

循環不息—— 解析海洋數據庫

李杰

Mathematics is the language in which God has written the universe.

（數學是上帝撰寫世界的語言）

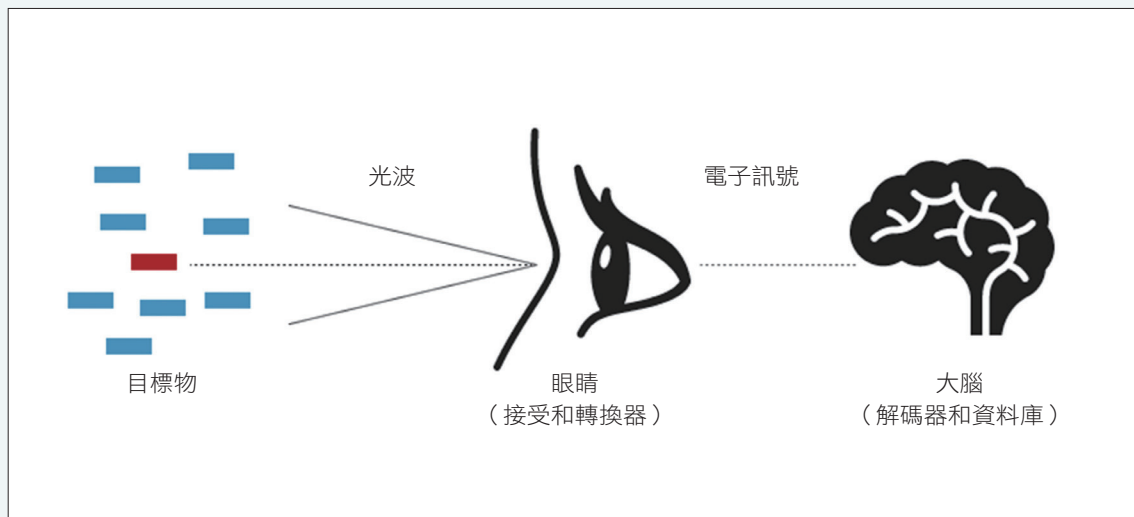
—— Galileo Galilei (1564–1642)

16、17 世紀間知名的數學天文學家伽利略曾說：「這世界是由許多數學符號和艱深的方程式所組成。」這些「數字」與我們日常的生活息息相關，甚至影響我們許多重要的決策和行為，如降雨機率超過 50% 時，我們就會考慮攜帶雨具出門；空氣品質指標（air quality index, AQI）若超過 101 時，因會對敏感族群不利，所以出門就應攜帶口罩；又如海邊出現了 5 公尺的長浪，進行水上活動的人就應提高警覺。但該如何客觀地解釋這些充斥於周遭的語言，並把它們寫進歷史的紀錄簿裡，則是個相當有挑戰性的工程。

說到「解釋」，其實和人體感受環境的方式有關。就像我們的眼睛能看到色彩，是因眼睛的感光細胞接收到光波後，把它轉成電子訊號，再透過神經網路傳輸至大腦以解釋這視覺資訊。以電腦科學來類比這個過程就是：光波轉換成的電子訊號經過傳輸系統傳達至中央處理器後，經解碼



透過神經網路傳輸至大腦，讓眼睛能看見色彩。
（圖片來源：種子發）



光波轉換成神經訊號示意圖；改編自網路文章 Design: Seeing Without Thinking (圖片來源 : <https://medium.com/@jonmyers/design-seeing-without-thinking-783d018bb82f>)

並進行影像處理與運算程序。而這些進入大腦的資訊會根據自身意識決定是否要儲存在經驗中（資料庫），以供未來使用。

那麼，在神祕的海洋世界中，科學家如何記錄那些看不見的數字，像是溫度、壓力、鹽度等，又以什麼樣的方式記錄和儲存這些數據呢？現在就來介紹海洋量測最基本的 3 種方式，分別是遙測、現場觀測，以及模式模擬。雖然，每種方式都各有優缺點，也有不同的資料格式和記錄的參數，但它們最終的目標都是為了能從海洋這個大數據庫中擷取重要的資訊。

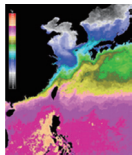
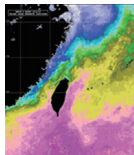
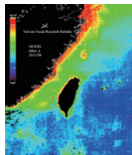
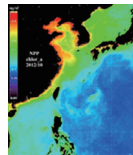
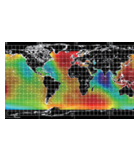
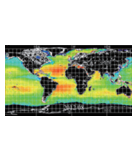
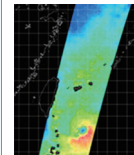
神奇的望遠鏡—遙測

