalisé dans le cadre d'un stage. Afin de mettre en pratique les notions accumulées lors de la formation et d'acquérir une expérience professionnelle supplémentaire, j'ai choisi de faire un stage. J'ai eu la chance de pouvoir intégré durant trois mois et demi le Global Resource Information Database (GRID) faisant partie du groupe d'information environnementale du Programme des Nations Unies pour l'Environnement (PNUE) pour travailler sur le projet 'Mapping and Assessing the Performance of Extractive Industries' (MAP-X), plateforme de données ouvertes, sous la supervision du Dr Pierre Lacroix.

1.1 Présentation de l'employeur

Créé en 1985, le GRID est un réseau mondial comprenant 15 centres gérés par la Division de l'Alerte Rapide et de l'Évaluation (DEWA) depuis son siège de Nairobi au Kenya. Son objectif est triple:

- faciliter l'accès et produire des données et informations environnementales de haute qualité pour supporter la prise de décision et l'établissement de politiques;
- soutenir le Programme des Nations Unies pour l'Environnement dans sa révision continue de l'état de l'environnement ainsi que des tendances;
- fournir des pré-alertes sur les problèmes et les menaces environnementales émergentes.

Le GRID-Genève est ainsi une plateforme unique fournissant des analyses et des solutions pour un large éventail d'enjeux environnementaux. Il répond principalement aux besoins de ses trois partenaires institutionnels: le PNUE, l'Office fédéral de l'environnement (OFEV) et l'Université de Genève (UNIGE) qui sont liés par un 'accord de partenariat' pluriannuel depuis 1998.

Le GRID-Genève est un centre clé du savoir-faire géo-spatial possédant des atouts dans les systèmes d'information géographique (SIG), la télédétection et l'analyse statistique, intégrée par le biais des infrastructures de données spatiales modernes et des applications web. En travaillant à l'interface entre l'information scientifique et la politique / prise de décision, le GRID-Genève contribue également à développer les capacités dans ses domaines d'expertise parmi le public cible, les pays ou d'autres groupes.

Le GRID-Genève, centre bilingue anglais / français, est localisé dans les locaux de la Maison Internationale de l'Environnement à Châtelaine et compte une quinzaine d'employés.

1.2 Présentation du projet MAP-X

Le projet MAP-X est une réponse de la Banque mondiale et du PNUE à la requête formulée par les pays du g7+. L'énoncé de la mission MAP-X indique que l'objectif du projet est de développer « *un partenariat à même d'améliorer la transparence dans le secteur de l'industrie extractive grâce à la consolidation d'informations pertinentes pour toutes les parties prenantes au sein d'une seule et unique plateforme de données ouvertes* ». Dans cette optique, la plateforme de données ouvertes MAP-X vise à assister les parties prenantes du secteur extractif grâce à la publication des données de l'Initiative pour la Transparence dans les Industries Extractives (ITIE) combinées avec d'autres données contextuelles fiables du secteur. Un mécanisme de rectification, de validation, de publication des données et un standard normatif sous-tendent le processus afin d'assurer l'interopérabilité et le référencement géospatial des données. Les initiatives nationales préexistantes pour la transparence et la gouvernance des ressources s'en trouvent renforcées, tandis que les services spécifiques fournis par le partenariat sont taillés sur mesure afin de répondre aux besoins de chaque pays tels que définis par les parties prenantes. MAP-X doit ainsi faciliter la collaboration entre les diverses agences gouvernementales et leurs partenaires respectifs grâce au partage des informations, à l'élaboration de normes standards pour les données et à la définition claire de procédures de partage, de sauvegarde et de publication de données fiables. Les plateformes, les procédés et les normes standards déjà existants constituent le point de départ de ce travail.

MAP-X est actuellement en phase de faisabilité en République démocratique du Congo (RDC), en Afghanistan et au Nigéria. En RDC deux missions ont été menées en octobre 2015 et en janvier 2016 afin de mettre en place des relations de travail avec les parties prenantes, de présenter la plateforme et ses fonctionnalités et d'initier un inventaire national de données en lien avec le secteur extractif – par exemple les concessions minières et les données publiées dans les rapports de l'ITIE – en vue de leur intégration dans la plateforme. Au cours de la seconde mission, l'équipe MAP-X a pu recueillir les commentaires des parties prenantes et leurs suggestions en vue d'améliorer les fonctionnalités de la plateforme et a acquis ainsi un meilleur aperçu des différentes sources de données, de leur disponibilité, des flux de traitement de ces données et des interactions entre les agences gouvernementales et les autres fournisseurs de données.

Une troisième mission en RDC a eu lieu du 12 au 16 septembre 2016. Les objectifs principaux de cette dernière ont été d'augmenter le niveau d'adhésion politique à l'initiative MAP-X, de continuer le renforcement de l'appropriation au niveau national et de sensibiliser les futurs utilisateurs de la plateforme à l'outil. Les résultats attendus de cette mission contribueront à clôturer la phase de faisabilité en vue du lancement national de la plateforme en RDC début 2017. L'expérience acquise à travers l'application et le test des technologies sera mise à profit au cours de la deuxième phase – le déploiement du pilote à l'échelle nationale – centrée autour d'un partenariat global et d'un ensemble de normes minimales en termes de données ouvertes. Si les tests en RDC s'avèrent concluants, la validation des fonctions proposées au regard des besoins des parties prenantes sera facilitée et la pertinence

d’un déploiement à l'échelle nationale en RDC sera démontrée. La réalisation de ces tests en collaboration avec l'ITIE pourrait également positionner MAP-X comme un outil utile à tous les pays membres.

1.3 Principales fonctions et catégories de données de MAP-X

Les fonctions principales du prototype MAP-X sont présentées ci-dessous:

- **Cartographie des concessions par type ou état d'activité:** MAP-X fournit des informations sur les concessions de ressources, soit par type, statut d'activité, date d'émission, ou propriétaire (figure 1) et chacune d’entre elles peut être interrogée afin d’accéder à des informations complémentaires. MAP-X comprend également une fonction de ‘curseur temporel’ pour visualiser les changements de ces variables au fil du temps.

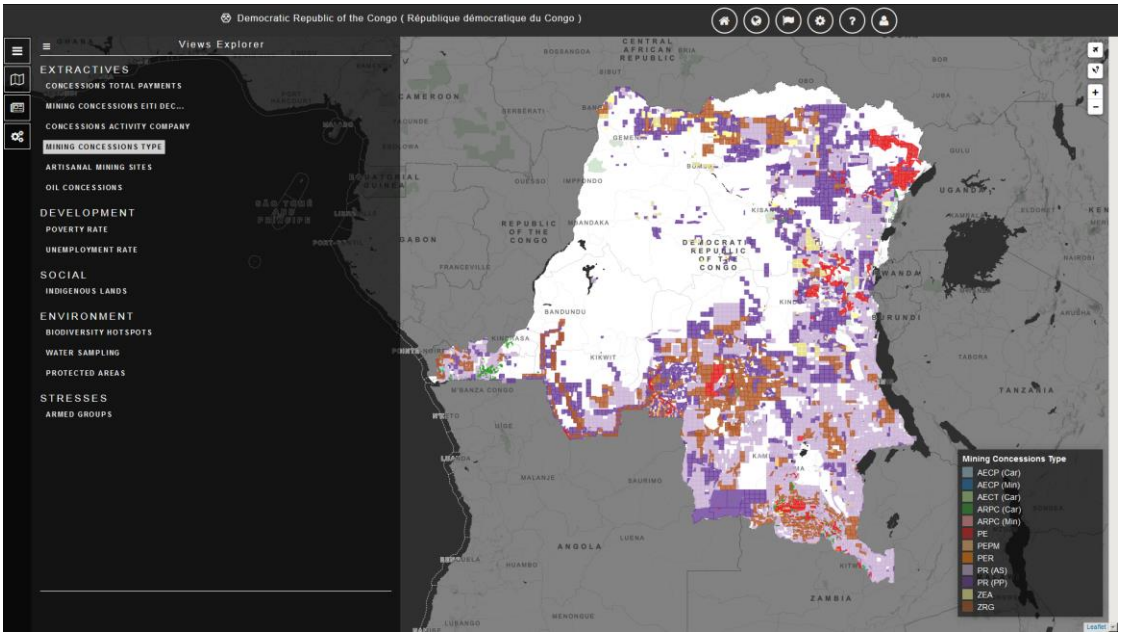

Figure 1: Concessions minières par type en RDC.

- **Visualisation des conflits d'utilisation des terres et de droits:** Les conflits entre les différents types d'utilisations des terres et de droits peuvent être illustrés en combinant les informations des concessions avec celles de l’utilisation des terres ou avec d'autres couches de données. MAP-X peut alors détecter automatiquement des zones de chevauchement entre différentes concessions, entre les concessions et d’autres formes d'utilisation des terres (par exemple, les zones agricoles et pastorales ou les zones protégées (figure 2)) ou entre les concessions et d’autres ayants droits (par exemple les territoires autochtones). MAP-X permettra également de générer des statistiques de chevauchement dans les zones d'intersection. En outre, cette fonctionnalité va montrer comment les différents risques opérationnels (par exemple, les risques naturels) se chevauchent et où les risques cumulatifs sont les plus intenses (par exemple, des cartes de chaleur).

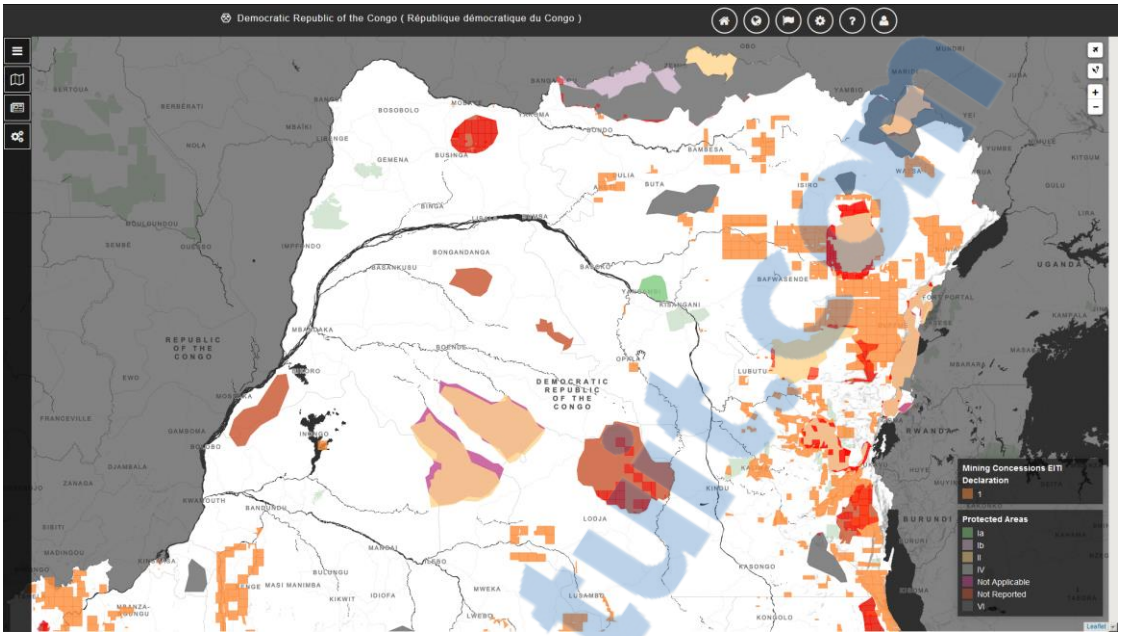

Figure 2: Chevauchements potentiels entre les limites des concessions minières et les zones protégées (zones surlignées en rouge).

- **Visualisation des exigences de déclaration de l'ITIE:** MAP-X fournit une plateforme et des outils pour visualiser et analyser toute une gamme de données de l'ITIE au niveau national tels les chiffres de production ou les effectifs des employés et sous-traitants des sociétés minières. Le volume déclaré des versements effectués par les concessions reçus par le gouvernement peut y être cartographié (figure 3). MAP-X peut également servir à archiver la documentation spécifique des concessions comprenant les contrats et les évaluations d'impact environnemental.

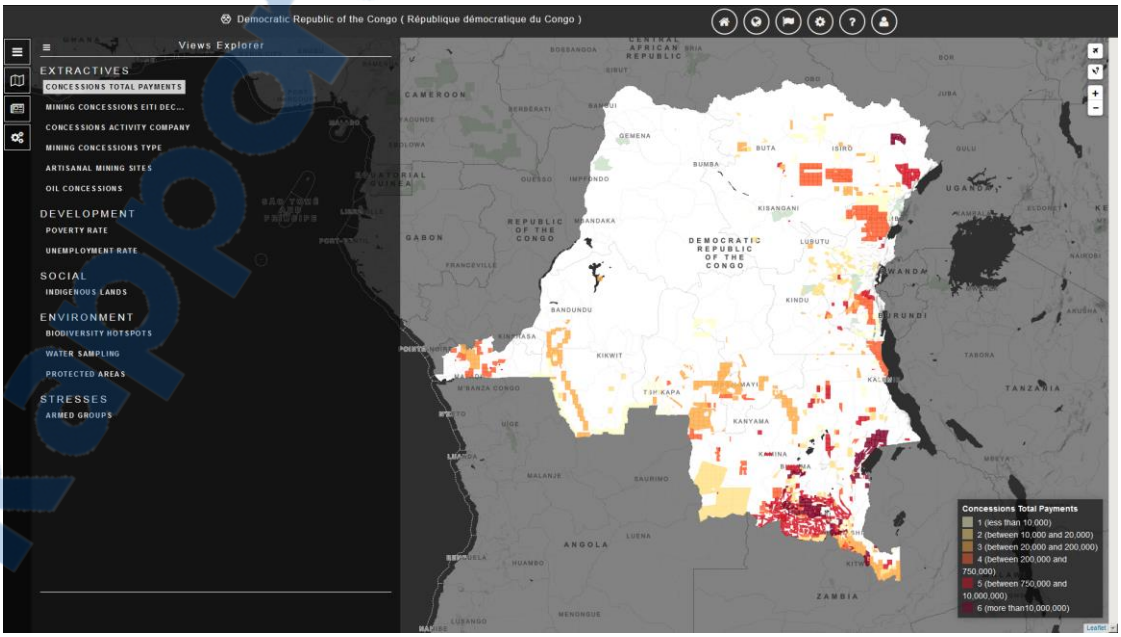

Figure 3: Volume et localisation des paiements effectués par les entreprises dans le rapport ITIE de 2013 en RDC.

