TECH SPRINT 2022
Mutualisation Confidentielle des Données
Appel à candidatures

16 mai 2022
Pôle Fintech-innovation

Data limite de candidature : 3 juin 2022
Envoi des candidatures : techsprint2022@acpr.banque-france.fr

1. Contexte, objectifs et principes du Tech Sprint

Contexte et objectifs

La lutte contre le blanchiment et le financement du terrorisme (LCB-FT) est un domaine d'exploration privilégié[1] pour les nouvelles technologies. La charge administrative des processus en cours, le nombre de « fausses alertes » déclenchées par les procédures actuelles, tout pousse les acteurs à chercher des solutions plus efficaces que les nouvelles technologies rendraient possibles.

Dans ce contexte, l'un des thèmes récurrents évoqués par les acteurs de la place est celui du partage ou de mise en commun d'informations (incluant les données de transactions ou concernant les clients). Toutefois, les projets basés sur le partage d'information peinent à se concrétiser.

Pourtant, et si l'on restreint le champ à la seule question de l'analyse avancée de données (ou de l'IA) appliquée à la détection des transactions suspectes, le partage ou la mutualisation d'information pourrait améliorer sensiblement les résultats obtenus actuellement dans le secteur. Le GAFI a consacré un rapport à ce thème – dont il a présenté les conclusions lors d'un webinaire ACPR-Telecom Paris[2] le 8 mars 2021. Il y affirme sa conviction que la mutualisation de données et l'utilisation de techniques d'analyse collaboratives permettrait d'améliorer l'efficacité de la LCB-FT[3].

Aussi l'ACPR a-t-elle décidé de lancer **une expérimentation** dont l'objectif est de prouver – ou d'infirmer – l'hypothèse selon laquelle la mutualisation de données permet d'améliorer la performance des systèmes de surveillance de transactions. Elle en a présenté les grandes lignes lors d'une réunion publique le 30 mars dernier. Une vue d'ensemble de ce projet d'expérimentation est donnée en annexe.

Dans le cadre de cette expérimentation, les questions soulevées par les exigences associées à l'objectif poursuivi amènent l'ACPR à organiser un événement, **le Tech Sprint ACPR 2022 sur la mutualisation confidentielle des données (Tech Sprint MCD)**, afin d'**évaluer les méthodes et techniques de maintien de la confidentialité et de l'intégrité de données mutualisées**. Le présent appel à candidatures s'adresse aux fournisseurs de technologie s'estimant en capacité de participer à cet événement.

Principes du Tech Sprint

Les participants au Tech Sprint seront invités à **concevoir et à mettre en œuvre une preuve de concept** de leur solution de mutualisation confidentielle des données, en utilisant des données et un scénario fictifs fournis par l'ACPR.

Les résultats de leurs travaux seront présentés en public le 13 septembre 2022, devant un panel d'experts techniques composé de représentants de l'ACPR et de la Banque de France ainsi que

---

1 Le thème avait d'ailleurs été sélectionné, avec la protection de la clientèle et la modélisation de risques, pour les ateliers IA menés par l'ACPR en 2019.

2 Les lundis de l'IA et de la finance avec l'ACPR : le partage de données pour l'IA en finance (telecom-paris.fr)

3 *STOCKTAKE ON DATA POOLING, COLLABORATIVE ANALYTICS AND DATA PROTECTION*, GAFI 2021. *"By pooling data and using collaborative analytics, financial institutions can better understand, assess, and mitigate money laundering and terrorist financing risks. This will result in a more dynamic, effective and efficient identification of these activities, and help the private sector comply with anti-money laundering and counter terrorist financing requirements in a timelier and less burdensome manner."*

d'établissements participant à l'expérimentation dans le cadre de laquelle s'inscrit le Tech Sprint. Cette présentation pourra prendre la forme d'une démonstration de la preuve de concept, accompagnée d'un exposé oral. Elle sera suivie d'une séance de questions-réponses.

Il sera en outre demandé aux participants :

- De fournir une documentation des travaux réalisés pour la preuve de concept **avant le 5 septembre**.
- De **répondre à toute question approfondie** posée par le panel d'experts techniques, après la restitution proprement dite, **entre le 13 et le 20 septembre**.

La participation au Tech Sprint est **gratuite et n'est pas rémunérée**.

Le Tech Sprint ACPR vise également à **faciliter la mise en contact des fournisseurs de technologie et des établissements participant à l'expérimentation**, en vue de l'exécution des protocoles expérimentaux sur données réelles prévue dans les étapes 4 à 6 du projet d'expérimentation (voir la vue d'ensemble de l'expérimentation en annexe). Pour autant, **la participation au Tech Sprint ne préjuge pas du choix final des établissements** de leur partenaire technologique pour ces phases ultérieures de l'expérimentation.

Les **candidats retenus** pour participer au Tech Sprint **s'abstiendront de se prévaloir d'une quelconque approbation de l'ACPR sur les technologies étudiées**, et plus généralement sur leur activité.

L'ACPR envisage de **restituer au marché**, sous forme anonymisée, les principaux enseignements qu'elle aura tirés de l'expérimentation, et notamment du Tech Sprint MCD. Elle se réserve également la possibilité de communiquer sur l'identité des organismes ayant participé à l'une des phases de l'expérimentation, et notamment le Tech Sprint MCD.

