

DELIBERATION PORTANT RENOUVELLEMENT DU DIPLOME D'UNIVERSITE DATA SCIENTIST
PORTE PAR L'ECOLE UNIVERSITAIRE DE PHYSIQUE ET D'INGENIERIE

LE CONSEIL DE LA FORMATION ET DE LA VIE UNIVERSITAIRE DE L'UNIVERSITE CLERMONT AUVERGNE, EN SA SEANCE
DU MARDI 09 AVRIL 2024,

Vu le code de l'éducation ;

Vu le décret n°2024-3 du 2 janvier 2024 modifiant le décret n°2020-1527 en date du 7 décembre 2020 portant création
de l'établissement public expérimental Université Clermont Auvergne (UCA) ;

Vu les statuts de l'Université Clermont Auvergne, notamment les articles 29 à 31 ;

Vu le règlement Intérieur de l'Université Clermont Auvergne ;

Vu la délibération du conseil d'administration du 16 mars 2021 portant élection du Président de l'université, Mathias
BERNARD ;

Vu le quorum atteint en début de séance ;

Vu la présentation de Françoise PEYRARD, Vice-Présidente en charge de la Formation ;

Après en avoir délibéré ;

DECIDE

de renouveler le Diplôme d'Université Data Scientist porté par l'Ecole Universitaire de Physique et d'Ingénierie, tel que
présenté en annexe.

Membres en exercice : 44

Votes : 31

Pour : 29

Contre : 2

Abstentions : 0

**Le Président de l'Université
Clermont Auvergne,**

Mathias BERNARD



DOSSIER DE VALIDATION DU-DIU Volet pédagogique

DU Data Scientist

- Renouvellement -

UFR de rattachement : EUPI
Établissements partenaires : LPC, IP, LIMOS, LMBP, CCIN2P3
Responsable du diplôme : Julien DONINI
Validation à compter du : 9 avril 2024
Validation accordée jusqu'à fin : 2026-2027
Délibération CFVU N°2024-04-09_14

A - Présentation du projet

I – Contexte général – Objectifs de la formation

Le Diplôme Universitaire Data Scientist propose depuis son ouverture en septembre 2018 des enseignements dans le domaine des data sciences à des professionnels et des étudiants. En six années de fonctionnement le DU a accueilli 121 apprenants : 18 inscrits en formation continue (FC), dont des salariés des entreprises Michelin, Sopra Steria, Casino et OpenStudio, et 103 étudiants en formation initiale (FI) ou reprise d'étude (RE).

Le terme data science recouvre un ensemble de connaissances et de compétences permettant d'exploiter de (grandes) quantités de données à l'aide d'outils d'analyse statistiques modernes. L'ensemble des méthodes employées, connues généralement sous les dénominations d'intelligence artificielle, fouille de données ou bien encore « big data » vise à extraire de systèmes complexes des informations permettant, entre autres, la visualisation, la classification et la modélisation des données. Des métiers spécifiques¹ sont apparus au cours des dernières années autour de ces thématiques, et la demande d'experts en traitement de données, les « data scientists », est en pleine croissance dans de nombreux domaines scientifiques et socio-économiques.

A l'occasion de ce deuxième renouvellement du DU Data Scientist une nouvelle UE dédiée à l'Informatique Quantique (Quantum Computing) est introduite dans la maquette. La formation complète repose sur 9 UE complémentaires se déroulant chacune sur une semaine de cours, pour un total de 202 heures d'enseignement. Les intervenants sont des chercheurs et enseignants-chercheurs de quatre laboratoires clermontois (LPC, LMBP, IP, LIMOS) et des ingénieurs d'un grand centre de calcul du CNRS (le CC-IN2P3, à Villeurbanne).

Les compétences globales proposées par cette formation sont les suivantes : maîtriser des concepts statistiques avancés, utiliser des techniques modernes de traitement des données, pratiquer des méthodes d'analyse prédictive et d'optimisation, implémenter des algorithmes d'apprentissage supervisé et non supervisé. L'ensemble de ces compétences est mis en œuvre depuis de nombreuses années dans les champs disciplinaires des chercheurs et enseignants-chercheurs qui portent cette formation et se traduisent dans les collaborations qu'informaticiens, physiciens et mathématiciens entretiennent aujourd'hui. Le Diplôme Universitaire offre des débouchés variés dans plusieurs secteurs, tels que :

