

資訊及通訊科技業 《能力標準說明》 能力單元

「數據科學」職能範疇

名稱	為確定內部或外部源，管理和實施不同的數據採集選項
編號	111137L6
應用範圍	管理和實施已識別數據源（內部和外部）的不同數據採集選項。這些選項包括收集新數據、轉換/轉換舊數據、共用/交換數據以及通過訂閱或數據即服務解決方案購買數據
級別	6
學分	6（僅供參考）
能力	<p>表現要求</p> <ol style="list-style-type: none"> 1. 了解不同的數據採集選項 <ul style="list-style-type: none"> • 了解數據描述、各種數據採集和整理技術（見備註1） • 評估和管理各種數據採集選項 2. 管理和實施不同的數據採集選項 <ul style="list-style-type: none"> • 使用不同的技術從所有已識別的內部和外部來源（結構化或非結構化數據）獲取數據，例如 ETL 批處理、流式攝取、抓取器、API和crawlers • 利用各種數據整理技術，例如複雜和多源數據的解析和算法 • 意識到獲取內部和外部數據的隱私和道德影響 3. 展會專業性 <ul style="list-style-type: none"> • 能夠： <ul style="list-style-type: none"> ○ 應用行業典範做法來管理和實施不同的數據採集選項 ○ 遵守機構的指導方針和程序以及任何（當地和國際）法律和法規要求（如果適用）
評核指引	<p>此能力單元的綜合成效要求為：</p> <ul style="list-style-type: none"> • 通過各種數據採集和整理技術收集適當的最新數據，以獲得有效、準確解釋，並適用於與已識別數據源相關的所需的數據
備註	<ul style="list-style-type: none"> • 1. 各種數據採集選項（針對第一方、第二方和第三方數據）包括（但不限於）新數據收集、遺留數據轉換/轉換、數據共享/交換、數據訂閱和/或數據即服務解決方案。這包括自動收集（例如，感測器衍生的數據）、手動記錄經驗觀察以及從其他來源獲取現有數據。