2. Description détaillée du Tech Sprint MCD

Calendrier

Le calendrier provisoire (sujet à d'éventuelles modifications marginales) du Tech Sprint MCD est le suivant :

- 16 mai : **publication de l'appel à candidatures.**
- 16 mai - 3 juin : **réponse à l'appel à candidatures.** Chaque participant intéressé fournit une description de sa solution en utilisant le modèle fourni dans le présent appel à candidatures (section 3).
- 3 - 13 juin : **processus de sélection des fournisseurs de technologie.** L'ACPR sélectionne les fournisseurs qui seront invités à participer au Tech Sprint MCD.
- 13 juin : **communication du cahier des charges de la preuve de concept.** L'ACPR envoie à chaque fournisseur sélectionné le cahier des charges incluant les exigences fonctionnelles et techniques pour la réalisation de la preuve de concept (incluant des liens de téléchargement des données fictives à utiliser).
- 13 juin - 13 sept. : **mise en œuvre de la preuve de concept.** Chaque fournisseur présélectionné réalise la preuve de concept durant ce délai de 3 mois.
    - 27 juin : **séance de questions-réponses sur le cahier des charges du PoC**. Tous les fournisseurs présélectionnés sont invités à une session de *briefing* organisée par l'ACPR et visant à répondre aux questions demeurant suite à leur lecture et analyse du cahier des charges.
    - 5 sept. : **documentation des travaux réalisés au PoC.** Chaque fournisseur sélectionné doit à cette date avoir communiqué à l'ACPR la documentation de ses travaux (sur la base d'un questionnaire qui lui sera fourni). Elle sera en particulier utilisée par le panel d'experts techniques (composé de représentants d'établissements participant à l'expérimentation d'ensemble, de l'ACPR et de la Banque de France) pour préparer l'appréciation des travaux lors de l'événement de restitution.
- 13 sept. : **événement de restitution des travaux réalisés pour la preuve de concept.** Chaque fournisseur présente sa solution et son implémentation de la preuve de concept, en incluant toute forme de preuve possible des arguments avancés, dans un délai imparti de 20 minutes, suivies de 10 minutes de questions-réponses.
  L'audience de cette restitution inclura les établissements participant à l'expérimentation, des représentants de l'ACPR et de la Banque de France, ainsi que des invités extérieurs issus du secteur financier, du monde académique, d'institutions publiques, ou de centres d'expertise technique. En outre, un panel d'experts techniques (composé de représentants d'établissements participants, de l'ACPR et de la Banque de France) procédera à une évaluation rigoureuse des capacités et des limitations techniques de chaque solution proposée.
- 13 - 20 sept. : **suivi de la restitution.** Chaque fournisseur ayant réalisé la preuve de concept se tient disponible durant cette période d'une semaine pour répondre aux questions des établissements participant à l'expérimentation et de l'ACPR concernant leurs travaux et plus généralement leur solution et les technologies mises en œuvre.

Le calendrier des **suites du Tech Sprint** est provisoirement :

- 13 - 30 sept. : choix des fournisseurs. Chaque établissement participant sera invité à choisir un fournisseur de technologies parmi ceux ayant concouru au Tech Sprint pour la mise en œuvre

des protocoles expérimentaux prévus par l'expérimentation (voir annexe). Chaque établissement participant reste toutefois libre de son choix et ne saurait être tenu d'exposer les motifs de sa décision – favorable ou défavorable – envers un candidat.

- 30 sept. - Q1 2023 : validation et mise en œuvre des protocoles expérimentaux. Les fournisseurs technologiques ayant conclu des partenariats avec les établissements participent avec eux aux étapes suivantes de l'expérimentation.
- Rapport sur le Tech Sprint MCD : un retour d'expérience basé sur les travaux du panel d'experts sera publié par l'ACPR, soit comme partie d'un rapport général sur l'expérimentation, soit sous la forme d'un rapport distinct.

Champ des technologies concernées

Le terme « mutualisation confidentielle de données » (MCD) vise à décrire aussi précisément que possible le type de solution technique visé par l'expérimentation. Le but principal est de **permettre le stockage de multiples jeux de données à caractère confidentiel, leur croisement, leur requêtage, et leur exploitation par des modèles prédictifs.** Dans le scénario d'application concerné par l'expérimentation, chaque jeu de données contient des données de transactions provenant d'un établissement particulier, bien que les technologies MCD pressenties soient génériquement applicables à tout type de scénario.

Le spectre des technologies MCD devrait s'avérer largement commun avec celui, d'acception récente mais consacré par les milieux industriels et académiques, des PET (*Privacy Enhancing Technologies*). Néanmoins un terme dédié est ici employé afin de ne pas limiter le champ de techniques pouvant être proposées par les candidats au Tech Sprint. **L'objectif recherché est en effet purement fonctionnel : satisfaire aux exigences de confidentialité**, de *privacy* et d'intégrité des parties prenantes à l'expérimentation et, en premier lieu, des établissements participants qui utiliseront des données commercialement sensibles et potentiellement des données personnelles.

Les solutions MCD considérées *a priori* comme pertinentes incluent la majorité des approches de type PET : confidentialité différentielle (*Differential Privacy*), chiffrement homomorphe (*Homomorphic Encryption*), calculs multipartites sécurisés (*Secure Multi-Party Computation*), enclaves matérielles (*Trusted Execution Environments*), ou toute combinaison de ces approches.

Sélection des candidats au Tech Sprint MCD

Toute société intéressée (qu'elle soit juridiquement rattachée à un établissement financier ou non) peut soumettre sa candidature avant le 3 juin au moyen du modèle fourni en section 3 de ce document.

Le tableau suivant résume les fonctionnalités, respectivement obligatoires et optionnelles, demandées aux solutions des fournisseurs de technologie MCD pour la présélection.

