The Alliance for Data Science Professionals

Certification Guidance and Process:

Advanced Data Science Professional

Introduction

All prospective applicants are advised to read carefully through this document before submitting an application with their selected Alliance member.

In this document we provide guidance notes and outline the process by which an individual can apply for **Advanced Data Science Professional** status, together with the standards against which applications are assessed.

Standard and Breadth of Knowledge

The Data Science Standard has five main Skill Areas, A-E. Evidence of how ethics and efficacy* are considered should be embedded in the information provided in each of the Skill Areas. The Skills Areas are as set out below.

Please see the Breadth of Knowledge section for guidance on Ethics & Efficacy

Skill Area	Evidential Requirements
A. Data Privacy and Stewardship *This skill relates to the security and protection of data, including design, creation, storage, distribution and associated risk.*	*1. Ensuring the protection of personal and sensitive data.* *2. Managing sensitive data.* *3. Data Stewardship and Standards.*
B. Definition, acquisition, engineering, architecture, storage, and curation. *This skill relates to the collection, manipulation and secure storage of data, the application of data management and analytical techniques.*	*1. Data Collection and Management.* *2. Data Engineering.* *3. Deployment.*
C. Problem definition and communication with stakeholders *This skill is about engaging stakeholders, demonstrating the ability to clearly define a problem and agree on solutions.*	*1. Problem definition.* *2. Relationship management.*

D. Problem solving, analysis, statistical modelling, visualisation. *This skill relates to the identification and presentation of solutions using a range of methods, tools and techniques, demonstrating the ability to analyse a problem and define and present options.*	*1. Identifying and applying technical solutions and project management approaches.* *2. Data preparation and feature modelling.* *3. Data Analysis and Model building.*
E. Evaluation and Reflection *This skill is about reflecting on performance and outcomes, identifying development needs and applying important principles associated with ethics and sustainability.* ***Note:*** *when completing your evidence for this Skill Area you can refer to evidence provided in the Skill Areas A-D, together with ensuring that ethical evaluation is reflected throughout Skill Areas A-D.*	*1. Project Evaluation.* *2. Ethical behaviour.* *3. Sustainability and Best Practices.* *4. Reflective Practice and Ongoing Development.*

See Appendix 1 for an expanded version, including types of suggested evidence and related skills.

When applying for the Advanced Data Science Professional, it is expected that all applicants:

- can provide evidence of **deep** understanding and knowledge in Skill Area **E** and at least **two** other Skill Areas.

  I have a deep understanding based on a mixture of training and experience of undertaking complex work activities. I am accountable, either in part or fully, for production and decisions made based on the work I do or oversee.

- can display an appropriate level of **general** knowledge and competence for the remaining **two** Skill Areas.

  I have a general understanding based on a mixture of training and experience of undertaking some of these work activities.

Level Descriptors

Level	Description
General	Has factual, procedural and theoretical knowledge and understanding. Can interpret and evaluate relevant information and ideas. Is aware of the area of work. Is aware of different perspectives or approaches within the area of work.
Deep	Has advanced practical, conceptual or technological knowledge and understanding of the field of work, enabling the applicant to create ways forward in contexts where there are many interacting factors. Understands different perspectives, approaches or schools of thought and the theories that underpin them. Can critically analyse, interpret and evaluate complex information, concepts and ideas.

Breadth of knowledge: Ethics & Efficacy

It is important that all professionals working within the field of Data Science have a clear understanding of the ethics which underpins the collection, management, use and communication of the data and the results with which they work. It is equally important that a Data Scientist takes responsibility for the assurance of the models they build. Assurance covers both the efficacy of the application and the ethical natures of its design and implementation. As such, these attributes are not something that can, or should, be assessed as one standalone criterion. Rather, when completing this application, you should wherever possible include reference to your knowledge and working practices relating to the appropriate **ethical** considerations such as:

- Data: collection, validity for use in the intended purpose, permission for usage, storage, security
- Model: development, testing (e.g. fairness, bias, error rates) usage (how could the model and results be used for an unintended purpose?) and transparency
- Communication: explanation of why the science is required; the results achieved and how misinterpretation of the results can be minimised.
- Relevant laws and permissions of usage for data (including legal rights of individuals, privacy and anonymity)

And **efficacy** considerations such as:
- Quality assurance of code and data
- Validation of model fit
- Robustness of the model and software implementation
- Ongoing monitoring of model implementation

It is important to note that the list is not exhaustive. It is here to serve as a guide to help you show the assessors you are aware of the professional expectations of those who work in this field. You should include any other areas of ethical and efficacy considerations you feel are important within your area of expertise.

