

Manager expert de solutions responsables en science des données (MS)

RNCP41673 · RNCP · Publiée

L'essentiel

Niveau de qualification	Niveau 7
Code(s) NSF	310, 326
Formacode(s)	31026, 32054, 32135, 31025
Code(s) ROME	M1403, M1802, M1805
Type d'enregistrement	Enregistrement sur demande
Date d'enregistrement	27/11/2025
Date d'échéance	27/11/2030

Certificateur(s)

INSTITUT POLYTECHNIQUE DE GRENOBLE (IPG) - INP GRENOBLE · SIRET : 19381912500017<https://ensimag.grenoble-inp.fr/>**Grenoble institut national polytechnique (Grenoble INP)****ETABLISSEMENT D'ENSEIGNEMENT SUPERIEUR CONSULAIRE GRENOBLE ECOLE DE MANAGEMENT ·****SIRET : 81738914100018**

Résumé de la certification

OBJECTIFS ET CONTEXTE

Le Manager-Expert de solutions responsables en science des données (MS) recouvre les métiers de Data Engineer et Data Scientist qui ont la spécificité de maîtriser les sciences de données (à l'intersection des mathématiques statistiques et de la programmation informatique) pour réaliser deux grands types de missions qui peuvent s'entrecroiser et s'alimenter : le traitement des données d'une part, et l'analyse des données d'autre part.

S'appuyant autant sur des fondamentaux scientifiques que sur l'ingénierie et les technologies de la data, il identifie une solution de traitement et d'analyse de données, constituant donc un outil à même de mettre à disposition une information pertinente à la prise de décision en entreprise. Il fournit une solution technoscientifique, dont la finalité essentielle de ses activités est de valoriser les données via les démarches et les outils des sciences des données, et ainsi fournir des informations fiables et pertinentes.

Le Manager-Expert de solutions responsables en science des données (MS)

occupe une place centrale et pivot dans les métiers de la data. Il constitue le pilier de l'approche de l'aspect opérationnel des solutions de traitement et d'analyse en sciences des données. Aujourd'hui, parce que les maturités data des organisations s'intensifient, les attentes en la matière ne sont plus exclusivement techniques. Elles s'élargissent : qualité des datas, valorisation des solutions, et prise en considération des composantes RSE nécessitent le développement d'un profil hybride. Le Manager-Expert de solutions responsables en science des données (MS) associe des compétences technoscientifiques de haut niveau et des compétences relevant des champs de la gestion et du management. Face aux spécificités des solutions qu'il met en œuvre et les débats qui les entourent, les solutions qu'il doit élaborer se doivent d'être de plus en plus responsables.

ACTIVITÉS VISÉES

Identification des opportunités des solutions en science des données pour une organisation et son modèle d'affaires

- Analyse du positionnement stratégique de l'entreprise et de sa gouvernance SI et data
- Veille systémique concernant les solutions en science des données
- Identification des questionnements éthiques

- Elaboration de recommandations et mise en exergue de points de vigilance concernant les opportunités et menaces relatives aux solutions data et IA

Pilotage d'un projet de solutions en science des données

- Validation du cahier des charges
- Définition de la méthodologie et des étapes du projet
- Gestion des risques
- Suivi du projet
- Animation des équipes et mobilisation des méthodes d'intelligence collective

Structuration de l'architecture et du déploiement d'infrastructures de gestion de données

- Choix d'une solution de traitement des données
- Sélection et traitement des données internes et externes
- Choix des méthodes et outils d'intégration et de stockage
- Choix des méthodes d'interrogation du Web
- Evaluation des risques de sécurité et de confidentialité

Conception des solutions en science des données visant une problématique stratégique ou organisationnelle

- Sélection de possibles solutions de traitement de données comparativement à un problème métier
- Définition des critères d'arbitrage techniques pour la solution de chaînes de traitement de données
- Sélection des méthodes statistiques ou d'apprentissages automatisés adaptées
- Estimation des coûts de fonctionnement

Caractérisation des créations de valeurs économiques et extra-économiques des solutions en science des données

- Propositions d'indicateurs de performances économiques et environnementales de la solution
- Arbitrages sur des enjeux de conformité et sécuritaires concernant la solution
- Définition d'interfaces et d'ergonomies
- Conduite des actions de communications écrites et orales de valorisation de la solution