Type de fonctionnalité	Fonctionnalités requises	Fonctionnalités optionnelles
Mutualisation confidentielle	Mise en commun ou partage automatisé avec garanties de confidentialité	Garanties d'intégrité tout au long du processus de mutualisation
Mutualisation confidentielle	API (incluant un support SQL) pour exécuter les requêtes de mutualisation	Support de langages de requêtage autres que SQL
Mutualisation confidentielle	Traçage et remontée des erreurs (pour déboguer les requêtes sur les données chiffrées ou enclavées)	Extraction d'un échantillon de données mutualisées (pour validation visuelle des résultats)
Intégration avec les modèles de détection LCB-FT	API d'intégration de modèle dans au moins un langage (tel que Python) ou protocole standard (tel que REST)	Implémentation de certains types de modèles de détection (à base de ML ou pas, de type graphe ou pas) sans avoir à intégrer un modèle externe
Intégration avec les modèles de détection LCB-FT		Apprentissage de modèles de ML sur données mutualisées
Intégration avec les modèles de détection LCB-FT	Inférence de modèle sur données mutualisées	
Intégration avec les modèles de détection LCB-FT	Évaluation des performances d'un modèle (via des métriques standard ou ad-hoc)	
Hébergement	Hébergement par le fournisseur de solution (on-premise, sur un cloud privé, ou sur un cloud public)	Autre option d'hébergement (dans le data center d'un établissement participant, ou d'un tiers de confiance)

L'ACPR examinera toutes les candidatures et sélectionnera une liste de fournisseurs de technologie candidats. Ce processus tiendra notamment compte des fonctionnalités requises et optionnelles remplies, de la capacité estimée de chaque fournisseur à réaliser l'expérimentation proposée, et d'autres facteurs incluant de façon non exhaustive : le degré d'innovation technique et scientifique de la solution proposée, son niveau de maturité et de robustesse, la preuve de valeur ajoutée découlant des déploiements déjà réalisés en production, le coût de la solution, etc.

L'ACPR n'est pas tenue de motiver ses décisions de retenir ou de ne pas retenir telle ou telle candidature.

Cahier des charges de la preuve de concept

Un document précisant le cahier des charges de la preuve de concept à réaliser pour le Tech Sprint MCD sera envoyé au plus tard le 13 juin à chaque fournisseur de technologie sélectionné.

Ce cahier des charges consistera principalement à demander une « preuve de concept » en implémentant une solution selon un scénario prédéfini utilisant des données fictives. L'un des buts de la preuve de concept est de convaincre un ou plusieurs établissements participants que la solution proposée peut, moyennant une quantité modérée de développements et ajustements supplémentaires (c'est-à-dire essentiellement l'intégration au SI des établissements selon l'étape 4), remplir les besoins de l'exécution du protocole expérimental retenu (c'est-à-dire l'étape 5, exécutée sur données réelles). Il s'agit donc littéralement de **fournir – en complément de la démonstration de la solution implémentée – une preuve**, lors de l'événement de restitution programmé le 13 septembre, que la solution opère selon les caractéristiques indiquées et satisfait aux garanties (de confidentialité, de performance, etc.) mises en avant par le fournisseur.

Les fournisseurs MCD participant au Tech Sprint disposeront de 3 mois (3 juin - 3 sept.) pour implémenter la preuve de concept. La définition du périmètre et du cahier des charges visera à éviter les tâches généralement les plus consommatrices en temps et en ressources pour les fournisseurs de technologie de ce type :

- Définition du cas d'usage : l'ACPR décrira un scénario hypothétique simplifié, qui permettra néanmoins de tester les fonctionnalité requises ou optionnelles.
- *Sourcing* des données : l'ACPR fournira les données d'entrée de la preuve de concept sous forme de jeux de données fictives.
- Déploiement et hébergement : afin d'éviter le processus long et ardu (aux plans juridique, technique et organisationnel) de déployer leur solution au sein de l'environnement informatique d'un établissement ou même dans le *data center* d'un tiers de confiance, les participants au Tech Sprint seront libres du choix de la cible de déploiement. On peut anticiper que certains déploieront et hébergeront leur implémentation de la preuve de concept dans leur propre *data center*, tandis que d'autres emploieront leur méthode de déploiement standard sur un cloud public (probablement sans contrainte additionnelle de sécurité car les données et modèles utilisés pour la preuve de concept n'auront aucun caractère sensible).

3. Candidature au Tech Sprint

Processus de candidature

Cette section précise les modalités permettant aux fournisseurs de solution MCD intéressés de soumettre leur candidature à l'ACPR, **au plus tard le 3 juin**.

Chaque fournisseur intéressé doit inclure dans sa candidature les coordonnées (nom, prénom, adresse email) d'un point de contact principal, en complément du modèle décrit ci-après rempli de façon aussi détaillée que possible.

Modèle de candidature

Ce modèle est composé d'un ensemble de questions et sous-questions, chacune devant faire l'objet d'une réponse en texte libre, aussi informative que possible. Le cas échéant, il pourra être demandé au fournisseur de technologie candidat de démontrer tout élément ou argumentaire inclus dans ses réponses.

Chaque réponse devra prendre en considération le format de l'expérimentation dans son ensemble. Outre la solution actuellement proposée par le fournisseur, les réponses devront ainsi préciser les réalisations possibles :

- d'une part, pour la preuve de concept du Tech Sprint (en supposant un ensemble minimal d'ajustements de la solution existante, tout en restant dans le délai de 3 mois imparti pour la preuve de concept) ;
- d'autre part, pour une éventuelle expérimentation sur données réelles (qui introduira par rapport à la preuve de concept des exigences supplémentaires de passage à l'échelle, de confidentialité, et d'intégration de systèmes).