Data Science – Level Definitions

Whilst there are two levels of certification associated with the Data Science standards, the standards will remain generic statements that can apply across a wide range of roles within the data science field.

The distinguishing features that define the levels are thus associated with the application of the standards and therefore the levels of competence.

The distinction between Data Science Professional and Advanced Data Science Professional level will be related to the following:

- **Responsibility** – the higher the level of registration the greater level of responsibility and accountability.

- **Decision making** – the level of authority to make decisions and the impact across the organisation.

- **Complexity** – this can be delivered within two spheres of complexity:

  - *Technical Complexity* – specifically associated with the technical skills applied.

  - *Organisational Complexity* – associated with skills and decision-making responsibilities that would apply across, and potentially beyond, the organisation.

- **Business impact** – relating to how far-reaching actions may apply and impact, and understanding how and on whom they impact.

Levels of Accountability and Responsibility

For those wishing to apply for the Advanced Data Science Professional, we expect an individual to provide suitable evidence within their application that they are fully accountable and responsible for their activities and that of others. Typically, these individuals would also demonstrate some additional levels of responsibility, awareness and understanding within their role/working activities **Applicants should demonstrate that they:**

- Are fully accountable for their own work and that of others, including ethical considerations
- Undertake a range of complex work activities that have a significant impact.

They should also demonstrate some of the following:

- Have responsibility for a significant function across the organisation.
- Have significant decision-making authority within their given area of expertise.
- Act as an advisor/consultant at a strategic level.
- Consider the impact across the business and, more widely, of actions undertaken based on their decisions.

Requirements & Flexibility

Dependent on the role, areas of strength and weakness within the standards may differ, for example:

- As people progress towards Leadership/Management positions so their responsibility will increase, but technical complexity may diminish.

- Others may not progress towards Leadership roles but will develop highly complex and valuable technical skills.

Therefore, it is **not** essential that an applicant meets **all** the criteria at the higher level, but that on balance, the **totality** of their evidence for each section meets the required level.

In addition, different roles at the same level will have differing levels of competence within the Skill areas defined in the standard. For example, a Data Engineer may have strong evidence against Skills Area B but less developed evidence against Skills Area D.

Taking this into consideration, it is expected that all applicants (at either level) can display an appropriate level of competence for Skill Area E. However, sections A, B, C and D may be weighted differently depending on the area of specialism.

At the Advanced Data Science Professional level, an applicant will be expected to deliver a deep level of competence for Skill Area E plus two other Skill Areas. They should also demonstrate a general knowledge of the remaining two Skill Areas.

Application Process

Applicants applying to be a certified Advanced Data Science Professional must do so via the competency-based route.

Please be aware that whilst different Alliance members may have differences in how an application is assessed, the information requested and assessed remains the same.

Applicants should be educated to at least Level 6 within the UK qualifications framework, in an appropriate subject and evidence should be provided of at least some formal training within data science, either as part of their degree or achieved by other methods. In addition to this, applicants must meet all other criteria listed below:

- typically have five years' relevant work experience
- provide at least two years' evidence of CPD
- Can evidence they meet the competencies and level of responsibilities of the Advanced Data Science Professional certificate via the chosen assessment method of the awarding society.

An individual applies, meeting all of the requirements of this route to the appropriate Alliance member	→	The application is reviewed by the assessors	→	The individual is notified of the outcome

Those individuals applying via the competency route will need to complete all sections of the application to an appropriate level.

- Personal information (if not known already)
- Academic/training history (including copies of transcripts and certificates)
- Competency-based and responsibility-based statements
- Details of relevant work experience
- A completed CPD document

Application form

The Application form for the Advanced Data Science Professional certificate is typically split into five sections. Please be aware that the information you provide and how this information is assessed may vary between different Alliance members, so please ensure you have reviewed the requirements of the member body with which you are applying.

Section 1	Personal Information	Name, address, memberships etc…	This section may be auto completed
Section 2	Academic/training history	Certificates Transcripts Name of institution/training provider Dates Relevant modules	Failure to provide evidence of both certificates and transcripts may result in applications being rejected without further scrutiny
Section 3	Competency & Responsibilitybased questions	Competency statements/Personal statements that demonstrate you've worked in one or more challenging, multifaceted roles where you've had full responsibility, working under broad direction with significant influence. Indicate how you exhibit vision and appreciation of the overall context of your role, and the role of a Data Science professional, in achieving successful results.	In addition to information provided within an application, you may be asked for references, others may cover this at an interview.
Section 4	Work experience	Brief details of work experience from the last five years, including tasks undertaken and level of responsibility.	
Section 5	CPD	A completed document highlighting how you maintain your professional development in accordance with the CPD guidelines of the appropriate membership body.	