- **Interopérabilité complète avec d'autres ensembles de données géospatiales**: MAP-X a la capacité d'extraire et d'afficher automatiquement des données provenant d'autres jeux de données géospatiales conformes à la norme Web Map Service

(WMS). Les utilisateurs peuvent ainsi importer divers jeux de données (par exemple, indicateurs socio-économiques, zones écologiquement sensibles ou occurrence des aléas naturels) afin de procéder à une analyse et un suivi plus approfondis.

- **Suivi des performances de sites spécifiques:** MAP-X comprendra des outils pour stocker, accéder et visualiser une gamme de données spécifiques aux sites (figure 4). Cela comprendra un accès aux contrats des concessions, un système de tableau de bord pour suivre les accords sur le partage des bénéfices et les griefs ainsi qu'un outil de suivi de l'échantillonnage environnemental et des recherches connexes. Les utilisateurs peuvent également définir une zone géographique d'intérêt et recevoir un message ou des mises à jour par courrier électronique lorsque de nouvelles concessions ou d'autres changements d'utilisation des terres sont détectés dans cette zone. MAP-X peut également donner accès à des images satellite haute résolution lorsqu'elles sont disponibles pour surveiller les changements d'utilisation des terres au fil du temps.

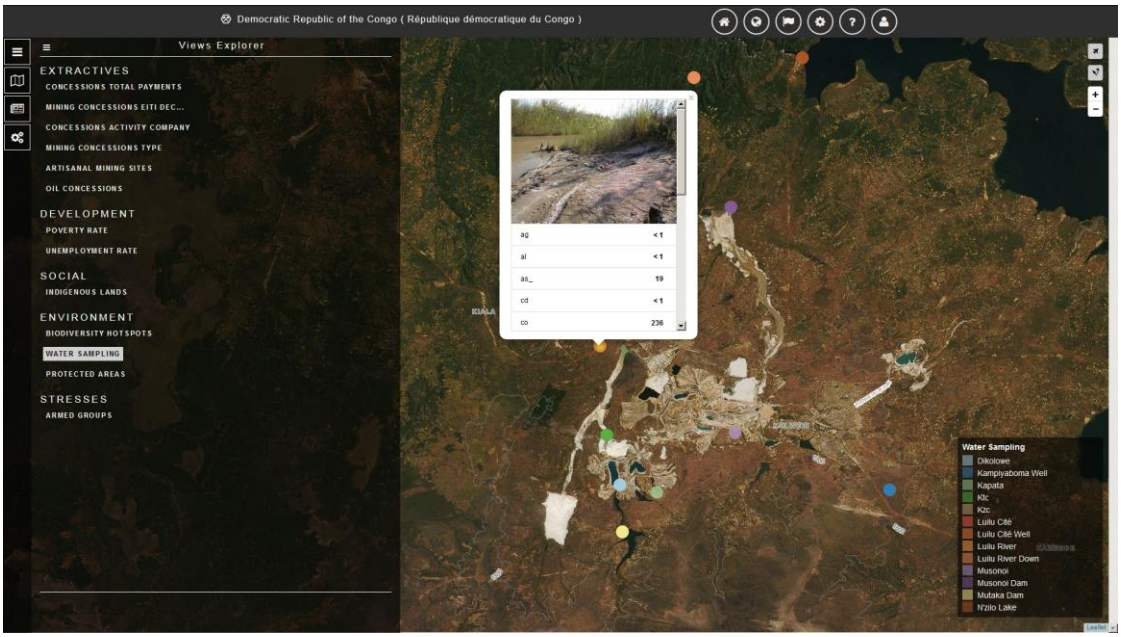

Figure 4: Résultats de la surveillance de l'environnement dans des emplacements sélectionnés pour la région du Katanga.

- **Story Maps:** MAP-X peut être utilisé pour documenter des études de cas spécifiques, des investissements en matière de responsabilité sociétale des entreprises (RSE) ou des griefs communautaires dans des story maps. A travers ces dernières, les utilisateurs peuvent facilement combiner des cartes et contextualiser des couches géographiques en utilisant du texte narratif, des images et du contenu multimédia. La fonctionnalité story map peut également servir à documenter la mise en œuvre des plans d'action spécifiques aux sites.

- **Système d'évaluation du degré d'intégrité des couches de données:** Ce système agira comme un tableau de bord pour les parties prenantes, leur permettant de suivre les progrès tout au long du processus de transformation des données et leur population au sein de la plateforme MAP-X. Il leur permettra également de comprendre l'étendue des informations disponibles à travers le cycle de vie d'un projet d'extraction. Ce

tableau de bord aidera également à comprendre les lacunes susceptibles d'empêcher la publication des couches de données de manière durable et, finalement, indiquera les actions nécessaires pour combler ces lacunes.

Les différentes catégories de données publiées dans MAP-X pour soutenir les secrétariats nationaux dans la mise en œuvre des exigences de l'ITIE sont les suivantes:

1. **Les données du secteur extractif**: essentiellement compilé au niveau national, ce groupe de données comprend la propriété réelle par société, les chiffres d'exportation et de production par société, le registre des licences, les concessions minières (cadastre minier) et les sites artisanaux. Ces données peuvent soutenir les exigences de l'ITIE telles que: 2.3 'Registre des licences', 2.5 'Propriété réelle', 3.2 'Production', 3.3 'Exportation' et 7.2 'Accessibilité des données'.

2. **Les données financières** sont principalement compilées à partir des rapports nationaux annuels de l'ITIE avant d'être liées aux concessions minières pour les afficher sur une carte. Les données financières sont désagrégées jusqu'au niveau des agences d'imposition (p. ex. DGDA, DGI, DGRAD, DRKAT, EP) et comprennent les paiements sociaux déclarés (obligatoires et sur une base volontaire). Les exigences de l'ITIE qui peuvent être supportées par ces données sont 4.1 'Divulgation exhaustive des taxes et des revenus', 6.1 'Dépenses sociales par entreprise extractive' et 7.2 'Accessibilité des données'.

3. **Les données de développement** comprennent p. ex. des indicateurs socio-économiques et de gouvernance au niveau national et mondial et informent le dialogue multipartite (exigence ITIE 7.1 'Débat public').

4. **Les données sociales** sont compilées au niveau national (terres indigènes, données de conflit, effectifs des employés et sous-traitants des sociétés minières) et soutiennent les exigences de l'ITIE 6.3 'Contribution du secteur extractif à l'économie' et 7.1 'Débat public'.

5. **Les données environnementales** peuvent être collectées à différentes échelles (mondiale, nationale, infranationale, locale). Elles peuvent aider les parties prenantes à surveiller leurs obligations contractuelles et éclairer le débat public.

6. **D'autres données contextuelles disponibles en tant que Web Map Services (WMS)** sont publiées afin d'informer le débat public. Elles couvrent l'ensemble des catégories précédentes et sont susceptibles d'obtenir le score maximum.

1.4 Contexte du stage

Comme mentionné dans la présentation du projet MAP-X, la plateforme vise à assister les parties prenantes du secteur extractif grâce à la publication des données de l'ITIE combinées avec d'autres données contextuelles fiables du secteur. Certaines données ont été rendues

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

accessibles depuis la plateforme par l'équipe MAP-X alors que d'autres, pertinentes pour le projet, ont déjà été identifiées ou doivent encore l'être, notamment pour les données de développement, sociales et environnementales. Un travail de préparation de ces dernières est nécessaire en vue de leur intégration dans MAP-X. Au besoin, il peut être réalisé dans un outil SIG, à l'aide d'un logiciel tableur comme Microsoft Excel ou d'un logiciel de traitement des données et d'analyse statistiques tel que R. En plus de la publication de nouvelles données contextuelles dans MAP-X, la préparation d'un didacticiel vidéo est planifiée et constituera un véritable atout pour l'équipe MAP-X car ce dernier offrira aux futurs utilisateurs de la plateforme (par exemple les parties prenantes du secteur extractif en RDC) un support pédagogique permettant de prendre en main l'outil de manière efficace.

1.5 Objectifs du stage

L'objectif principal du stage étant d'identifier, de déverrouiller et de traiter des données contextuelles pour étoffer le catalogue de données publiées dans MAP-X et ainsi soutenir l'analyse spatiale dans la plateforme, les tâches suivantes ont été définies afin d'être réalisées durant le stage:

- identification de services web pertinents pour le projet MAP-X (par ex. sous forme de services web);
- préparation de nouvelles données géospatiales en vue de leur intégration dans MAP-X, par exemple données de conflits, données environnementales (cas d'étude Ogoniland au Nigéria), indicateurs socio-économiques, etc. Dans certains cas, une automatisation du traitement de la donnée peut être nécessaire (par ex. production de scripts R);
- autres tâches techniques en lien avec l'intégration de nouvelles données et cartes dans la plateforme;
- test de la plateforme (interface utilisateur, expérience utilisateur);
- aide à la création d'un didacticiel vidéo destiné aux futurs utilisateurs de la plateforme.

2. Méthodologie

Les méthodes utilisées et traitements réalisés sur les données sont décrits dans cette section afin de permettre une reproduction rapide des tâches effectuées durant le stage par d'autres membres de l'équipe MAP-X. Une estimation du temps consacré à chacune d'entre elles (recherches, réalisation et documentation) est présentée dans le tableau 1:

Tableau 1: Estimation du temps total consacré à chacune des tâches réalisées durant le stage.

Identification de services web	10 jours
Déverrouillage des données de l'Annuaire Statistique 2014	5 jours
Intégration de données de conflit dans MAP-X	15 jours
Test de la plateforme: interface utilisateur et expérience utilisateur	5 jours
Création du didacticiel vidéo	2 jours

2.1 Identification de services web

Le moteur de recherche Google a été utilisé afin d'identifier de potentiels fournisseurs de données (l'Explorateur de Données Environnementales (EDE), la Plate-forme des données sur les risques globaux (PREVIEW) et le Centre des données socioéconomiques et des applications (SEDAC)) pour publier de manière dynamique de nouvelles données contextuelles dans MAP-X. Le but étant de ne pas fournir un 'endpoint' servant plusieurs dizaines de services cartographiques mais d'accéder directement à des services web pertinents en vue de leur publication dans l'explorateur de vue. Les catalogues identifiés ont été examinés et une liste de services web globaux d'intérêt a été établie. Les prérequis de publication dans MAP-X ont ensuite été évalués pour chacun d'entre eux, les services ne les satisfaisant pas étant automatiquement retirés de la liste. Les prérequis sont les suivants:

- la donnée est pertinente pour le secteur extractif et sa publication dans une infrastructure de données spatiales peut contribuer à acquérir ou renforcer la conformité aux exigences de l'ITIE ou aux normes de performance en matière de durabilité;
- la source des données est identifiée et reconnue comme telle par les parties prenantes;
- la donnée est soumise à des tests d'assurance qualité successifs et peut être considérée comme faisant autorité;
- la donnée possède une composante géospatiale afin qu'elles puissent être affichées sur une carte;

- la donnée est accompagnée de métadonnées appropriées ou il existe une personne de contact afin que l'accès en soit facilité;
- indication claire de la licence ou des droits d'auteur, y compris l'utilisation de données ouvertes et les restrictions d'octroi de licences spécifiques;
- la donnée est maintenue. Idéalement, elle doit être diffusée dans MAP-X de manière dynamique au moyen d'un lien automatique avec les bases de données nationales et/ou un mécanisme de mise à jour manuel dans MAP-X. Lorsque cela n'est pas possible, MAP-X peut également accueillir des données statiques.

Une sélection a été ensuite réalisée avec le Dr Lacroix afin de ne conserver qu'une dizaine de services web essentiels au soutien de l'analyse spatiale dans MAP-X et ceci pour deux raisons: ne pas surcharger l'interface graphique de la plateforme et ne pas noyer les utilisateurs dans une masse d'informations trop importante. Afin d'indiquer aux utilisateurs de la plateforme le degré d'intégrité de ces couches WMS, ces dernières ont été ajoutées au système d'évaluation du degré d'intégrité des données. Ce système comprend deux composants principaux:

1. un système d'évaluation du degré d'intégrité (la base de données) avec la liste des couches et une série de questions. Cette base de données n'est pas visible depuis l'interface graphique de la plateforme et sera mise à jour en continu;
2. une représentation graphique du degré d'intégrité (score) de chaque couche. Ce tableau de bord sera visible depuis l'interface graphique de la plateforme.