Le modèle de candidature est scindé en trois sections : caractéristiques fonctionnelles de la solution, caractéristiques techniques, et enfin le coût et la planification.

A. Caractéristiques fonctionnelles de la solution proposée

1) **Données d'entrée.** Indiquer toute contrainte imposée par votre solution :
   a) sur les types de données (catégorielles, numériques, textes courts) qui peuvent être traitées ;
   b) sur les schémas de données (par exemple support de tout modèle de base de données relationnelles ou limitation à un certain nombre de tables) ;
   c) sur les volumes de données.

2) **Capacités analytiques.** Pour chacune des fonctionnalités suivantes, indiquer si elle est proposée par la solution. Si oui, préciser le mode opératoire correspondant (c'est-à-dire quelles tâches manuelles ou automatiques sont nécessaires en comparaison du mode opératoire basique opérant sur les données en clair). Fournir aussi toute métrique possible sur la performance opérationnelle (c'est-à-dire comparer l'expérience utilisateur utilisant la solution à celle d'un traitement de données en clair).
   a) Exécution de requêtes SQL
      i) Chargement des données d'entrée
      ii) Jointure de données d'entrée de plusieurs établissements
      iii) Extraction d'un échantillon de données d'entrée (avant et après jointure)
   b) Implémentation de modèles de détection
      i) Création et paramétrage d'un modèle à base de règles sur données tabulaires

ii) Création et paramétrage d'un modèle à base de règles sur données en réseau
iii) Création et apprentissage d'un modèle à base de ML sur données tabulaires
iv) Création et apprentissage d'un modèle à base de ML sur données en réseau

c) Intégration de modèles de ML tierce-partie
   i) Chargement d'un modèle (non entraîné) dans le Périmètre de Confidentialité
   ii) Entraînement d'un modèle au sein du Périmètre de Confidentialité
   iii) Inférence d'un modèle sur un batch de données au sein du Périmètre de Confidentialité
   iv) Post-inférence, calcul et récupération de métriques de performance (*a minima* des statistiques descriptives)
   v) Post-inférence, extraction d'un échantillon de données (post-jointure) accompagnées de leurs prédictions
   vi) Téléchargement du modèle une fois entraîné
   vii) Quelles classes de modèles de ML sont supportées ? Quelles autres contraintes doivent être prises en compte sur les modèles ?

d) Autres types d'analyse (outre les analyses mentionnées ci-dessus)
   i) Quels types d'analyse sont disponibles ?
   ii) Dans quels langages de programmation et formats ?

3) Garanties de confidentialité. Quelles garanties démontrables de confidentialité (et plus spécifiquement de *privacy*) sont offertes par la solution sur les éléments sensibles suivants :

a) Sur les jeux de données chargés dans le Périmètre de Confidentialité, puis mutualisés avec d'autres jeux de données (en utilisant des jointures SQL ou tout autre Mécanisme de Mutualisation) et potentiellement fournis comme données d'apprentissage ou d'évaluation à un modèle de ML ? Par exemple :
   i) La solution permet-t-elle de générer des données synthétiques avec des garanties de Confidentialité Différentielle, afin de pouvoir entraîner un modèle de ML hors du Périmètre de Confidentialité ?

b) Sur les résultats de requêtes SQL ? Par exemple :
   i) Le taux de requêtes peut-il être limité ? Le type de résultats peut-il être restreint afin de limiter les requêtes au calcul de valeurs agrégées telles que des statistiques descriptives ?
   ii) Ces garanties sont-elles obtenues par Confidentialité Différentielle sur le jeu de données résultat ou par une autre méthode ?

c) Sur un modèle de ML une fois entraîné dans la solution ? Par exemple :
   i) La solution fournit-elle des garanties de Confidentialité Différentielle sur les paramètres d'un modèle de ML (c'est-à-dire les poids dans le cas d'un modèle de type régression linéaire) ?
   ii) Permet-elle de calculer des métriques de performance prédictive, et si oui lesquelles (taux de précision, taux de rappel, etc.) ?
   iii) Permet-elle de télécharger un échantillon de données de test accompagnées de leurs prédictions ?

4) **Modèle de menace.** Décrire aussi précisément que possible le modèle de menace (et d'attaques) dans lequel opère la solution.

5) Garanties d'intégrité. Quelles garanties d'intégrité sont offertes par la solution (afin de prouver qu'un code source donné a été exécuté, que les données chargées n'ont pas été altérées, etc.) vis-à-vis des éléments sensibles suivants ?

a) Sur les données hébergées au sein du Périmètre de Confidentialité (jeux de données d'entrée, données mutualisées) ?
b) Sur les requêtes SQL exécutées au sein du Périmètre de Confidentialité ?
c) Sur le code source utilisé pour créer puis entraîner un modèle de ML ?

B. Caractéristiques techniques de la solution proposée

6) **Hébergement de la solution.** Préciser toutes contraintes associées :
   a) à l'environnement d'hébergement (uniquement *on-premise* ou non, sur certains types de *cloud* uniquement, sur quel(s) OS hôtes, avec conteneurisation optionnelle ou obligatoire, avec orchestration optionnelle ou obligatoire, etc.)
   b) à l'espace de stockage (limites sur les données sources, *overhead* éventuel, etc.)
   c) au *monitoring* de l'environnement (possibilités de traçage *live*, d'audit ultérieur, etc.).

