Prestation Data Scientist - AAP Data Challenge

Cahier des charges - PA-26068

Date de début de consultation : 09/10/2026

Date limite de remise des offres : 30/10/2026 à 12h00

Les offres doivent être transmises à l'adresse suivante : ***https://www.marches-publics.gouv.fr****.ou à défaut par mail à* aymeric.bourdeau@chu-nantes.fr

Toute question peut être adressée à Aymeric Bourdeau – aymeric.bourdeau@chu-nantes.fr avant le 23/10/2026.

SOMMAIRE

I. Présentation du porteur de projet 2

II. Contexte et objet du besoin 2

III. Clauses générales 2

Article 1. Descriptif du besoin 2
Article 2. Répartition des missions 3
Article 3. Descriptif des missions et livrables attendus 3
Mission n°1 - Pilotage projet et benchmark des plateformes data challenge 3
Mission n°2 - Création de la base de données compétiteur 4
Mission n°3 - Développement d'un modèle Baseline 4
Mission n°4 - Préparation et configuration de la plateforme de data challenge 5
Mission n°5 - Gestion technique de la compétition, suivi du challenge et participation aux évènements 6
Mission n°6 – Support à l'élaboration d'un rapport final de valorisation 6
Article 4. Délais d'exécution 7
Article 5. . Conditions particulières d'exécution 8
5.1 *Données, confidentialité et sécurité* 8
5.2 Propriété et remise des livrables 8
5.3 Validation des livrables 8
5.4 Plateforme de Data Challenge 8
5.5 Sobriété numérique et déplacements 9
Article 6. Conditions financières et modalités de paiement 9
Article 7. Annexes 9
Article 8. Contenu de l'offre à remettre (offre technique et financières) 9
Article 9. Modalités de sélection 10

I. Présentation du porteur de projet

Nom : CHU de Nantes
Service demandeur : Département Innovation et Développement
Activités :
Correspondant : Anne-Sophie LE BONNIEC, chargée de valorisation données - chef de projet Data Challenge PED[IA]FEVER - annesophie.lebonniec@chu-nantes.fr
Financement n° RC26_0214
UF : 7093

II. Contexte et objet du besoin

Le CHU de Nantes s'inscrit dans une dynamique de valorisation croissante de ses données de santé. Dans ce contexte, des initiatives autour de l'intelligence artificielle sont en cours notamment pour appuyer l'optimisation des parcours patients ou prédire des facteurs de risque de développer certaines pathologies.

Le CHU de Nantes souhaite lance à compter du 2nd semestre 2026, une compétition de data science en pédiatrie. Il mettra à disposition les données issues de deux études multicentriques au préalable préparées et anonymisées. Elles seront la base sur laquelle les compétiteurs s'appuieront pour créer des modèles prédictifs des infections bactériennes sévères chez l'enfant.
Les résultats de la compétition seront valorisés via un rapport final ainsi que des publications. Les modèles et bases de données seront publiés en opensource/data à l'issue du challenge dans une démarche de science ouverte.

III. Clauses générales

Article 1. Descriptif du besoin

Le présent cahier des charges a pour objet de définir les attentes du CHU de Nantes dans le cadre de la prestation décrite ci-après.

La prestation a pour objectif d'aider le CHU de Nantes à organiser cette compétition en mettant en œuvre les étapes techniques préparatoires et en fournissant un support data science pendant l'évènement.
Plus particulièrement la ressource aura pour mission d'accompagner les équipes du CHU de Nantes dans la constitution des jeux de données nécessaires, le développement d'un modèle baseline ainsi que la configuration de la plateforme hébergeant la compétition. Il sera également mobilisé pour répondre aux compétiteurs durant l'évènement.
Enfin, il participera à l'élaboration du rapport final et fournira un support quant à la publication des données et du notebook sur arXiv.
Une présence locale sur le CHU de Nantes aux côtés de l'équipe projet, aux jalons critiques du projet est indispensable.

Article 2. Répartition des missions

La société partenaire s'engage d'une manière générale à effectuer les missions suivantes :

- Pilotage projet et benchmark des plateformes data challenge
- Création de la base de données compétiteur
- Développement d'un modèle Baseline
- Préparation et configuration de la plateforme de data challenge
- Gestion technique de la compétition, suivi du challenge et participation aux évènements
- Support à l'élaboration d'un rapport final de valorisation

Le porteur du projet assurera les missions suivantes :

- Coordination opérationnelle et scientifique du projet
- Mise à disposition des bases de données sources pour le challenge & anonymisation
- Gestion financière
- Préparation des événements de lancement et clôture du challenge
- Gestion des partenaires de l'événement
- Pilotage règlementaire et juridique
- Stratégie de communication et de valorisation des résultats du challenge

Article 3. Descriptif des missions et livrables attendus

Mission n°1 - Pilotage projet et benchmark des plateformes data challenge

Participation aux comités d'organisation et aux comités scientifiques pendant le déroulement du projet. Ce lot inclut également la phase de benchmark des différentes plateformes de data challenge disponibles.

