出國報告（出國類別：其它）

出席 2013 年「全球生物多樣性資訊機構第 20 屆理事會 GB20」會議報告

服務機關：行政院農業委員會林業試驗所
姓名職稱：王豫煌 特聘研究員
派赴國家：德國
出國期間：民國 102 年 10 月 03－12 日
報告日期：民國 102 年 12 月 24 日

目　　錄

摘要

GBIF 第 20 次理事會(GB20)係於 2013 年 10 月 8-10 日在德國柏林市舉行；在大會之前，GBIF 秘書處也舉辦了全球節點的相關活動，10 月 4-5 日為節點教育訓練研習會議(training workshop)，6-7 日為全球節點管理者會議(Global Nodes Meeting)。由我國委員會執行秘書中研院生物多樣性研究中心邵廣昭研究員、我國節點管理者林試所王豫煌博士，與特有生物保育研究中心林瑞興副研究員等三人代表出席。此次會議之主要結果包括：

1. 研習會中介紹利用 GBIF 提供的資料進行物種分布模式預測，以探討全球氣候變遷衝擊、檢討保護區規劃等實例，提倡使用 GBIF 資料應用於科學研究和政策擬定。
2. 公開發表 GBIF 全新的入口網平台與 API，並介紹網站平台的新功能及服務。
3. 針對秘書處提出新規劃的會員國會費方案及副會員每年至少亦能捐獻 5 千歐元之議案仍無共識。
4. 秘書處原擬在明年舉辦的第二屆全球生物多樣性學會議(2nd GBIC)，因遭區域節點委員會質疑 GBIF 未來工作重點是否仍需要推動以及經費分配之公平性問題，而決定 GBIC 會議再延後召開。
5. 選出第四屆新任 GBIF 理事會(Governing Board)成員：理事會新任主席為 Peter Schalk (荷蘭)，第一副主席為 Jorge Soberson (美國)，第二副主席為 Motomi Ito (日本)，第三副主席為 Claude-Anne Gauthier (法國)；預算委員會主席為 Walter G. Berendsohn (德國)，副主席為 Selwyn Willoughby (南非)；科學委員會主席為 Rod Page (英國)，第二副主席為 Arturo Ariño (西班牙)，第三副主席為 Jean Cossi Gangalo (貝南 Benin)。
6. 自 GBIF 開始成立即負責秘書處實際工作之主管 Hugo von Linstow 退休。
7. 中科院代表中國正式加入 GBIF 為其他副會員(Other Associate Participant)。
8. 下屆 GB21 舉辦地點在印度新德里。
9. 會後與日本國家節點討論，明年度亞洲區域節點會議可能將在日本召開。

壹、前言

全球生物多樣性資訊機構(Global Biodiversity Information Facility, GBIF)，為國際經濟合作發展組織(OCED)支持，於 2001 年 3 月正式成立之生物多樣性資訊共享國際組織，其秘書處設置於丹麥哥本哈根大學動物學博物館。GBIF 以理事會(GBIF Governing Board)為最高權力機構，其下設執行委員會(Excutive Committee)、預算委員會(Budget Committee)、學術委員會(Science Committee)、節點委員會(Node Committee)，分別掌管 GBIF 行政事務監督審查與重大政策擬定、推動。GBIF 目前之會員分為：投票會員(Voting member)、副會員(Associate member)，投票會員需依該國之國民所得比例繳交會費，具有會務與選舉投票權；副會員分兩類：國家副會員(Country Associate)、其他副會員(Other Associate)，副會員無需繳交會費，但亦無投票權，副會員雖無投票權但享有其他同等權益。會員資格以定期完成備忘錄簽署作為認定標準，依 2013 年 10 月止之統計：投票會員國 37 個、副會員國 15 個、其他副會員 38 個，共 90 個會員。

台灣以 Chinese Taipei 之名成為其他副會員(Other Associate Participants)，成立全球生物多樣性機構中華民國委員會，主任委員目前為中央研究院副院長陳建仁，執行秘書由中央研究院生物多樣性研究中心研究員邵廣昭擔任；並依 GBIF 之規定建置了 TaiBIF 節點作為國家生物多樣性入口網站，TaiBIF 設節點管理員，目前由林業試驗所特聘研究員王豫煌擔任。

2006 年以前 GBIF 每年召開兩次理事會，2007 年開始改為每年一次，2013 年理事會為第 20 屆(GB20)，於德國柏林召開；本次會議之安排為：10 月 4-5 日研習會議、6-7 日全球節點會議，均由節點管理員和節點工作相關人員參加；8-10 日大會則由會員代表團參加。

貳、目的

GBIF 近年來每年均舉辦一次理事會(Governing Board Meeting)，邀請投票會員(Voting Participants)及副會員(Associate Participants)之代表(最多 5 人)出席，一方面向會員們報告秘書處一年來的工作進展與成果，二來檢討目前之會務。今年會議在報告方面則包括有科學、區域節點、生物多樣性信息展望(GBIO)等；會務方面則有財務、會員結構、未來三年之工作計畫(2014-2016)等報告。由於今年適逢 GBIF 之官員(含正、副主席，科學、預算委員會)四年之任期已屆滿，故此次會議中亦須辦理選舉。台灣是 GBIF 之創始副會員，雖不必交繳年費，但每年在國科會及中研院之支持下，亦組團 3-5 名出席每年在不同城市所舉辦之理事會，及兩年一次的各國節點管理者會議。今年的會議原本規劃在哥倫比亞召開，後來因故而又改回到歐洲，在德國的柏林舉辦。去年的 GB19 會議是在挪威舉辦，由林試所的林朝欽研究員、王豫煌特聘研究員及特生中心林瑞興副研究員等三人出席。今年則由中研院生物多樣性研究中心邵廣昭研究員及林試所王豫煌特聘研究員、特生中心林瑞興副研究員代表出席。王豫煌及林瑞興亦出席在前三天(10 月 5-7 日)的節點研習會議及全球節點會議。此次會議舉行前的一個月，中國大陸終以中科院(Chinese Academy of Science)之名義簽署 MOU.加入 GBIF 為副會員，並由中科院植物所馬克平研究員一人代表出席此次會議；他也是 DIVERSITAS 大陸委員會的執行秘書，並負責整合及推動大陸動、植物的生物多樣性資料。今年度理事會議的另一項重要活動是分區域舉行各國節點首席代表與其節點負責人間之聯席會議，尋求各國國內與跨國間更密切之合作與蒐集資料，然後再合併作跨區域間之報告。

參、行程

10 月 3-4 日　　啟程

10 月 5 日　　參加節點研習會議(Training Workshop)

10 月 6-7 日　　參加全球節點會議(Global Nodes Meeting)

10 月 8-10 日　　參加第 20 屆理事會(GB20)

10 月 11-12 日　　返程

肆、會議內容與感想

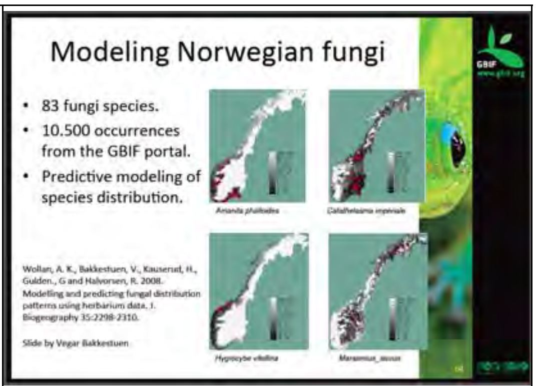
10 月 5 日研習會議
－提倡 GBIF 資料的使用(地點：Seminaris Campus Hotel Berlin)

09:00-10:30 提倡 GBIF 資料的利用，應用於關鍵的科學與政策領域
由 Dr. Patricia Koleff 主講，內容係針對達成愛知生物多樣性目標，如何使用 GBIF 所彙整全球生物多樣性分布的資料及其他國際資料平台的資料和分析工具，以墨西哥為例，應用於外來入侵生物的防除、生物多樣性的保育規劃及全球氣候變遷下的衝擊與調適(附件 1)。

	
研習會場	Dr. Patricia Koleff 說明墨西哥使用 GBIF 資料應用於生物多樣性保育與政策規劃。

11:00-12:30 資料分析最新的趨勢
由 Dr. Dag Endreson 介紹物種生態棲位與空間分布預測模式的觀念，使用 GBIF 提供的出現資料(presence only data)及其他平台(WorldClim)所提供的環境因子圖層資料，採用分析工具(例如 R 程式與 Dismo 套件)或科學工作流程軟體(Tavaner 或 Kepler)，首先進行資料清理、探索，再建立物種空間分布的預測模式，並加以驗證(附件 2)。

Dr. Dag Endreson 介紹物種生態棲位、分佈預測模式及相關的分析工具與資料來源。

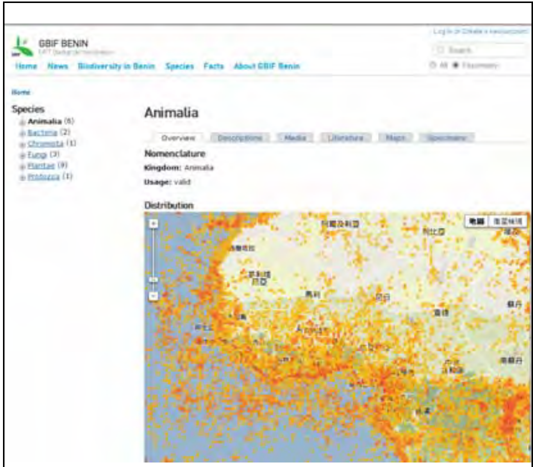
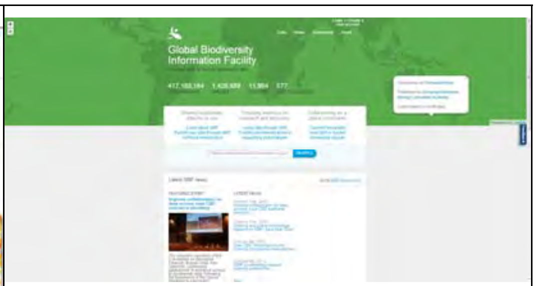


以 GBIF 提供 10,500 筆的出現紀錄預測挪威 83 種真菌的空間分佈。

13:30-14:30 GBIF 新節點工具、資料入口網及網路服務介紹

由 GBIF 資訊工程師 Daniel Amariles 介紹節點網站工具(NPT)及 GBIF 新入口網與相關應用服務(API)。NPT 是以 Drupal 開放原始碼網站管理系統為核心，整合了生命大百科(EOL)與 GBIF 彙整各國家節點生物多樣性的即時資訊，如物種的描述、多媒體資訊及物種名錄、生物出現紀錄、相關研究文獻等，可置於各國家節點網站的首頁或相關網頁，因此非常適合尚未建置網站或考慮更新網站系統的國家節點使用。GBIF 秘書處資訊團隊也呼應具備網站開發能力的國家節點也能參與 NPT 後續相關應用模組的開發與維護。今年度秘書處也徹底更新了 GBIF 的資料入口網，將所有國家節點發布至 GBIF 資料庫的生物多樣性資料進行即時的註冊(index[1])，採用處理巨量資料的 Hadoop 分散式架構[2]，彙整所有生物多樣性分布資料，進行即時資料整合處理，而查詢的資料亦可即時完整下載，無筆數限制；新的入口網也整合了 GBIF 秘書處、所有會員及最新活動等資訊，讓使用者以單一窗口查詢 GBIF 的資料及相關資訊。同時，秘書處也發佈了新的網路應用程式服務(API[3])，方便網站開發者或資料分析者，直接使用 HTTP 指令與相關參數，完整取用 GBIF 所發佈的資料，以利各國家節點生物多樣性資訊的整合，或大尺度的資料整合分析(請參考附件 3)。

[1] GBIF Registry Architecture http://www.gbif.org/infrastructure/registry
[2] GBIF Occurrences Architecture http://www.gbif.org/infrastructure/occurrences#componentArchitecture
[3] GBIF Webservice API http://dev.gbif.org/wiki/display/POR/Webservice+API

 NPT 展示網站 http://nptstartup.gbif.org/	 GBIF 全新資料入口網 http://www.gbif.org/

15:30-16:30 針對提倡資料使用進行意見與知識的交流

至目前為止，GBIF 的資料平台已累積了 1 萬 1 千 9 百多份資料集，包含 4.17 億筆物種出現紀錄。此一主題的意義是希望藉由參與研習的節點共同討論，交換經驗與知識，以找出有效利用 GBIF 巨量資料的最佳方法，支援資料分析，更進一步鼓勵使用 GBIF 提供的資料。上午的研習內容中已提及許多 GBIF 資料品質的問題，包含同物異名、座標錯誤或不精確、資料記錄重複、紀錄多集中於可及性高的地區等。儘管 GBIF 的資料諸多問題，但卻是目前世界上能彙整如此巨量的生物多樣性資料平台，其中必然包含了許多不精確或時空解析度不一的實際狀況，而這也正好提供了生物多樣性資料分析專家進行資料探索的好機會。使用者利用這些巨量資料，可先進行初步的資料視覺化，或依分類群、地理區系、時間等因素展示、過濾這些龐雜的資料，篩選出與分析目的吻合的趨勢或格局，再進一步進行分析。目前應可藉由許多利用 GBIF 提供的資料探討氣候變遷對生物多樣性衝擊的研究報告中，整理出清理、篩選 GBIF 巨量資料的建議方法，以提供後續資料使用者參考。然而，清理後的資料也應回饋至 GBIF 的資料平台，以促進資料的循環再利用。