7) Intégration. Préciser les possibilités d'intégration avec un TMS (système de surveillance de transaction) cible :
- Quels types de TMS sont supportés par la solution actuelle ?
- Quels types supplémentaires de TMS peuvent être supportés (dans les limites de temps et de ressources imposées par l'expérimentation ACPR) ?

8) Type de technologie. Quel(s) type(s) de technologie sont employés par la solution?
   a) HE/FHE (chiffrement homomorphe)
   b) SMPC (calculs multipartites sécurisés)
   c) DP (confidentialité différentielle)
   d) TEE (enclave matérielle) et si oui, sur quelle(s) plateforme(s)
   e) ZKP (*Zero-Knowledge Proof*)
   f) FL (apprentissage fédéré)
   g) Une combinaison de certaines de ces méthodes
   h) Une autre méthode.

9) Implémentation. La solution est-elle implémentée :
   a) en code ouvert (*open source*) ou fermé ?
   b) sur la base de technologie brevetée ou d'algorithmes du domaine public ?
   c) en utilisant quels algorithmes essentiels ?

C. Coût et planification de la solution proposée

10) **Calendrier.** Quel est le calendrier prévisionnel pour chacune des phases suivantes ?
   a) Implémentation de la preuve de concept pour le Tech Sprint (c'est-à-dire utilisation de la solution pour exécuter un scénario simplifié sur données fictives, le tout fourni par l'ACPR comme décrit en section 2)
   b) Expérimentation éventuelle sur données réelles (c'est-à-dire intégration de la solution dans le SI d'un établissement associé, puis son déploiement, enfin l'exécution du protocole expérimental sur les données fournies par cet établissement).

11) **Coût financier.** Quelle est votre première estimation du coût financier de l'adaptation de la solution à une éventuelle expérimentation sur données réelles ? (Note : la participation au Tech Sprint MCD, incluant l'implémentation de la preuve de concept, sera effectuée à titre gracieux par les fournisseurs de technologie intéressés).

12) **Ressources des partenaires**. Quelles ressources (techniques, métier, logistiques ou autres) sont attendues :
    a) De la part des établissements financiers associés dans une équipe de l'expérimentation ?
    b) De la part de l'ACPR en tant que facilitatrice de l'expérimentation ?

13) **Ressources internes.** À titre informatif, préciser quelles ressources seront allouées par le fournisseur de technologie :
    a) Pour l'implémentation de la preuve de concept au Tech Sprint MCD ?
    b) Pour adapter le cas échéant la solution actuelle à l'expérimentation sur données réelles, puis exécuter le protocole expérimental retenu ?

Traitement des données à caractère personnel

L'ACPR gère la liste des candidats au Tech Sprint dont la finalité est l'organisation du Tech Sprint objet du présent appel à candidature. Ce traitement est basé sur l'intérêt légitime et se conforme aux dispositions légales et réglementaires : la loi n° 78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés, ainsi que le Règlement Général sur la Protection des Données (Règlement UE 2016/679 du 27 avril 2016).

Les renseignements qui vous sont demandés à l'occasion de la candidature sont exclusivement réservés au traitement de la liste des candidats pour l'organisation du Tech Sprint et destinés à l'administration de l'ACPR. Dans ce cadre, elle collecte des données personnelles (nom, prénom du ou des représentants des sociétés candidates, adresse électronique). Ces données sont conservées pendant 2 ans maximum.

Seuls les agents du Pôle Fintech-innovation ainsi que les services de contrôle interne et d'audit interne ont accès aux informations vous concernant.

Vous disposez d'un exercice des droits : accès, rectification, effacement, opposition, que vous pouvez exercer auprès de l'ACPR par mel à l'adresse fintech-innovation@acpr.banque-france.fr.

Vous avez la possibilité de déposer une réclamation auprès de la CNIL. Les Coordonnées du Délégué à la Protection des Données sont : 1200-DPD-delegue-ut@banque-france.fr.

4. Prochaines étapes pour les fournisseurs intéressés

Cette section décrit, de façon générale, les étapes à suivre pour les fournisseurs de technologie souhaitant répondre à cet appel à candidatures.

- Déterminer, sur la base du présent document, **s'ils sont intéressés par l'expérimentation** et ont les ressources nécessaires pour répondre à l'appel à candidatures.
- Si oui, remplir le **modèle de candidature** (section 3) aussi précisément que possible afin que l'ACPR puisse déterminer s'ils représentent un candidat pertinent et fiable pour l'expérimentation.
- Si le répondant à l'appel à candidatures est sélectionné, participer au Tech Sprint MCD axé autour d'une preuve de concept de sa solution, conclu par un événement de restitution des travaux du Tech Sprint au cours duquel la solution pourra être présentée.

Quant aux **suites éventuelles au Tech Sprint**, les étapes seront les suivantes :

- Suite aux restitutions du Tech Sprint, les établissements participant à l'expérimentation pourront décider de **choisir un fournisseur de technologie**. Fournisseurs technologiques et établissements concernés définiront alors les détails de leur collaboration (accord contractuel, coût des travaux, etc.)
- Le fournisseur de technologie **collaborera avec l'équipe l'ayant intégré** sur la suite de l'expérimentation : finalisation du protocole expérimental et implémentation/adaptation de sa solution à ce protocole, exécution de l'expérimentation sur données réelles, et évaluation des résultats expérimentaux.