- Métiers de Data Scientist (data analyst, data miner)
- Chargé d'études statistiques
- Chargé d'études prospectives et d'optimisation
- Analyste en intelligence socio-économique
- Responsable gestion et analyse de données
- Métiers de la Recherche

II – Composante porteuse du projet

École Universitaire de Physique et d'Ingénierie (EUPI)

III – Autres composantes de l'UCA participant à la formation

Les enseignements sont dispensés par des enseignants-chercheurs et des chercheurs du Laboratoire de Physique de Clermont (LPC), de l'Institut Pascal (IP), du Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes (LIMOS) et du Laboratoire de Mathématiques (LMBP). La formation s'appuie sur les recherches conduites dans ces laboratoires, en particulier elle bénéficie de l'implication de ses intervenants dans de nombreuses expériences scientifiques (certaines de dimension internationale) et de leur expertise dans le domaine des data science: analyse de grande masses de données, traitement d'image, robotique, machine learning, méthodes et algorithmes statistiques, informatique, calcul et intelligence artificielle.

IV – Partenaires de la formation

IV.1 – Autres établissements de formation partenaires

Le Centre de Calcul de l'Institut National de Physique Nucléaire et de Physique des Particules (CC-IN2P3 à Villeurbanne) est un partenaire naturel sur lequel s'adosse la formation. Ce grand centre de calcul du CNRS dispose d'une expertise et d'un savoir-faire technologique de renommée internationale, en particulier dans le domaine du calcul intensif sur grille informatique (grid computing). Une UE « Data engineering » (UE 5, voir ci-dessous) a été conçue en partenariat avec les ingénieurs du CC-IN2P3 et se déroule dans les locaux du centre.

IV.2 – Autres partenaires du monde socio-économique et associatif

aucun

V – Responsable du DU

Nom, Prénom : DONINI, Julien

Statut : PU

Coordonnées : Université Clermont Auvergne - Laboratoire de Physique de Clermont LPC+
Campus Universitaire des Cézeaux, 4 Avenue Blaise Pascal
TSA 60026, CS 60026 - 63178 Aubière Cedex
julien.donini@uca.fr, tel: +33 4 73 40 73 02 (bureau), +33 6 43 70 71 43 (portable)

VI – Composition du conseil de perfectionnement

Personnel de l'Université Clermont Auvergne

- GUILLIN, Arnaud, PU
- BARRA Vincent, PU
- MONTEIL Stéphane, PU

Intervenants extérieurs à l'Université Clermont Auvergne

- ROUSSEAU David, Expert Machine Learning, Directeur de Recherche CNRS, Lab de l'Accélérateur Linéaire Orsay
- BEAUCOURT Léo, data scientist CGI

VII – Autres renseignements

Si renouvellement, date de la dernière demande : 17/04/2020

Première année universitaire concernée par la demande : 2024 - 2025

Niveau de la formation : 2^{ème} cycle

Public visé et niveau requis pour intégrer la formation :

	Public visé par la formation	Prérequis
FI	Étudiants en Master scientifique	Niveau Bac+4 cursus scientifique Niveau B2 en anglais
RE	Ingénieurs, titulaires d'un Master souhaitant faire une reconversion vers les data sciences	
FC	Salariés d'entreprises (ingénieur, cadre). Personnel EPST (CNRS, INRA), Université	Connaissances de base en informatique et d'un langage de programmation. Connaissances mathématiques (niveau Licence). Notions de probabilités et statistique Niveau B2 en anglais

Durée de la formation : 1 année universitaire, DU (5 UE) de 112 ou 122 h selon l'UE spécifique choisie.
volume horaire total des 9 UE proposées : 202 heures

Calendrier de la formation : septembre à juin. Sessions d'examens en novembre-janvier et mai-juin

Lieu de la formation : UCA, campus des Cézeaux, Aubière. UE 5 « Data Engineering » dispensée au Centre de Calcul de l'IN2P3 (21 Avenue Pierre de Coubertin, Villeurbanne)

Ouverture en formation initiale (FI/RE) : oui

Ouverture en formation continue (FC) : oui

Tarif 1 (tarif plein)	Tronc commun UE 1,2,3 et 4 (92h)	25 € / h – 2 300€	DU (4 UE tronc commun + 1 UE spéc.) 2 800 € ou 3 050 € selon l'UE spécifique choisie possibilité de compléter la formation avec d'autres UE spécifiques
	UE spécifiques UE 5 à 8 (20h / UE)	25 € / h – 500 € / UE	
	UE spécifique UE 9 (30h)	25 € / h – 750 €	
Tarif 2 (tarif UCAA inscrits dans l'année ou l'année précédente)			DU (forfait quel que soit le nombre d'UE choisies) 450 €