La carte de score est divisée en quatre composants: fiabilité de la donnée, accessibilité à la donnée (aspects techniques), ouverture de la donnée (aspects légaux) et durabilité de la donnée. La notation pour chacun de ces composants est basée sur une liste de questions qui contribue à évaluer l'intégrité de la donnée. Afin de tenir compte de la réponse à chacune des questions, une échelle cardinale basée sur des couleurs a été choisie: vert pour les scores élevés, orange pour les scores moyens et rouge pour les scores faibles. L'utilisation de trois couleurs (quatre dans la base de données) devrait être suffisante et maintient la méthodologie et la représentation claire et simple pour les utilisateurs de MAP-X. La carte de score est complétée de façon continue sur la base de consultations avec les fournisseurs de données, de recherches via Internet et de revues de la littérature.

Des recherches supplémentaires ont été effectuées afin d'intégrer de manière dynamique des jeux de données spécifiques dans MAP-X telles que les aires protégées ou la couverture de la canopée et d'identifier des services web d'intérêt pour MAP-X sur les plateformes 'Federal Geospatial Platform'[1] et 'A.P.E.S Portal'[2] (plus de détails dans l'annexe 1). Les services web sélectionnés ont été validés par le Dr Lacroix et ont ensuite été intégrés au système d'évaluation de l'intégrité des données.

[1] http://FGP-PGF.maps.arcgis.com/home/
[2] http://apesportal.eva.mpg.de/

2.2 Déverrouillage des données de l'Annuaire Statistique 2014

L'Annuaire Statistique 2014 de la République démocratique du Congo regroupant les statistiques démographiques, politiques, économiques et sociales du pays pour la période 1995-2014 a été réalisé par l'Institut National de la Statistique (INS) et le Programme des Nations Unies pour le Développement (PNUD) et succède à celui de 1995. Il comprend plus de 500 indicateurs répartis dans trois grands chapitres, à savoir un aperçu général de la RDC, les statistiques démographiques et sociales et les statistiques économiques.

Seule une liste non exhaustive d'indicateurs (tableau 2) a été sélectionnée afin d'être incorporée dans MAP-X.

Tableau 2: Liste des indicateurs de l'Annuaire Statistique 2014 de la RDC sélectionnés afin d'être incorporés dans MAP-X.

Indicateur	**Année(s)**	**Tableau(x)**	**Page(s)**
Répartition en pourcentage de la population par province selon le sexe et le groupe d'âge	2000-2014	2.10-2.24	75-89
Taux brut de scolarisation au pre-scolaire et par province	2006-2012	2.52	122
Taux brut de scolarisation au primaire par sexe et par province	2006-2012	2.62	131
Taux d'achèvement du primaire par sexe et par province	2006-2012	2.66	133
Taux brut de scolarisation pour le secondaire par sexe et par province	2006-2012	2.77	142
Taux d'achèvement du secondaire par province	2006-2012	2.79	143
Effectifs des étudiants de l'enseignement supérieur et universitaire par province et par sexe	2005-2011	2.89	153
Ratio filles/ garçons à l'ESU par secteur et par province	2007-2011	2.102	166
Indicateurs de santé par province	2009-2013	2.140-2.144	208-210
Répartition de l'approvisionnement en eau selon la province	2012	2.159	220
Taux de chômage au sens du BIT par tranche d'âge, par milieu de résidence et par province	2012	2.168	228
Taux de chômage au sens large par tranche d'âge, par milieu de résidence et par province	2012	2.171	230
Evolution du nombre d'abonnés à la REGIDESO par province	1996-2012	3.75	334

L'Annuaire Statistique a été transmis à l'équipe MAP-X sous la forme d'un document papier. Ce format nécessitant un long travail de numérisation à la main pour être incorporé dans la plateforme, la version PDF de l'Annuaire mise à disposition du public sur le site de l'INS[1] lui a été privilégié, facilitant ainsi le déverrouillage des données. L'Annuaire n'étant mis à jour que rarement, ce processus a été réalisé à la main, principalement dans Excel, et n'a ainsi pas été automatisé.

[1] http://www.ins-rdc.org/sites/default/files/Montage%20AnnuStat%20FINAL%202%20From%20VEROUILLE%20_0.pdf

2.2.1 Conversion et nettoyage des documents

Le logiciel gratuit, open source et multiplateforme *PDFsam Basic*[1] a été utilisé pour extraire individuellement les pages du fichier PDF correspondant aux indicateurs à intégrer dans MAP-X (tableau 2). Ces dernières ont ensuite été converties en fichier Excel à l'aide de l'outil online gratuit *online2pdf* [2]. Une fois la conversion effectuée, plusieurs tests ont été réalisés sur les fichiers Excel obtenus afin de vérifier si cette dernière était fiable. Aucune erreur n'a été repérée, l'outil *online2pdf* est donc approprié pour ce genre de traitement. Cependant, les fichiers Excel ont nécessité un nettoyage manuel important afin d'obtenir un contenu et une mise en forme cohérents et homogènes sur l'ensemble des données. En effet, de nombreux éléments n'étaient pas compatibles à l'intégration des indicateurs dans une base de données et ont dû être corrigés manuellement dans Excel:

- la présence de cellules fusionnées;
- l'orthographe des provinces variant selon les fichiers;
- les espaces pour séparer les milliers;
- les lignes et colonnes vides;
- la présence des sources sous les tableaux;
- les cellules et chaînes caractères en couleur.

Il est important de noter que les données présentées dans l'Annuaire Statistique 2014 n'ont pas été modifiées dans ce processus malgré l'observation de quelques erreurs dans les tableaux 2.10 et 2.16 à 2.20. En effet, dans ces tables, les données sont normalisées sur 100 % mais lorsque l'on additionne manuellement les lignes, nous pouvons observer que le total n'est pas toujours égal à 100%, des écarts allant jusqu'à ±15% ont été observés.

2.2.2 Mise en forme en vue d'une intégration dans la base de données MAP-X

Une fois les documents Excel nettoyés, ces tables ont été transformées afin d'obtenir une forme compatible et exploitable dans une base de données PostgreSQL. Il a été décidé de produire des tables où chaque ligne comprend une et une seule variable pour une province, un groupe et une année donnés (tableau 3). La fonction *melt()* du logiciel R (Version 3.3.0) a été utilisée afin de réalisé cet exercice. Les tableaux ainsi réalisés ont ensuite été fusionnés en un seul fichier Excel *cod__soc__variable_province* comprenant l'ensemble des indicateurs à intégrer dans MAP-X. Pour les indicateurs décrits par tranches d'âge, seule la colonne comprenant le total a été intégrée au fichier *cod__soc__variable_province* et ceci pour faciliter leur publication et représentation dans la plateforme.

[1] http://www.pdfsam.org/download-pdfsam-basic/
[2] https://online2pdf.com/

Tableau 3: Exemple de table transformée dans un format compatible avec la base de données PostgreSQL pour l'indicateur *répartition de l'effectif de la population par province et par sexe* pour l'année 2000.

province	year	group	variable	value
kin	2000	M	cod__soc__idx__pop__eff	3052
kin	2000	F	cod__soc__idx__pop__eff	3010
bco	2000	M	cod__soc__idx__pop__eff	1651
bco	2000	F	cod__soc__idx__pop__eff	1702
ban	2000	M	cod__soc__idx__pop__eff	2947
ban	2000	F	cod__soc__idx__pop__eff	3106
equ	2000	M	cod__soc__idx__pop__eff	2739
equ	2000	F	cod__soc__idx__pop__eff	2822
ori	2000	M	cod__soc__idx__pop__eff	3080
ori	2000	F	cod__soc__idx__pop__eff	3183
nki	2000	M	cod__soc__idx__pop__eff	1980
nki	2000	F	cod__soc__idx__pop__eff	2083
ski	2000	M	cod__soc__idx__pop__eff	1723
ski	2000	F	cod__soc__idx__pop__eff	1800
man	2000	M	cod__soc__idx__pop__eff	693
man	2000	F	cod__soc__idx__pop__eff	731
kat	2000	M	cod__soc__idx__pop__eff	3675
kat	2000	F	cod__soc__idx__pop__eff	3715
kor	2000	M	cod__soc__idx__pop__eff	2261
kor	2000	F	cod__soc__idx__pop__eff	2303
koc	2000	M	cod__soc__idx__pop__eff	1892
koc	2000	F	cod__soc__idx__pop__eff	1951

Un fichier de métadonnées des provinces, *cod__soc__meta_province,* et des variables, *cod__soc__meta_variable,* ont ensuite été créés afin de renseigner la donnée dans un format compatible avec la base de données de MAP-X. Les abréviations utilisées pour la dénomination des variables ainsi que leur correspondance ont été compilées dans un quatrième et dernier document *cod__soc__abbreviation.*

2.2.3 Structure des fichiers produits

La structure des fichiers produits est présentée ci-dessous avec une description détaillée de chacun de leurs champs.

cod__soc__abbreviation: table comprenant les abréviations utilisées pour la dénomination des variables ainsi que leur correspondance. La méthode utilisée afin d'élaborer les abréviations (codes) est décrite dans le tableau ci-dessous:

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

Nom du champ	Type	Description
abbreviation	texte	• **Pays**: code ISO 3166-1 *alpha-3*. Ex: RDC = cod. • **Mot unique***:* code de trois lettres correspondant aux trois premières lettres du mot. Ex: emploi = emp. • **Mot composé:** code de trois lettres: 1ère lettre du code = 1ère lettre du premier mot; 2ème & 3ème lettres du code = 1ère & 2ème lettres du second mot. Ex: non aménagé = nam. • **Age:** nombre indiqué suivi par la première lettre de l'indicateur temporel. Ex: 23 mois = 23m. La taille de ce code n'est donc pas obligatoirement de 3 caractères.
signification	texte	Signification de l'abréviation

cod__soc__variable_province: table regroupant les indicateurs démographiques, sociaux et économiques à intégrer dans MAP-X. Chaque ligne comprend une et une seule variable pour une province, un groupe et une année donnés.

Nom du champ	Type	Description
province	texte	Code de la province
year	numérique	Année
group	texte	Genre: • M: masculin • F: féminin • Total Secteur: • public • privé • total *Remarque:* les totaux ont été conservés lorsqu'ils ne pouvaient pas être calculés à partir d'autres indicateurs. *Remarque 2:* ce champ peut être vide.
variable	texte	Code de la variable
value	numérique	Valeur de la variable pour une province, une année et un groupe donnés *Remarque:* ce champ peut être vide

cod__soc__meta_province: table comprenant le code des différentes provinces de la RDC ainsi que leur dénomination en français. La traduction officielle du nom des provinces en d'autres langues pourra être ajoutée à cette table si à l'avenir MAP-X devenait multilingue. La géométrie des provinces, de type polygone, sera ajoutée à cette table afin de permettre une représentation des indicateurs dans MAP-X.

Nom du champ	Type	Description
province	texte	Code de la province
fre	texte	Nom de la province en français
geometry	polygone	Géométrie de la province

cod__soc__meta_variable: table comprenant les métadonnées des variables. L'attribut xlsx_link n'a pas été renseigné pour l'instant. Il est envisagé de l'utiliser afin de permettre à l'utilisateur de la plateforme de télécharger la table brute telle que présentée dans l'Annuaire Statistique 2014 de la RDC et d'utiliser ainsi le logiciel de son choix pour représenter les indicateurs sous forme de graphiques.

Nom du champ	Type	Description
variable	texte	Code de la variable
variable_desc	texte	Nom de l'indicateur statistique tel que nommé dans l'Annuaire Statistique 2014
unit	texte	Unité de l'indicateur statistique telle que précisée dans l'Annuaire Statistique 2014
source	texte	Source générale de l'indicateur statistique
source_detail	texte	Source de l'indicateur statistique telle que précisée dans l'Annuaire Statistique 2014
xlsx_link	texte	Liens vers un document Excel comprenant les tables brutes telles que présentées dans l'Annuaire Statistique 2014

2.3 Intégration de données de conflit dans MAP-X

Les données de conflit publiées par l'Armed Conflict Location & Event Data Project (ACLED) ont été sélectionnées pour être incorporées dans MAP-X. Ces dernières sont collectées dans deux jeux de données complémentaires: les 'données en temps réel' pour l'année 2016 recueillies et publiées sur une base hebdomadaire et la 'Version 6' contenant tous les conflits de 1997 à Décembre 2015. Afin d'intégrer ces données de manière dynamique dans MAP-X, un script R (3.3.0) a été développé. Il permet de:

1) télécharger les données de conflit depuis le site internet d'ACLED[1];
2) mettre en forme les jeux de données (format des dates, guillemets pour les chaînes de caractères);
3) fusionner les jeux de données;
4) filtrer les données par date de début, date de fin, pays et / ou type d'événement (batailles, autres événements violents, événements non violents ou émeutes / manifestations);
5) exporter les données filtrées en CSV et en GeoJSON qui est le format pris en charge par MAP-X;
6) exporter une carte des points chauds des données filtrées pour la RDC en GeoJSON.

Le choix du fournisseur de données de conflit, une description plus détaillée d'ACLED ainsi qu'un exemple d'utilisation du script sont présentés dans le document 'Integrating conflict data in MAP-X' disponible dans l'annexe 2.

[1] http://www.acleddata.com/

Rapport-Gratuit.com

2.4 Test de la plateforme: interface utilisateur et expérience utilisateur

La stabilité de MAP-X a tout d'abord été testée en utilisant les principaux navigateurs internet:

- sous Windows 7: Firefox 47.0, Opera 38.0, Chrome 51, Internet Explorer 11 (Microsoft ne supporte plus Internet Explorer au-delà de la version 11 mais en RDC, dans les ministères concernés par le développement de MAP-X, Internet Explorer est toujours en vigueur);
- sous Windows 10: Microsoft Edge.

Tout comportement anormal ou bug identifié a été reporté par le biais de la page Github[1] de la plateforme ou directement signalé au développeur.