所謂遙測，就是用間接的方式獲取目標物狀態的方法，其優點是在短時間內就能獲取目標物大範圍的資料，因此對於即時觀測的作業，遙測技術是個很強大的利器。

海洋界遙測技術的重要利器是離地表幾百公里上方的人造衛星，較知名且服役中的人造衛星有：可記錄海水顏色顏色的 Aqua-1 和 Terra-1、可記錄海表面水溫的 MetOp-A 等。

一般大眾都知道，人造衛星通過不同波長的電磁波來接收目標物的訊號，但為什麼是藉由電磁波接收呢？其實就跟生物一樣，海水會因水體性質的改變（如溫度變高），而對不同波長的電磁波有不同的反射量。就像優養化的海域中海水的顏色會較偏綠，因此波長 495 至 570 奈米的電磁波會比其他可見光波長強，人造衛星只要準確地捕捉到這些能量就可以了。其後會把這個電磁波訊號轉化為有意義的數據，並藉由不同的格式儲存（例如 HDF5、NetCDF 等）在記憶體中，再傳回陸地上的接收站。

遙測資料庫簡介

種類	水溫 SST		水色		海面高	鹽度 SSS	風場
衛星名稱	NOAA 系列	MetOp-A	Aqua-1 Terra-1	VIIRS (NPP)	Jason-2	Aquarius	Ascat
資料來源	海大接收站	中央氣象局接收站	http://Oceancolor.gsfc.gov.nasa	http://Oceancolor.gsfc.gov.nasa	ftp://ftp.nodc.noaa.gov/pub/data.nodc/jason2/	http://Oceancolor.gsfc.gov.nasa	ftp://podaac-ftp.jpl.nasa.gov/allData/
解析度	1km	1km	1km	750m		100km	12.5km
資料格式	Terascan® data format	Terascan® data format	HDF-eos	HDF5	NetCDF	HDF5	Netcdf
資料時間	2000 年～迄今	2011 年 8 月～迄今	2000 年 3 月～迄今	2012 年～迄今	2013 年～迄今	2013 年～2015 年 5 月	2012 年～迄今
累計資料量	1.1TB	987GB	1.85TB	1.6TB	214GB	111GB	158GB
產出頻率	每日更新 (延遲一天)	每日更新 (延遲一天)	每日更新 (延遲一天)	每日更新 (延遲一天)	每日更新 (延遲二天)	每日更新 (延遲四天)	每日更新 (延遲二天)
影像圖							

海浪滔滔—現場觀測

現場觀測則跟遙測的概念相反，通常是直接到目標物的所在地去量測數據，因此資料的時間與空間的解析度較遙測高，也可直接測量海水面下的目標參數。當科學家想了解小尺度（像是 1 公里內）或海水垂直方向的溫度變化時，首選的方式就是攜帶具可信度的溫度計，搭著研究船直接到現場布放測量。

由於出海實地調查是很艱辛的任務，除了受天氣影響外，還要克服暈船的難過，能勇於接受這些挑戰便成了一個稱職的海

洋研究員的基本條件。然而科學家另發展了一套自動化的錨碇觀測系統，以解決因海象太差而無法出海作業的困擾，同時能達到長期觀測的目的。

錨碇系統的運作架構是先固定一串搭載著測量儀器的觀測浮標，使其能長期地在目標位置上記錄特定參數的時間序列變化。而隨著科技的發展，錨碇系統甚至可配合衛星通訊的技術，把浮標上記錄到的氣象資料封包成資料檔，再傳送至陸地的接收站或資料庫保存，甚或使用者手中，除可節省人力和時間的成本外，也方便了科學家能即時分析錨碇觀測站的數據。

上帝的盒子—模式模擬

模式是以描述環境現象的理論方程式組（如各種的能量守恆方程式），來模擬自然系統運作的數學模型。通常模式的建置需要幾個元素：初始條件、邊界條件、模型方程式（或稱理論方程式）。