Nombre minimal d'inscrits (FI-RE/FC) pour ouverture de la formation : 4 stagiaires au tarif 1 ; il est nécessaire de s'assurer de l'autofinancement de la formation avant l'ouverture

Capacité d'accueil maximale : 20. La capacité d'accueil est limitée par le nombre de postes informatiques disponibles et le nombre d'encadrants pour les séances de travaux pratiques et tutoriels

Période d'inscription / de candidature : les candidatures et inscriptions sont ouvertes jusqu'au 1^{er} septembre précédent la rentrée de la formation. Les candidatures et inscriptions aux UE proposées individuellement à la FC peuvent suivre un calendrier spécifique à chaque UE

Contact Scolarité : Sophie FOURNIER, 04 73 40 70 05, Sophie.FOURNIER@uca.fr

B – Description de la formation

I – Organisation générale de la formation

Une partie des enseignements de la formation est mutualisée avec le Master Physique Fondamentale et Applications (Master PFA) de l'UCA, également porté par l'EUPI.

Les UE forment un ensemble auto-cohérent et peuvent aussi être proposées séparément à la FC. Chaque UE est dispensée sur une période courte (généralement 5 jours consécutifs) afin de satisfaire les besoins de la FC.

La formation est composée des parties suivantes :

- Un Tronc Commun (UE 1 à 4, 92h) constitué de quatre UE entièrement mutualisées avec la 2^{ème} année du Master Physique Fondamentale et Applications. Les enseignements ont lieu au Campus des Cézeaux. Les enseignements sont répartis sur les deux semestres.
- Quatre UE spécifiques (UE 5 à 8, 20h chacune), se déroulant au Campus des Cézeaux.
- Une UE spécifique (UE 9, 30h) réalisée en collaboration avec le Centre de Calcul de l'IN2P3. Les enseignements se déroulent sur 5 jours consécutifs au CC-IN2P3 (Villeurbanne) au cours du 2nd semestre.

Il n'y a pas de différences dans le contenu des UE entre la FI-RE et la FC : les enseignements sont communs à l'ensemble du public suivant la formation et comportent les mêmes cours, séances de travaux dirigés et tutoriels.

II – Maquette de la formation et Modalités de Contrôle des Connaissances (MCC)

Chaque UE laisse une place significative aux projets et sessions pratiques sur machine à partir des serveurs Linux des salles informatiques de l'UCA (salles SCI) ou du CC-IN2P3 pour l'UE5. Les exercices et sessions pratiques sont réalisés essentiellement à l'aide de l'environnement interactif open-source Jupyter Notebook et du service web de gestion et d'hébergement de logiciels github.

L'évaluation des acquis se fera lors d'un examen terminal (examen écrit ou travaux pratiques).

Intitulé UE	Nombre d'heures				Contrôle des connaissances					
	Total	CM	TD	TP	Théorique		Pratique		Période d'examens	Coefficient de l'UE
					Nature et durée de l'épreuve	Coefficient	Nature et durée de l'épreuve	Coefficient		
UE 1 : Data Analysis with Python 1) Programmer à l'aide du langage Python. 2) Analyser des données efficacement avec numpy 3) Représenter des données avec matplotlib	25			25			Projet, TP		novembre – janvier	1
UE 2 : Advanced Statistics 1) Maîtriser des concepts statistiques avancés 2) Estimation de paramètres, test d'hypothèse 3) Inférence Bayésienne	20	15	5		Écrit 1h30				novembre – janvier	1
UE 3 : Quantum Computing 1) rappels de mathématique et d'informatique classique 2) Introduction à l'informatique quantique 3) Programmation des processeurs quantiques	27	16,5	1,5	9			Projet, TP		novembre – janvier	1
UE 4 : Machine Learning 1) Concepts de base du machine learning 2) Méthodes de régression et de classification 3) Introduction aux réseaux de neurones	20	10		10			Projet, TP		novembre – janvier	1
UE 5 : Data Mining 1) Fouille de données et Big data 2) Pre-processing, réduction de dimensionnalité 3) Classification non supervisée	20			20			Projet, TP		novembre – janvier	1
UE 6 : Statistical tools (R language) 1) Utilisation des package de traitement de données 2) Statistiques avec le langage R 3) Étude de cas pratiques	20	10		10			Projet, TP		mars – juin	1