MAP-X étant en phase de prototypage, la plateforme n'a jusqu'à présent eu que peu de retours d'utilisateurs sur l'accessibilité, les fonctionnalités et la facilité d'utilisation de la plateforme. Suite à l'utilisation intensive de cette dernière durant le stage, les idées de développement et d'améliorations de MAP-X et de ses outils ont été compilées dans un document qui a ensuite été transmis à l'équipe en charge du développement. Une attention particulière a été portée à l'ergonomie et au bon fonctionnement du cœur de la plateforme, à savoir l'explorateur de vue ('view tools', 'view meta data'), la boîte à outil ('overlap analysis', 'polygon of interest', 'create a view', etc.), le lecteur de story map et le profil du pays (WDI, RGI).

2.5 Création du didacticiel vidéo

Une série de vidéos a été réalisée afin d'illustrer l'utilisation de MAP-X et de ses fonctionnalités aux participants de l'atelier qui sera tenu pour différents ministères lors de la troisième mission MAP-X en RDC (12 au 16 septembre 2016). Le programme gratuit et open-source ShareX[2] (v. 11.2.1) a été choisi pour enregistrer les manipulations produites à l'écran pour sa simplicité d'utilisation, son format d'enregistrement, sa stabilité et ses réglages complets. Les paramètres sélectionnés pour les enregistrements sont les suivants:

- résolution: 1600 x 900;
- images par seconde: 30;
- codec vidéo: x264 (mp4);
- CRF: 10;
- préréglage: Ultra fast.

Le contenu des différentes vidéos produites lors du stage est similaire à celui des vidéos réalisées pour la 7ème Conférence mondiale de l'ITIE qui a eu lieu à Lima, Pérou, du 24 au 25

[1] https://github.com/fxi/map-x-shiny/issues
[2] https://getsharex.com/

février 2016. Cependant, une nouvelle version de MAP-X a été déployée durant l'été 2016, le didacticiel vidéo a ainsi été réalisé dans une interface graphique renouvelée et les dernières fonctionnalités intégrées à la plateforme ont également été présentées telles que les catégories d'utilisateur ou la sélection du pays dans la barre de navigation.

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

3. Résultats

3.1 Identification de services web

Les services web identifiés lors du stage et répondant aux prérequis de publication dans MAP-X sont listés dans le tableau 4. Les quatre premières colonnes de ce dernier présentent les caractéristiques techniques des jeux de données identifiés, à savoir leur nom, le thème leur étant associé, leur étendue spatiale et leur format. Pour être conforme à la Norme ITIE[1], les pays mettant en œuvre l'ITIE doivent satisfaire toutes les exigences décrites dans cette dernière. La cinquième et sixième colonne du tableau 4 présentent les exigences de la Norme ITIE soutenues par chaque jeu de données (cette information sera disponible dans la représentation graphique du système d'évaluation du degré d'intégrité des données) ainsi qu'un exemple concret du support au processus de l'ITIE.

Pour des raisons de mise en page et de clarté, le système d'évaluation de l'intégrité des données permettant de contrôler la fiabilité, l'accessibilité, l'ouverture et la durabilité de ces services web est présenté en annexe numérique (\Systeme_d_evaluation_du_degre_d_integrite_des_couches_Stage_TP_2016.xlsx). Ce dernier sera continuellement mis à jour par l'équipe MAP-X et sera visible du public dans la plateforme.

Tableau 4: Liste des services web identifiés au vu d'une intégration dans MAP-X.

Jeu de données	Thème	Étendue spatiale	Format	Quelles sont les exigences de l'ITIE soutenues?	Description / exemple
Barrages	Développement	Globale	Image	7.1 "Débat public"	Informer le dialogue MP, p. ex. en montrant des réserves d'eau à proximité des zones minières
Empreinte humaine	Développement	Globale	Image	7.1 "Débat public"	Informer le dialogue MP, p. ex. en superposant les biomes avec la couche des concessions minières
Grille de la densité de population 2015	Développement	Globale	Image	7.1 "Débat public"	Informer le dialogue MP, p. ex. en superposant le jeu de données avec la couche des concessions minières
Couverture du sol GlobCover	Développement	Globale	Image	7.1 "Débat public"	Informer le dialogue MP, p. ex. en superposant les terres agricoles avec la couche des concessions minières
Conditions environnementales appropriées aux grands singes	Environnement	Régionale/ Nationale	Image	7.1 "Débat public"	Informer le dialogue MP, p. ex. en superposant le jeu de données avec la couche des concessions minières
Points chauds de la biodiversité	Environnement	Globale	Polygone	7.1 "Débat public"	Informer le dialogue MP, p. ex. en superposant le jeu de données avec la couche des concessions minières
Corridors de stock de carbone végétal	Environnement	Régionale/ Nationale	Polygone	7.1 "Débat public"	Informer le dialogue MP, p. ex. en superposant le jeu de données avec la couche des concessions minières

[1] https://eiti.org/sites/default/files/migrated_files/french_eiti_standard.pdf

Aires protégées	Environnement	Globale	Polygone	7.1 "Débat public"	Informer le dialogue MP, p. ex. en superposant le jeu de données avec la couche des concessions minières
Couverture de la canopée et perte de couverture arborée	Environnement	Globale	Image	7.1 "Débat public"	Informer le dialogue MP, p. ex. en superposant le jeu de données avec la couche des concessions minières
Risque de mortalité multi-aléas (PREVIEW)	Environnement	Globale	Image	7.1 "Débat public"	Informer le dialogue MP, p. ex. comprendre les risques clés en se basant sur des cartes

La 'Federal Geospatial Platform' étant exclusivement consacrée au Canada, elle ne possède pas de données pertinentes pour une publication dans MAP-X. Cependant, une liste comprenant les indicateurs socioéconomiques et les couches liées au secteur extractif présents sur la plateforme a été créée et complétée de quelques captures d'écran afin d'illustrer les couches principales. Ces documents sont disponibles en annexe numérique dans le dossier: \SDI_Canada.

3.2 Déverrouillage des données de l'Annuaire Statistique 2014

Les données de l'Annuaire Statistique 2014 ont été déverrouillées avec succès et les quatre fichiers produits selon la méthode présentée dans la section 2.2 ont été transmis au développeur pour qu'il puisse les intégrer dans la base de données PostgreSQL de MAP-X. Les fichiers en question sont disponibles sous forme d'annexes numériques dans le dossier: \Annuaire_Statistique_2014. A ce jour, les indicateurs statistiques n'ont pas encore été incorporés dans la plateforme.

3.3 Intégration de données de conflit dans MAP-X

Le script R produit pour télécharger, traiter, mettre en forme, filtrer et exporter les données de conflit en formats compatibles avec MAP-X (csv et GeoJSON) est présenté dans l'annexe 3. Il est composé de deux fichiers:

1. *processing_ACLED_data.r*: script a exécuté avec R. Il comprend toutes les variables définies par l'utilisateur (filtres, nom des données en sortie…);
2. *helper_functions.r*: comprenant l'ensemble des fonctions développées et appelées dans le script *processing_ACLED_data.r*.

Le script permet à l'utilisateur de filtrer les données de conflit par:

- pays;
- date de début;
- date de fin;

- type d'événement (batailles, autres événements violents, événements non violents ou émeutes / manifestations).

Afin de produire un exemple concret de publication des données ACLED dans MAP-X, le script a été exécuté pour sélectionner tous les conflits ayant eu lieu en RDC entre le 11 septembre 2015 et le 10 septembre 2016. Les fichiers obtenus sont disponibles en annexe numérique dans le dossier: Script_ACLED\Exemple. Une vue a été produite dans MAP-X à l'aide du fichier *ACLED_RDC_2015-09-11_to_2016-09-10.geojson* où chaque conflit est représenté sous forme de point (figure 5). L'option 'timeslider' ayant été sélectionnée lors de l'exécution du script, la fonction curseur temporel est ainsi utilisable sur cette vue.

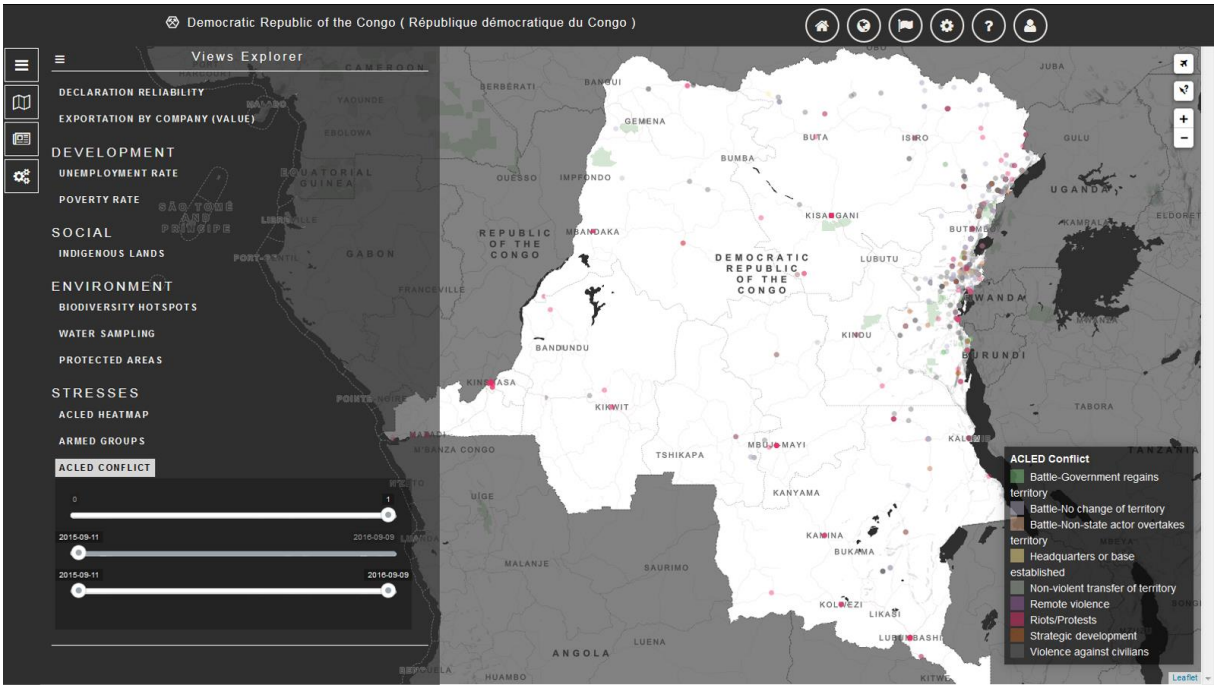

Figure 5: Conflits ayant eu lieu en RDC entre le 11 septembre 2015 et le 10 septembre 2016.

3.4 Tests de la plateforme: interface utilisateur et expérience utilisateur

Dans un premier temps, la stabilité de MAP-X a été évaluée en utilisant les navigateurs internet usuels. Seuls quelques problèmes mineurs ont été observés sous Firefox et Microsoft Edge. Ces derniers ont été signalés au développeur de la plateforme qui les a corrigés immédiatement.

Quelques remarques sur l'environnement graphique et les fonctionnalités de MAP-X ainsi que des suggestions d'amélioration résultent de l'utilisation intensive de la plateforme durant le stage. Ces dernières sont listées ci-dessous:

Général

- La touche F5 implique une déconnexion de la plateforme et une perte des couches sélectionnées et des paramètres d'affichage.

- La plateforme n'est pas homogène d'un point de vue de la langue. Elle est pour l'instant en anglais sauf les rapports des entreprises qui sont en langue française.

Page d'accueil

- Le sous-titre sur la page d'accueil est 'Mapping and Assessing the Performance of Extractive Industries' alors que sur les autres pages, il est écrit 'Mapping and Assessing the Performance of Extractive Industries in Emerging Economies and Fragile States'.
- Une fois enregistré, l'utilisateur se retrouve face à une page essentiellement vide qui ne met pas en valeur la plateforme. Il serait intéressant d'ajouter sur cette page un texte introductif sur le projet ou un petit guide d'utilisation de la plateforme complété de quelques captures d'écrans ou vidéos.

Carte

- Le panneau de gauche est clair et bien hiérarchisé rendant ainsi son utilisation facile et agréable. Cependant, pour plus de clarté, il faudrait ajouter les titres des paramètres réglables dans 'view tools' (p.ex. opacité, curseur temporel…).
- Lorsque l'on affiche une couche possédant un *curseur temporel*, la nature de l'information présentée n'est pas claire. S'agit-il des données de 1994 ou celles de 1994 à 2040? Pour plus de clarté, un petit guide d'utilisation du *curseur temporel* pourrait être intégré dans MAP-X.
- La fonction *curseur temporel* pourrait être améliorée. Il serait intéressant de pouvoir 'joindre' ensemble quelques couches pour que le déplacement dans le temps se fasse simultanément sur chacune d'elles. L'utilisateur pourrait ainsi se focaliser sur l'information affichée à l'écran et non sur les menus. Une fenêtre pourrait apparaître contenant la liste des couches 'jointes' et un *curseur temporel* général. Le panneau de gauche pourrait ainsi être fermé et l'utilisateur pourrait simplement travailler dans cette nouvelle fenêtre.
- La fonction *polygone d'intérêt* n'est pas toujours très intuitive. Pour effacer, le réflexe de l'utilisateur est d'appuyer sur la touche *delete* du clavier mais cette action ramène ce dernier à la page d'accueil et il perd ainsi tout ce qu'il a fait sur MAP-X.
- On ne peut pas refaire fonctionner la fonction *polygone d'intérêt* sans modifier la géométrie du polygone.
- Permettre aux utilisateurs de créer une vue avec les résultats de l'outil *overlap analysis* et / ou leur permettre de télécharger les résultats.
- Afficher un message pour avertir les utilisateurs qu'ils n'ont pas sauvegardé leur story map lorsqu'il quitte l'outil *edit* sans l'avoir fait.

