



CHARTE DU DATALAB D'ANTICOR

Préambule

La création du Datalab d'Anticor s'inscrit dans l'action historique de l'association : la lutte contre la corruption et la promotion de l'éthique en politique. Dans un contexte où le droit pour les citoyens de demander des comptes aux agents publics de leur administration est devenu un droit constitutionnel, il est manifeste que la donnée publique est un enjeu de société. Son accès est, pour les citoyens, une condition de leur contrôle et de la responsabilisation des acteurs publics. Il est donc primordial de créer des outils pour accéder à cette donnée, pour la comprendre et pour l'exploiter. Ainsi, le Datalab répond à la nécessité de créer des outils innovants et efficaces au service du contrôle citoyen.

Cette charte définit les objectifs, le fonctionnement et les exigences du Datalab, tout en inscrivant son action dans le respect des valeurs fondamentales d'Anticor et en étroite collaboration avec l'association. Le Datalab est aussi attaché, dans son action, à mettre en œuvre une démarche scientifique afin de développer des outils rigoureux et pertinents, également respectueux des exigences légales en matière d'utilisation de données.

Article 1 - Objectifs du Datalab

L'objectif principal du Datalab est de créer un espace de recherche et d'analyse à la frontière entre la science des données et l'analyse juridique. Le Datalab vise à mettre en valeur, d'une part, la riche expérience de nombreux bénévoles d'Anticor dans le contrôle de l'activité des responsables publics ainsi que l'expertise juridique de l'association et, d'autre part, l'immense potentiel d'analyse des sciences des données contenues dans des documents et informations publics.

Les objectifs du Datalab sont multiples (les types de projets et cas d'usage sont explicités plus longuement en Annexe 1) :

- Produire des études de qualité, sur la base de données publiques, grâce à l'utilisation de techniques avancées telles que l'apprentissage automatique, l'analyse statistique, et la visualisation des données ;

- Développer de nouveaux outils permettant le contrôle citoyen et notamment des algorithmes de détection d'anomalies pour identifier des irrégularités dans les données publiques et des défaillances dans l'action publique ;
- Soutenir l'action contentieuse d'Anticor, notamment en automatisant la collecte et l'analyse de preuves, la vérification des informations et la production de rapports détaillés ;
- Valoriser la transparence et sensibiliser le public aux enjeux liés à la corruption en utilisant des techniques de visualisation interactive pour rendre les informations compréhensibles et accessibles au grand public.

Enfin, le Datalab souhaite promouvoir une **culture de collaboration et de rigueur scientifique** en associant des experts de divers horizons - dont des chercheurs, des praticiens du droit ou d'économie et des spécialistes en analyse de données. Cette approche collaborative permet de bénéficier d'une expertise pluridisciplinaire et de garantir la pertinence des actions menées.

Le Datalab encourage la participation de ces experts à travers des partenariats avec des universités, des centres de recherche, et des institutions spécialisées dans la lutte contre la corruption. Il peut par exemple, avec l'autorisation des instances dirigeantes d'Anticor, organiser des ateliers, des conférences ou des séminaires pour échanger des idées et des connaissances avec ces experts ou diffuser ses actions auprès du public.

Sous réserve de la validation préalable du conseil d'administration d'Anticor, le Datalab peut accueillir des étudiants et des chercheurs en stage, leur offrant une expérience pratique et leur permettant de contribuer aux projets en cours. Cette collaboration permet de renforcer les compétences du Datalab et de garantir la rigueur scientifique de ses analyses.

Article 2 - Organisation du Datalab

Le Datalab est dirigé par un comité composé d'au moins trois personnes nommées pour trois ans par le Conseil d'administration d'Anticor. Ce comité (ci-après le "Comité de gouvernance") coordonne les différents projets du Datalab et décide de ses orientations stratégiques.

La double responsabilité du Datalab sur le plan de l'analyse des données et de la question juridique est essentielle pour assurer la rigueur et la pertinence de ses actions. Aussi, les membres comprennent des spécialistes du droit et de la science des données dont au moins un représentant des instances dirigeantes d'Anticor (Conseil d'administration ou bureau). Les premiers membres du Comité de gouvernance sont M. Julien PEREZ, Mme Emma TAILLEFER et Mme Élise VAN BENEDEEN.

Le Comité de gouvernance assure la gestion opérationnelle et la coordination des projets. Il vérifie la concordance entre les différents projets et le respect des objectifs fixés par la présente charte. Il décide de la publication, le cas échéant restreinte, de certains projets selon leur sensibilité. Les décisions sont prises de manière collégiale. Si aucun accord n'est trouvé, les arbitrages sont soumis à la décision du bureau de l'association.

Le Comité de gouvernance établit au premier trimestre de chaque année civile un bilan d'activités adressé au conseil d'administration d'Anticor.