10 月 6 日全球節點會議第一日(地點：Freie Universität Berlin)

09:00-09:45 開幕致詞與執行秘書報告

首先由主辦單位開幕致詞及節點委員會主席介紹全球節點會議目的，隨後由執行秘書 Donald Hobern 針對全球生物多樣性資訊學展望(GBIO[4])及 GBIF 配合達成愛知生物多樣性目標(Aichi Targets)的工作規劃策略與方向進行報告。GBIO 是 2012 年在哥本哈根邀集全球百位生物多樣性資訊研究專家學者齊聚討論未來 20 年生物多樣性資訊研究發展重點研討會的總結成果報告。主要內容是將會議討論結果

[4] Global Biodiversity Informatics Outlook http://www.biodiversityinformatics.org/

彙整為文化(culture)、資料(data)、證據(evidence)與理解(understanding) 4 個層面的主要架構，強調開放資料與資訊共享文化的建立、豐富多面向資料的收集與流通、提供工具促進資料與資訊的整合再利用、運用整合的資訊產生新知識以支援決策和大眾教育。GBIF 在此架構之下，以諮詢、溝通策略尋求合作夥伴，強化國家節點健全發展，並協同軟體開發者開發改進資訊整合工具，以達成愛知目標(附件 4)。

09:45-10:30 執行秘書報告討論
所有節點管理者針對執行秘書的工作報告內容、秘書處對 GB19 節點會議回應、2013 年工作進度與 2014 年工作規劃內容進行討論。

11:00-12:00 國家節點和秘書處問答與討論
所有節點管理者對上午議程內容與秘書處進行討論。其中核心議題在於，秘書處一方面希望發展強化區域國家節點能力的策略，但在未來有限的經費資源下卻縮減區域節點會議和教育訓練的活動經費，轉而投入更多資源在舉辦第二屆全球生物多樣性資訊學研討會(GBIC)，此作法遭受多數節點代表的反對。

12:00-12:30 下午分組討論議題介紹
針對下午分組討論資料品質檢核、節點網站平台工具開發、再利用現有平台工具的節點經驗等三項議題進行簡要說明。

13:30-15:00 分組討論
參與 NPT 節點工作平台開發規劃之討論。參與的節點提出此平台未來應提供多語系介面以應因不同國家的使用需求，並能夠保存多樣化的資料集(datasets)；為利於使用推廣，也應建立詳細的參考手冊。要推動這些發展工作，在秘書處有限的經費資源下，具有開發能力的節點應積極參與合作。

15:30-17:00 分組討論結果報告與綜合討論
資料品質檢核、節點網站平台工具開發、再利用現有平台工具三項議題分組討論的結果報告。對於資料品質改善，討論建議以完整的 metadata 描述資料集，並採用永久的資料集辨識碼(persistent IDs)、專有詞彙(controlled vocabulary)和資料附註(data annotation)，並能夠連結其他類型的資料，以提升 GBIF 資料的可用性(usability)。NPT 的開發雖採用 Open source 的網站管理系統 Drupal 為核心，但進一步的開發應用，也需要各節點的參與和投入資訊專業人力研發。未來各節點建置網站平台需要物種分布繪製、學名分類架構彙整及其他資料的整合，可以先參考運用澳洲 ALA、加拿大、瑞典、挪威、法國節點網站，及 GBIF 新網站與 API 的功能，避免重複開發。

17:00-17:30 本日會議成果摘述
今日的會議重點在於討論 GBIF 的會員能做些甚麼，GBIF 又能為會員作些甚麼。參與的會員代表強烈的反應出 GBIF 雖然採取區域策略，但卻嚴重的忽略了區域的需求；GBIO 的策略看似能為 GBIF 帶來寬廣的視野，但是，如何能與 GBIF 採取區域化的實際做法達成平衡，借助與區域節點的合作，提升各國家節點的能力，進而提升 GBIF 的整體貢獻，則有待更進一步的釐清。

	
執行秘書 Donald Hobern 工作報告	分組討論綜合報告

10 月 7 日全球節點會議第二日

09:00-09:30 資料流通與使用之國際政策
由秘書處報告目前 GBIF 與其他生物多樣性國際組織如生物多樣性公約(CBD)、地球觀測群之生物多樣性觀測網(GEO BON)及生物多樣性與生態系服務跨政府平台(IPBES)的分工合作關係。GBIF 可提供生物多樣性分布資料與相關的資訊服務給這些國際組織，共同促進生物多樣性保育與全球入侵生物防治等重大議題的跨國合作。

09:30-10:30 區域方法、策略之扼要檢討
秘書處報告節點(node)與生物多樣性資訊機構(national biodiversity information facility, BIF)的功能定義。GBIF 2012-2016 年的策略計畫中，期望在 2016 年底前實現在每個區域和國家建立永久性生物多樣性資訊機構(BIF)的目標，協助每個國家生物多樣性資料的流通與公開，以輔助政策制定、自然保育與經營管理、科學研究和大眾教育。由於各國經濟政治狀況差異很大，須採用不同的發展策略以達到以一個通用的目標架構，建立各國家或區域層級的生物多樣性資訊機構，但也必須採用一些發展指標用以評估各 BIF 的發展狀況。

11:00-11:30 討論區域方法策略

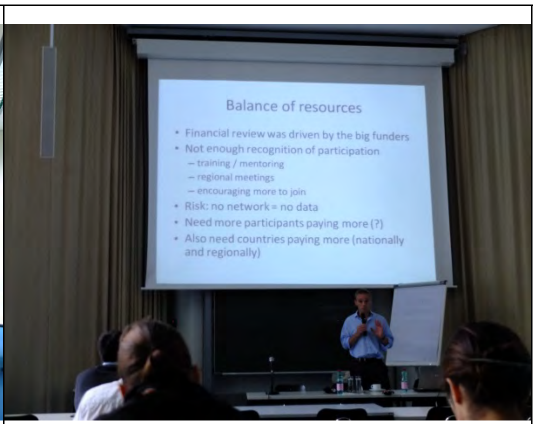
GBIF 期望透過以通用的架構，在 2016 年底前建立各國家永久性的生物多樣性資訊機構，此規劃目標在一些生物多樣性公約締約國或 GBIF 的正式會員國較可能達成；對於非締約國或非正式會員國，甚至包含締約國或正式會員國，因為缺乏主導國家生物多樣性保育或環境保護的政府機關代表，無法提供穩定經費支持永久性的機構設置，GBIF 應主動聯繫各國政府相關的高層單位，協助推動永久性國家生物多樣性資訊機構的設置，以做為國家生物多樣性資料整合與公開存取的基礎資訊建設，提供保育、經營管理、決策、科學研究與大眾教育之用。

11:30-12:30 下午議題分組討論介紹
針對下午分組討論 GBIF 2014-2016 工作計畫節點觀點、合作溝通平台、建立通用的節點發展計畫等議題進行簡要說明。

13:30-15:00 議題分組討論
參與合作溝通平台的討論，主要係針對現有的 Community Site 此一協作平台使用意願低落的原因進行討論與建議改善的項目。參與討論的節點普遍認為現有新平台的操作不夠直覺，也缺乏舊平台備份保存個人信件及其他資料的功能。台灣也提出一些建議，針對亞洲區域節點之間的溝通聯繫最需要的是文件協作、檔案分享、視訊會議等功能，這些功能在 Community Site 都無法滿足，但是卻可以使用 Google 提供完整的服務；若 community site 純粹只是要發布一些節點的訊息、文件，並不需要再建立一個網站平台，而是應該整合到新的 GBIF portal 中，以避免過往讓人感覺 GBIF 內部資訊分散的印象；區域間的協同合作可以善加利用 Google 提供的網路整合應用服務。

15:30-16:00 分組討論結果報告
工作計畫節點觀點、合作溝通平台、建立通用的節點發展計畫等議題討論結果報告。

16:30-17:00 摘述、前瞻及綜合整理 GB20 大會節點會議成果報告
會議最後由節點委員會主席 Steve Wilkins 彙整今日各會員節點的建議事項和討論結果，準備於 10 月 8 日舉行的理事會大會報告。

	
分組討論綜合報告	節點委員會主席 Steve Wilkins 報告

19:00 Ice Breaker

主辦單位在柏林植物園內的溫室內舉辦來歡迎所有與會的國家代表和節點管理員，藉此場合相互認識及交談。酒會由園長 Thomas Borsch 及 GBIF 主席 Daly 及 Freie 大學副校長 Monika Schafer-korting 致歡迎詞，另有安排導覽介紹全歐第三大的溫室之建設及展示內容。

10 月 8 日理事會(第一日)

GB20 會議在柏林市中心的 Auditorium Friedrichstrasse (四樓)之會議廳舉行。

9:00-10:00 開幕式

1. 由德國主辦單位致歡迎詞後，由 GBIF 主席 Joanne Daly 主持及報告此次出席概況，共有 26 個國家投票會員，6 個國家副會員及 10 個其他國際組織或合作計畫之副會員出席。
2. 同意秘書處所安排的議程，及確認 GB19 之會議紀錄及結論，包括摘要。
3. GBIF 主席之報告，因 Joanne Daly 今年任滿，故她亦就其任內的重要工作成果做一報告，包括歷經執行長由 Nick King 改聘為 Donald Hobern 及讓 GBIF 之網站穩定化，完成目前及未來工作計畫之編撰及審議，在 2011 年準備新的 MoU 與各會員重簽，將收集之資料由 1.9 億筆增加到 4.16 億筆，新的整合及發佈工具如 IPT 之發佈。當然也有一些尚待努力完成的工作，包括原始資料與後設資料(metadata)，將 GBIF 提供資料的目的由原來支援政策決策之用改回到給學術研究之用；GBIF 在 GNA 中所扮演的角色，如何與基因體資料聯結，爭取巨生物多樣性國加入為會員等。
4. 報告 MoU 目前各會員簽署之狀況及執委會對投票權的看法(備忘錄)

10:30-12:00 GBIF 未來的管理與財務狀況

5. GBIF 未來的財務狀況

由於全球之經濟衰退亦影響到各會員年會費之如期與如數之交繳。由於 GBIF 繳費最多的前七名國家其年費之總和即占了 GBIF 年度總預算的 70%，但這七國在 2011 年只付了 68%，2013 年迄今只付了 44%，情況並不佳。財務永續工作小組(含主席瑞典 Nileson，及其提名的 Roskok(美)、Naka(日)、Haeuser(德)、Stevenson(英))指出除經濟不景氣因素外，GBIF 之能見度、架構之品質、沉默的使用者眾、及未經常與 GBIF 之創始人定期聯繫等之其他因素均有關；執委會(EC)所做的建議包括要減小各會員間年費之差距。至於，對所有會員均需繳交年費的建議，11 位 EC 委員投票結果是 5 位同意、3 票考慮中、2 票建議延後、1 票反對，後來在請大家發言時，若干開發中國家之副會員代表多表反對。最後主席宣布留到明天再繼續討論。

6. 未來 GBIF 理事會型態是否改變之議題
   接下去是執委會、科委會及節點委員會三個委員會的工作報告

   13:30-14:45 科學與會務運作議題報告 I

7. 執行委員會報告
   此次會議 10 月 4-5 日之訓練班共有 18 位參加，10 月 6-7 日之全球節點會議則共有 55 位參加。新開發之工作 NPT 已完成 40 個國家的審視，有 10 個國家表示有興趣採用。今年出版之資料論文共 25 篇；分別在印度及西班牙舉辦 NPT 之課程；近來已與 Nature Scientific Data 期刊合作；Biofresh 促進資料發佈，支持 TDWG 工作坊之召開等；已成功完成 identifiers, metadata, quality, license, citation 等平台之建置；正在進行 Vocabulary development tools。過去一年內亦與 CBD 及 IPBES 有密切的聯繫與合作，包括 CBD 執委會參與 GIBIF 之活動(如 GB10) ；將參加在蒙特婁召開之 SBSTTAS 會議；受邀在 IPBES 工作計畫會議中作能力建設之報告等等。目前從其他機構所募集之經費包括 EUBON(4 年 290k 歐元)、EMODNet2(3 年 40K 歐元)、CBD(7.5 日 50k 歐元)、Eye on Earth(1 年 325k 歐元)等

8. 科學委員會報告
   由主席 Krishtalka 報告，他強調資料對科學研究的重要，把地球上 300 年來的資料數位化整合，可用模式來做預測，才能將科學變得 powerful，但 GBIF 提供之資料需可供使用者公開查詢下載使用，所產生的學術成果需忠於資料；而應用到施政時，施政者須相信科學研究結果之論述。他並舉肯亞 10.7 萬筆鳥類原始資料為例，如剔除錯別字，地點錯誤及缺地理座標後之有效資料只剩 2.75 萬筆，後來經過 GBIF 之各種工具之偵錯及修訂後，所有資料均已完全可用。目前 GBIF 雖已蒐集全球 4 億筆資料，但仍缺乏南美、非洲、中國、印度、東南亞等生物多樣性豐富地區之資料。其他蒐集資料之「熱區」，包括非 GBIF 會員國家，古生物學及長期氣候變遷之資料。此外，GBIF 未來亦勢必朝向蒐集與整合遙測資料、基因體資料，與佔地球生物多樣性 50%的微生物及濾過性病毒等(缺乏種名)資料。其報告十分精彩，但後來也有會員指出報告中忽略了海洋生物多樣性的重要