- Le rapport des entreprises manque de lisibilité comme le démontre la figure 6 tirée de celui de Tenke Fungurume mining:

    1. La police bleue des onglets n'est pas du plus bel effet.
    2. La police grise des légendes ne se démarque pas assez du fond noir de la page.
    3. La légende dynamique du graphique en barres est coupée dans certain cas.

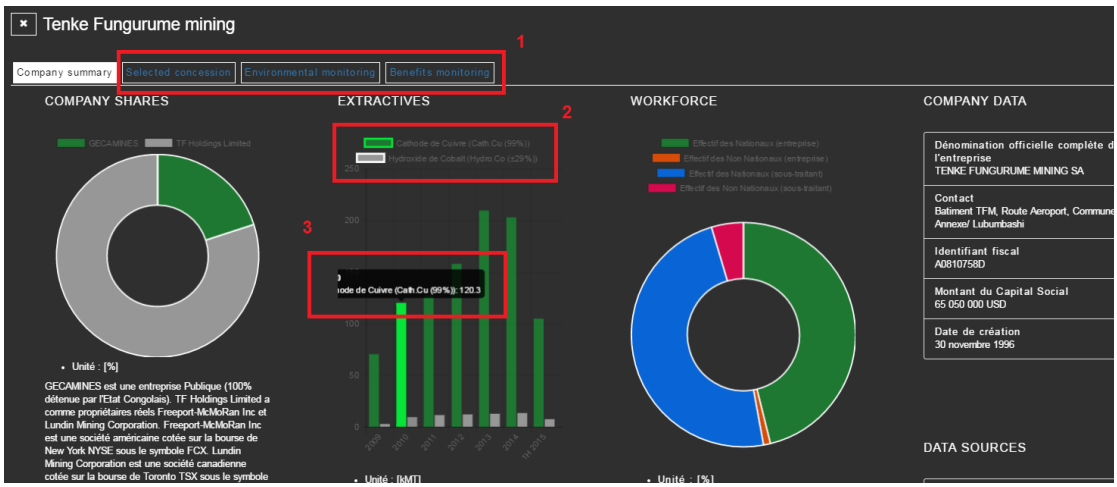

Figure 6: Capture d'écran du rapport de Tenke Fungurume mining.

- Quand on affiche de nombreuses couches, les légendes se superposent et finissent par sortir de l'écran. L'utilisateur ne peut ainsi pas accéder à toutes les légendes. Il faudrait peut-être les mettre dans un panneau sur la droite de la page où il serait possible de les faire défiler verticalement. Une alternative serait de pouvoir accéder à la légende d'une couche à l'aide d'un bouton 'view legend' qui serait disposé entre 'view meta data' et 'view tools'.

Descriptif du pays

- Pour les indicateurs de la Banque Mondiale n'ayant pas de donnée disponible, il serait intéressant de les barrer ou de les griser dans la liste déroulante pour plus de clarté.

- La mise en page pourrait être revue comme présentée sur la figure 7:

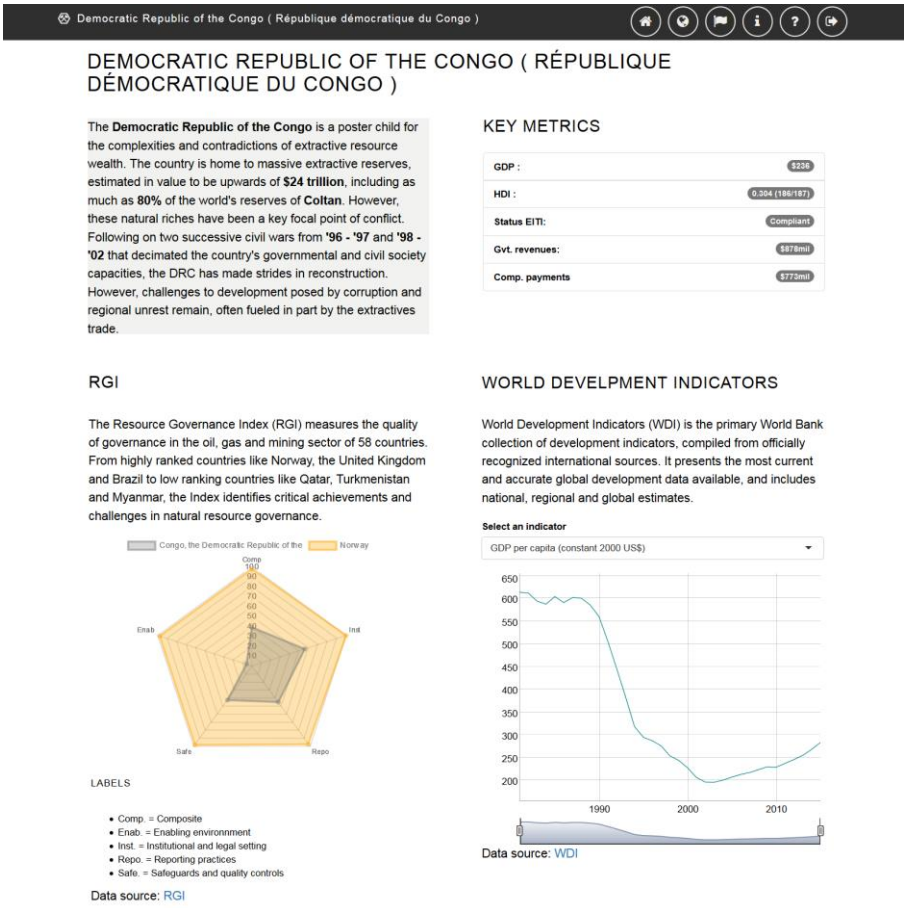

Figure 7: Mise en page alternative pour le descriptif des pays.

Aux remarques précédentes s'ajoute une série de suggestions de développement qui pourraient combler certains manques ressentis lors de l'utilisation de la plateforme ou simplement amener un gain à MAP-X. Ces suggestions sont présentées ci-dessous:

Général

- Ajouter une page contact où il est proposé aux utilisateurs de donner leur avis sur la plateforme et de partager leurs idées pour l'améliorer.
- Ajouter une page regroupant l'ensemble des données présentes sur la plateforme, leurs sources ainsi que leurs conditions d'utilisation et de redistribution.
- Il serait intéressant de faire un wiki comprenant un lexique compilant tous les acronymes utilisés sur la plateforme et une description succincte des organisations et concepts clés.
- Obliger le navigateur à ouvrir un nouvel onglet lorsque l'on clique sur un hyperlien (p.ex. les sources dans le profil du pays).

Carte

- Ajouter une échelle sur la carte et un outil 'règle' pour mesurer manuellement les distances entre divers éléments représentés dans MAP-X.
- Ajouter un bouton à chaque couche pour revenir à l'affichage par défaut.
- Avoir la possibilité d'afficher simultanément plusieurs services WMS dans la plateforme.
- Pourvoir régler l'ordre d'affichage des couches que cela soit entre les couches de MAP-X ou entre ces dernières et des services WMS.

Descriptif du pays

- Ajouter un résumé des données extractives à l'échelle nationale (production, exportations, réserves, revenus, emploi…).
- Avoir la possibilité d'afficher plus d'un indicateur de la Banque Mondiale en même temps.

3.5 Création du didacticiel vidéo

Une série de onze vidéos a été réalisée afin de présenter l'environnement graphique de MAP-X et d'introduire les fonctionnalités de la plateforme aux utilisateurs. Chaque vidéo est centrée sur l'utilisation d'une fonction et est disponible en annexe numérique dans le dossier: \Videos. Le contenu de chaque vidéo est brièvement présenté dans le tableau 5.

Tableau 5: Liste et contenu des vidéos composant le didacticiel de MAP-X.

Nom et durée de la vidéo	**Contenu de la vidéo**
0. Login et interface *Durée: 01:44*	▪ Connexion via email ▪ Navigation dans les menus de l'aide intégrée ▪ Introduction au profil des pays ▪ Navigation dans les indicateurs de la Banque Mondiale et l'Indice de gouvernance des ressources naturelles
1. Superposer des couches et analyse de superposition *Durée: 03:07*	▪ Affichage de diverses couches ▪ Affichage des métadonnées d'une couche ▪ Modification de l'opacité des couches ▪ Introduction à la fonction 'Overlap Analysis'
2. Rapports de l'ITIE *Durée: 01:37*	▪ Accès aux rapports de l'ITIE de la compagnie Twangiza ▪ Consultation des rapports 2013 & 2014 dans Excel

3. Imagerie satellite autour de la concession de Tenke *Durée: 01:03*	▪ Affichage d'une image satellite en fond de plan ▪ Changement d'image satellite depuis la boîte à outils
4. Information spécifique sur la compagnie Tenke *Durée: 01:13*	▪ Consultation du rapport de la compagnie Tenke
5. Fonction de curseur temporel *Durée: 01:26*	▪ Utilisation des deux curseurs de la fonction 'curseur temporel'
6. Interopérabilité WMS *Durée: 02:17*	▪ Illustration de l'interopérabilité entre les plateforme PREVIEW et MAP-X ▪ Ajout d'un service WMS depuis la boîte à outils
7. Ajouter des données et configurer des vues *Durée: 02:45*	▪ Introduction à la fonction 'Create a View' ▪ Affichage de la vue produite, de ces métadonnées et utilisation des outils 'view tools' (opacité, curseur temporel)
8. Campagne de mesures de contamination autour de Kolwezi *Durée: 01:05*	▪ Affichage de la couche 'Water Sampling' ▪ Identification et affichage des photos et de la contamination ▪ Affichage d'une image satellite en fond de plan
9. Story map creator *Durée (partie a): 00:58* *Durée (partie b): 02:36*	▪ Visualisation d'une story map ▪ Création d'une nouvelle story map ▪ Ecriture d'une histoire avec intégration d'une couche de MAP-X
10. Polygone d'intérêt *Durée: 02:07*	▪ Création d'un polygone avec la fonction 'Polygone d'intérêt' ▪ Obtention des attributs d'une couche et récupération du résultat par email ▪ Visualisation du résultat dans Github ▪ Téléchargement des données en GeoJSON depuis Github et intégration dans un projet QGIS

Rapport-Gratuit.com

4. Discussion et perspectives

Les recherches effectuées afin d'identifier des services web pertinents pour le projet MAP-X ont dans un premier en été très fructueuses, une quarantaine de services web ayant été sélectionnée. Au final, seule une dizaine de services a été conservée et ceci pour deux raisons: un nombre important de services ne répondaient pas aux prérequis de publication dans MAP-X (chapitre 2.1) et une sélection trop importante aurait inondé la plateforme d'informations rendant son utilisation moins ergonomique.

L'objectif fixé était de n'avoir que des 'feux verts' dans le système d'évaluation de l'intégrité des données (\Systeme_d_evaluation_du_degre_d_integrite_des_couches_TP_2016.xlsx) afin de publier sans encombre les données dans MAP-X. On peut observer que les données identifiées en WMS sont essentiellement en vert d'où la pertinence de se focaliser sur ce type de données lors des recherches. Sur les dix services sélectionnés deux nécessitent encore un accord écrit des ayants droit afin d'être publiés sur la plateforme. Il s'agit des jeux de données 'barrages' et 'empreinte humaine' dont le fournisseur est le Center for International Earth Science Information Network (CIESIN).

Lors de la sélection des services web, il a été supposé que ces derniers seront utiles aux utilisateurs finaux de MAP-X. Il sera donc nécessaire de faire valider ces sources de données auprès ces derniers lors des ateliers organisés avec l'équipe MAP-X. Les utilisateurs pourront ainsi indiquer à cette dernière la manière d'utiliser ces services dans leur travail quotidien.

La prochaine étape pour l'équipe MAP-X concernant ces jeux de données sera leur publication dans la plateforme de manière dynamique en utilisant les 'endpoints' identifiés lors des recherches.

Les indicateurs de l'Annuaire Statistique 2014 (tableau 2) ont été déverrouillés manuellement en utilisant principalement Excel. Ce choix résulte de deux observations: la première est que l'annuaire n'est pas mis à jour régulièrement et la seconde est qu'un changement de structure entre deux versions de l'annuaire est plus que probable. Le développement d'un script pour automatiser le déverrouillage des données de l'annuaire n'était alors que peu pertinent. Lorsque l'annuaire sera mis à jour dans quelques années, un travail similaire à ce qui a été effectué dans le cadre de ce stage sera alors nécessaire pour intégrer les indicateurs dans MAP-X.

Cependant, si une automatisation de ce processus venait à être envisagée par l'équipe MAP-X, les données fournies par l'INS devraient satisfaire les prérequis suivants:

- l'annuaire statistique est transmis à l'équipe MAP-X dans un format lisible par une machine, idéalement en csv;
- les indicateurs possèdent un identifiant unique;
- les données sont homogènes et cohérentes sur l'ensemble des indicateurs;

- la mise en forme des données est compatible avec les logiciels de base de données usuels.

A ce jour, aucun standard de développement n'a été mis en place au sein de l'équipe MAP-X. Par conséquent, un système de nomenclature des fichiers, variables, etc. a été créé lors du déverrouillage des indicateurs de l'Annuaire Statistique. Il serait intéressant que ce système soit développé puis utilisé comme standard de nomenclature à l'échelle de la plateforme permettant ainsi d'obtenir une base de données homogène et cohérente. L'utilisation d'un tel standard rendra également la base de données de MAP-X plus compréhensible pour l'équipe de développement qui devrait s'agrandir prochainement. De plus MAP-X étant toujours en phase de prototypage, la majeure partie des données restent encore à être publiée dans la plateforme facilitant ainsi la mise en place d'un tel standard.