COMPÉTENCES ATTESTÉES

- Identifier, comparativement au périmètre d'activité en charge, les éléments clés du diagnostic stratégique comme ceux de la gouvernance du système d'information de l'organisation considérée – qu'elle soit privée, publique ou hybride -, de sorte à définir les objectifs de veille à conduire sur les dimensions technologiques comme non technologiques pertinentes (réglementaires, écologique, éthique, géopolitique..) et ainsi inventorier des opportunités et risques nécessaires aux futurs arbitrages concernant des solutions data et IA
- Organiser la veille – ou compléter une veille existante - selon les objectifs définis, en inventoriant les sources d'informations - formelles et informelles - pertinentes et fiables au regard des domaines à couvrir, en sélectionnant les outils d'exploration à mobiliser, en spécifiant les rythmes d'actualisation et les modalités de restitution de manière à structurer et rendre efficace le processus de veille
- Investiguer, en termes de veille, non seulement les évolutions en termes de modalités de traitement et d'analyse, afin de répertorier de nouveaux modèles, algorithmes, langages et techniques informatiques, mais aussi le domaine des bases de données disponibles - qu'elles soient externes (dont l'open data) ou internes à la structure -, pour améliorer, compléter, préciser les résultats d'analyses et leur niveau de pertinence en vue d'identifier les opportunités de leur intégration dans la chaîne d'analyse des données (processus data)
- Explorer un axe de veille orienté sur les dernières avancées technologiques et scientifiques (quantiques et habilitantes, 6 et 7G, etc...) afin d'anticiper les évolutions des pratiques et des expertises datas avec les possibles incidences sur les besoins en compétences et la pérennité de certains choix technoscientifiques
- Evaluer les opportunités et risques écologiques des solutions en science des données envisagées tant sur les phases de production, de modes de gestion de données, et d'usages qui en découleraient afin d'intégrer l'impact écologique comme dimension d'arbitrage
- Répertorier les bonnes pratiques et les questionnements en termes d'éthiques afin de s'assurer que les solutions en science des données repérées pour de possibles applications futures respectent les principes de responsabilité sociétales et ne risquent pas d'être à l'origine d'une prise de risques pour l'organisation
- Effectuer des recommandations de solutions datas pertinentes et viables pour l'organisation en proposant une synthèse récapitulative des principales opportunités et menaces des options répertoriées tant dans leurs dimensions techniques (SI et Data), organisationnelles mais aussi économiques, environnementales, sociétales, et éthiques afin de proposer les alternatives possibles et les éléments d'arbitrages à la Direction
- Etablir le cahier des charges (note de cadrage) fonctionnel et technique de la demande client – interne ou externe - à partir d'une analyse des besoins, reposant sur une approche adaptée des processus métiers, de manière à le

reformuler, le valider de sorte à définir le dimensionnement du projet, les technologies à mettre en place, les démarches nécessaires, sa faisabilité (technique, ressources, et « timing »), notamment au regard de l'infrastructure existante du système d'information

- Définir, à partir du cahier des charges, des technologies pressenties, et du caractère innovant ou pas de la solution à élaborer, la méthodologie de projet adéquate classique (agile, waterfall, stage gate process, TLS théorie des contraintes Lean Six Sigma,...) ou innovante (design thinking, lean startup,...), pouvant être fonction du projet, des pratiques et des exigences de l'organisation et du secteur, de manière à optimiser la réalisation des livrables du projet en termes de faisabilité, qualité, coûts et délais
- Décliner, si le projet est innovant, le processus de l'innovation – dit « pipeline de l'innovation » - nécessitant des phases spécifiques amont (brainstorming, créativité, prototypage, enquête terrain,...) ainsi que les multiples boucles d'itération concernant les différentes étapes de validités de sorte à proposer une solution à même d'offrir le caractère innovant idoine souhaité et conforme au besoin à satisfaire
- Identifier, dès les premières phases du projet, les phases de la chaîne de valeur du process générant les impacts écologiques les plus négatifs, de manière à soutenir une éco-conception (réduction de l'obsolescence, IA frugale, plan de nettoyage du Dark data ...) la plus pertinente sur la solution en science des données en cours d'élaboration
- Répertorier les différents types de risques (technologiques, réglementaires, concurrentiels, ...) pouvant impacter le projet afin d'en définir leur criticité et les actions correctives possibles en mobilisant les méthodologies en contexte opérationnel de type adéquates (AMDEC,...)
- Organiser, que ce soit pour un projet innovant ou non, les phases de déroulement d'un projet classique selon la méthodologie définie articulant les étapes –souvent itératives- de collecte, d'exploration, d'exploitation et de mise en production des données et les différentes actions nécessaires (organigramme, organisation planification des tâches, chemin critique, définitions des livrables, « work package »)... afin de garantir une réponse satisfaisante, en temps et en heure, au cahier des charges des clients internes et externes
- Animer les équipes et les parties prenantes, dans le respect des valeurs éthiques et inclusive, au travers notamment d'actions de communication cadencées, structurées, mobilisant les tiers requis, usant des outils et des modalités adéquats, de manière à assurer la transparence et la fluidité de l'information propice à l'efficacité de l'organisation du travail et au climat de confiance
- Promouvoir l'usage des méthodes d'intelligence collective (mobilisation des profils de personnalités, inclusion, brainstorming, event storming, ...) afin de renforcer l'engagement des parties prenantes, d'améliorer l'appréhension de la diversité des points de vue, de stimuler la créativité, et d'optimiser la prise de décision, contribuant ainsi à l'efficacité du projet et à la qualité du livrable final
- Définir, en relation avec d'autres experts des sciences des données, une solution de traitement de données concernant leur acquisition, stockage et mise à disposition, en déployant les différentes étapes de la constitution d'un pipeline de traitement de données compatible avec le SI existant, sur la base d'une analyse de celui-ci, de l'identification des axes de stratégie et de politique Données et SI de l'organisation, comme des besoins métiers afin de permettre une mobilisation des données générant de la valeur
- Identifier, en considérant les interdépendances avec les autres problématiques et enjeux de l'entreprise, les données pertinentes pour la solution envisagée, en sélectionnant à la fois les données déjà repérées et disponibles en interne, comme des données externes mobilisables, via des achats, des sources disponibles en open data, ou des partenariats, afin d'accroître la pertinence des analyses qui en seront réalisées
- Évaluer la qualité nécessaire des données afin de décider de la mise en œuvre ou non de leur nettoyage (par modification, par imputation, par suppression,...), de sorte à ultimement optimiser la pertinence des résultats obtenus par les analyses effectuées sur ces données
- Intégrer les données identifiées dans le système d'information de l'entreprise en arbitrant sur la ou les méthodes adaptées (ETL et ELT), les langages (Python, JavaScript,...), les outils [bases de données relationnelles, bases de données NoSQL - MongoDB, Neo4J,...], l'utilisation de systèmes distribués de manière à avoir accès facilement et efficacement à l'ensemble des données requises
- Mettre en place des processus d'interrogation du Web -sémantique ou non- en tirant partie du fonctionnement sous-jacent des moteurs de recherche textuel ainsi que de la structuration des données disponibles (pages web, web sémantique,...) dans le but d'extraire de l'information exploitable
- Mener des évaluations concernant le respect de la sécurité et la confidentialité -issues notamment des réglementations nationales et internationales dont le RGPD-, des infrastructures utilisées, en mettant en œuvre des phases de test, une démonstration des vulnérabilités aux parties prenantes de manière à prodiguer les recommandations nécessaires et leur déploiement le cas échéant