及其不足。

9. 節點委員會報告

協助建立國家或區域組織節點是 GBIF 之使命及重要工作，也是使 GBIF 有其價值之理由。相對的，各節點亦有義務及責任要蒐集資料，但目前還有許多資料尚不能蒐集整合及公開。近年來在各國節點之建立，GBIO 及區域性節點建設工作之推動上已有不錯的進展，特別是全球性之能力建設及訓練，例如，IPT 之推動與教育訓練。但在財務方面較為拮据，區域性之支持在各區間多寡亦不一致。成立區域性節點的策略，自 2008 年在 GB 會議中通過後，節點委員會也在 2011 年正式成立，2008、2010、2012 年間共召開了近 20 次之會議。目前有非洲、歐洲、南美、北美及亞洲幾個區域性節點，各節點所發展之方向及工作計畫均各有不同，如歐洲聚焦在水生外來種、南美以技術為主、北美發展網絡系統、亞洲則以國家名錄及魚類名錄之整合為主，但目前尚缺 Oceania 之節點。每位節點管理者都有自己專職的工作，可以投入這項工作的時間與資源較不充足，且有些國家仍有建設能力不足和經費短絀等問題。GBIF 規劃的國家節點架構是期望國家首席代表(Head of Delegation, HoD)代表國家正式的支持，與國家節點負責人能共同合作，整合資源，推動國家生物多樣性資料的彙整與公開，促進生物多樣性資料的再利用，應用於保育研究、政策制定和大眾教育。但是，此願景在許多國家推動起來有實質的困難，因此，節點委員會呼籲各國政府代表能與 GBIF 攜手合作，建立永久性的國家生物多樣性資訊機構做為一個國家生物多樣性保育的基礎建設，共同朝向生物多樣性公約在 2020 年之前達成愛知目標而努力。

15:30 節點管理者與各國首席代表之會議

由日本代表 Dr. Tsuyoshi Hosoya 主持，出席的共有日、韓、台、中、ACB 等 6 個節點的代表出席，就三項主要的問題做調查：(1)是否有足夠的資料來推動區域優先工作；(2)區域優先工作為何，是否符合 IPBES, GEO BON 之需求；(3)是否能找到經費來辦理區域活動，除了印度及 ACB 有困難外，其餘均可與 IPBES、CBD 及 GEO BON 等接軌，除韓國及台灣外，均有直接接觸。至於資料庫整合之工作則在今年三月在東京筑波舉辦的東亞瀕危種/入侵種/物種名錄整合工作坊已有實質的成果，整合之報告已完成在審定修改中。

18:30 參訪柏林自然史博物館

主辦單位招待所有與會代表參訪館內典藏及若干展廳之導覽活動(分成蜘蛛、鳥及甲殼類三組)，並在該館之恐龍骨骸展示廳內晚宴。

科學委員會主席報告

節點委員會主席報告

亞洲區域國家首席代表與節點會談

參訪柏林自然史博物館

10 月 9 日理事會(第二日)

09:00-10:30 科學與會務運作議題報告 II

10. 生物多樣性資訊學策略方向─生物多樣性資訊學展望的建言及挑戰，由 GBIF 秘書長 Hobern 就去年 7 月在丹麥所召開之 GBIC 會議所完成之報告做扼要之介紹，如下：

GBIF 為了達成《愛知目標》，於 2012 年 7 月與 EOL 及 BOL 等國際組織合作邀請全球一百位學者專家在丹麥召開了第一屆《全球生物多樣性資訊學研討會(GBIC)，希能提供生物多樣性研究、保育及管理、監測的架構，促進對地球上生物多樣性更深入的瞭解，同時提升對其進行保護和持續利用的認識。GBIC 之報告 ”全球生物多樣性資訊展望 (GBIO)”目前已正式出版。GBIO 主要內容為建議在下列四個關鍵而又相互連接的領域採取行動：(1) 建立一種***文化***以分享技能、政策、及共通的資料標準，並建立資料分享、資料存儲和存檔的的獎勵機制；(2) 從各種有效的來源獲取多種格式的生物多樣性***資料***，並使其能被快速、常設性的取用及重複利用；(3) 提供工具，從資料中提取有關意涵，轉換成科研及政策所需的***證據***；(4) 將上述證據用於模型及視覺化工具，激發對生物多樣性及人類影響的***認識和瞭解***，同時決定未來資料整合的優先順序。上述每個領域都包含五個

單元，其成果可以將環境、氣候及社會資料連接起來，進一步支援生物多樣性的指標體系和評估流程，以及不同決策方案下的後續發展。落實 GBIO 能夠對 2011-2020 年《生物多樣性戰略計畫》的五大戰略目標做出貢獻。GBIO 建議：***政策制定者***需要修訂相關的法令及規定以支持資料共用和再利用；***經費提供者***應該調整經費審定標準以確保專案對 GBIO 架構元件的貢獻；***生物多樣性組織***應該將 GBIO 納入其任務；**資料管理者**需要保證資料能夠永久地、自由、開放地取用；***科研人員***應該將 GBIO 元件的目標寫入他們的專案申請中；***生物多樣性資訊學專家***應該給出能夠支援 GBIO 的元件和目標的標準、技術、協定及工具；***公眾成員***則可以參與和生物多樣性有關的科研專案（例如"公民科學"的公眾監測企劃），並爭取對生物多樣性科學研究和相關政策的支持以鼓勵自由、開放地存取資料。

11. GBIO 架構下 GBIF 2014-2016 之工作計畫

請大家就 2014-2016 已先草擬的 GBIF 工作計畫書提出意見、問題及討論，正式之決議會留到第三天再做表決。

11:00-12:30 全球及國家節點生物多樣性研究基礎設施

12. 國家節點報告及新節點之正式發佈

今年是請哥倫比亞(SiB)及日本(JBIF)之國家生物多樣性節點做示範性報告，隨後則為 GBIF 新網頁的正式啟用之簡介。兩國的節點均有在會場外張貼壁報更詳細地介紹其網站，其中 JBIF 之國家策略除整體目標外，還有下列六點策略目標，值得我們參考。

整體目標：推動日本國內生物標本採集及其資訊之應用，並加速提供國際社群使用

A. 提升生物多樣性資訊之公眾意識
B. 促進博物館作為具生物多樣性典藏功能之能力
C. 提升民眾與政府機關對生物多樣性資訊重要性之認知
D. 加強日本之國家節點在 GBIF 社群中之能見度
E. 促進與相關計畫之合作
F. 朝擔任亞洲區域活動之主事者來努力

13:00-17:30 GBIF 科學研討會[5]

在科委會主席致詞後，即正式開始半天共 6 場之報告，最後有安排一小時的 cocktail 供與會者與講者間的討論。六位講者的報告內容摘述如下：

A. 「為何物種會出現在牠們的分布之處」(由本屆 Ebbe Nielsen 獎得主 Miguel Bastos Araújo 報告)

這是一篇剛發表在 Global Change Biology 的文章，他用了 ANN, CTA, GAM, GLM 等各種生態分布預測模式去推測物種分佈，並與其實際觀測分布做比較，結論是用同一組資料套用不同模式可能會有不同的預測結果，也可能會預測錯誤。換言

[5] GBIF 2013 Science Symposium Programme http://www.gbif.org/resources/2243

之，一般均認為物種分佈是受到生物因子、非生物之物理因子及其遷移所影響，究竟何者重要不易判斷，但此一專題對生物地理學、群聚生態學及作氣候變遷之預測卻十分重要。

B.　「PANGAEA─地球及環境科學之資料分布：研究資料進入學界」(由 Michael Diepnbroek 報告

PANGAEA 及 ICSU 之 World Data Center，目的在蒐集、處理、長期儲存及發佈地理參照之資料。目前已收集約 50 萬筆資料，完全免費公開。

C.　「利用 GBIF 資料從事現地(*in situ*)及異地(*ex situ*)的保育計畫」(由 Julian Ramirez─Villegas 報告)

利用南美資料預測在氣候變遷的目前及未來的情景下預測野生作物生物多樣性之改變後，哪些會在保護區外或內，此為了解未來氣候變遷對農業之重要遺傳種原之衝擊及調查，並據以提出保育之規劃。

D.　「結合古生物分布模式及統計地理親緣學來探尋新熱帶草原及季節性乾林之歷史」(由 Rosane G. Collevatti 報告)

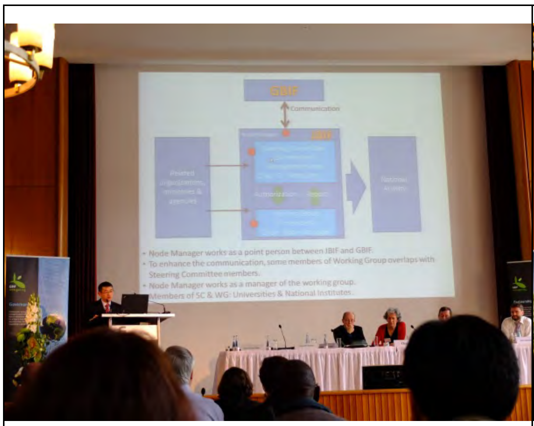
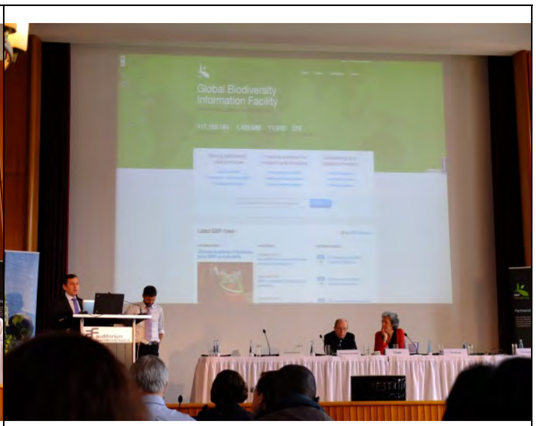
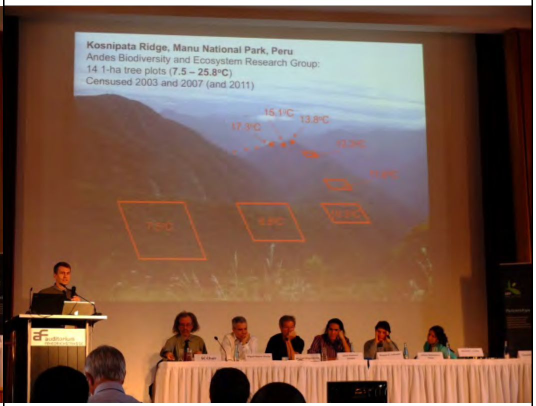
用 GBIF 及 PMIP/CMIP 之資料庫，以 ENM 模式來探討林相之成因，包括由目前遺傳多樣性之類型來推估草原在過去冰河期(LGM)有分布範圍之收縮而造成多重庇護所的存在。

E.　「”Wallacean Shortfall”之線上解決方案：GBIF 對物種分布範圍知識之貢獻」(由 Lillana Ballesteros─Mejia 報告)

”Wallacean Shortfall”是指由於物種地理分布資料不足，故對生物多樣性科學及了解全球變遷之影響是一項很大的障礙，因此需要加強資料募集及整合(博物館,NGO,GBIF)加強採樣，並以模式為基礎的整合。她比較 GBIF 資料與人工傳統方式在博物館典藏資料間之優劣，而發現 GBIF 之資料雖多，但多半是常見種及有重複，此乃因稀有種在採樣不足時會有偏少現象，亦即資料之代表性應要比資料量之多寡來得重要。

F.　「逃離熱浪：熱帶森林可否改變其分布來調適氣候之變化」(由 Kenneth J. Feeley 報告)

他在秘魯境內安地斯山脈從高海拔到中海拔去監測植物相長期之分布變化，發現平均每年會上移 9m，他在 Costa Rica 再重複一次試驗仍得到一年會上移 2m 之結果，但目前仍缺低海拔之資料，但用 GBIF 資料來分析發現氣候變遷對低海拔植物之遷移並不明顯，且會範圍縮減。平均 5%物種會改變分布，15%會擴張，30%會減縮其分布(即上移)。

	
日本國家節點介紹	GBIF 發表新網站平台
	
2013 年 Ebbe Nielsen 獎得主 Dr. Miguel Araújo 專題演講”為什麼物種會出現在它出現的地方？”	Dr. Kenneth J. Freeley 專題演講”逃離熱浪－熱帶森林能否改變其分佈以因應氣候變遷?”