Les données de conflit ACLED qui seront prochainement incorporées dans MAP-X à l'aide du script développé durant le stage ne font pas l'unanimité pour l'étude des conflits armés (Eck, 2012). Au vu des besoins nécessaires pour l'intégration des données de conflit dans MAP-X (cf. annexe 2.2), ACLED était la seule source pertinente et gratuite disponible. La lecture du matériel mis à disposition du public sur le site de ACLED[1] démontre que de nombreuses critiques visant leurs données (biais urbain, groupes armés non identifiés, décalages temporels, etc.) semblent injustifiées.

Les événements de ACLED étant codés de manière journalière par localisation, acteur et type d'événement, l'utilisation de la fonction 'curseur temporel' permettra aux utilisateurs de MAP-X de visualiser l'évolution spatiale et temporelle des conflits à l'échelle régionale ou locale. L'utilisation de listes déroulantes sur des attributs spécifiques du jeu de données permettra en outre aux utilisateurs d'étudier les tendances de conflits pour un acteur spécifique ou un type d'événement. Ces deux fonctions fonctionneront en parallèle faisant de MAP-X un outil puissant pour l'analyse des conflits et de leurs impacts potentiels sur le secteur extractif.

Lors de ce stage, la plateforme a été utilisée de manière intensive et sa stabilité a été évaluée sur les navigateurs internet usuels. Dans l'ensemble, MAP-X est stable et très peu de bugs ont été repérés. Son interface est claire et son utilisation facile pour toute personne ayant quelques bases en géomatique. Cependant, la fonction du 'curseur temporel' mériterait d'être documentée dans la plateforme car son utilisation ne coule pas de source. Les remarques et suggestions de développement présentées à l'équipe MAP-X sont dans la plupart des cas des compléments non essentiels au bon fonctionnement de la plateforme. Plusieurs idées mises en avant dans ce rapport ont d'ailleurs déjà été envisagées par l'équipe MAP-X (page de contact, wiki, système d'évaluation de l'intégrité des données, etc.) et devraient être développées pour une intégration dans une future mise à jour de la plateforme.

[1] http://www.acleddata.com/research-and-publications/working-papers/

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

La plateforme MAP-X étant en constante évolution, son interface et ses fonctionnalités sont sujettes aux changements. Le didacticiel vidéo produit pour l'atelier tenu pour le Ministère du Plan deviendra alors en partie obsolète lorsque les prochains changements surviendront et certaines vidéos devront être à nouveau réalisées. Il serait intéressant de faire des recherches sur l'automatisation de la production de vidéo afin d'évaluer la faisabilité d'un système autonome et dynamique gérant les modifications de la plateforme.

Sur le plan personnel, ce stage m'a tout d'abord permis de mettre en application les connaissances acquises lors de mon parcours académique et du Certificat complémentaire en Géomatique mais également de développer des connaissances plus avancées dans le domaine des Infrastructures de Données Spatiales et en programmation R.

Lors de mon travail sur l'intégration de nouvelles données contextuelles dans MAP-X, j'ai été amené à réaliser des tâches très variées et à utiliser une large gamme d'outils informatiques constituant ainsi un des atouts majeurs de mon stage. En effet, ce travail n'a pas été restreint à la seule mission de recherche et de traitement de données, j'ai également participé à certaines actions quotidiennes de l'équipe MAP-X et ai vu ainsi de nombreuses facettes de la gestion d'un projet international. Les tâches suivantes ont été ainsi réalisées:

- recherches bibliographiques (données de conflit, grands singes, corridors de stock de carbone végétal, 'readiness score card');
- rédaction, traduction et relecture de plusieurs documents clés;
- prise de contact avec le Max Planck Institutes;
- test des scripts pour déverrouiller les rapports ITIE développés par l'équipe Map-X;
- préparation de l'atelier de Kinshasa de septembre 2016.

Le Dr Lacroix m'a donné le temps et le matériel nécessaires pour me renseigner sur la plateforme MAP-X, ses partenaires et ses objectifs et m'a ainsi permis de participer activement au développement de cette dernière.

M. Moser et le Dr Alexandris m'ont donné de précieux conseils techniques et m'ont initié à la gestion de versions en utilisant le logiciel Git couplé aux sites internet Github et Gitlab, à knitr et au langage Markdown.

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

5. Conclusion

Les objectifs fixés pour le stage ont été atteints et le travail réalisé a ainsi permis d'identifier et de déverrouiller des données contextuelles essentielles au projet MAP-X qui seront intégrées prochainement dans la plateforme. Le script développé pour intégrer automatiquement les données ACLED dans la plateforme est fonctionnel en l'état mais peut être perfectionné. Dans un futur proche, la publication de ce dernier dans un dossier distant sur Github permettra de continuer son développement et offrira aux internautes intéressés par l'étude des conflits armés la possibilité de l'utiliser et de participer également à son développement. Les tests effectués sur MAP-X ont montré une très bonne stabilité de la plateforme sur les navigateurs internet usuels et seules quelques suggestions d'amélioration et de développement résultent de l'utilisation de MAP-X durant ces trois mois et demi de stage. Ces dernières seront prises en considération lors des prochaines mises à jour de la plateforme et constitueront une plus-value à l'expérience d'utilisation de la plateforme.

Pour conclure ce mémoire, j'ai pris énormément de plaisir à travailler au GRID au sein d'une équipe dynamique sur un projet de grande envergure. Etant géologue de formation, travailler sur le milieu extractif m'a vraiment intéressé et m'a permis d'acquérir des connaissances sur l'impact économique, environnemental et social de ce milieu, sujet qui n'a été que légèrement abordé lors de mes cours à l'université.

Bibliographie

Annuaire Statistique 2014, République Démocratique du Congo, Ministère du Plan et Révolution de la Modernité, Institut National de la Statistique, juillet 2015

Baccini, A. G. S. J., Goetz, S. J., Walker, W. S., Laporte, N. T., Sun, M., Sulla-Menashe, D., ... & Samanta, S. (2012). Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change*, *2*(3), 182-185.

Cunningham, E. (2013). ACLED Country Report: DR-Congo.

Eck, K. (2012). In data we trust? A comparison of UCDP GED and ACLED conflict events datasets. *Cooperation and Conflict*, *47*(1), 124-141.

Jantz, P., Goetz, S., & Laporte, N. (2014). Carbon stock corridors to mitigate climate change and promote biodiversity in the tropics. *Nature Climate Change*, *4*(2), 138-142.

McLennan, M. R., & Hockings, K. J. (2016). The Aggressive Apes? Causes and Contexts of Great Ape Attacks on Local Persons. In *Problematic Wildlife* (pp. 373-394). Springer International Publishing.

Raleigh, C., & Dowd, C. (2016). Armed Conflict Location and Event Data Project (ACLED) codebook.

Annexe 1: Integrating great apes and carbon stock corridors data in MAP-X

Thomas Piller (UNEP/GRID-Geneva)
August 2016

1) Environmental concerns

The Democratic Republic of the Congo (DRC) is home to several species of great ape whose conservation is a major environmental concern. Their conservation status established by The International Union for Conservation of Nature (IUCN) varies from endangered to critically endangered. Table 1 summarizes the different great apes living in the DRC, their estimated population and conservation status.

Table 1: estimated population and conservation status for the DRC great apes.

	Population (est.)	Conservation Status
Central Chimpanzee	70'000 - 116’500	Endangered
Eastern Chimpanzee	200'000 - 250’000	Endangered
Bonobo	50’000	Endangered
Mountain Gorilla	880	Critically Endangered
Eastern Lowland Gorilla	2'000 - 10’0000	Endangered

According to the World Wide Fund (WWF)[1], the decrease in the population of the great apes is mainly due to:

- loss of habitat to human activities such as agriculture, mining and logging. “It has been predicted that by 2030, less than 10 % of African […] great ape habitat will remain undisturbed from human activities” (McLennan & Hockings, 2016).
- hunting for bushmeat
- diseases
- civil wars that make conservation difficult if not impossible for many great ape populations in Africa.

In addition to the aforementioned conservation aspects, the great apes have several characteristics that should not be ignored by the extractive sector. McLennan & Hockings (2016) report that according to their species, cohabitation between great apes and humans is not always easy. Indeed, with the decrease in habitat, chimpanzees tend to steal food in villages and crops making local population displeased. In rare cases, attacks of chimpanzees and gorillas towards human were recorded and some of which were lethal. Furthermore, the

[1] http://wwf.panda.org/what_we_do/endangered_species/great_apes/

transmission of diseases between apes and humans is possible and can lead to major epidemic like Ebola.

Besides having a large variety of great apes, the DRC has the second largest stock of carbon (22.0 PgC) in forests in the world (Baccini et al., 2012). To protect this stock of carbon, we must not only focus on the preservation of protected areas which has been the dominant strategy for the past decades. As noted by Jantz et al. (2014), "avoiding deforestation by preserving carbon stored in vegetation between protected areas provides an opportunity to mitigate the effects of land use and climate change on biodiversity by maintaining habitat connectivity across landscapes". Those carbon stock corridors were mapped by Jantz et al. (2014) and 59% of them are at least as dense in vegetation carbon stock as the protected areas they connect.

Therefore, the addition of great apes and carbon stock corridors data in MAP-X will allow the analysis of key environmental risk factors for the extractive sector and incorporate biodiversity conservation in the multi-stakeholder dialogue.

2) Available data

Great apes and carbon stock corridors data are available through the Max Planck Institute for Evolutionary Anthropology geoserver [1] and are listed in table 2:

Table 2: available great apes and carbon stock corridors services for MAP-X integration.

Dataset name	Layer name	Available services	Description	Publication year
Central Chimpanzee Range	REDD:range_centralchimpanzee	WMS WFS	Polygon layer showing geographique range of Central Chimpanzee	2008
Eastern Chimpanzee Range	REDD:range_easternchimpanzee	WMS WFS	Polygon layer showing geographique range of Eastern Chimpanzee	2008
Bonobo Range	REDD:range_bonobo	WMS WFS	Polygon layer showing geographique range of Bonobo	2008
Mountain Gorilla Range	REDD:range_mountaingorilla	WMS WFS	Polygon layer showing geographique range of Mountain Gorilla	2008
Eastern lowland (or Graueri's) Gorilla Range	REDD:range_easternlowlandgorilla	WMS WFS	Polygon layer showing geographique range of Eastern lowland (or Graueri's) Gorilla	2008
SEC* Central Chimpanzee 1990 & 2000	REDD:pt_troglo90 REDD:pt_troglo20	WMS WCS	Raster layer showing the occurrence probability of Central Chimpanzee	2012
SEC* Eastern Chimpanzee 1990 & 2000	REDD:pt_schwein90 REDD:pt_schwein20	WMS WCS	Raster layer showing the occurrence probability of Eastern Chimpanzee	2012
SEC* Bonobo 1990 & 2000	REDD:pt_paniscus90 REDD:pt_paniscus20	WMS WCS	Raster layer showing the occurrence probability of Bonobo	2012

[1] http://primatdbext.eva.mpg.de:8080/geoserver/

SEC* Mountain Gorilla 1990 & 2000	REDD:pt_beringei90 REDD:pt_beringei20	WMS WCS	Raster layer showing the occurrence probability of Mountain Gorilla	2012
Continent Corridors	REDD:per_country_corridors	WMS WFS	Corridors optimizing pathways between tropical protected areas via high vegetation carbon stock areas over the entire pantropical region	2014

* SEC: suitable environmental conditions

3) Using the data for the DRC pilot

The bonobo is the only great ape whose habitat is exclusively located in the DRC. Figure 1 shows the suitable environmental conditions (SEC) for the bonobo and its range. Figure 2 shows carbon stock corridors located in the DRC. One must observe that multiple carbon corridors are located in the bonobo range and therefore their preservation goes together.

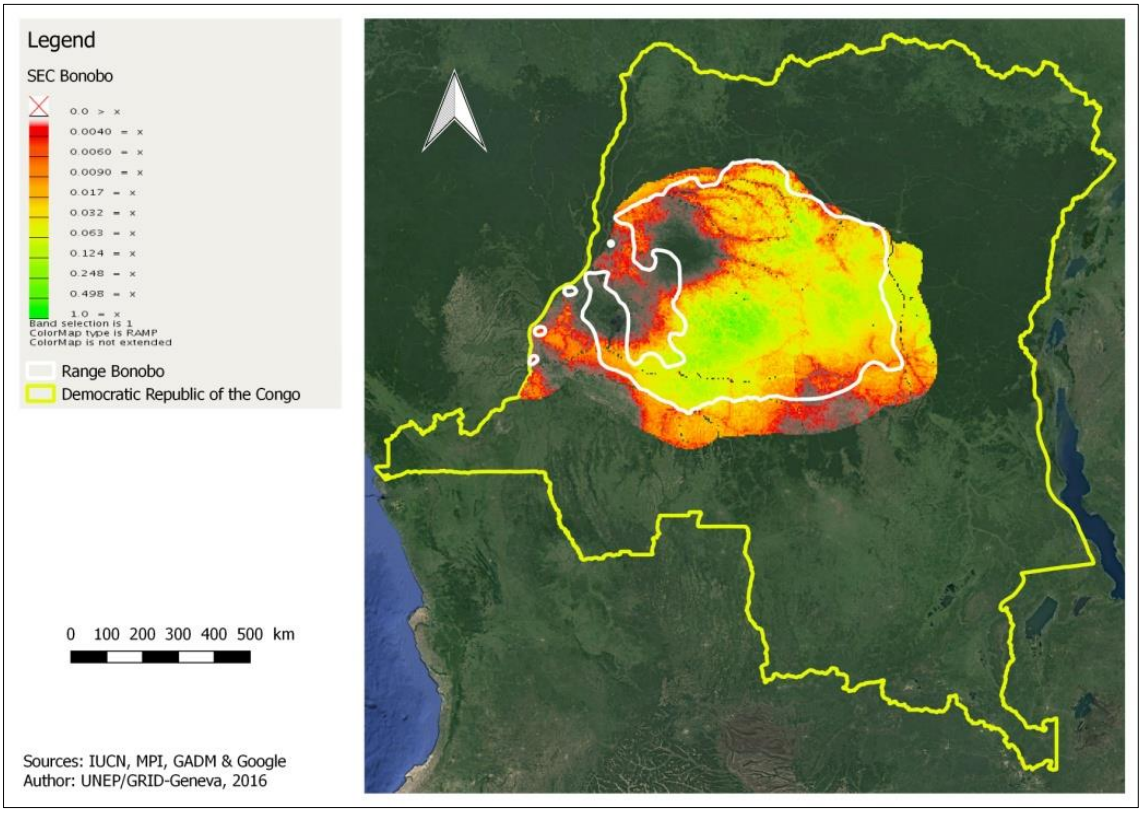

Figure 1: SEC for bonobo in the DRC.