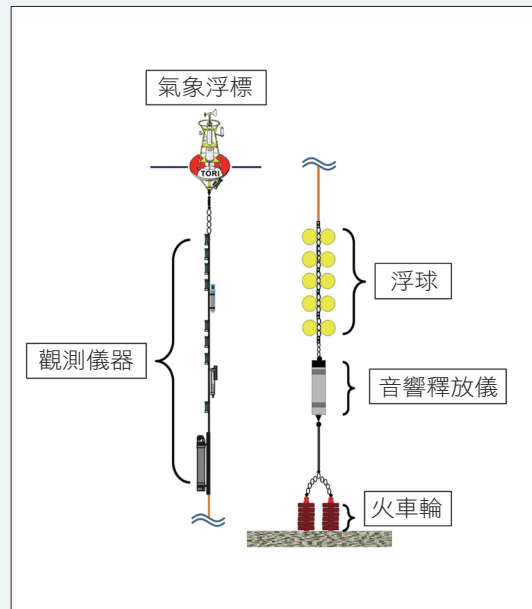
以燒開水為例，初始條件就像是水的起始溫度；邊界條件則是鍋子邊緣的溫度分布；再配合基本的熱力學方程式（模型方程式），就可以模擬水在加溫過程中的溫度變化。而在現實世界中，初始條件的輸入數值通常會依使用者所欲觀測的現象，或藉由現場所量測的數據來定義，至於邊界條件則與觀測區域的環境有關。

目前比較知名的模式套裝軟體有美國的 HYCOM（HYbrid Coordinate Ocean Model）、POM（Princeton Ocean Model），以及海洋中心所發展的 TaiCOM 和 TOPS 海象模擬模式。這些軟體都可以根據使用者的需求，提供不同時空尺度下海洋環境中物理和生地化參數的變化。

資料保存的祕方

上述所提到的觀測方式，在經年累月地記錄下往往會產生非常大量的數據，例如：一個研究航次可能就會收集到 1 GB（giga bytes）的海流資料。因此傳統上，原始資料會先複製到便宜且攜帶方便的磁帶中備份保存。而資料在保存的程序上除了從儀器本身下載至磁碟的方式，和前述所提到的備用磁帶外，有時還會把資料儲存在異地當作備援，以防止原儲存空間發生天然或人為災害後，科學家會有失去這批珍貴資料的風險。

到目前為止，我們已學習如何以儀器記錄海洋中多元的數據，並存放至硬碟中。



錨碇串示意圖

但就像收藏家一樣，當科學家的搜集物品變多時，總需要有個獨立的空間來安置這些收藏品。而數據庫就像一個可以塞滿大量數據的圖書館，可幫助我們在日後使用、提取和儲存的作業更為便利。

數據的寶庫

一間架構完整的圖書館除了須擁有豐富的藏書外，本身也要有強大的歸納和搜尋系統，讓讀者在尋找特定的書籍（數據）時，可以有效率地取得。因此資料的統整、分類、搜尋、入庫（輸入）和外借（輸出）等作業，對於資料庫管理系統的建置相當重要。

資料庫運作的類型可分為關聯性資料庫和非關聯性資料庫兩種。前者運作的規則是利用統計學中的交集、聯集、差集等方式，來擷取所需求的數據。而後者剛好相反，其概念是「有什麼，就塞什麼」，