- Conceptualiser un problème professionnel en problématique de traitement de données, en définissant, une fois le ou les cas d'usage bien clarifié(s), si la tâche est descriptive ou prescriptive, les méthodes à utiliser, les analyses à mener et les livrables à produire, de sorte à répondre à une sollicitation des Directions de l'organisation, visant à améliorer l'efficacité du SI et de ses données, l'aide à la décision, l'élaboration de nouvelles offres
- Etablir un premier inventaire des types de solutions en science des données, adapté au système data existant, en termes de : données, modalités de stockage, modèles statistiques et d'algorithmes d'apprentissage automatique -machine learning, deep learning-, de sorte à adapter au mieux la pertinence de la solution à la nature des données devant être analysées
- Evaluer l'intérêt de l'introduction de nouvelles méthodes de data science (IA, Deep learning, technologie de seconde génération quantique, ...) en mobilisant les connaissances des fondements de la science des données pour questionner la pertinence des options techniques envisagées et l'intégration de ces dernières au SI existant
- Mobiliser les méthodes statistiques afin d'analyser les données disponibles et évaluer les réponses aux questions posées de manière à éviter l'influence des valeurs extrêmes, erronées, manquantes, ou des situations de sur-apprentissage
- Élaborer un modèle d'apprentissage automatisé en sélectionnant un algorithme (Deep learning, KNL,...) et la mise en place de jeux de données d'apprentissages, de tests et de validations pour entraîner le modèle et valider ses capacités prédictives
- Intégrer dans les solutions techniques envisagées les coûts de fonctionnement des structurations des données et des modalités de traitement de manière à en optimiser l'impact budgétaire et à en réduire l'impact environnemental
- Elaborer une estimation de la valorisation économique de la solution data, en mobilisant par exemple le Business Model Canvas –BMC- permettant d'identifier la création de valeur et sa structuration afin de stabiliser la pertinence et la pérennité de la solution, et plus généralement sa contribution aux performances économiques et extra-économiques de l'entreprise
- Définir, outre la valorisation économique, différentes catégories d'indicateurs de performances économiques et non économiques permettant de piloter et de spécifier les résultats attendus en termes d'optimisation des ressources (Retour sur investissement –ROI-, coût total de fonctionnement), en termes d'efficacité opérationnelle, en termes de conformité, de satisfaction client,.. de manière à préserver et à atteindre les objectifs de la solution visée
- Justifier de la prise en considération des risques et enjeux de conformité (droit, légal) et sécuritaires dans les arbitrages concernant les différentes composantes de la solution responsable de manière à offrir une solution data sécurisée, fiable et conforme aux lois en vigueur sur la science des données et leur exploitation chez leurs clients (interne – externe)
- Préconiser une solution de visualisation des données utilisant les standards actuels, les règles relevant du fonctionnement de la perception visuelle et de la sémiologie graphique, de manière à la rendre inclusive, d'être à même d'en évaluer l'efficacité afin de permettre une restitution qualitative des informations pertinentes aux différentes équipes métiers concernées et d'en améliorer l'impact
- Communiquer à l'oral ou à l'écrit de manière efficiente, pertinente, percutante et juste sur une solution data vers des spécialistes et non spécialistes (Flyer, Pitch, présentation Flash BMC, document technique,) de manière à assurer l'engagement et l'intérêt pour la solution tant au niveau des équipes parties prenantes à l'élaboration de la solution, les futurs usagers, les sponsors du projet comme les Directions, actrices de la décision et de promotion