Le renouvellement du Comité de gouvernance est assuré par appel à candidatures sous la responsabilité du Conseil d'administration d'Anticor, qui peut également mettre fin à la mission d'un ou plusieurs membres du Comité de gouvernance à tout moment et sans motif.

Le Comité de gouvernance pourra être appuyé par une équipe de data scientists, en charge de la conception des analyses et du développement technique appliquant les standards méthodologiques rigoureux requis par la présente charte ainsi que par une équipe de juristes veillant à la rigueur juridique de l'ensemble des méthodes et analyses mises en place. Ces équipes pourront transmettre des propositions et avis au Comité de gouvernance.

Article 3 - Données personnelles et confidentialité

Le Datalab se conforme strictement aux lois et réglementations en vigueur, en particulier celles relatives à la protection des données personnelles, telles que le Règlement Général sur la Protection des Données (RGPD). Le respect de la confidentialité est une priorité, surtout lorsque des informations sensibles sont en jeu. Ainsi, si le Datalab est amené à connaître d'informations provenant de lanceurs d'alertes, il s'engage à assurer la protection et la sécurité des informations sensibles, en suivant des protocoles de confidentialité stricts.

Une politique de confidentialité rigoureuse s'applique également à toutes les données traitées par le Datalab. En cas d'information sensible ou de données liées à des procédures contentieuses, le Datalab s'engage à protéger ces informations de manière stricte, en suivant des protocoles de sécurité avancés pour éviter toute fuite ou usage non autorisé.

Les bénévoles qui participent aux travaux du DataLab signent un accord de confidentialité.

Article 4 - Propriété intellectuelle

Les travaux du Datalab sont régis par les dispositions relatives aux œuvres collectives (article 113-2¹ et suivants du Code de la propriété intellectuelle).

Les œuvres produites par le Datalab sont encadrées par la **licence publique générale GNU Affero (GNU AGPL)**. Cette licence autorise les usages commerciaux mais impose l'exigence suivante : si une personne exécute un programme modifié sur un serveur et laisse d'autres utilisateurs communiquer avec lui, ce serveur doit aussi permettre de télécharger le code source correspondant à la version modifiée en fonctionnement.

Le Comité de gouvernance se réserve le choix de ne pas publier une partie des travaux du Datalab ou bien de le publier sous une autre licence.

¹ Une œuvre collective (art. L.113-2 al. 3 du CPI) est une œuvre créée à l'initiative d'une personne physique ou morale qui la divulgue sous son nom et à laquelle plusieurs auteurs participent. La contribution de chaque auteur se fond dans l'ensemble, sans qu'il soit possible d'attribuer à chacun un droit distinct sur l'ensemble. L'œuvre collective est, sauf preuve contraire, la propriété de la personne sous le nom de laquelle elle est divulguée (art. L.113-2 al. 3 du CPI). Cette personne est investie des droits d'auteur (article L.113-5 du CPI)

Article 5 - Principes éthiques et standards de qualité

Dans tous ses travaux, le Datalab applique des **principes d'intégrité et de transparence** : les pratiques d'analyse et de traitement des données sont menées de manière à éviter tout biais et à préserver l'objectivité des résultats. Une attention particulière est portée à l'éthique des données, en garantissant que leur usage respecte les personnes concernées et en limitant les traitements aux seules informations publiques ou explicitement autorisées par la loi.

La rigueur méthodologique est au cœur des activités du Datalab. Elle est essentielle pour garantir que les résultats obtenus par le Datalab sont non seulement précis et fiables, mais également opérationnels : c'est la raison pour laquelle chaque projet est conduit selon des protocoles bien définis, qui garantissent la reproductibilité des analyses et la validité des résultats.

La qualité des données est un pilier fondamental des standards de méthodologie du Datalab. Les sources de données sont sélectionnées pour leur fiabilité et leur pertinence. Le Datalab s'appuie sur des bases de données publiques, des rapports officiels, et d'autres sources vérifiées pour garantir que les informations utilisées sont exactes et à jour.

La transparence est un principe clé dans les activités du Datalab. Pour assurer la transparence, chaque étape des processus de traitement des données est documentée de manière claire et exhaustive. Cela inclut la description des méthodes de collecte des données, des techniques d'analyse utilisées, et des outils statistiques ou algorithmiques employés. Cette transparence permet aux parties prenantes de comprendre et de vérifier les résultats obtenus. Elle facilite également la collaboration et le partage des connaissances, en permettant à d'autres experts de reproduire et de valider les analyses du Datalab. Cette documentation permet également de vérifier la validité des conclusions et de reproduire les analyses si nécessaire, ce qui est crucial pour les dossiers juridiques où la précision et la fiabilité des preuves sont primordiales.