5. Glossaire

Apprentissage Fédéré (FL ou *Federated Learning*): une famille possible de techniques de MCD. Voir la page Wikipedia.

Calculs Multipartites Sécurisés (SMPC ou *Secure Multi-Party Computation*): une famille possible de techniques de MCD. Voir la page Wikipedia.

Chiffrement Homomorphe (FHE ou *Fully Homomorphic Encryption*): une famille possible de techniques de MCD. Voir la page Wikipedia.

Confidentialité Différentielle (DP ou *Differential Privacy*): une famille possible de techniques de MCD. Voir la page Wikipedia.

Établissement (financier) : toute entité du secteur financier ayant des obligations en LCB-FT et volontaire pour participer à l'expérimentation ACPR. Ces entités sont attendues majoritairement parmi les catégories établissement de crédit, établissement de paiement, et organisme d'assurance. Différentes catégories d'établissement auront différents cas d'usage à évaluer, sur différents types de données, et définiront donc différents protocoles expérimentaux. Cependant différents établissements au sein d'une même catégorie pourront aussi opter pour divers types de protocoles : par exemple l'étude d'un modèle de détection à base de *Random Forest* (classe de Machine Learning) sur le segment banque de détail dans une équipe, et l'étude d'un modèle à base de règles sur le segment TPE/PME dans une autre équipe d'institutions bancaires.

Enclave matérielle (TEE ou *Trusted Execution Environment*): une famille possible de techniques de MCD. Voir la page Wikipedia.

Machine Learning (ML) ou apprentissage automatique : voir la page Wikipedia**.** Dans le contexte de ce document, tout modèle de détection utilisé en lutte contre le blanchiment ou le financement du terrorisme et relevant d'une des techniques suivantes : régression linéraire ou logistique, *random forest*, *gradient boosted trees*, réseaux de neurones, ou toute combinaison de ces techniques.

Mécanisme de mutualisation : toute méthode (et son implémentation) utilisée par une équipe d'établissements participants pour mutualiser leurs jeux de données respectifs. Afin d'évaluer au mieux le potentiel d'amélioration associé à la mutualisation, l'ACPR encourage les établissements participant à l'expérimentation à explorer les mécanismes de mutualisation les plus sophistiqués. Cela pourra inclure par exemple ceux basés sur un grand nombre de requêtes de SQL, y compris des jointures exactes et du *matching* flou (*fuzzy matching*), à l'opposition de techniques de mutualisation plus traditionnelles (c'est-à-dire l'union de jeux de données au schéma identique, ou le partage d'informations minimalistes tels que des indicateurs binaires de suspicion).

Mutualisation confidentielle de données (MCD) : comme décrit en section 2 du présent document, ce terme (forgé par l'ACPR pour l'expérimentation) vise à décrire toute technologie permettant le stockage, la mutualisation, le requêtage de jeux de données sensibles, ainsi que leur utilisation par des modèles de détection en LCB-FT. « Jeux de données sensibles » sont ici entendus comme tout type de données – mais avant tout transactionnelles – auxquelles les parties prenantes de l'expérimentation ACPR associent des exigences de confidentialité (y compris *privacy* lorsque des données à caractère personnelles y sont incluses) et éventuellement d'intégrité. Le champ des technologies MCD recouvre largement celui des PET (*Privacy Enhancing Technologies*) ; néanmoins un terme dédié, distinct de PET, est employé pour l'expérimentation afin de ne pas restreindre le spectre des techniques qui pourront être proposées par les fournisseurs de technologie candidats au Tech Sprint.

Périmètre de confidentialité : pour les besoins de décrire une solution MCD, ce périmètre se réfère en général à une zone de stockage et d'analyse de données, physique ou logique (donc

d'implémentation matérielle et/ou logicielle), qui fournit les garanties de confidentialité recherchées. Par exemple dans le cas de chiffrement de bout en bout, ce périmètre comprendrait l'ensemble du protocole expérimental (incluant les zones de stockage intermédiaire mais aussi tous les flux des données matérialisés pour ce protocole) en aval du chiffrement initial des données sensibles.