Blocs de compétences

RNCP41673BC01 — Identifier des opportunités de solutions en science des données pour une organisation et son modèle d'affaires

LISTE DE COMPÉTENCES

- Identifier, comparativement au périmètre d'activité en charge, les éléments clés du diagnostic stratégique comme ceux de la gouvernance du système d'information de l'organisation considérée – qu'elle soit privée, publique ou hybride -, de sorte à définir les objectifs de veille à conduire sur les dimensions technologiques comme non technologiques pertinentes (réglementaires, écologique, éthique, géopolitique..) et ainsi inventorier des opportunités et risques nécessaires aux futurs arbitrages concernant des solutions data et IA
- Organiser la veille – ou compléter une veille existante - selon les objectifs définis, en inventoriant les sources d'informations - formelles et informelles - pertinentes et fiables au regard des domaines à couvrir, en sélectionnant les outils d'exploration à mobiliser, en spécifiant les rythmes d'actualisation et les modalités de restitution de manière à

structurer et rendre efficace le processus de veille

- Investiguer, en termes de veille, non seulement les évolutions en termes de modalités de traitement et d'analyse, afin de répertorier de nouveaux modèles, algorithmes, langages et techniques informatiques, mais aussi le domaine des bases de données disponibles - qu'elles soient externes (dont l'open data) ou internes à la structure -, pour améliorer, compléter, préciser les résultats d'analyses et leur niveau de pertinence en vue d'identifier les opportunités de leur intégration dans la chaîne d'analyse des données (processus data)
- Explorer un axe de veille orienté sur les dernières avancées technologiques et scientifiques (quantiques et habilitantes, 6 et 7G, etc...) afin d'anticiper les évolutions des pratiques et des expertises datas avec les possibles incidences sur les besoins en compétences et la pérennité de certains choix technoscientifiques
- Evaluer les opportunités et risques écologiques des solutions en science des données envisagées tant sur les phases de production, de modes de gestion de données, et d'usages qui en découleraient afin d'intégrer l'impact écologique comme dimension d'arbitrage
- Répertorier les bonnes pratiques et les questionnements en termes d'éthiques afin de s'assurer que les solutions en science des données repérées pour de possibles applications futures respectent les principes de responsabilité sociétales et ne risquent pas d'être à l'origine d'une prise de risques pour l'organisation
- Effectuer des recommandations de solutions datas pertinentes et viables pour l'organisation en proposant une synthèse récapitulative des principales opportunités et menaces des options répertoriées tant dans leurs dimensions techniques (SI et Data), organisationnelles mais aussi économiques, environnementales, sociétales, et éthiques afin de proposer les alternatives possibles et les éléments d'arbitrages à la Direction

MODALITÉS D'ÉVALUATION

- Proposition d'un processus de veille concernant les solutions en science des données, comprenant
 - 1/ une synthèse des éléments clés du diagnostic stratégique de l'organisation considérée et de la gouvernance concernant le système d'information et data
 - 2/ la définition des objectifs de veille concernant les solutions, les bases de données, et les nouvelles technologies comme les dimensions nécessaires à couvrir
 - 3/ l'inventaire des principales sources d'informations pertinentes à explorer en matière de solutions en science des données
 - 4/ les modalités de recherche, de sélection, les outils de veille à mobiliser

Cas réel ou fictif

Pour 1/ Production écrite collective

Pour 2, 3 et 4 : Production écrite individuelle et collective

- Note de recommandations des opportunités et risques de solutions en science des données pour l'organisation comprenant

1/ la proposition argumentée de :

- solutions en science des données (dont d'IA) pertinentes
- les possibles sources de bases de données pertinentes à exploiter
- des nouvelles technologies prometteuses comparativement aux objectifs répertoriés

2/ des focus concernant la proposition argumentée sur les 3 domaines ci-dessus en termes d'impact/ d'enjeux environnementaux et éthique, en particulier dans leurs dimensions de menaces

Cas réel ou fictif

Production écrite et orale, individuelle et collective

RNCP41673BC02 — Piloter un projet de solutions en science des données

LISTE DE COMPÉTENCES

- Etablir le cahier des charges (note de cadrage) fonctionnel et technique de la demande client – interne ou externe - à partir d'une analyse des besoins, reposant sur une approche adaptée des processus métiers, de manière à le reformuler, le valider de sorte à définir le dimensionnement du projet, les technologies à mettre en place, les démarches nécessaires, sa faisabilité (technique, ressources, et « timing »), notamment au regard de l'infrastructure existante du système d'information
- Définir, à partir du cahier des charges, des technologies pressenties, et du caractère innovant ou pas de la solution à élaborer, la méthodologie de projet adéquate classique (agile, waterfall, stage gate process, TLS théorie des contraintes Lean Six Sigma,...) ou innovante (design thinking, lean startup,...), pouvant être fonction du projet, des pratiques et des exigences de l'organisation et du secteur, de manière à optimiser la réalisation des livrables du projet en termes de faisabilité, qualité, coûts et délais
- Décliner, si le projet est innovant, le processus de l'innovation – dit « pipeline de l'innovation » - nécessitant des

phases spécifiques amont (brainstorming, créativité, prototypage, enquête terrain,...) ainsi que les multiples boucles d'itération concernant les différentes étapes de validités de sorte à proposer une solution à même d'offrir le caractère innovant idoine souhaité et conforme au besoin à satisfaire