Assessment of Applications

Once you have provided both information and evidence within the five sections below, applications will be passed to the relevant assessors, who need to be able to clearly identify where and how you have met the Standard at the level of responsibility, awareness and understanding required.

Section 1	Personal Information
Section 2	Academic/training history
Section 3	Competency & Responsibility-based questions
Section 4	Work experience
Section 5	CPD

Whilst it is expected that most of those individuals applying at this level would have met a number of the criteria via their academic and professional training, Assessors will be able to identify any skills and knowledge gaps that have been met via a mixture of work experience, professional competencies and CPD. For a full list of suggested evidence in addition to academic and professional training, please see an extended view of the Data Science Standard in Appendix 1.

Example application

The competency-based Route:

Section 1	Personal Information	Sam Smith
Section 2	Academic/training history	Degree – BSc in Mathematics & Statistics Certificates - attached Transcripts - attached Name of institution/training provider – The Open University Dates – qualified May 2018 Relevant modules – M348 Applied Statistical Modelling
Section 3	Competency & Responsibilitybased questions	In this section, we're looking for evidence relating to the three Skill Areas selected, that demonstrates you've worked in one or more challenging, roles where you've had specific responsibility(ies), working under broad direction with significant influence. Indicate how you critically evaluate and exhibit understanding of the overall context of your role, and the role of a Data Science Professional, in achieving successful results. ***Note: be sure to include ethical issues that you have had to consider, resolve or act upon within your evidence.***
Section 4	Work experience	Employer 1 - Data Analyst August 2019 – Present Level of responsibility –I had responsibility for daily tasks and led an analytics team that analysed and cleaned data for use by our research team Brief description of my role: XXXXX
Section 5	CPD	Please see a CPD activity summary attached from the two years, including 120 hours of CPD.

This is an example of an application form template, please note the ordering and wording may vary between Alliance members

Please note that based on the information provided and other certificates the individual may be applying for at the time of submitting this application, the reviewing body may wish to request more information that can include, but is not limited to:

- Shorter application questions, followed by an interview
- Longer application questions only
- A mixture of both of the above
- Personal statements
- Short assessment/accredited certificate

Appendix 1– Extended version of the Standard

Skill Area	Evidential Requirements	Types of evidence
A. Data Privacy and Stewardship This skill relates to the security and protection of data, including design, creation, storage, distribution and associated risk.	1. Ensuring the protection of personal and sensitive data.	i. Assess risks and enact data protection policies and procedures. ii. Ensure safe and secure management of sensitive data, models and infrastructures. iii. Apply appropriate data controls, such as encryption, (pseudo)anonymisation, and synthetic data. iv. Risk management around environment and infrastructure.
	2. Managing loss of sensitive data.	i. Act with integrity, giving due regard to legal and regulatory requirements. ii. Be aware of the actions that should be taken to respond to potential data loss in line with organisational, legal and regulatory procedures.
	3. Data Stewardship and Standards.	i. Incorporates the FAIR Guiding Principles for scientific data management and stewardship into practices, where appropriate and practicable. ii. Identify opportunities for efficient and creative reuse of data. iii. Understand the relationship between technical standards and regulation/governance, and their benefits for interoperability and knowledge sharing.
B. Definition, acquisition, engineering, architecture, storage and curation. This skill relates to the collection, manipulation and secure storage of data, the application of data management	1. Data Collection and Management.	i. Source and access data appropriate for the problem. ii. Critically analyse the availability of appropriate data and resources to meet project requirements. iii. Critically evaluate and synthesise data. iv. Ensure data provenance processes are followed v. Identify data characteristics (volume, velocity and variety). vi. Identify infrastructure requirements for data storage and analysis. vii. Show familiarity or experience with tabular and non-tabular data (e.g. unstructured and streaming data).