Une grille comportant les principaux critères de sélection devra être créé et complété par le prestataire en collaboration avec l'équipe projet.
À charge du prestataire :

- Participation à 1 comité d'organisation par mois
- Participation à 1 comité scientifique par mois
- Benchmark des solutions de data challenge (nov 2026)
- Participation au choix de la plateforme du challenge

À charge du CHU de Nantes :

- Organisation des comités et des différents points projets
- Choix définitif de la plateforme du challenge
- Disponibilité des parties-prenantes tout au long du projet

Livrables :

- Matrice d'évaluation des solutions de data challenge complétée

Calendrier de mobilisation : fin octobre à fin novembre 2026 pour le benchmark, puis participation aux comités selon les besoins du projet.

Mission n°2 - Création de la base de données compétiteur

Cette étape vise à réaliser une analyse des données et mettre en œuvre les opérations de préparation (suppression des doublons, corrections des formats, gestion des valeurs vides, création/suppression de colonnes) pour la mise en œuvre du data challenge, en lien avec la Clinique des données du CHU de Nantes.

Pour ces opérations, la librairie Python PANDAS pourra être exploitée. Les données seront traitées en local sur les ordinateurs de la salle numérique de la Fabrique de l'innovation en santé (Bâtiment GINA)

À charge du prestataire :

- Sélection des données pertinentes pour le data challenge
- Analyse des jeux de données sélectionnés
- Préparation des données en vue de la compétition
- Création de la documentation liée aux données (anglais/ français)

À charge du CHU de Nantes :

- Garantir l'accessibilité, la disponibilité et la qualité des données aux échéances attendues.

Livrables :

- Jeux de données constitués, prêt à être intégrés sur la plateforme de data challenge choisie
- Documentation jeux de données

Calendrier de mobilisation : Novembre à Avril 2027

Mission n°3 - Développement d'un modèle Baseline

Développement d'un modèle qui servira de baseline pour la compétition. Il permettra notamment de vérifier que la tâche est réalisable, établir une référence initiale, contribuer à la définition des métriques d'évaluation des modèles, tester la chaîne du challenge (leaderboard, soumission des résultats, etc)., et mesurer l'impact de l'anonymisation sur la capacité prédictive des données. Le modèle sera entraîné sur des données pseudonymisées et anonymisées pour s'assurer du maintien des performances du modèle et de la non dégradation des données anonymisées.

À charge du prestataire :

- Analyse exploratoire des données
- Sélection et entrainement itératif d'un modèle adapté au data challenge sur données pseudonymes/ anonymes
- Contribution à la définition des métriques de la compétition

À charge du CHU de Nantes :

- Disponibilité des parties-prenantes pertinentes pour questions/réponses
- Mobilisation du comité scientifique dans la définition des métriques d'évaluation
- Anonymisation des jeux de données

Livrables :

- Modèle baseline entraîné
- Code développé pour l'entraînement du modèle
- Note présentant les tests sur les deux versions des données.

Calendrier de mobilisation : Janvier à Mars 2027

Mission n°4 - Préparation et configuration de la plateforme de data challenge

La mise en œuvre de la compétition sera réalisée par le prestataire. Cela consistera à configurer la plateforme sélectionnée avec les éléments de communication, la documentation et les jeux de données prévus. Il sera également important de tester la compétition en mode sandbox (se mettre à la place d'un participant) pour en valider le bon fonctionnement. Ce processus itératif permettra ainsi de s'assurer du bon niveau de maturité de la compétition avant son lancement officiel.

À charge du prestataire :

- Intégration du design souhaité dans les limites des options proposées par la plateforme
- Configuration de la plateforme de challenge retenue avec les jeux de données adéquats
- Construction et traduction de la documentation de la compétition
- Tests et validation du paramétrage de la plateforme

À charge du CHU de Nantes :

- Disponibilité des parties-prenantes pertinentes pour questions/réponses
- Sélection et mise à disposition de la plateforme dans les délais adéquats

Livrables :

- Plateforme configurée pour l'ouverture du challenge
- Rapport de tests de la plateforme
- Documentation technique de la compétition en multilingue

Calendrier de mobilisation : Mars /Avril 2027

Mission n°5 - Gestion technique de la compétition, suivi du challenge et participation aux évènements

Gestion technique de la compétition : cette phase consistera à animer les échanges sur la plateforme au travers des forums et sections de questions/réponses pendant la durée du challenge.