- Identifier, dès les premières phases du projet, les phases de la chaîne de valeur du process générant les impacts écologiques les plus négatifs, de manière à soutenir une éco-conception (réduction de l'obsolescence, IA frugale, plan de nettoyage du Dark data ...) la plus pertinente sur la solution en science des données en cours d'élaboration
- Répertorier les différents types de risques (technologiques, réglementaires, concurrentiels, ...) pouvant impacter le projet afin d'en définir leur criticité et les actions correctives possibles en mobilisant les méthodologies en contexte opérationnel de type adéquates (AMDEC,...)
- Organiser, que ce soit pour un projet innovant ou non, les phases de déroulement d'un projet classique selon la méthodologie définie articulant les étapes –souvent itératives- de collecte, d'exploration, d'exploitation et de mise en production des données et les différentes actions nécessaires (organigramme, organisation planification des tâches, chemin critique, définitions des livrables, « work package »),... afin de garantir une réponse satisfaisante, en temps et en heure, au cahier des charges des clients internes et externes
- Animer les équipes et les parties prenantes, dans le respect des valeurs éthiques et inclusive, au travers notamment d'actions de communication cadencées, structurées, mobilisant les tiers requis, usant des outils et des modalités adéquats, de manière à assurer la transparence et la fluidité de l'information propice à l'efficacité de l'organisation du travail et au climat de confiance
- Promouvoir l'usage des méthodes d'intelligence collective (mobilisation des profils de personnalités, inclusion, brainstorming, event storming, ..) afin de renforcer l'engagement des parties prenantes, d'améliorer l'appréhension de la diversité des points de vue, de stimuler la créativité, et d'optimiser la prise de décision, contribuant ainsi à l'efficacité du projet et à la qualité du livrable final

MODALITÉS D'ÉVALUATION

- Elaboration du cahier des charges (ou note de cadrage) à partir de la transformation de l'expression du besoin client précisant le besoin, l'objectif, le périmètre, les contraintes, les ressources, les principaux critères de validation et de réussite

Cas réel ou fictif

Production collective, écrite et orale

- Rapport de clôture d'un projet comprenant :
 - Les objectifs initiaux et leur atteinte
 - Les décisions clés et arbitrages effectués (notamment concernant les choix de méthodologie, le séquençage, l'analyse de risques, le plan de communication)
 - Les écarts entre prévisions et réalisations
 - Les enseignements et bonnes pratiques à retenir
 - Les recommandations pour les futurs projets

Cas réel ou fictif

Production orale collective avec supports

- Note synthétique de recommandations d'écoconception d'une solution en sciences de données sur l'ensemble de son cycle de vie comprenant :
 - l'analyse du cycle de vie des données
 - les phases les plus pertinentes à optimiser

Cas réel ou fictif

Production écrite, individuelle et collective

Note de recommandations OU Retour d'expérience concernant le choix/la mise en œuvre de méthodes d'intelligence collective dans un contexte de projet portant sur une solution data et IA à même de favoriser un environnement de travail sain

Cas réel ou fictif

Production écrite individuelle

RNCP41673BC03 — Structurer l'architecture et le déploiement d'infrastructures de gestion de données

LISTE DE COMPÉTENCES

- Définir, en relation avec d'autres experts des sciences des données, une solution de traitement de données concernant leur acquisition, stockage et mise à disposition, en déployant les différentes étapes de la constitution d'un pipeline de traitement de données compatible avec le SI existant, sur la base d'une analyse de celui-ci, de l'identification des

axes de stratégie et de politique Données et SI de l'organisation, comme des besoins métiers afin de permettre une mobilisation des données générant de la valeur

- Identifier,

en considérant les interdépendances avec les autres problématiques et enjeux de l'entreprise, les données pertinentes pour la solution envisagée, en sélectionnant à la fois les données déjà repérées et disponibles en interne, comme des données externes mobilisables, via des achats, des sources disponibles en open data, ou des partenariats, afin d'accroître la pertinence des analyses qui en seront réalisées

- Évaluer la qualité nécessaire des données afin de décider de la mise en œuvre ou non de leur nettoyage (par modification, par imputation, par suppression,...), de sorte à ultimement optimiser la pertinence des résultats obtenus par les analyses effectuées sur ces données

- Intégrer les données identifiées dans le système d'information de l'entreprise en arbitrant sur la ou les méthodes adaptées (ETL et ELT), les langages (Python, JavaScript,...), les outils [bases de données relationnelles, bases de données NoSQL - MongoDB,