UE 7 : Time series 1) Modélisation de tendance et saisonnalité, désaisonnalisation 2) Modélisation et prévision par processus ARMA 3) Applications avec le logiciel R	20	8	4	8			Projet, TP		mars – juin	1
UE 8 : Deep Learning 1) Entraînement de réseaux de neurones profonds 2) Architecture de réseaux pour l'analyse d'images 3) Sessions pratiques: autoencodeurs, RNN	20	10		10			Projet, TP		mars – juin	1
UE 9 : Data engineering 1) Architecture, calcul et stockage 2) Bases de données 3) Infrastructure de données distribuées	30	15		15			Projet, TP		mars – juin	1
Total	202	84,5	10,5	107						

III – Règles de délivrance du diplôme

La validation des connaissances dans chaque module est sanctionnée par un examen final. Une attestation de fin de formation est délivrée pour chaque UE validée, si l'apprenant en fait la demande.

L'apprenant doit choisir au minimum 5 UE sur les 9 proposées, et valider l'ensemble des UE choisies pour obtenir son diplôme. Cette validation peut éventuellement s'étaler sur plusieurs années consécutives dans le cadre de la Formation Continue.

Les UE se compensent entre elles. En cas d'échec à la première session, une seconde session est prévue.

Le DU ne comporte pas de stage. Néanmoins les étudiants n'étant pas inscrits en Master peuvent réaliser un projet tutoré avec immersion en entreprise d'une durée de 50h (établissement de conventions de partenariat relatif à un projet tutoré). Les projets seront suivis par l'équipe enseignante du DU (à raison de 5h de suivi par stagiaire).

IV – Compétences acquises pendant la formation

Compétences	UE 1	UE 2	UE 3	UE 4	UE 5	UE 6	UE 7	UE 8	UE 9
Maîtriser des concepts statistiques avancés		X		X	X	X	X	X	
Pratiquer des outils d'analyse de données	X	X		X	X	X	X	X	X
Programmer avec le logiciel R						X	X		
Programmer en Python	X	X		X	X			X	X
Utiliser un Notebook de type Jupyter pour analyser et représenter les données	X	X		X	X			X	X
Appliquer des méthodes d'analyse prédictive		X		X	X	X	X	X	X
Savoir construire des modèles statistiques		X		X	X	X	X	X	X
Appliquer des algorithmes de régression et de classification		X		X	X	X	X	X	X
Concevoir et entraîner un réseau de neurones supervisé				X				X	
Savoir implémenter un réseau de neurones sur une architecture GPU				X				X	X
Savoir utiliser des techniques de fouille de données					X				X
Comprendre les problématiques liées au calcul	X							X	X
Savoir choisir le stockage le plus adapté									X
Connaître les concepts de base des séries temporelles							X		
Savoir modéliser des séries temporelles							X		
Savoir utiliser les méthodes d'informatique quantique			X						
Savoir évaluer les perspectives d'une implémentation expérimentale			X						
Savoir élaborer un algorithme quantique			X						
Savoir utiliser un ordinateur quantique à distance			X						

V – Équipe pédagogique

L'équipe pédagogique élargie comprend des enseignants-chercheurs de l'UCA (LPC, IP, LIMOS, LMBP), des chercheurs et ingénieurs CNRS et des intervenants extérieurs (CC-IN2P3, notamment).