10 月 10 日理事會(第三日)

09:00-12:30 財務報告

13. 預算委員會報告

2012 年經費收支狀況，截至 2012 年底尚有 48%的會費未收齊。2012 年預算收入與支出為 3,144,860 和 3,642,400 歐元，預估不足 497,540 歐元；實際收入、支出和結餘分別為 2,252,712、3,427,777 和-1,175,065 歐元。在預期主要貢獻經費的投票會員國無法即時或全數繳清會費的狀況下，必須調整 2014 年的工作計畫經費支出，2014 年預算收入為 3,050,000 歐元，計畫支出為 3,215,000 歐元，調整後的預算仍有嚴重的短絀。為解決長期收支不平衡的問題，委員會必須討論新的收費制度。投票會員國若無法準時繳交會費，應盡可能告知可繳清的確切時間；若情況允許，亦希望投票會員國能在 2013 年底前先行支付 2014 年的會費。

14. 選舉

今年度大會為第四屆理事會改選；首先進行主席投票選舉，雖然，主席候選人僅有一位荷蘭代表參選，卻經過三度投票才達三分之二的門檻，產生新任主席；副主席與各委員會的主席與副主席也順利選出。第四屆新任 GBIF 理事會 (Governing Board)主席為 Peter Schalk (荷蘭)，第一副主席為 Jorge Soberson (美國)，第二副主席為 Motomi Ito (日本)，第三副主席為 Claude-Anne Gauthier (法國)；預算委員會主席為 Walter G. Berendsohn (德國)，副主席為 Selwyn Willoughby (南非)；科學委員會主席為 Rod Page (英國)，第二副主席為 Arturo Ariño (西班牙)，第三副主席為 Jean Cossi Gangalo (貝南 Benin)。

13:30-15:00 財務議題討論

15. 繼續討論財務問題

    上午討論 2014 年預算規劃，預定節縮區域節點會議和教育訓練活動的經費提供舉辦第二屆 GBIC 會議之爭議。日本代表認為 2012 年舉辦之 GBIC 會議，歷經一年才於 2013 年理事會正式發表 GBIO 報告，而 GBIC 與 GBIO 之內容需要經過一段時間的消化，並將其中的訊息轉達國內的政府部門、學術界和社會大眾，這至少也需要一年的時間。因此，在經費來源短絀的財務狀況下，建議第二屆 GBIC 會議至少應延後一年舉辦，2014 年預算經費的運用還是應該著重在區域策略的推行。

16. 通過 GBIF 之文件

    討論執行委員會、科學委員會、預算委員會及節點委員會的修訂執掌規範文件。最後各項文件均無異議順利通過。

17. 未來 GB 會議

    因今年度僅有印度主動提出主辦 2014 年 GBIF 理事會大會，故明年 GB 大會將於印度新德里舉行。

18. 臨時動議

    在第四屆理事會及各委員會主席、副主席順利產生，各項會議內容充分討論後，於 GBIF 創立以來即擔任秘書處主管的 Hugo von Linstow 正式宣布退休。最後，由 Hugo 和卸任的各理事會主席致詞感謝大家的祝福，並表達對 GBIF 未來的期望。

19. 閉幕

預算委員會報告

GBIF 秘書處資深主管 Hugo von Linstow 退休致詞

伍、建議

1. 根據去年 GB19 會議之結果原本以為 GB20 會討論出未來 GBIF 會員結構及是否副會員(AP)亦須交繳會費等重大議題做出決議，但此次會議的討論大家之意見仍然分歧，故在此次會議並未獲得結論，包括每次各會員參加 GB 會議人數是否限縮。此外，GBIF 舉辦相當成功的第一屆 GBIC 會後原本秘書處希望要兩年開一次，後因 GBIF 經費不足及區域性節點委員會亦要求繼續支助其發展，故明年擬舉辦之第二屆 GBIC 會議將會延期。換言之，GBIF 雖然這十多年來在資料蒐集上已有很好的成果，資料被使用及引用之次數亦逐年增長，但仍多少會受到目前全球經濟不景氣的狀況而影響其投票國年費穩定之收入，而必須朝向其他國際組織及團體來募集經費。總之，如要副會員繳錢，若錢不多時，有些副會員尚可能繳交，但有些落後國家或非國家的副會員可能會因無法繳納而被迫退出，又有繳年費之副會員是否會有投票權亦應討論。台灣目前的狀況最好是靜觀其變，如果真要繳交會費，則相信在每年交 1,000-10,000 歐元之範圍內應可獲得中研院或國科會來編列預算支持，比照 ICS 下之學術團體(目前參加 DIVERSITAS，我國每年交繳 2 萬歐元)。
2. 中國大陸此次係以中科院(Chinese Academy of Science)之名義加入 GBIF 成為副會員，而不用"China"之名義加入，相信是因為中國大陸亦不願意在加入投票會員後需繳交每年五十萬歐元之昂貴年費，而且還有要繳交資料之壓力；故乃在新秘書長 Donald Hobern 數度前往北京交涉後，可能有感於新任秘書長之積極與熱誠之態度及大陸學界之期盼下，而同意以中科院名義加入。此次與會代表馬克平研究員與我國已有多年之合作關係，座位也依字母順序排在一起。故相信中國大陸之加入對雙方之與會權益應不致有任何影響與改變。倒是由日方主導之亞洲區域節點之活動，中方可能不會完全配合；例如，馬研究員在此次亞洲區域性節點會議中有強調何謂「區域性」，要先定義清楚，且主張所採用之資料格式或網路系統，應考慮採用中科院生物多樣性委員會在 2013 年 3 月在廣東肇慶已組織的「亞洲生物多樣性保護與信息網絡：ABCDNet」為基礎，該組織已涵蓋中、印度、韓、印尼、東盟生物多樣性中心(ACB, http://www.aseanbiodiversity.org)、國際山地綜合發展中心(ICIMOD)、IUCN 等組織。此一組織尚未含括日本及台灣在內，未來之發展值得關注。
3. GBIF-ROC 委員會主辦 11 月 20-22 日在台灣召開的一項「生物多樣性與生態研究開放資料國際研討會」，GBIF 秘書長 Donald Hobern 此次應邀來台參加，其他講者還包括 JBIF、NEON、DataONE、EU-BON 等國際組織或合作計畫之成員，對於在國內推動資料整合分享及將台灣的成果推展到國際，並加強與這些國際合作計畫之密切合作，有實質深遠的助益。
4. 開放資料和巨量資料的時代已然來臨。GBIF 秘書處在今年度的 GB 大會上發表新的資料整合提供平台(data portal)和網路應用服務(API)，顯示過去大量取

用 GBIF 彙整的生物多樣性資料的瓶頸已消失。雖然，許多 GBIF 的資料使用者批評 GBIF 資料的品質不佳，但是，資料品質參差不齊是巨量資料必然存在的問題，且目前世界上也沒有一個組織或平台能夠提供如此巨量的生物多樣性分布資料，反觀此一問題倒是成了探索生物多樣性資料科學的最佳資料來源。資料科學研究者可使用各種資料探勘的技術，於巨量資料中提取重要的資訊，再製作成不同的資料產品。同樣的，GBIF 彙整的資料，可以藉由不同資料視覺化和分析技術，針對不同分類群或時間、地域，萃取出重要或可用的資訊，進一步提供與其他環境或生態資料的整合分析。重點在於生物多樣性研究者本身是否具有操作巨量資料、運用資料探勘技術的能力。國內生物多樣性和生態學領域的研究者，及大學和研究所師生，對開放資料、巨量資料尚不熟悉的狀況下，整個國際的研究趨勢已發展得如此迅速，因此建議國科會應鼓勵大學及研究所開辦運用國際開放資料平台，進行巨量科學資料分析的課程；並獎勵生物多樣性、生態及資訊科學跨領域運用國際資料平台進行整合分析之研究，以提升我國生物多樣性和生態資訊學研究發展的能量。

5. rOpenSci[6] 基於 GBIF 發布的 API 和 Web Service 發展了 R 統計軟體的工具模組 rgbif[7]，方便使用者在 R 的分析環境中直接取用 GBIF 提供的資料進行整合分析。建議未來可以諸如 R、Tervana、Kepler 等 open source 分析工具為基礎，舉辦使用國際公開資料平台(GBIF, WorldClim, PANGAEA…)的生物多樣性和生態資料整合分析研習營，以提升我國運用國際資料平台進行大尺度生物多樣性研究的能力。
6. 理事會大會結束後，與日本節點管理者 Dr. Tsuyoshi Hosoya 和國家代表 Ms. Fumiko Nakao 商談明年舉辦亞洲區域節點會議的可能性。因為 GBIF 經費來源短絀，2014 財務規劃縮減區域活動的經費，可能無經費支援各區域兩年一次節點會議的召開。日本節點與國家代表確認日本環境部將有部分經費支持亞洲區域節點會議在日本舉行，但此經費可能無法支援所有亞洲會員赴日參加會議的旅費，因此請求台灣、韓國等經濟條件較為優渥的國家能自行負擔旅費，而將經費用於資助亞洲區域經濟條件較為弱勢的國家或組織節點派員參加。第四屆亞洲區域節點會議於 2012 年 6 月在台北舉行，會議的結果確立了亞洲節點 2012-2016 年共同合作的目標為區域瀕危物種名錄、外來入侵生物名錄及魚類資料庫的彙整；此次會議凝聚了亞洲區域節點的合作共識，並擬定了中、長期的策略和目標，所有節點對台灣舉辦第四屆區域節點會議均深表讚賞。因此，明年度若亞洲節點有共識在日本舉行第五屆區域節點會議，檢討合作成果進度，並進一步討論其他合作細節，則期望國科會能贊助我國派員參加會議的旅費，以促進亞洲區域生物多樣性資料整合和協同研究的發展。

[6] rOpenSci http://ropensci.org/
[7] rgbif R-package http://ropensci.org/packages/rgbif.html

附錄、教育訓練與節點會議簡報檔案

1. Promoting data use II: Use in key scientific and policy areas
2. Latest trends in data analysis
3. New GBIF Tools II: 2013 Portal and NPT Startup
4. GBIF Secretariat update

Promoting data use II
Use in key scientific and policy areas

Room Cambridge. Trainer:

Patricia Koleff

CONABIO

Contents:

- Introduction
- KEY CASE I: Invasive Alien Species
- KEY CASE II: Conservation planning
- KEY CASE III: Climate change impacts and adaptation

Aichi Biodiversity Targets

http://www.cbd.int/sp/targets/

CBD

Address the underlying causes of biodiversity loss by mainstreaming biodiversity across government and society

Reduce the direct pressures on biodiversity and promote sustainable use

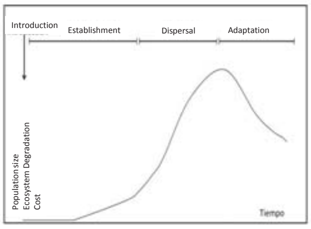
To improve the status of biodiversity by safeguarding ecosystems, species and genetic diversity

Enhance the benefits to all from biodiversity and ecosystem services

Enhance implementation through participatory planning, knowledge management and capacity building

CBD

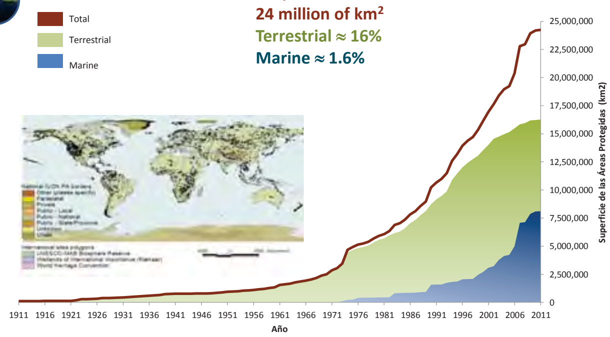
Article 8(h) of the CBD establishes that the parties shall "Prevent the introduction of, control or eradicate those alien species which threaten ecosystems, habitats or species."



(Source: Modified from Reise et al., 2006).

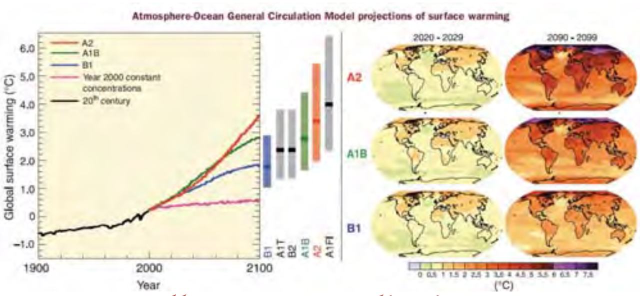
Source: Virginia Gwein. News in Focus: **Tsunami triggers invasion concerns.** 7 March 2013. *Nature* 14(495):13-14.

157,897 Protected areas
24 million of km²
Terrestrial ≈ 16%
Marine ≈ 1.6%



IUCN and UNEP-WCMC (2012) The World Database on Protected Areas (WDPA): February 2012. Cambridge, UK: UNEP-WCMC

Climate is changing



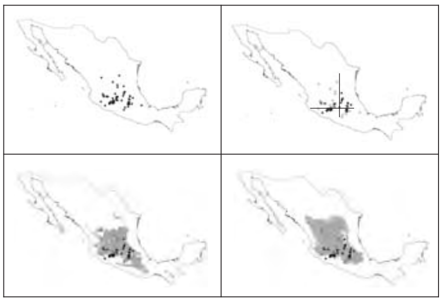
How well can we predict impacts on biodiversity?

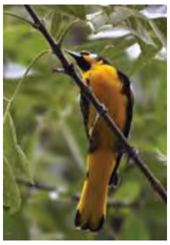
Facing the survey sampling bias (e.g. effect of roadside)

- For some taxa, at certain scales and from certain perspectives information is wide.
- For other scales, taxonomic groups or perspectives it is insufficient.