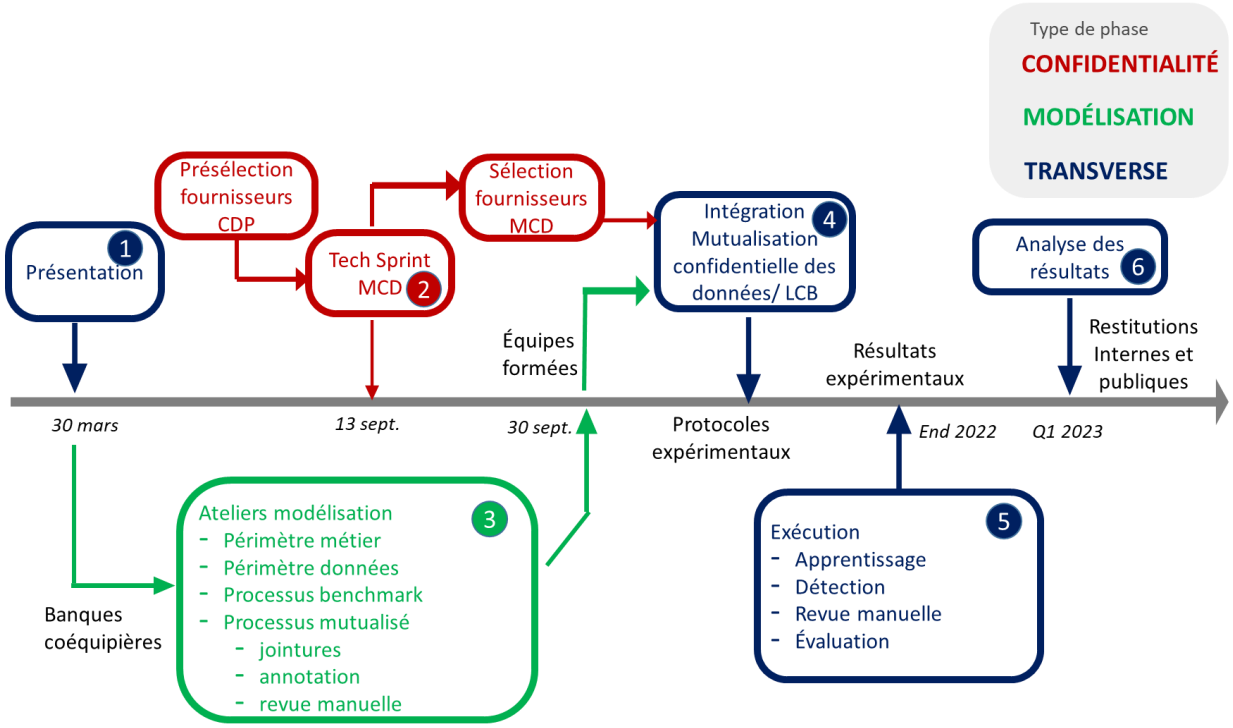
Zero-Knowledge Proof (ZKP) : une famille possible de techniques de MCD. Voir la page Wikipedia.

6. Annexe - Vue d'ensemble de l'expérimentation

Cette section fournit le contexte pertinent à cet appel à candidatures sous la forme d'une vue d'ensemble (déjà présenté lors de la réunion de lancement du 30 mars) de l'expérimentation.

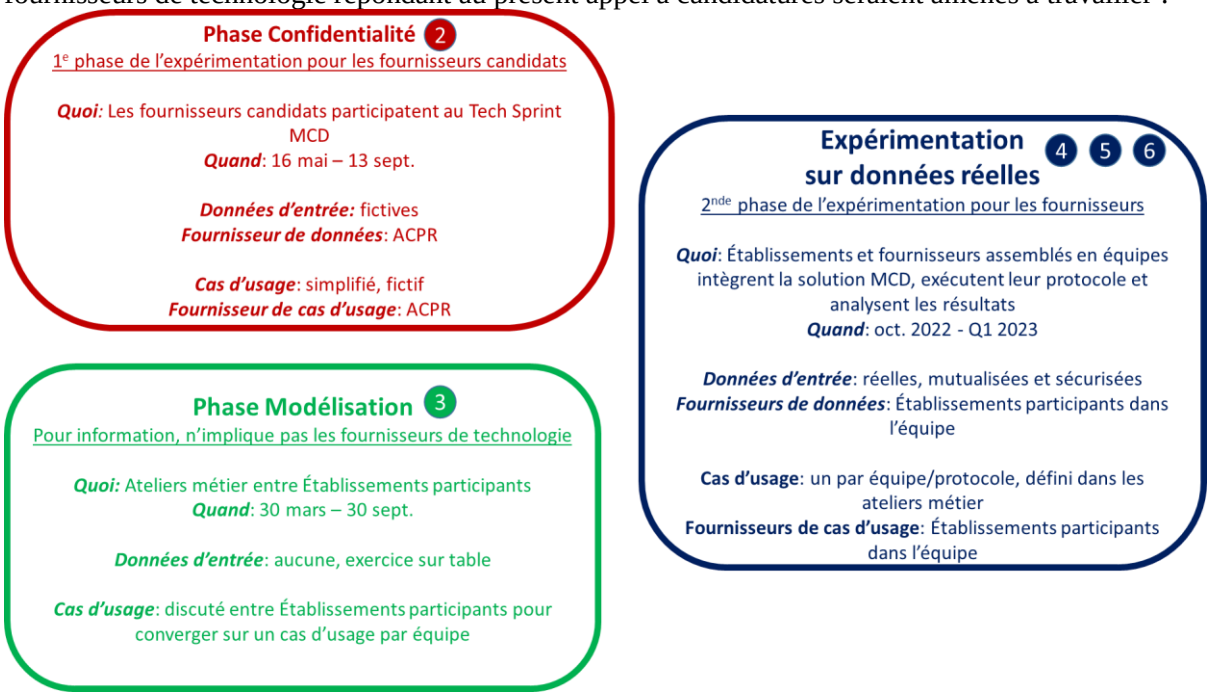
Feuille de route

Le diagramme suivant résume la feuille de route de l'ensemble de l'expérimentation. Chacune des six étapes principales est décrite plus en détail dans la suite de cette section.



Feuille de route des fournisseurs de technologie

Le diagramme suivant résume les scénarios (données d'entrée et cas d'usage) sur lesquels les fournisseurs de technologie répondant au présent appel à candidatures seraient amenés à travailler :



Réunion de présentation (étape 1)

La réunion de présentation a eu lieu le 30 mars, devant une audience composée d'une part d'établissements financiers, d'autre part de fournisseurs de technologie MCD ayant déjà exprimé un intérêt pour le projet et disponibles pour cet événement. L'ACPR, en tant qu'organisatrice, a présenté les objectifs de l'expérimentation, la feuille de route générale, et chacune des six principales étapes avec un niveau de détail modéré. Une séance de questions-réponses a permis de clarifier les points les plus sensibles.