Sauf décision contraire, les résultats, une fois validés par le Comité de gouvernance et les instances dirigeantes d'Anticor, sont publiés selon les standards de communication d'Anticor, et les méthodes employées sont partagées avec les parties prenantes pour faciliter la compréhension et l'utilisation des résultats par d'autres acteurs de la lutte contre la corruption.

Article 6 - Collaboration et ouverture aux communautés scientifiques

Le Datalab d'Anticor encourage activement la collaboration avec des universités, des centres de recherche et des experts en science des données. Ces projets doivent être validés par le Conseil d'administration d'Anticor et peuvent prendre plusieurs formes :

- Projets de Recherche Conjoint : Le Datalab et les institutions partenaires peuvent collaborer sur des projets de recherche appliquée, en combinant les compétences académiques et les besoins pratiques d'Anticor. Ces projets peuvent inclure des études sur les mécanismes de corruption, des analyses de données publiques, et des évaluations d'impact des politiques publiques.

- Stages et Formations : le Datalab peut accueillir des étudiants et des chercheurs en stage, leur offrant une expérience pratique et leur permettant de contribuer aux projets en cours. En retour, ces stagiaires apportent des idées nouvelles et des compétences techniques avancées.
- Conférences et Séminaires : Le Datalab organise et participe à des conférences et des séminaires académiques, où les chercheurs peuvent présenter leurs travaux et échanger des idées. Ces événements favorisent le partage des connaissances et la diffusion des résultats de recherche.

Article 7 - Collaboration et ouverture aux communautés d'experts en science des données

Le Datalab vise à créer des liens avec les communautés des *data scientists*, facilitant leur participation par le biais de contributions externes et d'initiatives de *crowdsourcing*. Cette approche collaborative permet d'enrichir les méthodes d'analyse et d'encourager une participation citoyenne dans la lutte contre la corruption :

- Contributions externes : Le Datalab encourage les *data scientists* à contribuer à ses projets en partageant leurs compétences et leurs connaissances. Ces contributions peuvent inclure le développement de nouveaux algorithmes, l'amélioration des méthodes d'analyse, et la création de visualisations interactives.
- Initiatives de *crowdsourcing* : Le Datalab peut lancer des initiatives de *crowdsourcing* pour mobiliser la communauté des *data scientists* et des citoyens intéressés par la lutte contre la corruption. Ces initiatives peuvent inclure des concours de data science, des hackathons, et des projets collaboratifs en ligne. Par exemple, un hackathon peut être organisé pour développer des outils de veille et d'alerte, ou pour analyser des ensembles de données spécifiques.
- Plateformes de collaboration : Le Datalab utilise des plateformes de collaboration en ligne, telles que GitHub, pour faciliter la participation des *data scientists*. Ces plateformes permettent de partager des codes, des données, et des résultats d'analyse, et de collaborer en temps réel sur des projets communs.

Article 8 - Suivi, évaluation et amélioration continue

Le Datalab d'Anticor adopte une approche proactive de suivi et d'évaluation. Des indicateurs de performance sont établis pour mesurer l'impact des projets sur les activités d'Anticor et identifier les axes d'amélioration. Les projets sont régulièrement évalués pour vérifier leur alignement avec les objectifs initiaux et ajuster les stratégies si nécessaire.

Le Datalab recueille aussi des retours d'expérience auprès de ses utilisateurs et partenaires, afin d'adapter et d'enrichir les outils développés en fonction des besoins exprimés. Cette approche d'amélioration continue permet au Datalab de rester à la pointe de l'innovation et de la pertinence dans la lutte contre la corruption.

Article 9 - Diffusion et Communication

La diffusion des résultats obtenus par le Datalab fait partie intégrante de sa mission. Un plan de communication, validé par les instances dirigeantes d'Anticor, assure la promotion des activités du Datalab et permet de rendre ses résultats accessibles à un large public. La transparence est essentielle, et chaque rapport ou étude est publié dans un format compréhensible, avec des explications contextuelles pour le grand public.

Le Datalab coordonne également ses actions avec les médias pour garantir que les informations diffusées reflètent fidèlement les conclusions et les analyses produites. L'objectif est d'informer le public, de sensibiliser les citoyens et de rendre compte des activités d'Anticor de manière claire et accessible.

En outre, le Datalab adopte une politique de publication scientifique rigoureuse. Les résultats des recherches sont soumis à des revues académiques et des conférences spécialisées pour garantir leur validité et leur reconnaissance par la communauté scientifique. Les publications scientifiques sont réalisées en respectant les normes éthiques et les standards de qualité, afin de contribuer à l'avancement des connaissances dans le domaine de la lutte contre la corruption. Cette approche permet également de renforcer la crédibilité des analyses et des conclusions du Datalab, en les soumettant à l'évaluation par les pairs.