Survey localities for birds in "Atlas de las Aves de México" and main roads and highways in the Yucatan Peninsula, Mexico (Lira-Noriega, 2006).

Species distribution modeling

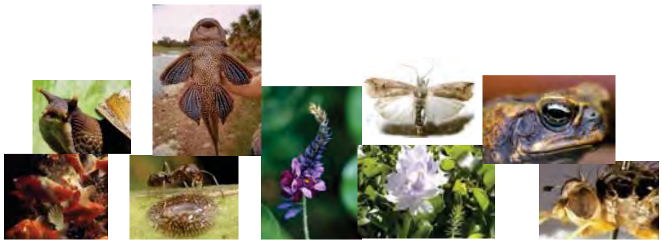


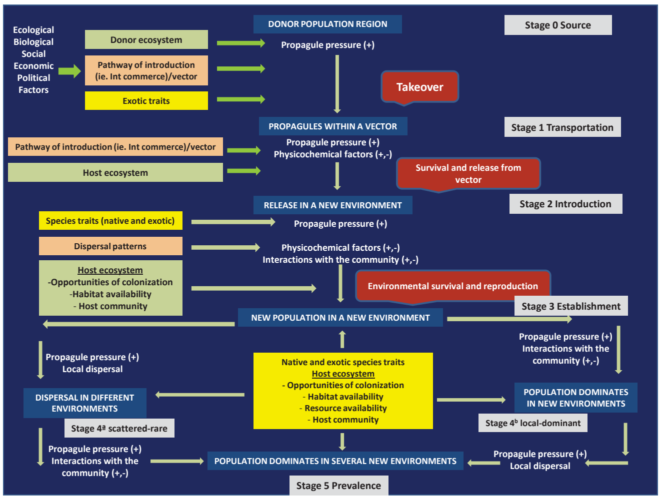


Getting from a set of occurrence data (survey localities) to a species distribution model of *Icterus abeillei* based on ecological niche modeling, considering variables of temperature, precipitation and topography, and using the Genetic Algorithm for Rule-set Prediction (GARP, Stockwell & Noble, 1992; Stockwell & Peters, 1999; DesktopGarp: http//www.lifemapper.org/desktopgarp/).

Key Case I. Invasive Alien Species

- General considerations
- How to build an official list of IAS
- A case study





Official invasive species list for Mexico

- 2010 Changes in the Legislation regarding Invasive Alien species
  - Identify Invasive species for Mexico
  - Regulate Imports and within the country
  - Establish prevention, control, management and eradication programes
  - Prohibits the introduction of IAS to Protected areas

What we need

- Identify Invasive species that
  - Are in Mexico (where, status etc.)
  - Are not in Mexico (Pathways, Risk of entrance)
- Additionally we need to detect
  - Species with unknown status in Mexico (research)
  - Native species that are introduced to new areas

Key information needed

1.- General information
2.- Distribution
3.- Environment
4.- Natural history
5.- Invasion background
6.- Impacts and management
7.- Invasiveness
8.- Risk analysis

Official invasive species list for Mexico

Full risk analysis
Potential distribution/climate match model

1. Reported as invasive elsewhere
2. Relationship with invasive taxons
3. Vector of other invasive species
4. Risk of introduction
5. Risk of establishment
6. Risk of dispersion
7. Sanitary impacts
8. Socio-economic impacts
9. Environmental impacts
10. Impacts to biodiversity

415 species evaluated so far

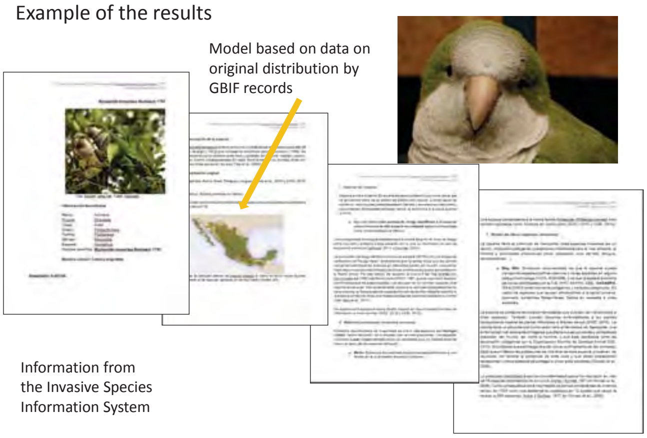
Sources of information on AIS

- Scientific literature
- Expert advice
- Online databases: GBIF, GISD, ISC-CABI, GISIN, NBIS, USFWS, ISSG, Mobot-Tropicos, Fishbase, Algaebase, The Plant base, GRIN USDA, WoRMs...
- General internet searches
- Peer review of all the information sheets

Additionally:
- **Research projects**
- **Networks of experts:** Constant collaboration with other sectors: government, academia, NGOs, international organizations

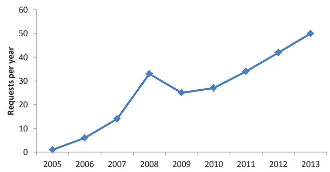
Example of the results

Model based on data on original distribution by GBIF records



Information from the Invasive Species Information System

Information requests on IAS



Information requested more often
- Distribution
- Species lists
- Impacts
- Control strategies

Users
- Governement
- Media
- Academia
- Foreign organisations
- General public

To date
- > 350 species identified as alien invasive species
- >1000 exotic
- > 200 native/invasive
- > 50,000 records for 384 species
- 157 detailed information sheets
- 415 rapid screenings
- 30 research projects

Case study
Risk assessment of *Bombus impatiens*

- Native of Eastern United States
- Imported to Mexico as a pollinator for green house agriculture
- 2000-2009 Conflicting and unconfirmed reports of sightings in the wild in different areas of Mexico
- Ministry of Agriculture requested help in determining the situation of the species as it has huge economic implications with tomato growers

Importance of updating the information

- In 2008 CONABIO did a Potential distribution model using GBIF data, however, in spite of the reports of escaped individuals, the data on the natural distribution available at the time, did not produce any results that would indicate that the species could become established in Mexico, so the imports were allowed to continue

- In 2010 Bumblebee specialists, producers and the Ministry of Agriculture create a working group to assess the situation of *B impatiens* in Mexico.
- The geographic information of the alleged sightings was sent to CONABIO
- We reviewed GBIF and found that new data on the species distribution had been added, documenting the species range towards Florida in the South.
- A new PDM was developed and the new results showed that the areas of the alleged sightings were a perfect match.

The species is not yet established in the wild, and we are currently working on a project to assess the status of native bumblebees and their susceptibility as greenhouse pollinators, to replace the exotic species and promote sustainable and conservation af native species. This will have to be discussed with the producers...

Key Case 2. CONSERVATION

Systematic Conservation Planning for spaces and species

- Basics concepts
- Examples of Mexico
- Final remarks

Systematic conservation planning **consists of the use of specific protocols to identify priority areas and separate them from processes which threaten their persistence.**

(Margulles and Sarkhar 2000)

- **The approach supports the decision-making process, where to establish new protected areas and direct conservation efforts towards more effective conservation.**

Overall planning approach in eleven stages

(Margules and Pressey 2000, Cowling and Pressey 2003)

1) Identify stakeholders for the planning region
2) Compile, assess, and refine biodiversity and socio-economic data for the region
3) Identify biodiversity surrogates for the region
4) Establish conservation targets and goals
5) Review the existing conservation area network
6) Prioritize new areas for potential conservation actions
7) Assess prognosis for biodiversity within each newly selected area
8) Refine networks of areas selected for conservation action
9) Examine feasibility
10) Implement a conservation plan
11) Periodically reassess the network

The concept of biodiversity must be made operational through the use of surrogates that refers to a sub-set of species, communities, or ecological systems (species or other taxa, landscape features, vegetation types,, etc..) that can be quantified (geographic distribution, populations, etc..).

- **Surrogates represent the conservation elements, which are selected to comprehensively represent the biodiversity of a given area.**
- **Goals has to be set for the representation of these surrogates**

(Margules and Sarkar 2009)

Setting conservation targets

¿How much is enough?

- 1982, World-Parks-Congress: less than 10% of a country or region.
- 1987, Brundtland Commission: 12% of an area or region.
- 25-75% of an areas is adequate to represent all species and ecosystems (Noss and Cooperrider 1994).
- 50% of a region (Soulé and Sanjayan 1998).
- 20-30% of an habitat (Fahring, 1997, Andrén 1994), below this threshold the effects of habitat fragmentation become very deleterious and lead to the extinction of species and alteration of ecological processes.
- Scientists recommend 10, 12, 15% or higher percentages (Margules y Sarkar 2009); 10-100% (Maiorano *et al.* 2006), 30-40% of each ecosystem.

Systematic conservation planning

Area design can take into account:

- **Detailed spatial information of biogeographic distribution**
- **Spatial economy**
- **Threat factors**
- **Socio-economic data**

(Sarkar et al. 2006. *Annu. Rev. Environ. Resour.*)

Prioritization of new potential areas for conservation actions

Biological criteria
- Threatened species or ecosystems
- Endemism
- Biogeographical distribution
- Threats to biodiversity
- Complementarity

Social and economical criteria
- Cost of land acquisition
- Threat factors
- Population

Desgin criteria
- Form
- Size
- Dispersion
- Connectivity
- Replication

During the planning process you can include the following criteria:
- **Minimize coincidence with highly impacted areas**
- **Reduce economic costs / social**
- **Implement spatial criteria: connectivity, area configuration.**

CONABIO

Species distributional data

information is available from broad-scale distributions to point localities

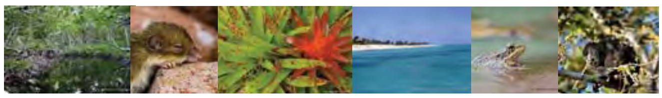
Examples for online databases

- Global Biodiversity Information Facility (GBIF) http://data.gbif.org/species/
- IUCN Red List of Threatened Species www.iucnredlist.org/
- TROPICOS database http://www.tropicos.org/
- SNIB http://www.conabio.gob.mx/remib/doctos/remibnodosdb.html?
- WildFinder www.worldwildlife.org/wildfinder/index.cfm
- Catalogue of Life http://www.catalogueoflife.org/col/search/all
- Species 2000 www.sp2000.org/
- World Biodiversity Database http://wbd.etibioinformatics.nl/bis/index.php

Uses of primary species-occurrence data

Some examples of Mexico

- GAP analyses (*i.e.* Terrestrial priority sites)
- Priority sites for biological corridors
- Priority sites for primate species



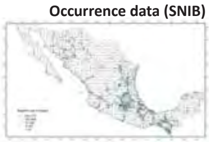
Identifying terrestrial priority sites

5 workshops to establish and define:

- scale of analysis (planning unit 256 km^2)
- **biodiversity surrogates (1450 cartographic maps for analysis)**
- criteria to define conservation targets (targets: 5 - 99%)
- threats and pressures(19 cartographic maps)
- thresholds and values of different parameters
- optimization program (MARXAN, 10,000 runs; 1 million iterations)

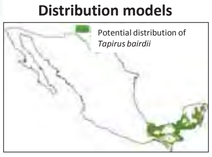
Basic inputs

Occurrence data (SNIB)



Data cleaning

Distribution models



Models were generated for most of the described species of Mexico:
- 83.7% amphibians and
- 86.4% reptiles (Ochoa-Ochoa, Flores-Villela et al. 2006)
- 86.0 % birds (Navarro y Peterson 2007)
- 87.3% mammals (Ceballos *et al.* 2006)

Species richness maps

Maps were generated for all vertebrates and endemisms.

Basic data cleaning (GAP analysis)

A) Nomenclatural and Taxonomic Error
- Identification certainty (synonyms)
- Spelling of names
  - Scientific names
  - Author names

B) Spatial Data
- Data Entry
- Georeferencing

Advanced data cleaning

A) Nomenclatural and Taxonomic Error
- Identification certainty (synonyms)
- Spelling of names
  - Scientific names
  - Common names
  - Infraspecific rank
  - Cultivars and Hybrids
  - Unpublished Names
  - Author names
  - Collectors' names

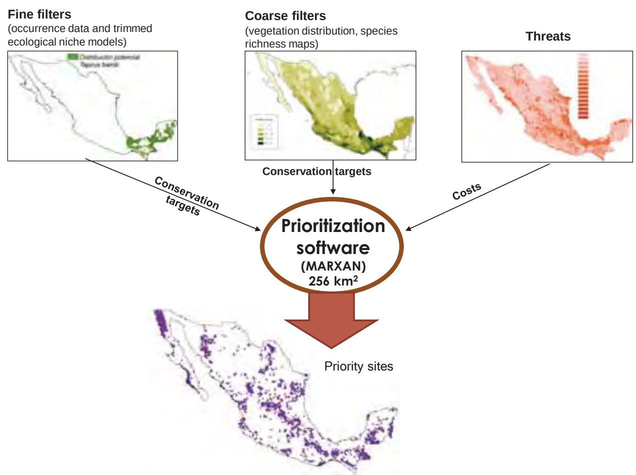
B) Spatial Data
- Data Entry
- Georeferencing

C) Descriptive Data
D) Documentation of Error
E) Visualisation of Error

Selected biological variables *	Data type
Critical vegetation types (68)	Polygon
Plant families (12)	Occurrence data, adjusted to fit the distribution according to grid resolution
Threatened plant species NOM-059-2001 (152)	Trimmed potential distribution models (endemic and not endemic)
Threatened tree species NOM-059-2001 (39)	Trimmed potential distribution models
Threatened agave species (23)	Trimmed potential distribution models
Resident birds (273)	Trimmed potential distribution models
Reptiles (424)	Trimmed potential distribution models
Amphibians (208)	Trimmed potential distribution models
Mammals (242)	Trimmed potential distribution models
Species richness (9)	Areas of highest richness of endemic and not endemic groups

1,450 maps were considered for the analysis out of 2,546

*9 out of 10 variables were obtained from primary ocurrence data.