資料庫的建置可以有效率地取得數據（圖片來源：種子發）

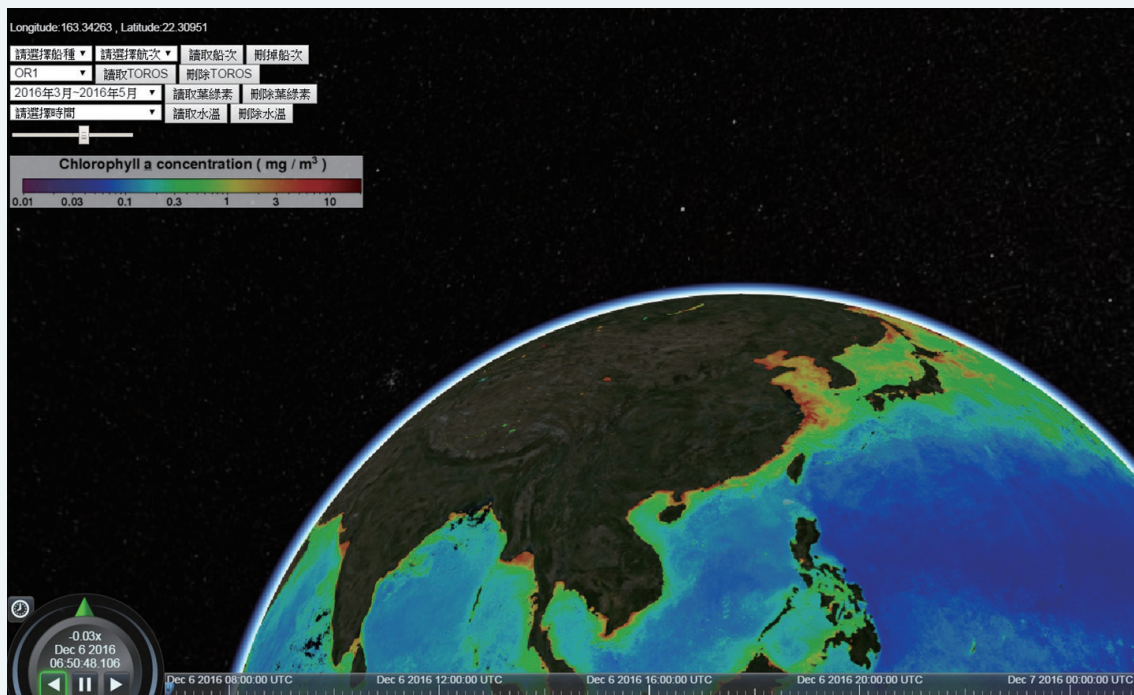
不需要分析輸入數據的格式和型態，只需給予一個空間，把要儲存的東西放進去就行了。

關聯性資料庫是由兩個或兩個以上的資料表格所組成，每個表格都包含許多欄位，每個欄位也各自擁有獨立的屬性，呈現的方式就像是常用的 Excel。它叫做「關聯性」資料庫，是因為表格之間產生關聯的方式，是藉由彼此相同的欄位來鍵結。其優點是節省了因重複輸入而浪費的儲存空間，也降低因多次輸入而誤植的機率，並使資料達到一致性。

舉個例子來說，每個人各自的住家地址都會依照縣市、區域、路、巷、弄，以及號的方式依序排列，但如果把地址排列起來，就會發現縣市和區域不斷地重複出現，甚至還有寫錯的狀況。這時如果擁有關聯性的資料庫表格，就可以有效地解決前述的尷尬，還能節省工時。



關聯性資料庫表格的示意圖



葉綠素 3D 示意圖

為了能夠在資料庫中有效地搜尋和擷取數據，資料屬性表的建置相當重要。資料分析學家必須先研究資料本身的「特性」，找出所有資料的共通點及相似處來建立儲存表格的格式（像是文字格式、數字格式、日期格式等）。除此之外，在建立資料的屬性表時，也必須整理及合併表格中重複定義的欄位（例如出生年和歲數可以整合），以簡化資料表。

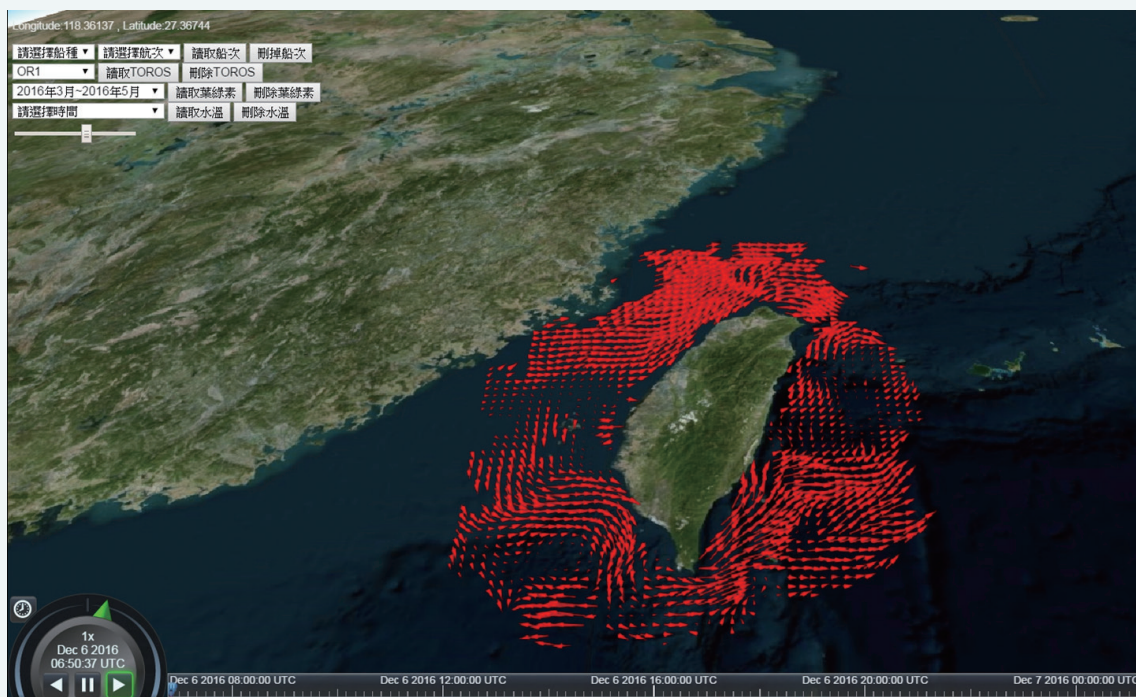
就像圖書館的書需經過審核才能入庫一般，數據進入資料庫前也必須進行品管。國際知名的全球溫鹽剖面計畫（Global Temperature Salinity Profile Project，簡稱 GTSP）就利用了 5 個資料檢測的方式給數據不同的標籤（Flag），提供使用者資料品質的參考。這有點像是電子公司的檢測產線，每個工程師各自檢查電子產品的不