Nom Prénom	Etab. de rattachement	Statut	Qualité	N°UE	Matière	Volume horaire (CM/TD/TP)
MADAR Romain	LPC	CNRS	Chargé de recherche	1		25h TP
DESCHAMPS Olivier	UCA (LPC)	MCU		2		15h CM/5h TD
DONINI Julien	UCA (LPC)	PU		3		10h CM/10h TP
FALIH Issam	UCA (LIMOS)	MCU		4		20h TP
GADRAT Sébastien	CC-IN2P3	CNRS	Ingénieur recherche	5	Calcul scientifique, Python et Notebooks	3h CM / 3h TP
AIDEL Osman	CC-IN2P3	CNRS	Ingénieur recherche	5	Bases de données	2h CM / 2h TP
BOUVET David	CC-IN2P3	CNRS	Ingénieur d'étude	5	Ferme de calcul	2h CM / 2h TP
SUTER Frédéric	CC-IN2P3	CNRS	Chargé de recherche	5	Notebooks, Jupyter et Anaconda	2h CM / 2h TP
TORTAY Loïc	CC-IN2P3	CNRS	Ingénieur d'étude	5	Architecture et stockage	2h CM / 2h TP
HERNANDEZ Fabio	CC-IN2P3	CNRS	Ingénieur recherche	5	Encadrement	2h CM / 2h TP
RIGAUD Bertrand	CC-IN2P3	CNRS	Ingénieur recherche	5	Encadrement	2h CM / 2h TP
MBAYE Papa	Missioneo	Ext.	Docteur en maths	6		10h CM/10h TP
AZZAOUI Nourddine	UCA (LMBP)	MCU		7		8h CM/4h TD/8h TP
CHATEAU Thierry	UCA (IP)	PU		8		10h CM
TEULIERE Céline	UCA (IP)	MCU		8		10h TP
VULPESCU Bogdan	LPC	CNRS	Ingénieur recherche	9		9h TP
KRIEGLER Christoph	UCA (LMBP)	MCU		9		4,5 CM/1,5h TD
MALPUECH Guillaume	IP	CNRS		9		10h CM
SOLNYSHKOV Dmitry	UCA (IP)	PU		9		2h CM

Nombre d'heures assurées par des personnels UCA : 108h = 49,5h CM + 10,5h TD + 48h TP (UE 2, 3, 4, 5, 7, 8)

Nombre d'heures assurées par des intervenants extérieurs à l'UCA : 94h = 35h CM + 59h TP (UE1, 3, 6, 9)

VI – Composition du jury

Un jury est nommé pour l'année universitaire par le Président de l'Université sur proposition de la composante qui porte la formation.

Le jury proposé est composé de personnes prises parmi l'équipe pédagogique du DU. Il s'agit de :

- MADAR Romain (LPC, CR)
- DESCHAMPS Olivier (LPC, MCU)
- DONINI Julien (LPC, PU)
- GADRAT Sébastien (CC-IN2P3, CNRS)
- MBAYE Papa (Extérieur)
- AZZAOUI Nourddine (LMBP, MCU)
- TEULIERE Céline (IP, MCU)
- VULPESCU Bogdan (LPC, CNRS)
- KRIEGLER Christoph (LMBP, MCU)
- SOLNYSHKOV Dmitry (IP, PU)
- MALPUECH Guillaume (IP, CNRS)

VII – Modalités d'admission

L'admission (FI-RE et FC) repose sur un dossier de candidature détaillé : CV, lettre de motivation, relevé de notes des années précédentes pour la FI-RE, descriptif des compétences en informatique et mathématiques pour la FC. Des entretiens individuels pourront également être mis en place si nécessaire.

VIII – Mobilité à l'international

sans objet

IX – Pour les formations de santé : Développement Professionnel Continu (DPC)

sans objet

Avis du Conseil de Gestion de la composante : favorable à l'unanimité le 16/02/2024

Approbation du Conseil de la Commission de la Formation et de la Vie Universitaire le 09/04/2024

Validation accordée pour une durée de 3 ans, jusqu'à fin 2026 - 2027

Signature du Vice-Président Formations, en charge de la CFVU



Compte-rendu du Conseil de l'EUPI

Séance du 16/02/2024

MEMBRES PRESENTS : Mmes BADAUD-GARDY Frédérique, GREIVELDINGER Christelle PALADIAN Françoise et PLANCHE Céline à/c de 10h30 et MM. BEAUCOURT Léo, LEFEVRE Régis, MARTIN Franck, jusqu'à 10h, NDIAYE Amadou, ORLOFF Jean à/c de 10h00, VIEY Alexis et TRASSOUDAINÉ Laurent, président sans voix délibérative.

MEMBRES EXCUSES : Mmes GENESTY Christelle et MM. GREVOUL Stéphane et SARRY Laurent.

PROCURATION de M. MENECIER Sébastien à Mme BADAUD-GARDY Frédérique à/c de 10h00

INVITES : Mme FOURNIER Sophie et M. PERISSE Frédéric

Laurent TRASSOUDAINÉ ouvre la séance à 9h30 puis présente l'ordre du jour.

3. Evolution de l'offre de formation

a) Evolution et renouvellement du DU Data Scientist (Julien Donini)

Le renouvellement du DU doit être soumis au CFVU du 09/04.

Julien DONINI dresse le bilan de cette formation qui a ouvert en 2017.

Laurent TRASSOUDAINÉ soumet le renouvellement du DU au vote du conseil.

RESULTATS : POUR à l'unanimité.