Marxan
Informing Conservation Decisions Globally
http://www.uq.edu.au/marxan/

The software was developed to aid in the design of reserve systems which is made up of a selection of planning units which will satisfy a number of ecological, social and economic criteria and targets in the smallest possible area.

(Ball and Possingham 2000, Possingham *et al.* 2000)

Inputs

- Conservation features
- Conservation targets for each surrogate
- Threat factores
- Matrix of features per planning unit
- Boundary file

Objective function

(used by the iterative algorithm and by simulated annealing)

$$\sum_{Sites} \text{Cost} + BLM \sum_{Sites} \text{Boundary} + \sum_{ConValue} CFPF \times \text{Penalty} + \text{Cost Threshold Penalty}(t)$$

Cost: measure of the cost, area, or opportunity of the reserve system

BLM (Boundary Length Modifier): importance given to the boundary length relative to the cost of the reserve system

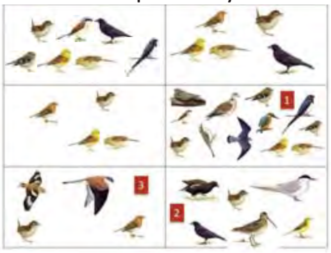
CFPF (Conservation Feature Penalty): penalty given for not adequately representing a conservation feature

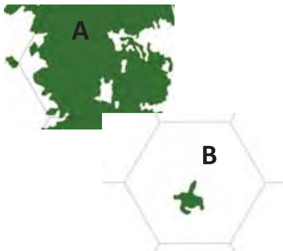
Penalty: additional value associated with each underrepresented conservation feature

Cost Threshold Penalty: penalty applied to the objective function if the target cost is exceeded.

Important considerations

Complementarity **Scale**



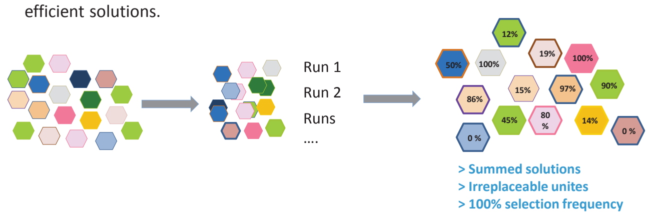


Marxan preferably selected sites with greater vegetation coverage (A) in order to meet conservation targets in the smallest possible area. However, if the conservation element has a restricted and fragmented distribution, the algorithm may select sites with little cover (B).

Irreplaceability

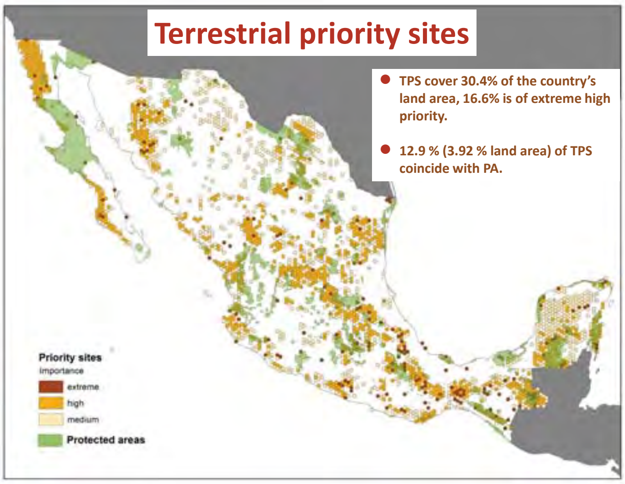
Pressey (1994): The probability that a site is required to meet the established conservation goals. Values vary from 0-100% (0-1); values represeninf 100% values is essential to achieve the goals for one or more elements for what they are called irreplaceable.

Game and Grantham: A unit with 100% selection frequency may or may not be irreplaceable in the strict sense, but it will contribute significantly to provide efficient solutions.



Terrestrial priority sites

- TPS cover 30.4% of the country's land area, 16.6% is of extreme high priority.
- 12.9 % (3.92 % land area) of TPS coincide with PA.



Other examples using GBIF-data & MARXAN software

Example #2

Priority sites for biological corridors

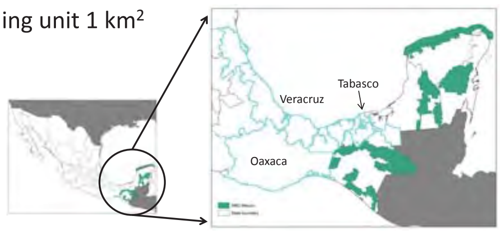
Mesoamerican Biological Corridor (MBC)

The MBC is established in Mexico in the states of Chiapas, Quintana Roo, Yucatan and Campeche, and comprises 5 corridors. The **main objectives** are: maintain biological diversity, reduce fragmentation, improve the connectivity of the landscape and ecosystems, and promote sustainable production processes that improve the quality of life of the local human populations who utilize, manage and conserve biodiversity.

The aim of this study was to assess a conservation area network to guide the establishment of a biological corridor in the states of Tabasco, Oaxaca, Veracruz.

Method

- **Phases1: preparation of input data**
  → **selection of species**
- Phase 2: analysis with the optimization algorithm (ConsNet v1.01) which established the prioritization and selection of municipalities.
- Planning unit 1 km^2



Input data: species

Gathering species data: potential distribution maps of plants and terrestrial vertebrates with potential distribution in the states.

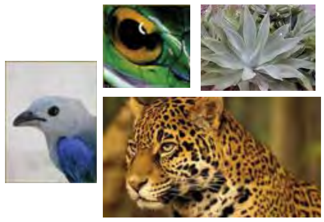
Group		# species used for Tabasco	# species used for Oaxaca	# species used for Veracruz
Terrestrial vertebrates	Birds	97	174	151
	Amphibians	3	66	54
	Mammals	28	74	68
	Reptiles	19	148	106
	Total	147	462	379
Plants	Agaves	-	12	9
	Trees	-	24	23
	Plants	-	88	80
	Total	-	124	112

Conservation targets

Example: Oaxaca and Veracruz

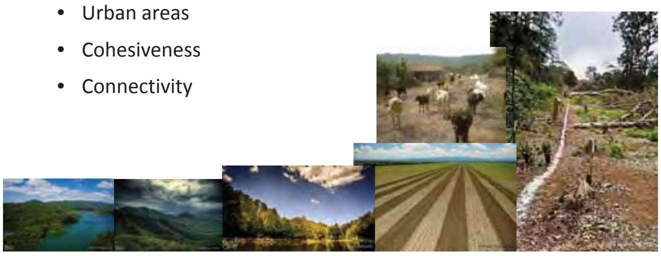
Endemic		Rareness (fourth quartile divieded in four)				NOM-059				IUCN			CITES		Total
Yes	No	4	3	2	1	E	P	A	Pr	Cr	En	Vu	I	II	
20	10	20	16	13	10	25	25	15	0	15	10	5	10	5	90

Total	Conservation target
85 – 64	40
63 – 42	30
41 – 22	10
< 21	5



Other criteria used for site selection

- Protected areas
- Vegetation cover (primary and secondary arboreal)
- Areas designated for agriculture, livestock and forestry
- Urban areas
- Cohesiveness
- Connectivity



ConsNet

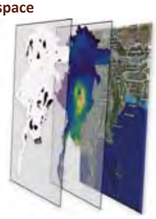
Advanced Software for Systematic Conservation Planning

http://uts.cc.utexas.edu/~consbio/Cons/consnet_home.html

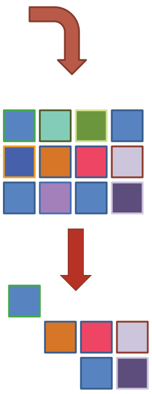
Comprehensive software package for the design and analysis of conservation area networks to represent biodiversity.

ConsNet is built on the Modular Abstract Self-Learning Tabu Search (MASTS) framework , a metaheuristic that relies on memory structures to organize and navigate the search space intelligently (Ciarleglio 2008).

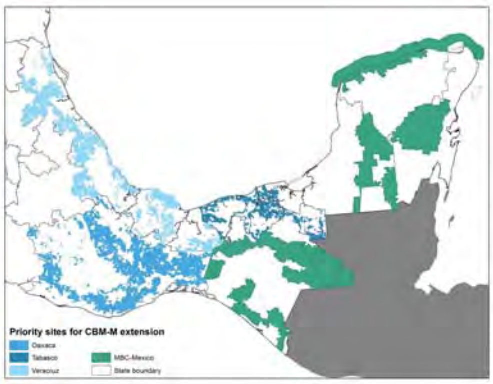
- Inclusion of multiple criteria
- Different targets for different criteria
- Design criteria
- Different search methods



- The study region is partitioned into "cells."
- For each cell, there are data on:
  - expected presence or abundance of surrogates
  - potential costs (or benefits) of placing each cell under a conservation plan, and the spatial properties of the cell.
- A conservation area network is assembled as a collection of cells which best meets the goals of the planner (Margules and Sarkar 2007).



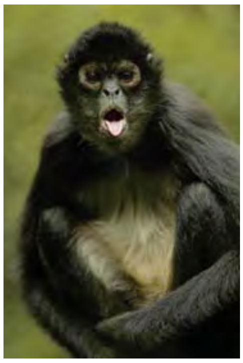
Proposal for extension of Mesoamerican Biologicla Corridor in Tabasco, Oaxaca and Veracruz



Example #3

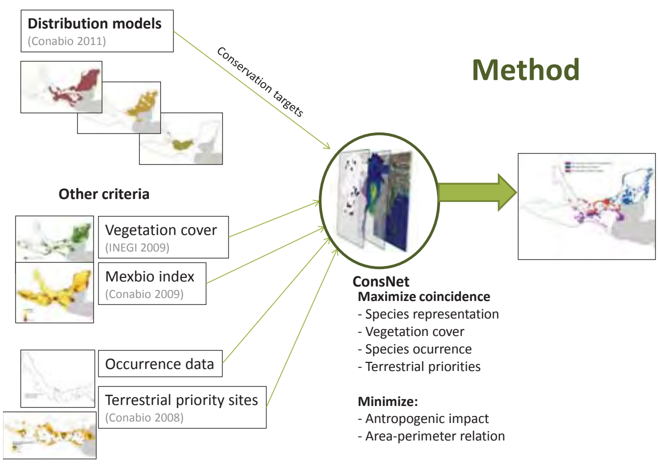
Priorities for primate conservation in Mexico

Mexican primates: a priority for conservation actions



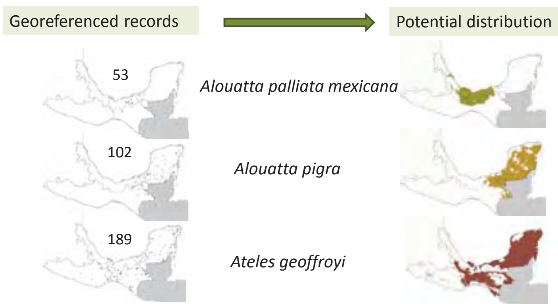
- Their populations are seriously threatened within and outside protected areas.
- They play a key role during the process of regeneration of tropical ecosystems.
- They can bring economic benefits in rural communities (ecotourism) and encourage the creation of private and community protected areas.
- As emblematic and recognized flagschip species, they are useful e.g. for environmental education campaigns.

Method



Distribution modelling

- Ecological niche models
- Identification of areas where environmental conditions are similar to those where presence of species have been confirmed.
- Generation of maps using MaxEnt at a spatial resolution of 1 km².

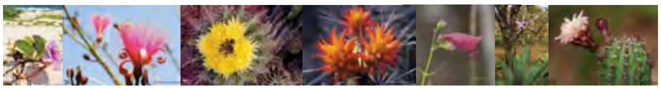


http://www.biodiversidad.gob.mx/pais/pdf/prioridades_primates.pdf

Examples of using GBIF data for conservation purpose

Conservation assessments

- ❖ Projects require extensive amounts of species-occurrence data in order to come up with a meaningful result.
- ❖ Information of biodiversity has become the most important consideration for conservation area selection.
- ❖ Key concepts to consider during the priority setting process: complementarity, replication, representativeness and irreplaceability.



Do we need more data???

The need to discover and describe species has never been more urgent.