Neo4J,...], l'utilisation de systèmes distribués de manière à avoir accès facilement et efficacement à l'ensemble des données requises

- Mettre en place des processus d'interrogation du Web -sémantique ou non- en tirant partie du fonctionnement sous-jacent des moteurs de recherche textuel ainsi que de la structuration des données disponibles (pages web, web sémantique,...) dans le but d'extraire de l'information exploitable

- Mener des évaluations concernant le respect de la sécurité et la confidentialité -issues notamment des réglementations nationales et internationales dont le RGPD-, des infrastructures utilisées, en mettant en œuvre des phases de test, une démonstration des vulnérabilités aux parties prenantes de manière à prodiguer les recommandations nécessaires et leur déploiement le cas échéant

MODALITÉS D'ÉVALUATION

- Définition et justification du pipeline à partir d'une problématique de mise en place d'une solution de stockage et de traitement d'un jeu de données existants (cas intégré ou fragmenté) traitant de :

- le choix d'un modèle de données en vue du stockage des données

- l'intégration des données (internes ou/et externes à l'organisation) incluant l'évaluation de la qualité et leur nettoyage potentiel

Cas réel ou fictif

Production écrite individuelle

- Diagnostic de sécurité et de potentielles failles intégrant :

- l'identification et l'analyse d'une faille de sécurité en lien avec le thème proposé

- l'identification du service ou du programme compromis, en précisant la vulnérabilité, ses possibles exploitations par un attaquant potentiel

- la description de l'architecture du système concerné

- les recommandations pour limiter l'impact de la faille et empêcher son exploitation, en s'appuyant sur les bonnes pratiques de sécurité

- la proposition de mesures préventives pour les équipes de développement

- L'illustration par un scénario d'entreprise de la criticité de la faille et des actions à intégrer dans la politique de sécurité

- La définition d'une expérimentation visant la démonstration de la vulnérabilité, en créant un environnement de test et en détaillant une procédure d'exploitation

Production écrite individuelle

Cas réel ou fictif

RNCP41673BC04 — Concevoir des solutions en science des données visant une problématique stratégique ou organisationnelle

LISTE DE COMPÉTENCES

- Conceptualiser un problème professionnel en problématique de traitement de données, en définissant, une fois le ou les cas d'usage bien clarifié(s), si la tâche est descriptive ou prescriptive, les méthodes à utiliser, les analyses à mener et les livrables à produire, de sorte à répondre à une sollicitation des Directions de l'organisation, visant à améliorer l'efficacité du SI et de ses données, l'aide à la décision, l'élaboration de nouvelles offres

- Etablir un premier inventaire des types de solutions en science des données, adapté au système data existant, en termes de : données, modalités de stockage, modèles statistiques et d'algorithmes d'apprentissage automatique -machine learning, deep learning-, de sorte à adapter au mieux la pertinence de la solution à la nature des données

devant être analysées

- Evaluer l'intérêt de l'introduction de nouvelles méthodes de data science (IA, Deep learning, technologie de seconde génération quantique, ...) en mobilisant les connaissances des fondements de la science des données pour questionner la pertinence des options techniques envisagées et l'intégration de ces dernières au SI existant
- Mobiliser les méthodes statistiques afin d'analyser les données disponibles et évaluer les réponses aux questions posées de manière à éviter l'influence des valeurs extrêmes, erronées, manquantes, ou des situations de sur-apprentissage
- Élaborer un modèle d'apprentissage automatisé en sélectionnant un algorithme (Deep learning, KNL,...) et la mise en place de jeux de données d'apprentissages, de tests et de validations pour entraîner le modèle et valider ses capacités prédictives
- Intégrer dans les solutions techniques envisagées les coûts de fonctionnement des structurations des données et des modalités de traitement de manière à en optimiser l'impact budgétaire et à en réduire l'impact environnemental

MODALITÉS D'ÉVALUATION

- Mise en place d'un modèle de Machine learning :
- Acquisition et analyse des données
- Filtrage et sélection des caractéristiques
- Choix d'un algorithme adapté à partir de la connaissance des données
- Evaluation de la qualité des résultats

Cas réel ou fictif

Production écrite individuelle

- Evaluation de l'impact énergétique et environnemental d'une solution en science des données.

Cas fil rouge

Cas réel ou fictif

Production écrite collective

RNCP41673BC05 — Caractériser des créations de valeurs économiques et extra-économiques des solutions en science des données

LISTE DE COMPÉTENCES

- Elaborer une estimation de la valorisation économique de la solution data, en mobilisant par exemple le Business Model Canvas –BMC–

permettant d'identifier la création de valeur et sa structuration afin de stabiliser la pertinence et la pérennité de la solution, et plus généralement sa contribution aux performances économiques et extra-économiques de l'entreprise

- Définir,

autre la valorisation économique, différentes catégories d'indicateurs de performances économiques et non économiques permettant de piloter et de spécifier les résultats attendus en termes d'optimisation des ressources (Retour sur investissement –ROI–, coût total de fonctionnement), en termes d'efficacité opérationnelle, en termes de conformité, de satisfaction client,.. de manière à préserver et à atteindre les objectifs de la solution visée