and analytical techniques.	2. Data Engineering.	i. Source and access data appropriate for the problem. ii. Construct data sets, potentially drawing from multiple disparate sources using data linkage. iii. Perform data profiling and characterisation to understand the surface properties of the data. iv. Handle missing data, through principled inclusion/exclusion criteria and imputation methods. v. Take a systematic approach to data curation and the application of data quality controls. vi. Identify the most appropriate solutions (e.g. cloud vs on-premise) in response to business and project needs.
	3. Deployment.	i. Plan the deployment of data products with their end-users. ii. Develop monitoring and maintenance processes. iii. Deliver secure, stable and scalable data products to meet the needs of the organisation, e.g. Application Programming Interfaces (APIs), derivative datasets, dashboards, reports and do so according to modern software development best practices. iv. Design and deliver data products that meet appropriate accessibility standards for their users.
C. **Problem definition and communication with stakeholders** This skill is about engaging stakeholders, demonstrating the ability to clearly define a problem and agree on solutions.	1. Problem definition.	i. Identify and elicit project requirements. ii. Determine success criteria and frame these in the context of the business. iii. Clearly articulate the problem statement. iv. Identify and critically evaluate assumptions. v. Recognise and quantify biases and identify solutions to manage and mitigate these. vi. Assess risk. vii. Demonstrate sector/domain knowledge and knowledge of how data science can deliver value to these sectors/domains.

	2. Relationship management.	i. Communicate in an effective manner for diverse audiences, including technical colleagues, subject matter experts and leadership.
		ii. Effectively manage the expectations of diverse stakeholders with conflicting priorities to mediate equitable solutions. iii. Use relevant communication techniques (written, oral or visual), appropriate for the audience. iv. Build appropriate and effective business relationships. v. Show experience in human factors considerations with respect to data-driven solutions.
D. Problem solving, analysis, statistical modelling, visualisation. This skill relates to the identification and presentation of solutions using a range of methods, tools and techniques, demonstrating the ability to analyse a problem and define and present options.	1. Identifying and applying technical solutions and project management approaches.	i. Identify viable solutions based on requirements and data available. ii. Identify and provide guidance to technical and non-technical stakeholders on the most appropriate solution. iii. Apply appropriate technical and project management methodologies appropriate for the organisation and project.
	2. Data preparation and feature modelling.	i. Identify appropriate solutions, including statistical and machine learning approaches and demonstrate an understanding of the assumptions, strengths and weaknesses of the selected approaches. ii. Identify and evaluate appropriate evaluation metrics, including computational performance and accuracy. iii. manipulate data with due regard for differences in characteristics. iv. Creation and evaluation of new data features.

	3. Data Analysis and model building.	i. Apply appropriate solutions, including statistical and machine learning approaches. Demonstrate competence in a modern programming language. ii. Use appropriate analysis platforms and tools. iii. Adopt a systematic approach to exploratory data analysis to embrace and manage ambiguity and uncertainty. iv. Critically analyse data and analytical results. v. Adopt appropriate methods to visualise data and communicate complex findings.
E. Evaluation and Reflection This skill is about reflecting on performance and	1. Project Evaluation.	i. Ongoing monitoring of project performance and outcomes. ii. Identify and feed forward lessons learned. iii. Participate in and lead collaborative project evaluations, e.g. retrospectives.
outcomes, identifying development needs and applying important principles associated with ethics and sustainability. Note: when completing your evidence for this Skill Area you can refer to evidence provided in the Skill Areas A-D, together with ensuring that ethical evaluation is reflected throughout Skill Areas A-D.	2. Ethical behaviour.	i. Identify and manage the risks of erroneous and biased data. ii. Act with integrity with respect to legal and regulatory requirements. iii. Uphold principles of ethical and safe use of data and AI technologies. iv. Implement data use procedures to ensure sensitive data is only used for its agreed purpose. v. Implement data retention strategies in line with regulatory and legal requirements.
	3. Sustainability and Best Practices.	i. Evidence of incorporating the principles of open science and/or reproducible research within the organisation, and perhaps beyond. ii. Competence in programmatic approaches to undertaking data science work. iii. Apply the scientific method in delivering solutions. iv. Ensure high technical standards, in line with software development best practices; for example, software testing, version control, Continuous Integration and Continuous Delivery. v. Apply automation to promote reproducibility analyses

4. Reflective Practice and Ongoing Development.	i. Learn from experience through selfassessment of one's own responses to practice situations. ii. Identify learning opportunities to maintain knowledge and skills in the relevant area of data science. iii. Take ownership of ongoing professional development. iv. Contribute to knowledge-sharing across their organisation and/or the wider community. v. Contribute to the management and empowerment of the broader team. vi. Engage with the latest developments across industry and academia and incorporate these into solutions.