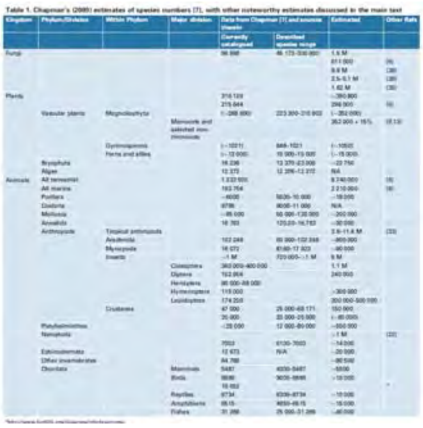
(Primm *et al.* 1995; Dirzo and Raven 2003)

Complete inventories are critical:

1) to properly estimate the magnitude of species loss.

2) to accurately formulate conservation assessments and plans.

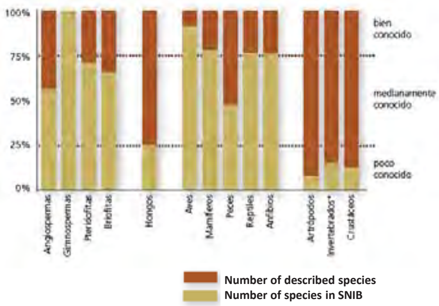
- **How many species are there?**
- **How many species are known?**



Completeness (aprox.)
97% for mammals
80–90% flowering plants
79% fish
67% amphibians
30% arthropods
<4% nematodes

May (2011) suggest another 480 years to describe all the species on Earth.

- **How many species are in Mexico?**
- **How many species are known?**



Escobar, F. *et al.* 2009. *Evaluación de las capacidades de conocimiento*

Consequences of poor-quality or sparse data
(and potentially subjected to high uncertainty)

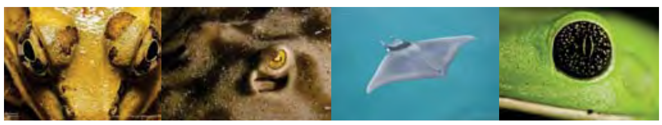
- ➢ **Poor decision-making**
- ➢ **Misuse of the limited resources available**
- ➢ **Failure of conservation practice**

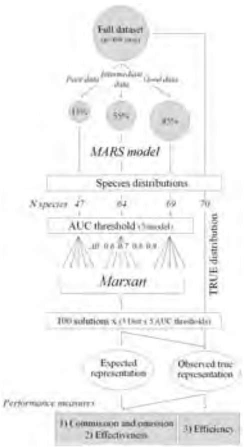
"Understanding the limitations and consequences of uncertainties in input data is a key issue in developing robust conservation recommendations from systematic planning." (Hermoso *et al.* 2013)

BUT: do we really need more data?

a) Is new data acquisition an adequate strategy to deal with errors in conservation planning outcomes derived from the use of poor datasets?

b) Is there a significant improvement in conservation planning outcomes by constraining the data used to those species with low uncertainty?



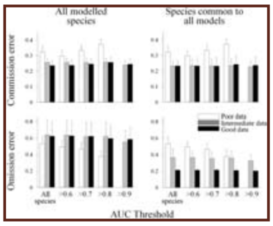


Four different predictive models were built:
- one on the complete dataset (true distribution model)
- three subsets that simulated different data availability scenarios.

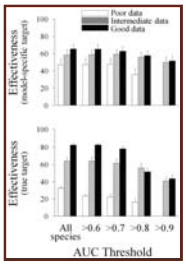
To test the effect of species uncertainty on conservation outcomes, independent analyses for different subsets of species were run, using the Area Under the ROC Curve to filter species with increasing certainty.

Results from Marxan were compared against the true distribution to obtain estimates of three different performance measures: 1) commission and omission errors, 2) effectiveness, and 3) efficiency.

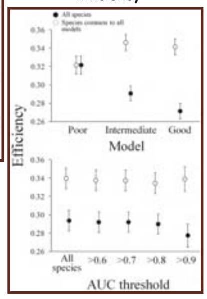
Commission and omission errors



Effectiveness



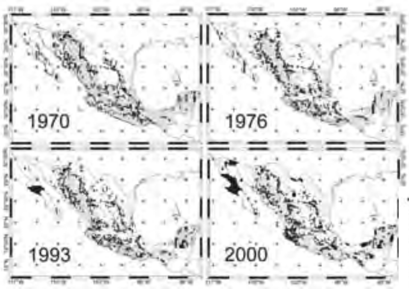
Efficiency



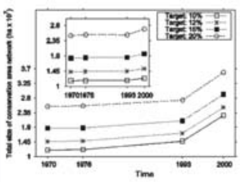
Concluding remarks

- Sparse data might be suitable for portraying the spatial patterns of biodiversity surrogates used for conservation planning.
- Data addition led to an increase in effectiveness as more species were adequately represented within priority areas but at the expense of reducing efficiency.
- Given that data acquisition is limited by high cost and time required, **improve the quality of data for those species with the highest level of uncertainty (rare species) when acquiring new data.**

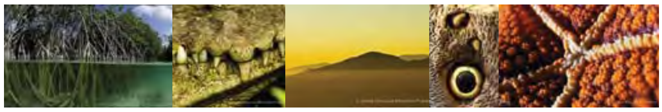
Conservation assessments
Species AND Threats



Fuller *et al.* 2006. The cost of postponing biodiversity conservation in Mexico. *Biol. Conserv. 134*



It requires a larger area in the year 2000 (41-89%, deposing the targets) to represent the habitat of mammals than in 1970.



"Tools for systematic area-based planning and plans have proliferated, but very few have yet led to changes in land use on the ground."

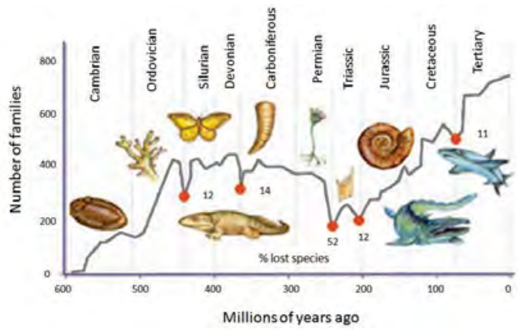
(Balmford and Cowling 2006)

Publications in Spanish in:
http://www.biodiversidad.gob.mx/pais/vaciosyom.html

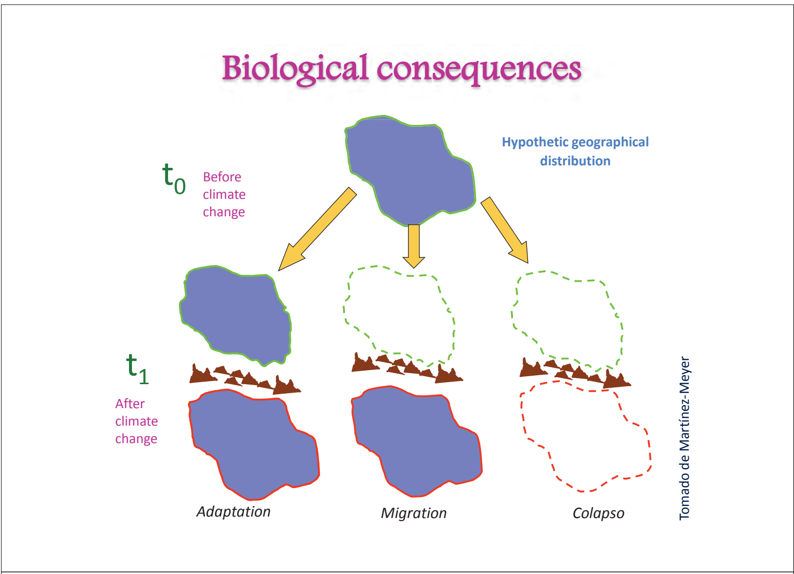
KEY CASE III: Climate change impacts and adaptation

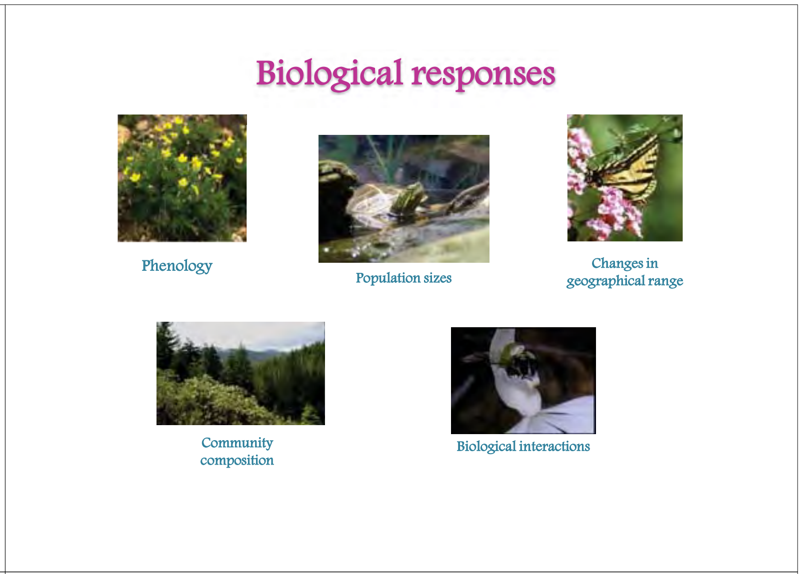
- Background
- Projections
- Effects on biodiversity
- Adaptation

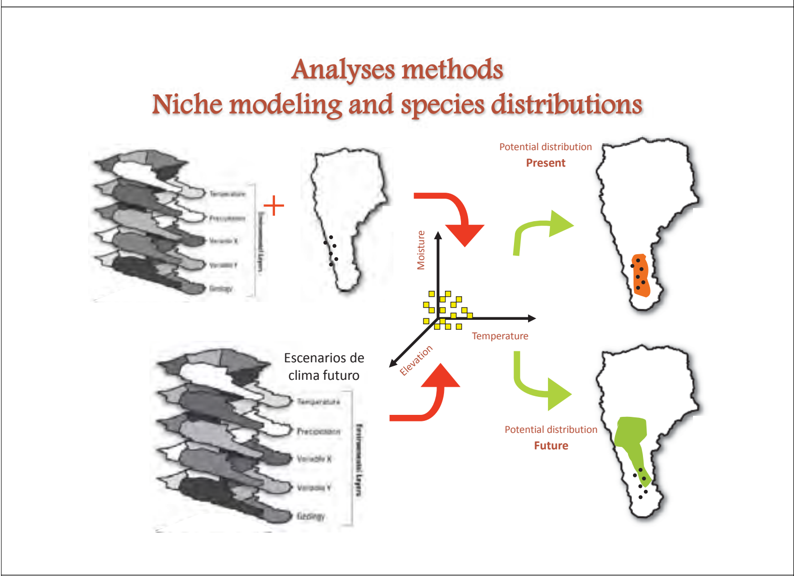
Effects of climate change on Earth

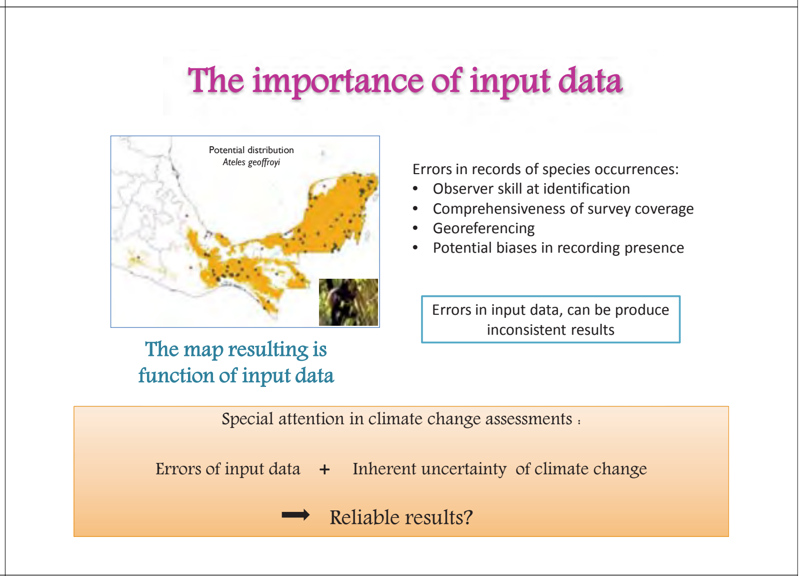


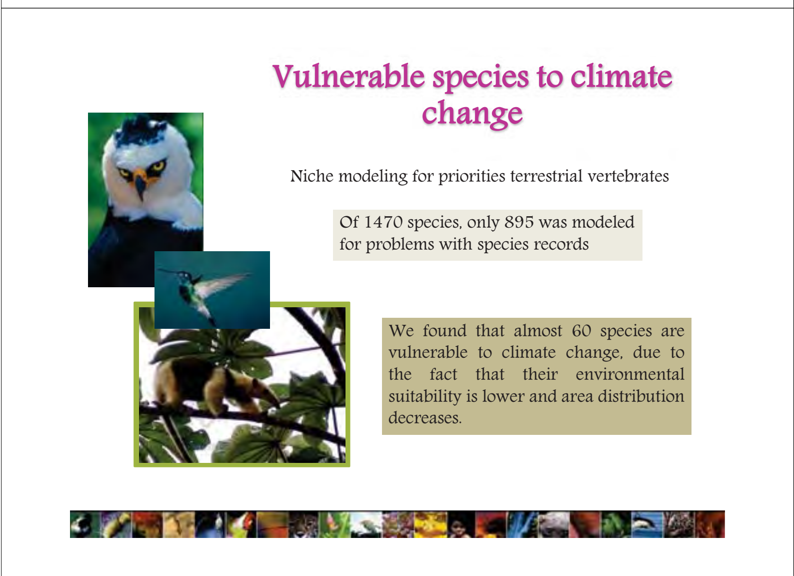
Currently we know few of the species that have existed on the Earth!