- Justifier de la prise en considération des risques et enjeux de conformité (droit, légal) et sécuritaires dans les arbitrages concernant les différentes composantes de la solution responsable de manière à offrir une solution data sécurisée, fiable et conforme aux lois en vigueur sur la science des données et leur exploitation chez leurs clients (interne – externe)

- Préconiser une solution de visualisation des données utilisant les standards actuels, les règles relevant du fonctionnement de la perception visuelle et de la sémiologie graphique, de manière à la rendre inclusive, d'être à même d'en évaluer l'efficacité afin de permettre une restitution qualitative des informations pertinentes aux différentes équipes métiers concernées et d'en améliorer l'impact

- Communiquer à l'oral ou à l'écrit de manière efficiente, pertinente, percutante et juste sur une solution data vers des spécialistes et non spécialistes (Flyer, Pitch, présentation Flash BMC, document technique,) de manière à assurer l'engagement et l'intérêt pour la solution tant au niveau des équipes parties prenantes à l'élaboration de la solution, les futurs usagers, les sponsors du projet comme les Directions, actrices de la décision et de promotion

MODALITÉS D'ÉVALUATION

- Selon le contexte et le type de solutions Data et IA considérés, proposition d'une analyse de la création de valeur et/ou d'indicateurs de performances clé (KPI) de pilotage et de caractérisations de la solution :

1. Pour une analyse de la création de valeur

Proposition d'une analyse de la création de valeur en mobilisant un outil adapté (type Business Model Canvas (BMC), Kano Model & JTBD, Balance score card,..) abordant

- La justification de l'outil choisi
 - La mise en œuvre conforme de l'outil
 - Les principaux résultats obtenus et leur analyse critique
2. Pour la proposition d'indicateurs clés de performances (KPI)
- Sélection des KPI clés concernés par la solution et les impacts attendus
 - Proposition des indicateurs de nature autre que les KPI spécifiques à la solution data
 - Estimation et appréciation des évolutions attendues/observées des KPI
 - Synthèses des plus-values et première analyse de la performance de la solution

Cas réel ou fictif

Production écrite, individuelle et collective

- Selon la solution et l'acuité des problématiques et enjeux de conformité et sécuritaire la concernant, note d'identification des éléments de conformité et/ou sécuritaire actuels et potentiels pris en considération dans l'arbitrage des caractéristiques de la solution définie

Cas réel ou fictif

Production écrite, individuelle et collective

- Proposition argumentée d'une visualisation d'un ensemble de données fourni.

Cas réel ou fictif

Production individuelle

- Réalisation d'une action de communication écrite ou orale portant, selon la solution et le contexte, soit sur :
 - un pitch de la solution
 - un flyer à destination d'un public spécifié
 - une présentation orale de la solution devant un public d'experts ou de non experts,
 - un document de synthèse technique dédié à la gestion des connaissances sur la solution

Mise en situation réelle ou fictive

Production écrite ou orale, individuelle ou collective

Secteur d'activité et type d'emploi

SECTEURS D'ACTIVITÉS

Le certifié exerce ses fonctions dans toutes les tailles d'entreprises, que ce soient des grands groupes, des PMT/PMI, des start-up, et dans une moindre mesure des TPE, principalement en France. Il évolue à travers tous types de secteurs d'activités, aussi bien dans le secteur industriel (automobile, transports, fabrication d'équipements industriels, agroalimentaire) que dans celui des services (activités financières et d'assurance, commerce/distribution, activités de conseil).

TYPE D'EMPLOIS ACCESSIBLES

Data scientist, data engineer, consultant data science, data solutions engineer, analyste scientifique des données, data miner, expert en sciences des données, ingénieur data scientist, expert big data, ingénieur big data, ingénieur données, responsable de pôle

RÉGLEMENTATIONS D'ACTIVITÉ

Les activités ne sont pas spécifiquement assujetties dans leur ensemble à une réglementation particulière. L'exercice du métier n'est pas conditionné à la détention d'une habilitation particulière. Cependant, le professionnel doit notamment tenir compte des réglementations spécifiques liées à la protection et à la confidentialité des données.

Marché du travail

Indicateurs au niveau des familles professionnelles (FAP) des métiers visés, agrégés au niveau national — France Travail / DARES (open data).

Tension de recrutement (DARES 2024)	Très élevée (5/5) Moyenne pondérée par l'emploi ; les familles ci-dessous s'échelonnent de 3 à 5 sur 5.
Facteurs principaux	Manque de main-d'œuvre disponible, Lien formation-emploi
Salaire médian (mensuel)	2 912 €