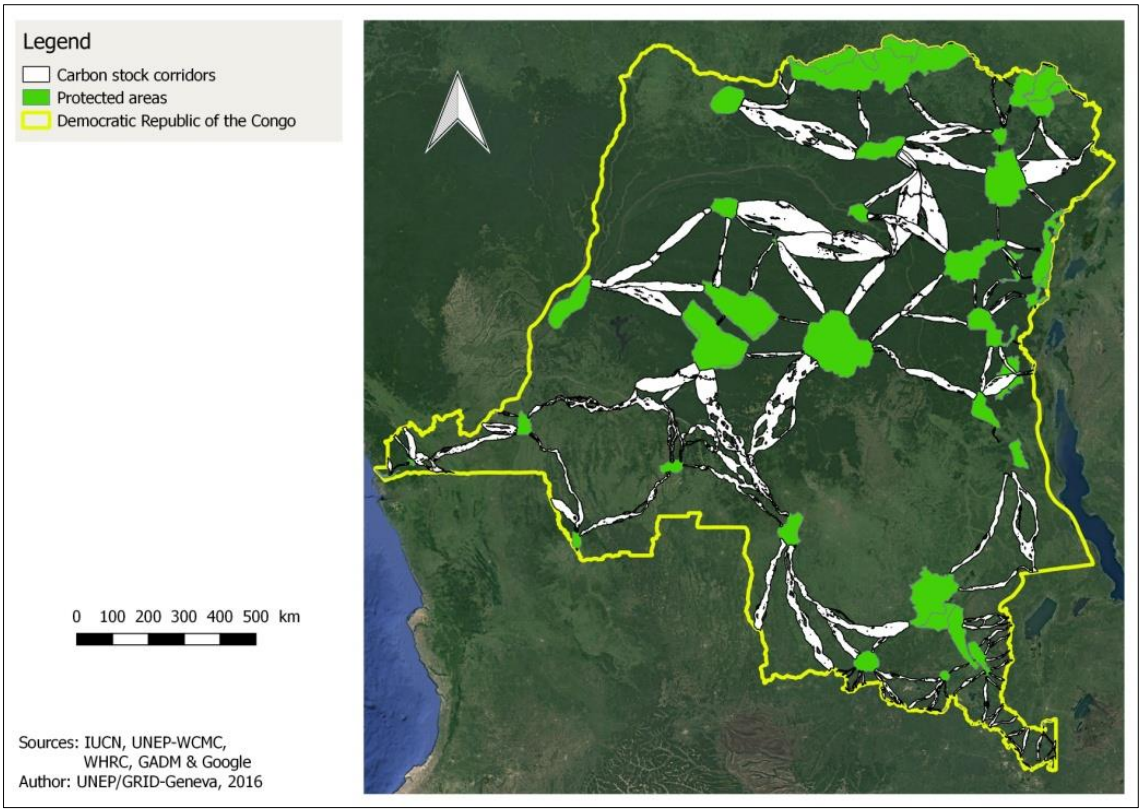

Figure 2: carbon stock corridors in the DRC.

4) Perspectives and future work

The great apes habitat is not limited to the DRC. When MAP-X will come out of its pilot phase, data at the level of Africa should be integrated whether for the great apes or carbon stock corridors to have a comprehensive view of actions needed to conserve biodiversity.

Annexe 2: Integrating conflict data in MAP-X

Thomas Piller (UNEP/GRID-Geneva)
August 2016

1) Problem statement

One of the aims of the MAP-X platform is to present the key risk factors for the extractive sector at the environmental level through analysis of multi-hazard risk or protected areas and at the human level through the occurrence of armed conflict. To study the latter, several open databases are available to the public on the internet, the main ones being:

- Armed Conflict Location & Event Data Project (ACLED)
- Global Database of Events, Language, and Tone (GDELT)
- Uppsala Conflict Data Program Georeferenced Event Dataset (UCDP GED)
- Integrated Crisis Early Warning System (ICEWS)

MAP-X team had to investigate these providers to choose the most relevant database to be integrated into the platform.

2) List of requirements for integration in MAP-X

In order to support spatial analysis, conflict data shown in MAP-X must be:

- recent
- updated regularly
- ideally available in real time
- open
- free of charge

On the aforementioned data providers, only ACLED[1] and GDELT[2] meet these requirements and are candidate databases for MAP-X.

Both databases are based on events reported in the press and media. GDELT publishes data on a global scale while ACLED focuses on sixty developing countries of Africa and Asia. Although both providers offer real time data, their update frequency differs. GDELT database is updated on a daily basis while ACLED updates its database once a week on Monday.

[1] http://www.acleddata.com/
[2] http://gdeltproject.org/

Rapport-Gratuit.com

Publication of data differs between the two providers. Indeed, ACLED database counts events while that of GDELT counts the number of citations of an event in the press and media. If many sources cite the same event, this event will appear multiple times in GDELT database. This will require a rigorous process for removing duplicates and complexify the use of GDELT data. Furthermore, data from ACLED are coded manually and three reviewers check the data before their publication to ensure their validity, accuracy and relevance while GDELT database is coded automatically by a machine. Many conflicts in GDELT have missing information. There are even events where none of the actors involved is specified. ACLED database does not seem to be affected by this type of coding error. Thus ACLED appears to be a more reliable than GDELT.

For all these reasons, ACLED was preferred to GDELT for integration in MAP-X.

As described in Raleigh & Dowd (2016), "ACLED collects and codes reported information on developing world political violence, with a focus on civil and communal conflicts, violence against civilians, remote violence, rioting and protesting. ACLED covers violent activity that occurs both within and outside the context of a civil war, particularly violence against civilians, militia interactions, communal conflict and rioting." Moreover, ACLED codes a range of "non-violent events that are crucial junctures in periods of political violence (e.g. rallies, recruitment drives, peace talks, high-level arrests)" (Raleigh & Dowd, 2016).

ACLED conflict and protest data for African states are collected in two complementary datasets: "Realtime data" for 2016 conflicts, which is collected and published on a weekly basis and "Version 6" which contains all conflicts from 1997 to December 2015. Both of them are structured in the same manner: events are always recorded by day and location, actor and event type.

As described in the country report published on ACLED (Cunningham, 2013), "The Democratic Republic of Congo (DRC) is the second most violent country in the ACLED dataset when measured by the number of conflict events; and the third most fatal over the course of the dataset's coverage (1997 - September 2013)." Version 6 of the dataset (1997 - 2015) provides about 9'830 events and 73'649 related fatalities. A significant increase in violence since 2011 is noticeable (figure 1) mainly due to increased conflicts in North Kivu and in South Kivu (Cunningham, 2013). Real time data for 2016 shows the same trend with nearly 500 events registered since the beginning of the year until early July.

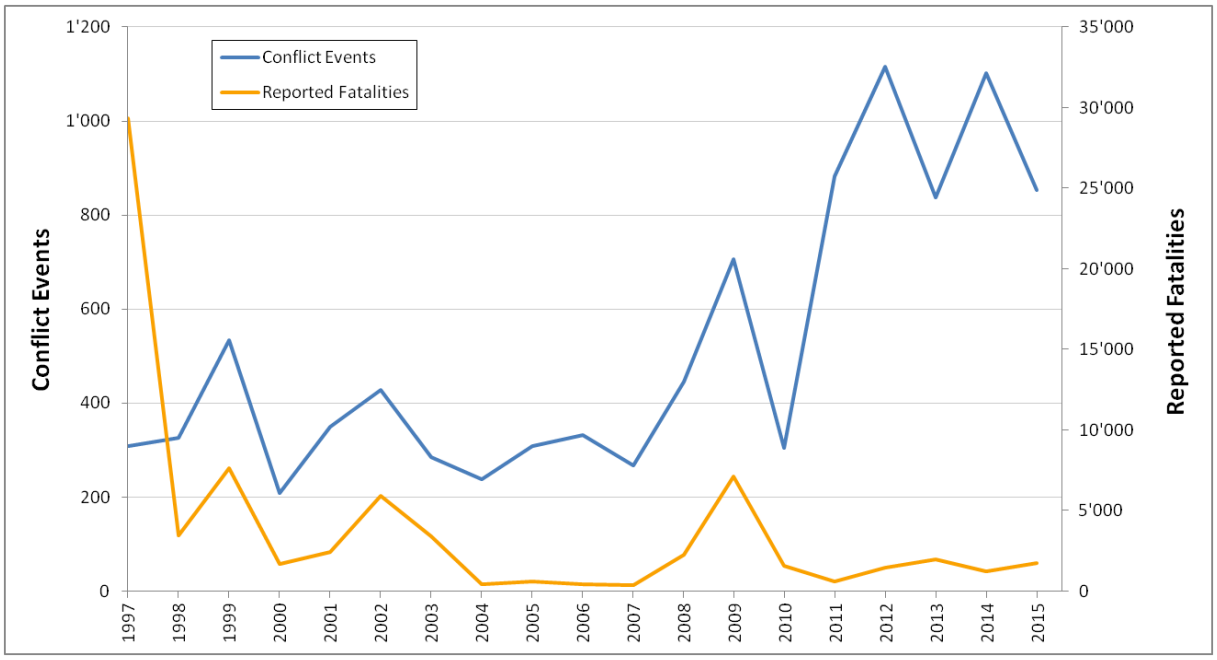

Figure 1: number of events and reported fatalities recorded by year in ACLED in DRC.

3) MAP-X integration

To integrate ACLED data in real time in MAP-X, an R script was developed. The steps are the following:

1) downloading the Version 6 and Realtime data (2016) datasets from ACLED website
2) formatting the datasets (date, double quote in character type column)
3) merging the datasets
4) filtering the data by start date, end date, country and/or type of event (battles, other violent events, non-violent events or riots/protests)
5) exporting the filtered data in CSV and GeoJSON, which is the input MAP-X format
6) exporting a heat map of the filtered data for DRC in GeoJSON

4) Results for the DRC pilot

For the purpose of the DRC pilot, the script was used to extract all events taking place in the *country* between 2015/01/01 and 2016/07/23. There are altogether 1’389 events recorded on this period. Table 1 shows the number of events and the related facilities by type (Battle-Government regains territory, Battle-No change of territory, Battle-Non-state actor overtakes territory, Headquarters or base established, Non-violent transfer of territory, Remote violence, Riots/Protests, Strategic development and Violence against civilians) while a map (external to

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

MAP-X) shows the location of these events (figure 2). We can observe that many of the events took place in the eastern part of the country where armed groups are present.

Table 1: number of events and related facilities by type

in ACLED in DRC between 2015/01/01 and 2016/07/23.