À charge du prestataire :

- Animation des échanges techniques et du forum sur la plateforme pendant la compétition , mise à jour d'une Q&A partagée, identification des difficultés rencontrées par les participants et proposition d'ajustement à l'équipe projet
- Participation à l'événement de lancement pour présenter les objectifs techniques, les données mises à disposition, la baseline et les modalités d'évaluation, répondre aux questions des équipes.
- Contribuer à l'analyse des résultats et participer à l'événement de clôture et de remise des prix : présentation des performances, comparaison avec la baseline, analyse des approches et discussion de leurs limites.

À charge du CHU de Nantes :

- Organisation des évènements de lancement et de clôture du challenge
- Disponibilité des parties-prenantes pertinentes pour questions/réponses

Livrables :

- Supports de lancement et de restitution
- FAQ actualisée
- Synthèse technique des résultats

Calendrier de mobilisation : Avril à Juin 2027 (selon date de l'événement de clôture et remise des récompenses)

Mission n°6 – Support à l'élaboration d'un rapport final de valorisation

La compétition donnera lieu à l'écriture d'un rapport final présentant les résultats, perspectives de valorisation et modalités de partage des données en Open Data. Ce rapport sera ainsi porté par le prestataire en collaboration avec les membres du projet pertinents.
Un travail collectif avec le personnel médical ainsi que les lauréats visera à la publication des données et des algorithmes en open data/open source & une publication sur la plateforme arXiv.

À charge du prestataire :

- Rédaction du rapport final
- Publication des données, algorithmes et contribution à la rédaction d'un notebook sur arXiv

À charge du CHU de Nantes :

- Disponibilité des parties-prenantes pertinents pour questions/réponses

- Participation à l'élaboration du rapport final et validation.

Livrables :

- Rapport final contenant les résultats, perspectives de valorisation & modalités de partage
- Données, algorithmes des lauréats publiés en open data/ sources sur les plateformes identifiées
- Notebook publié sur ArXiv

Calendrier de mobilisation : Juillet à décembre 2027

La prestation sera soumise aux stipulations du CCAG FCS (Arrêté 30/03/2021 portant approbation du Cahier des Clauses Administratives Générales applicables aux marchés publics de fournitures courantes et services), qui prévalent sur les conditions générales de vente du titulaire.

Article 4. Délais d'exécution

Les livrables sont attendus pour chaque mission :

- Mission n°1 : Matrice comparative des plateformes, avec recommandation motivée, participation aux comités prévus. **Délai d'exécution** : Remise des matrices comparative des plateformes de challenge fin novembre 2026.
- Mission n°2 : Préparation des jeux de données et documentation rédigée en français et en anglais. **Délai d'exécution** : Version de travail fin mars 2027 et version prête pour intégration dans la plateforme fin avril 2027
- Mission n°3 : Code, modèle entraîné, métriques proposées et courte note présentant les tests sur les deux versions des données. **Délai d'exécution** : Fin mars 2027, sous réserve de la mise à disposition des jeux nécessaires.
- Mission n°4 : Plateforme configurée, documentation de compétition et compte rendu des tests en conditions de participation. **Délai d'exécution** : Fin mars 2027, avant l'ouverture du challenge, après validation des données et mise à disposition de la plateforme.
- Mission n°5 : Animation technique du forum et des questions/réponses, participation au lancement et à la clôture, bilan des incidents et questions traités. **Délai d'exécution** : Avril à juin 2027 pendant la durée effective de la compétition, puis bilan dans les 15 jours suivant l'événement de clôture.
- Mission n°6 : Rapport final et éléments techniques dont la diffusion aura été autorisée par le CHU de Nantes. **Délai d'exécution** : projet de rapport à l'automne 2027 et version finale fin décembre 2027.
- Le candidat joint à son offre un planning détaillé de réalisation des différentes missions et de remise des livrables, précisant les principales étapes, jalons et échéances. Le planning proposé par le titulaire et accepté par le CHU de Nantes sera intégré aux pièces contractuelles du marché. Les échéances prévues au présent cahier des charges constituent les échéances

contractuelles de référence. Tout événement susceptible d'entraîner un retard devra être signalé sans délai au CHU de Nantes, accompagné des mesures correctives proposées.

Article 5. . Conditions particulières d'exécution

5.1 Données, confidentialité et sécurité

Les dispositions relatives à la protection des données à caractère personnel sont définies dans l'annexe RGPD. Dans le cadre de l'exécution des prestations, le titulaire assure un traitement de type 1, tel que défini dans cette annexe.

Les données mises à disposition par le CHU de Nantes sont utilisées exclusivement pour les besoins du présent marché. Toute copie, extraction, transfert, hébergement ou utilisation des données en dehors des environnements autorisés par le CHU de Nantes est interdite, sauf accord préalable du CHU. Le titulaire ne pourra notamment utiliser ces données pour l'entraînement ou l'amélioration de modèles, outils ou services ne relevant pas de l'exécution du présent marché.