Tech Sprint MCD (étape 2)

Calendrier provisoire : mai-sept. 2022 (13 sept. pour la journée de restitution des travaux)

Participants : tous les fournisseurs de technologie MCD ayant répondu au présent appel à candidatures. Les établissements financiers participant à l'expérimentation assisteront à la restitution et seront représentés dans le panel d'experts techniques afin d'intégrer, le cas échéant, des fournisseurs de technologie dans les équipes mettant en œuvre les protocoles expérimentaux.

Descriptif : l'objectif de cette étape est d'évaluer les méthodes et techniques de maintien de la confidentialité et de l'intégrité de données mutualisées ; cette étape devrait permettre également de susciter des partenariats entre établissements (qui pourraient avoir déjà constitué des équipes de deux ou plus) et fournisseurs de technologie. Les fournisseurs MCD seront ainsi invités à créer une preuve de concept de leur solution, en précisant comment celle-ci peut être adaptée au cadre expérimental proposé.

Ateliers de modélisation LCB-FT (étape 3)

Calendrier provisoire : mars-sept. 2022

Participants: tous les établissements financiers participants, assemblés en équipes de deux ou plus, ainsi que l'ACPR comme facilitatrice.

Descriptif : un programme de travail sur les sujets métier (LCB-FT) de l'expérimentation sera mené en parallèle du programme technologique incluant le Tech Sprint MCD. Le format en sera tout différent : il s'agira d'une série d'ateliers de modélisation visant à préparer l'expérimentation proprement dite. Ces ateliers rassembleront les établissements participants en équipes. Les fournisseurs de technologie n'y prendront aucune part, aussi cette étape n'est-elle pas détaillée dans ce document.

Le but essentiel de cette étape est de définir un « protocole logique »[4] pour l'expérimentation, c'est-à-dire un protocole qui puisse être articulé sur le papier, sans requérir ni travaux d'implémentation ni traitement de données réelles. L'implémentation viendra en effet seulement à compter de l'étape 4, une fois les fournisseurs de technologie choisis par chaque établissement participant.

Intégration des travaux LCB-FT et MCD (étape 4)

Calendrier provisoire : oct.-nov. 2022

Participants : chaque équipe (établissements participants + fournisseurs de technologie MCD le cas échéant) travaillant sur la définition et l'implémentation de son protocole expérimental, l'ACPR facilitant les travaux en fonction des besoins et de ses propres disponibilités.

Descriptif : les programmes de travail LCB-FT et MCD (étapes 2 et 3) ayant convergé, c'est-à-dire une fois que chaque établissement financier participant aura choisi son fournisseur de technologie, cette étape consistera pour chaque équipe à atteindre un double accord. D'une part, chacune définira son mode opératoire pour l'expérimentation : les « protocoles logiques » étant finalisés, leurs conditions matérielles d'exécution seront à préciser (y compris la question de l'hébergement de chaque solution MCD et des données qu'elle traitera), de même que le calendrier du reste de l'expérimentation. D'autre part, un accord juridique entre les membres de chaque équipe sera formalisé dans une convention multipartite spécifiant les responsabilités de chaque partie (en termes de gestion des données et des modèles, de répartition des coûts de l'expérimentation, etc.)

Ici aussi, l'ACPR agira comme facilitatrice de ces deux objectifs : afin d'assurer la cohérence entre les protocoles des différentes équipes et vis-à-vis des objectifs généraux assignés à l'expérimentation, l'Autorité validera les protocoles et les feuilles de temps proposés.

Enfin, cette étape aboutira au développement de la solution (logicielle, matérielle, ou hybride) sur laquelle chaque protocole expérimental sera exécuté – en d'autres termes le « protocole physique » correspondant au « protocole logique » défini à l'étape 3. Ces travaux d'implémentation devraient inclure d'une part les composants LCB-FT issus des ateliers de modélisation, d'autre part le code d'intégration entre la solution MCD et ces composants.

Exécution des protocoles expérimentaux (étape 5)

Calendrier provisoire : déc. 2022

Participants : chaque équipe (établissements participants + fournisseurs de technologie MCD sélectionnés) exécutant son protocole expérimental, l'ACPR facilitant les travaux en fonction des besoins et de ses propres disponibilités.

4 En employant une analogie avec les modèles de données logiques en bases de données relationnelles.

Descriptif. Cette étape suivra la finalisation des protocoles expérimentaux. Chaque équipe exécutera son protocole, ce qui inclura typiquement les phases suivantes :

- mutualisation des données selon le mécanisme retenu ;
- Machine Learning dans le cas des modèles de détection à base de ML ;
- inférence (en mode batch) en utilisant d'une part le processus de benchmark, d'autre part le processus sur données mutualisées ;
- revue manuelle d'un échantillon de sorties produites par le processus sur données mutualisées ;
- calcul des métriques d'évaluation du protocole.

Analyse des résultats (étape 6)

Calendrier provisoire : Q1 2023

Participants: chaque équipe lors des séances de restitution détaillées ; les équipes réunies pour le partage d'informations entre les différents protocoles ; enfin, une séance publique pour la communication externe des résultats de l'expérimentation.

Descriptif : la restitution des résultats de l'expérimentation sera réalisée à 3 niveaux distincts :

- pour chaque équipe, une analyse post-mortem détaillée de l'expérimentation ;
- un événement interne incluant l'ensemble des participants (établissements et fournisseurs de technologie) ;
- un événement public, suivi d'un rapport de synthèse, afin de communiquer toute information partageable avec le secteur.