	Number of events	Fatalities
Battle-Government regains territory	64	82
Battle-No change of territory	395	986
Battle-Non-state actor overtakes territory	11	30
Headquarters or base established	7	0
Non-violent transfer of territory	43	0
Remote violence	8	3
Riots/Protests	358	78
Strategic development	108	0
Violence against civilians	395	1'214
Total	*1'389*	*2'393*

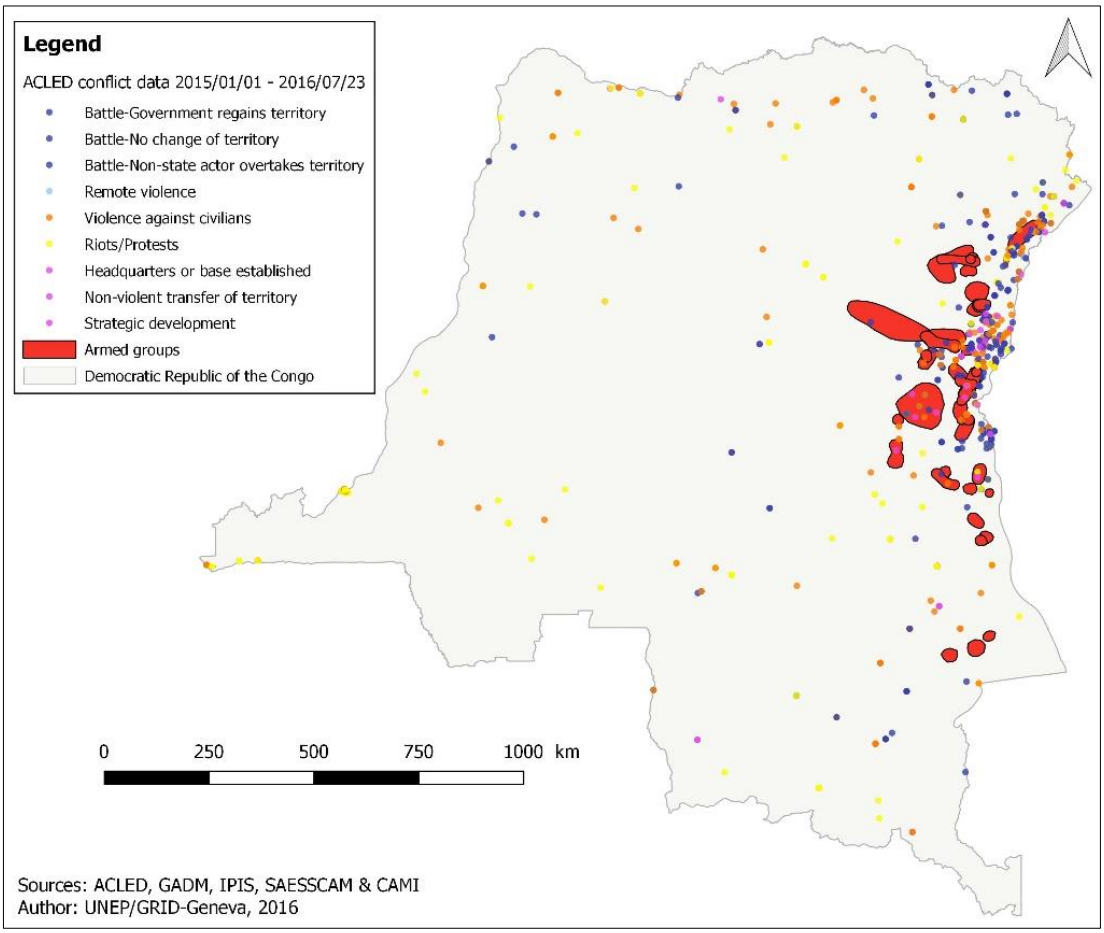

Figure 2: ACLED events taking place in DRC between 2015/01/01 and 2016/07/23.

5) Perspectives and future work

In order to allow users to take full advantage of the ACLED data in MAP-X, several functions already integrated in the platform will be used:

- By using drop-down lists on specific attributes, MAP-X users will be able to study the trends of conflict related to a specific actor or event type.
- As the events are coded on daily basis, using the time slider will allow one to visualize the temporal and spatial evolution of conflict at the regional or local scale.

These two main functions will operate in parallel making MAP-X a powerful tool for the analysis of conflicts and their potential impacts on the extractive sector.

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

Annexe 3: ACLED R script

processing_ACLED_data.r

```
## Automatic processing of ACLED data
## Author: Thomas Piller
## Date: 2016/09/14
## Version: 1.01

# As the data of the previous week are published on Monday on ACLED
# website, the execution of the script works only from Tuesday to Saturday.

# By using the commented line (22) rather than line 21,
# you can download ACLED data until the week before a specified date.


library(geojsonio)

source("helper_functions.R")

# Options to be set by the user
opt <- list(
  outputDir = "output"
)


# Execution of the function date.sat
dateSat <- date.sat(Sys.Date())
# dateSat <- date.sat(as.Date("2016-08-26", "%Y-%m-%d"))
# Allows to download ACLED data until the week before the specified date


# Parameters of ACLED URLs
url.prefixe <- paste0("http://www.acleddata.com/wp-content/uploads/",
                      dateSat$year, "/")
realtimefile.prefixe <- "ACLED-All-Africa-File_"
acledversion.filename <- "ACLED-Version-6-All-Africa-1997-2015_csv_dyadic"


# URLs of ACLED zipped csv for Africa (real time data & version 6 data)
realtime.url <- paste0(url.prefixe, dateSat$month, "/",
                       realtimefile.prefixe, dateSat$year, "0101-to-",
                       dateSat$year, dateSat$month, dateSat$day,
                       "_csv.zip")
acledversion.url <- paste0(url.prefixe,"01/", acledversion.filename,
                           ".zip")


# Execution of the function df.tot
df_tot <- df.tot(realtime.url,
                 acledversion.url,
                 sep = ",",
                 eventdate.col = "EVENT_DATE",
                 eventdate.format = "%d/%m/%Y",
                 outputDir = opt$outputDir,
                 timeslider = TRUE)
```

```r
# Grouping ACLED conflict type
conflict.battle <- c("Battle-No change of territory",
                     "Battle-Non-state actor overtakes territory",
                     "Battle-Government regains territory")
conflict.otherviolent <- c("Violence against civilians",
                           "Remote violence")
conflict.nonviolent <- c("Headquarters or base established",
                         "Strategic development",
                         "Non-violent transfer of territory")
conflict.riotsprotests <- "Riots/Protests"


# Input parameters for the df.filter function
start.date <- dateSat$dateSat - 365
end.date <- dateSat$dateSat
country <- "Democratic Republic of Congo"
conflict.type = c(conflict.battle, conflict.otherviolent,
                  conflict.nonviolent, conflict.riotsprotests)


# Execution of the function df.filter
df_filtered <- df.filter(df_tot,
                         eventdate.col = df_tot$EVENT_DATE,
                         start.date,
                         end.date,
                         country.col = df_tot$COUNTRY,
                         country,
                         eventtype.col = df_tot$EVENT_TYPE,
                         conflict.type)


# Execution of the function export.df
file_name <- paste0("ACLED_RDC_", start.date, "_to_", end.date)
export.df(df_filtered, file.name = file_name, outputDir = opt$outputDir)


# Execution of the function spa.pts
spa_pts <- spa.pts(df_filtered, CRS = "+proj=longlat +datum=WGS84 +no_defs
                        +ellps=WGS84 +towgs84=0,0,0")


# Export of ACLED filtered data in GeoJSON
geojson_write(spa_pts, file = file.path(opt$outputDir, paste0(file_name,
                  ".geojson")), overwrite = TRUE, encoding = "UTF-8")


# Execution of the function ddl.rds for DRC borders
poly.url <- "http://biogeo.ucdavis.edu/data/gadm2.8/rds/COD_adm0.rds"
poly.column <- c("OBJECTID", "ISO", "NAME_ENGLISH")
poly <- ddl.rds(poly.url, poly.column)


# Execution of the function create.grid
poly.grid <- create.grid(poly, resolution.km = 30)


# Execution of the function pip.count
pip_count <- pip.count(spa_pts, poly.grid)
```

```
# Export of the heat map in GeoJSON
geojson_write(pip_count, file = file.path(opt$outputDir, paste0(file_name,
                "_Heatmap.geojson")), overwrite = TRUE, encoding = "UTF-8")
```

helper_functions.r

```
library(lubridate)
library(RCurl)
library(plyr)
library(rgdal)
library(raster)
library(spatialEco)


#' function that identifies the saturday preceding the input date
#'
#' @param date  input date

date.sat <- function(date){
  w.day <- wday(date)
  if (w.day >= 3){
    results <- list()
    results$dateSat <- date-w.day
    results$year <- format(date-w.day, "%Y")
    results$month <- format(date-w.day, "%m")
    results$day <- format(date-w.day, "%d")
    return(results)
  } else {
    'error'
    stop("It's too early to run the script")
  }
}


#' function that downloads, unzips and reads a zipped csv file
#'
#' @param url   url of the zipped csv file
#' @param sep   field separator character
#' @param outputDir output directory

ddl.file <- function(url, sep, outputDir){
  if(url.exists(url) == TRUE){
    temp <- tempfile()
    download.file(url, temp)
    con <- unzip(temp, exdir = outputDir)
    df <- read.csv(con, header = TRUE, sep = sep, stringsAsFactors=FALSE)
    unlink(temp)
    return(df)
  } else {
    'error'
    stop("Invalid url")
  }
}


#' function that applies gsub to all character type columns in a data frame
#'
#' @param pattern   character string containing a regular expression to be
matched in the given character vector
#' @param replacement   a replacement for matched pattern
```

```
#' @param file  name of the file on which the gsub function is applied

f.gsub <- function(pattern, replacement, file) {
  file2 <- file
  c_character <- adply(1:ncol(file2), 1,
function(x)(is.character(file2[,x])))
  c_val <- as.numeric(array(c_character[which(c_character$V1),1]))
  for (i in 1:length(c_val)){
    file2[,c_val[i]] <- gsub(pattern, replacement, file2[,c_val[i]])
  }
  return(file2)
}


#' function that creates a global data frame from ACLED data
#'
#' @param realtime.url  url of ACLED realtime data
#' @param acledversion.url  url of ACLED version 6 data
#' @param sep   field separator character used in ACLED csv
#' @param eventdate.col name of the column that contains the event date
#' @param eventdate.format  event date format used in ACLED csv
#' @param outputDir output directory
#' @param timeslider if TRUE, add mx_date_start and mx_date_end columns to
#'                      allow compatibility with the time slider function
#'                      in MAP-X

df.tot <- function(realtime.url, acledversion.url, sep, eventdate.col,
                      eventdate.format, outputDir, timeslider = FALSE){
  df_rt <- ddl.file(realtime.url, sep, outputDir)
  df_rt <- f.gsub('"', '\"', df_rt)
  df_ver <- ddl.file(acledversion.url, sep, outputDir)
  df_ver <- f.gsub('"', '\"', df_ver)
  df_rt[,eventdate.col] <- as.Date(df_rt[,eventdate.col], eventdate.format)
  df_ver[,eventdate.col] <- as.Date(df_ver[,eventdate.col],
                                    eventdate.format)
  column.names <- names(df_ver)
  df_tot <- rbind(df_ver[,column.names], df_rt[,column.names])
  if(isTRUE(timeslider)){
    df_tot$mx_date_start <- as.numeric(as.POSIXct(df_tot[,eventdate.col]))
    df_tot$mx_date_end <- df_tot$mx_date_start
    }
  return(df_tot)
}


#' function that filters ACLED data
#'
#' @param df.tot    name of ACLED global data frame
#' @param eventdate.col name of the column that contains the event date
#' @param start.date    date from which the data is selected
#' @param end.date  date until which the data is selected
#' @param country.col   name of the column that contains the country
#' @param country   country(ies) for which the data is selected
#' @param eventtype.col name of the column that contains the event type
#' @param conflict.type event type(s) for which the data is selected

df.filter <- function(df.tot, eventdate.col, start.date, end.date,
                     country.col, country, eventtype.col, conflict.type){
  df.fil <- df.tot[eventdate.col >= start.date &
                    eventdate.col <= end.date &
                    country.col %in% country &
```

```
                      eventtype.col %in% conflict.type,]
  rownames(df.fil) <- seq(length=nrow(df.fil))
  return(df.fil)
}


#' function that exports a data frame in a csv file with quotes applied on
#' character type columns
#'
#' @param df.filtered   input data frame
#' @param file.name name of the output file
#' @param outputDir output directory

export.df <- function(df.filtered, file.name, outputDir){
  c_character <- adply(1:ncol(df.filtered), 1,
                       function(x)(is.character(df.filtered[,x])))
  quote_val <- as.numeric(array(c_character[which(c_character$V1), 1]))
  write.csv(df.filtered, file = file.path(outputDir,
            paste0(file.name,".csv")), quote = quote_val, row.names = F,
            fileEncoding = "UTF-8")
}


#' function that creates a SpatialPointsDataFrame from a data frame with
#' CRS specified
#'
#' @param df.filtered   input data frame
#' @param CRS   CRS applied to the SpatialPointsDataFrame

spa.pts <- function(df.filtered, CRS){
  coordinates(df.filtered) <- ~LONGITUDE+LATITUDE
  proj4string(df.filtered)  <- CRS
  df.filtered@data$IDCount <- 1
  return(df.filtered)
}


#' function that downloads and reads a rds file
#'
#' @param url   url of the rds file
#' @param column.names  name of the column(s) to keep in the rds file

ddl.rds <- function(url, column.names){
  if(url.exists(url) == TRUE){
    temp <- tempfile()
    download.file(url, temp, mode = "wb")
    rds <- readRDS(temp)
    rds <- rds[,column.names]
    unlink(temp)
    return(rds)
  } else {
    'error'
    stop("Invalid url")
  }
}


#' function that creates a grid from a SpatialPolygonsDataFrame
#'
#' @param poly  spatial extent of the output grid
#' @param resolution.km spatial resolution of the output grid (in
```

```
#'                                kilometers)

create.grid <- function(poly,resolution.km){
  grid <- raster(extent(poly))
  res(grid) <- (360/40007.863)*resolution.km
  proj4string(grid)<-proj4string(poly)
  gridpoly <- rasterToPolygons(grid)
  poly.grid <- intersect(poly, gridpoly)
  poly.grid <- poly.grid[, -which(names(poly.grid) %in% c("layer"))]
  poly.grid@data$OBJECTID <- 1:nrow(poly.grid)
  return(poly.grid)
}


#' function that evaluates the number of point by cell
#'
#' @param spa_pts   input SpatialPointsDataFrame
#' @param poly.grid input grid

pip.count <- function(spa_pts,poly.grid){
  pts.poly <- point.in.poly(spa_pts,poly.grid)
  count.pts <- tapply(pts.poly@data$IDCount, pts.poly@data$OBJECTID ,
                      FUN=sum)
  count.pts <- data.frame(OBJECTID=names(count.pts),
                          Count=as.vector(count.pts))
  poly.count <- merge(poly.grid, count.pts, by="OBJECTID")
  poly.count$Count[is.na(poly.count$Count)] <- 0
  return(poly.count)
}
```

Rapport-Gratuit.com