Diplôme le plus fréquent	Diplôme de niveau bac+5 ou plus (doctorat, master ou équivalent) en communication et information (41 % des actifs en emploi)
Familles professionnelles (FAP)	L4X82 — Techniciens et agents de maîtrise chargés d'études socio-économiques · tension 4/5 (Élevée) L5X93 — Chargés d'études socio-économiques · tension 3/5 (Modérée) M1X80 — Techniciens d'étude et de développement en informatique · tension 4/5 (Élevée) M2X90 — Ingénieurs et cadres d'étude, recherche et développement en informatique et télécom · tension 5/5 (Très élevée) M2X93 — Experts et consultants en systèmes d'information · tension 5/5 (Très élevée)
Projets de recrutement (BMO 2026)	34 487
Jugés difficiles à recruter	50 %
Emplois saisonniers	1 %
Familles professionnelles (FAP)	L4X82 — Techniciens et agents de maîtrise chargés d'études socio-économiques L5X93 — Chargés d'études socio-économiques M1X80 — Techniciens d'étude et de développement en informatique M2X90 — Ingénieurs et cadres d'étude, recherche et développement en informatique et télécom M2X93 — Experts et consultants en systèmes d'information

Voies d'accès

MODALITÉS D'ACCÈS

Formation initiale, Formation continue, Apprentissage, Contrat de professionnalisation, VAE

PRÉREQUIS À L'ENTRÉE EN FORMATION

Sont recevables les candidatures d'étudiants titulaires d'un des diplômes suivants :

- Titre d'ingénieur diplômé conférant le grade de master
- Diplôme d'une école de management, privée ou consulaire, conférant le grade de master
- Diplôme de 3ème cycle habilité par les autorités universitaires (Diplôme national de master, DEA, DESS, ...) ou diplôme professionnel de niveau BAC + 5
- Diplôme ou attestation de validation d'un niveau équivalent M1 pour des candidats ayant au moins 3 ans d'expérience professionnelle en lien avec la formation visée
- Diplôme de M1 ou équivalent, pour des auditeurs justifiant d'au moins trois années d'expérience professionnelle
- Titre inscrit au Répertoire national des certifications professionnelles (RNCP) niveau 7
- Diplôme étranger équivalent aux diplômes français exigés ci-dessus.

Conditions d'accès dérogatoires :

1. Dans la limite de 40 % maximum de l'effectif de la promotion suivant la formation Mastère Spécialisé concernée, sont recevables après une procédure VAPP (Validation des Acquis Professionnels et Personnels), les candidatures de personnes justifiant à minima de 5 années d'expérience professionnelle pour lesquelles les activités exercées ont un lien avéré avec les compétences professionnelles visées par la formation.

2. Par dérogation pour 30 % maximum du nombre d'étudiants suivant la formation Mastère Spécialisé concernée, sont recevables les candidatures d'étudiants titulaires d'un des diplômes suivants à orientation techno-scientifique :

- Diplôme ou attestation de validation d'un niveau équivalent M1 sans expérience professionnelle ou ayant moins de trois ans d'expérience professionnelle en lien avec la formation visée
- Diplôme de licence (L3) ou grade de Licence ou titre inscrit au RNCP niveau 6 justifiant d'une expérience professionnelle de 3 ans minimum, en lien avec la formation visée.

Le pourcentage total des dérogations prévues au a) et au b) ci-dessus ne doit pas excéder 40%.

Ce programme requiert d'avoir abordé des aspects scientifiques lors de ses études ou d'avoir une 1ère expérience dans le domaine du traitement de la donnée (pré-requis en informatique et statistique).

COMPOSITION DU JURY — FORMATION INITIALE, FORMATION CONTINUE, APPRENTISSAGE, CONTRAT DE PROFESSIONNALISATION

- Le Président : Les Directeurs académiques/des programmes GEM/Grenoble INP ou leurs représentants
- Les Directeurs des programmes MS, MS-Titre ou Titre ou leurs représentants
- Le Responsable académique ou un membre de la faculté non enseignant dans le programme, représentant l'un des Départements de Formation et de Recherches respectivement de GEM et de Grenoble INP Ensimag
- Les membres extérieurs à l'organisme de formation, étant a minima 2 représentants du secteur professionnel et comptant pour 55% des voix.

COMPOSITION DU JURY — VAE

Le jury compte un minimum de 3 personnes :

- Le Responsable académique de la VAE
- Le Responsable du programme de la certification visée (sauf si le RP assure un enseignement dans un processus d'accompagnement ou formatif proposé dans le cadre de la VAE)
- Un ou des représentant(s) professionnel(s) qualifié(s) du domaine de la certification, comptant pour au moins 50% des voix du jury

Base légale

Date de décision	27/11/2025
Durée de l'enregistrement	5 ans
Date d'échéance de l'enregistrement	27/11/2030
Date de dernière délivrance possible	27/11/2034

Statistiques

Année	Certifiés	Dont VAE	Insertion 6 mois	Métier 2 ans
2023	14	0	85 %	—
2022	21	0	100 %	80 %
2021	19	0	100 %	88 %

Liens avec d'autres certifications

CORRESPONDANCES (PAR BLOCS)

RNCP35197, RNCP36398, RNCP36739, RNCP36921, RNCP37172, RNCP37422, RNCP37624, RNCP37750, RNCP39586, RNCP39775, RNCP40167

Pour plus d'informations

Fiche officielle (France Compétences)

<https://www.francecompetences.fr/recherche/rncp/41673/>