同部位，然後總結物品不良的狀態來決定是否要淘汰它，或當作 NG（no good）商品處理。

數據現身—視覺化的世界

科學家夜以繼日地蒐集存在於海洋環境中的各種數據，就是希望能讓政府單位、學術機構，或一般民眾在擬定政策或採取行動時可以有客觀的參考。因此，把記錄到的數據以視覺化展示，就成為解析海洋大數據中很重要的學問。

同樣的數據對於不同的使用者可能有不同的關心議題。舉個例子來說，同樣都是海水表面的溫度，但一般泳客關注的可能是今日的海水狀況是否適宜下水游泳；但對於海洋科學家來說，海水的最高、最



海水表面葉綠素（Chlorophyll a）濃度

低，甚至是平均溫度，都是研究海洋環境非常重要的因子。像是今日的水溫是否已超過珊瑚能適應的最高溫度；記錄到迅速降低的海水溫度是否代表又有了湧升流的出現；又或海水的平均溫度相較於歷史數據，是否有異常的趨勢等。

除此之外，隨著科技的進步，不僅能在網路上查詢到觀測站記錄的近即時資料，還能把衛星影像搭配球型世界地圖以 3D 視覺化的方式展示。再加上互

動式網頁的出現，使用者能更便利地獲得所需的數據和資訊，客觀地做出重要的決策。

為了達到上述的目標，海洋中心近幾年來不斷地整合現場觀測、模式預報，以及衛星遙測等不同領域的資料，並彙整到海洋環境資料庫（Marine Environmental Database, <http://med.tori.narlabs.org.tw/>）中，讓社會大眾可以用更直覺且便利的方式認識這位孕育萬物的母親。

隨著科技的進步，不僅能在網路上查詢到觀測站記錄的近即時資料，還能把衛星影像搭配球型世界地圖以 3D 視覺化的方式展示。

MED 海洋環境資料庫 Marine Environmental Databank

網站導覽 | 國研院 | 海洋中心 文字大小 小 中 大

關於MED ↓ 資料展示 ↓ 資料服務 ↓ 聯絡我們

歡迎光臨!!

我們是台灣海洋科技研究中心，長期以來深入海洋探索人類感到最神秘的世界，我們把珍貴資料數位化，建置了海洋環境資料庫 (Marine Environmental Databank, MED) 收集、儲存與管理海洋中心各分項計畫所收集之海洋資料，提供近岸與遠洋觀測系統、衛星遙測、岩心以及船測等多元資料整合供應與加值應用服務。

see more

2014/7/19 研究五號0042航次回收底拖網作業

船測資料

收集展示海洋中心參與的研究船探測航次資料，包含探測紀錄、採樣作業與沿航跡的水文、海流與氣象等資料。

TOROS資料

高頻雷達測流系統具有觀測範圍廣大、近即時、高空間解析及成本相對低以及全天候觀測之優點，為當代有效監控國家領海海洋表面海流、海象

遙測資料

以衛星對海洋進行間接和遠距離的觀測，是進行大尺度海洋觀測、生態動力特性及生化通量研究的利器。

錨碇觀測

深海錨碇長期觀測網針對南海與臺灣東部西北太平洋海域進行海氣象、海流與水文的長期觀測，並提供近即時資料展示。

海洋環境資料庫網站

另外，這個網頁也建立了海洋資料申請的服務平台，提供海洋環境的保育和規畫更完整的參考資訊，讓自然界與人之間彼此能更和諧地交流。

李杰
國家實驗研究院台灣海洋科技研究中心