Article 10 - Mise à jour de la charte

La charte du Datalab Anticor sera révisée périodiquement pour assurer son adéquation aux évolutions technologiques, légales, et méthodologiques. Une révision biannuelle est prévue pour ajuster les pratiques du Datalab en fonction des retours d'expérience et des nouvelles normes de la science des données.

Les modifications sont soumises au comité de gouvernance et aux instances dirigeantes d'Anticor pour validation, garantissant ainsi une continuité dans les standards d'éthique, de qualité, et de transparence adoptés par le Datalab.

Annexe 1 - Types de projets et cas d'usage

Le Datalab coordonne divers types de projets dans le but d'aider Anticor à réaliser ses missions. Ces projets sont conçus pour répondre à des besoins spécifiques et pour maximiser l'impact des actions de l'association dans la lutte contre la corruption, de l'analyse de données ouvertes à la visualisation interactive, en passant par la conception d'outils de veille et d'alerte, et le soutien aux actions contentieuses. Chacun de ces projets est conçu pour contribuer à la lutte contre la corruption de manière efficace et rigoureuse.

i. Analyse de Données Ouvertes

L'analyse de données ouvertes constitue l'une des activités principales du Datalab. L'objectif est de détecter des signaux de manquements à la probité dans des domaines clés tels que la gestion des collectivités publiques, l'administration publique et la commande publique. Pour ce faire, le Datalab utilise des techniques avancées de traitement et d'analyse des données, telles que l'apprentissage automatique, l'analyse statistique, et la visualisation des données. Ces analyses permettent d'identifier des anomalies, des irrégularités, et des comportements suspects qui pourraient orienter Anticor vers des irrégularités.

Par exemple, le Datalab peut analyser des dépenses publiques pour détecter des dépenses injustifiées ou des transactions suspectes. En utilisant des algorithmes de détection d'anomalies, le Datalab peut identifier des dépenses qui dépassent les seuils habituels ou qui ne correspondent pas aux budgets alloués. De même, l'analyse des marchés publics est susceptible de révéler des pratiques de favoritisme ou des conflits d'intérêts. En examinant les données des appels d'offres et des attributions de contrats, le Datalab peut détecter des schémas récurrents, tels que l'attribution répétée de contrats à la même entreprise et constituer automatiquement un dossier qui sera soumis à l'expertise juridique d'Anticor, puis le cas échéant, à l'appréciation du conseil d'administration.

ii. Projets de Visualisation Interactive

Des projets de visualisation interactive sont également développés afin de rendre les informations compréhensibles et accessibles au grand public. La visualisation des données est un outil pertinent pour communiquer des informations complexes de manière claire et engageante. Le Datalab crée des tableaux de bord interactifs, des cartes géographiques, et des graphiques dynamiques qui permettent aux utilisateurs d'explorer les données de manière intuitive. Ces visualisations peuvent être utilisées pour illustrer des tendances, des anomalies, et des résultats d'analyses, facilitant ainsi la compréhension des enjeux liés à la corruption.

Par exemple, une carte interactive des marchés publics peut montrer la répartition géographique des contrats attribués, permettant aux utilisateurs de voir quelles régions reçoivent le plus de contrats et quelles entreprises sont les plus actives. Un tableau de bord peut afficher les dépenses publiques par catégorie et par période, permettant aux utilisateurs de suivre l'évolution des dépenses au fil du temps et d'identifier des anomalies ou des tendances inquiétantes. Ces visualisations interactives sont conçues pour être accessibles à

un large public, y compris les citoyens, les journalistes et les décideurs politiques, afin de sensibiliser et d'informer sur les enjeux de la corruption.

iii. Outils de Veille et d'Alerte

En complément, le Datalab conçoit des outils de veille et d'alerte, qui surveillent en temps réel les données publiques pour identifier d'éventuelles anomalies ou tendances inquiétantes. Ces outils utilisent des algorithmes de détection d'anomalies et des systèmes d'alerte automatisés pour signaler des comportements suspects ou des irrégularités. Par exemple, un outil de veille peut surveiller les transactions financières en temps réel et envoyer des alertes en cas de dépenses anormales ou de transactions suspectes. Cet outil peut être configuré pour détecter des transactions qui dépassent certains seuils ou qui ne correspondent pas aux budgets alloués. De même, un outil d'alerte peut surveiller les marchés publics et signaler des attributions suspectes de contrats. Cet outil peut analyser les données des appels d'offres et des attributions de contrats pour détecter des schémas récurrents ou des irrégularités dans le processus d'attribution. Ces outils de veille et d'alerte permettent de réagir rapidement aux signaux de corruption et de prendre des mesures préventives ou correctives, renforçant ainsi la capacité d'Anticor à détecter et à prévenir les pratiques de corruption.