Tout recours à un service externe, à une solution d'intelligence artificielle générative, à une plateforme de calcul ou à un espace de stockage susceptible d'entraîner un transfert ou un traitement des données en dehors de l'environnement autorisé par le CHU devra faire l'objet d'un accord préalable du CHU.
Le titulaire informe sans délai le CHU de Nantes de tout incident de sécurité, perte, accès non autorisé ou divulgation susceptible de concerner les données ou informations mises à sa disposition.

5.2 Propriété et remise des livrables

Les livrables remis dans le cadre du marché comprennent notamment les jeux de données préparés, la documentation, le modèle baseline, le code source développé, les métriques, les notebooks et les éléments techniques nécessaires à la réutilisation des travaux. Le titulaire identifie dans son offre les éventuels composants, bibliothèques ou logiciels tiers utilisés et précise leurs conditions de licence.
Les modalités de propriété intellectuelle et de réutilisation des livrables sont celles prévues par les documents contractuels du marché. Toute publication ou diffusion des données, codes, modèles ou autres éléments issus de la prestation est soumise à la validation préalable du CHU de Nantes et au respect des conditions de diffusion applicables.

5.3 Validation des livrables

Les livrables sont soumis à validation du CHU de Nantes. Lorsque le CHU formule des observations portant sur la conformité du livrable aux exigences du présent cahier des charges, le titulaire procède aux corrections nécessaires dans le cadre du prix du marché.

5.4 Plateforme de Data Challenge

La plateforme retenue sera sélectionnée et mise à disposition par le CHU de Nantes, sauf stipulation contraire dans l'offre retenue et les pièces contractuelles. Les éventuels coûts de licence, d'abonnement ou d'hébergement de la plateforme doivent être clairement identifiés dans l'offre financière du candidat.

5.5 Sobriété numérique et déplacements

Compte tenu de la nature principalement numérique de la prestation, le titulaire met en œuvre, dans la mesure compatible avec les exigences techniques, de sécurité, de confidentialité et de performance, des mesures visant à limiter l'impact environnemental de la prestation : limitation des impressions, limitation des duplications et stockages inutiles, recours aux échanges dématérialisés et aux réunions à distance lorsque cela est pertinent, et limitation des traitements informatiques non nécessaires. Les déplacements sont regroupés autant que possible et les modes de transport à moindre émission sont privilégiés lorsque les conditions de la prestation le permettent.

Article 6. Conditions financières et modalités de paiement

Les prix forfaitaires indiqués dans l'offre financière sont fermes pendant toute la durée d'exécution du marché et ne sont pas révisables.

Le règlement des prestations est effectué par acomptes et règlements partiels définitifs, selon les modalités prévues par les documents contractuels du marché et au fur et à mesure de l'admission des prestations.

Les demandes de paiement sont transmises sous forme électronique via le portail CHORUS PRO, conformément à la réglementation applicable. Pour le CHU de Nantes, les références de facturation sont : SIRET 26440013600471, service RECHERCHE, engagement RC26_0214.

Article 7. Annexes

Le dossier comprend les annexes suivantes :

- RGPD.MODELE.ModeleAnnexeRGPD.2026-07-03a (document informatif)
- Annexe financière - Prestation Data science V3 **(à compléter)**
- Annexe RSE - Prestation data science_data challenge 2026 **(à compléter)**
- Macroplanning Data challenge (document informatif)

Article 8. Contenu de l'offre à remettre (offre technique et financières)

Les candidats sont invités à fournir une offre technique synthétique et une offre financière comprenant :

- Une brève note de compréhension du besoin,
- Une proposition de méthodologie de l'accompagnement
- Une proposition de planning de livraison des livrables au vu des délais proposés par le CHU
- Une proposition financière détaillée pour chaque mission conformément à l'annexe financière
- Modalités de sélection

Article 9. Modalités de sélection du prestataire

Les offres seront analysées selon les critères suivants :

- **Critère 1 :** Qualité de l'offre technique au regard du descriptif fourni - 55 %
- **Critère 2 :** Prix de la prestation - 40 %
- **Critère 3** : Prise en compte des enjeux environnementaux liés aux déplacements des personnes affectées à l'exécution du marché– 5 %
- Le candidat complète le questionnaire environnemental / RSE joint au présent cahier des charges. Les réponses sont appréciées au regard des mesures effectivement applicables aux personnes mobilisées pour l'exécution du marché.

L'acheteur pourra attribuer le marché sur la base des offres initiales. Toutefois, il se réserve la possibilité de recourir à la négociation.