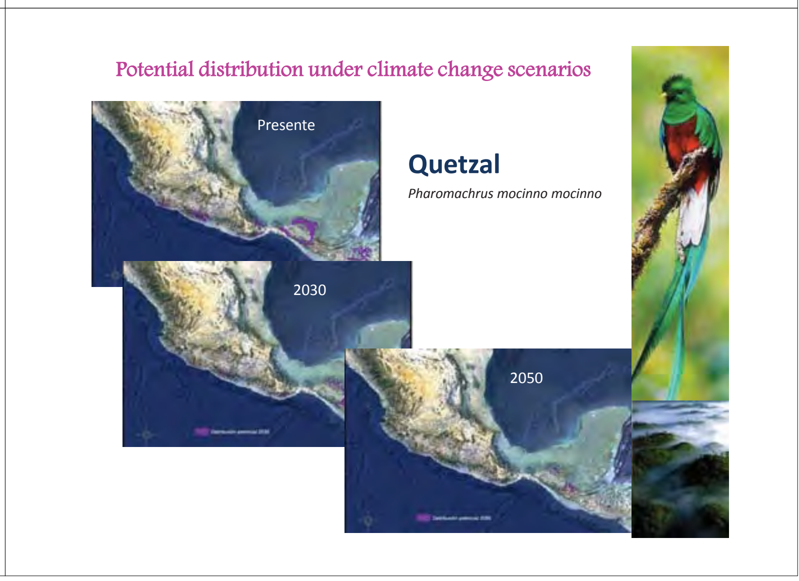

Biological consequences
Hypothetic geographical distribution
t0
Before climate change
t1
After climate change
Adaptation
Migration
Colapso
Tomado de Martínez-Meyer


Biological responses
Phenology
Population sizes
Changes in geographical range
Community composition
Biological interactions


Analyses methods
Niche modeling and species distributions
Potential distribution
Present
Moisture
Temperature
Elevation
Escenarios de clima futuro
Potential distribution
Future


The importance of input data
Potential distribution
Ateles geoffroyi
Errors in records of species occurrences:
• Observer skill at identification
• Comprehensiveness of survey coverage
• Georeferencing
• Potential biases in recording presence
Errors in input data, can be produce inconsistent results
The map resulting is function of input data
Special attention in climate change assessments :
Errors of input data + Inherent uncertainty of climate change
Reliable results?


Vulnerable species to climate change
Niche modeling for priorities terrestrial vertebrates
Of 1470 species, only 895 was modeled for problems with species records
We found that almost 60 species are vulnerable to climate change, due to the fact that their environmental suitability is lower and area distribution decreases.

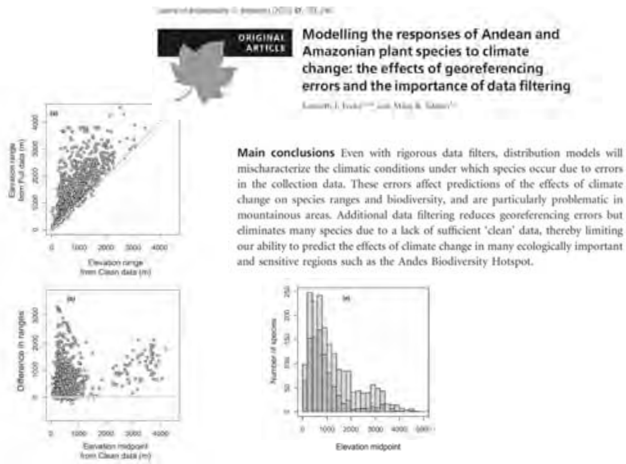

Potential distribution under climate change scenarios
Presente
Quetzal
Pharomachrus mocinno mocinno
2030
2050

Assemble approach for reduce uncertainty

Consists in assign as potential distribution area, those cells positively predicted by methods and models most widely used.

100 models were generated for each taxon:
- 4 Niche modeling algorithms
- 3 Global climates models
- 4 climate scenarios
- 2 times in the future and one in present.

9 maps for each taxon:
- One for the present
- 2 times in the future (2030 and 2050) with four stages (A1, A2, B1 and B2)


ORIGINAL ARTICLE

Modelling the responses of Andean and Amazonian plant species to climate change: the effects of georeferencing errors and the importance of data filtering

Main conclusions Even with rigorous data filters, distribution models will mischaracterize the climatic conditions under which species occur due to errors in the collection data. These errors affect predictions of the effects of climate change on species ranges and biodiversity, and are particularly problematic in mountainous areas. Additional data filtering reduces georeferencing errors but eliminates many species due to a lack of sufficient 'clean' data, thereby limiting our ability to predict the effects of climate change in many ecologically important and sensitive regions such as the Andes Biodiversity Hotspot.

Adaptation?

Challenges for the conservation of biodiversity

- Review the recent conservation strategies facing Climate Change (i.e. protected areas, biological corridors)
- Design new strategies taking in count (considering) the possible effects of Climate Change in Mexican Biodiversity
- What is necessary to know?:
  - Vulnerable Species and habitats
  - To what are vulnerable?
  - Other threats
  - Actions to reduce their vulnerability

Adaptation strategies for biodiversity

Protection FOR biodiversity

- Assisted migration?

More protected areas? or
Biological corridors

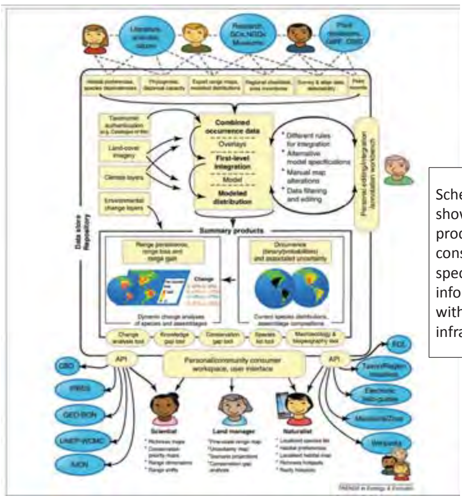
Conservation of biodiversity for adaptation

Protection FROM biodiversity

- Mangroves

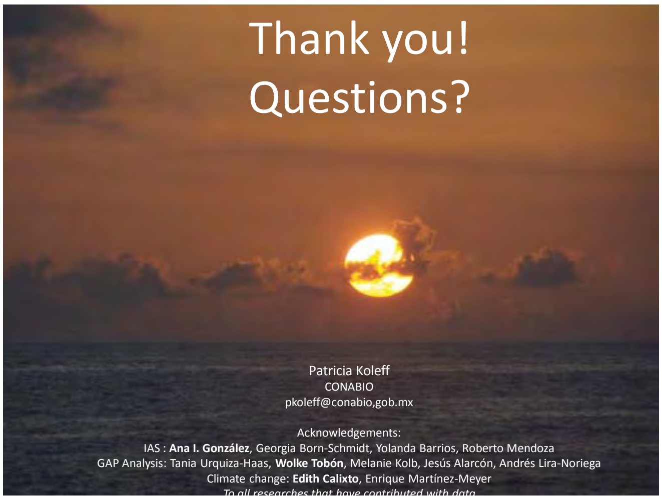
- Forest and emissions reducing

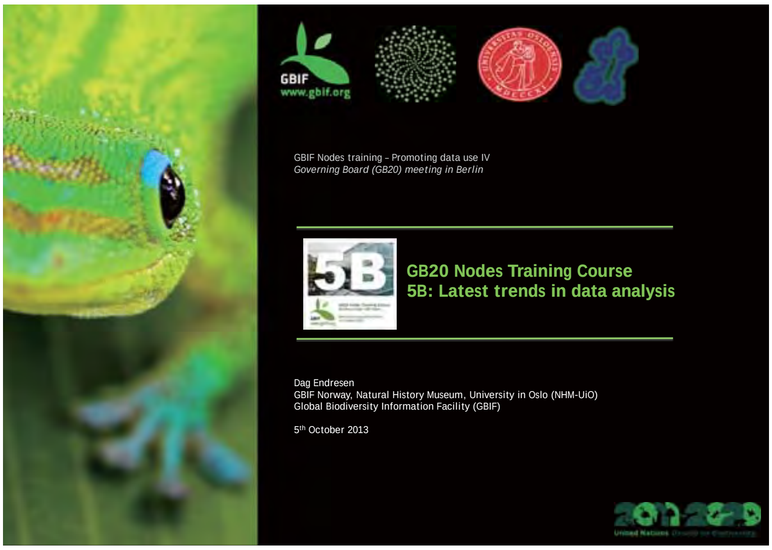
- Healthy ecosystems to face climate change

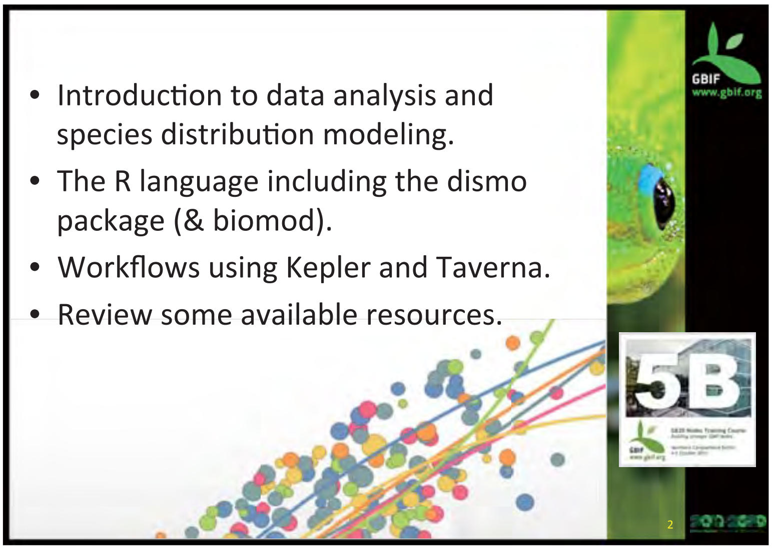


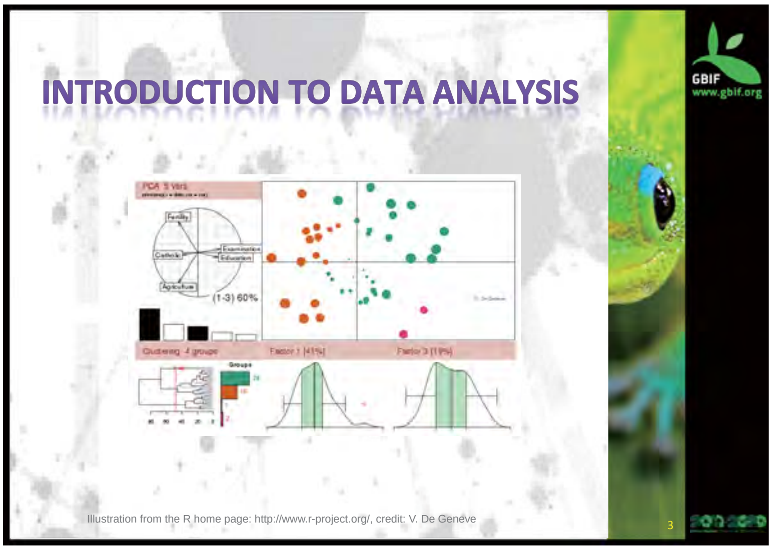
Schematic diagram showing how producers and consumers of species distribution information interact with the envisioned infrastructure.

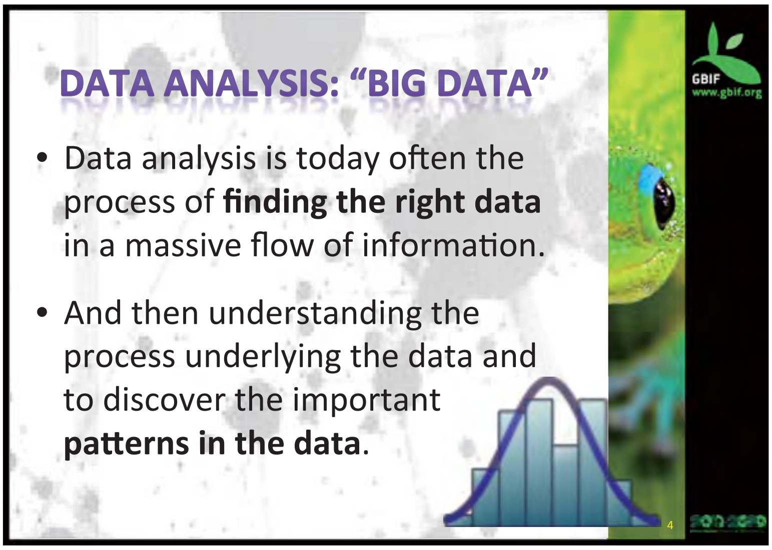
Jetz *et al.* 2012

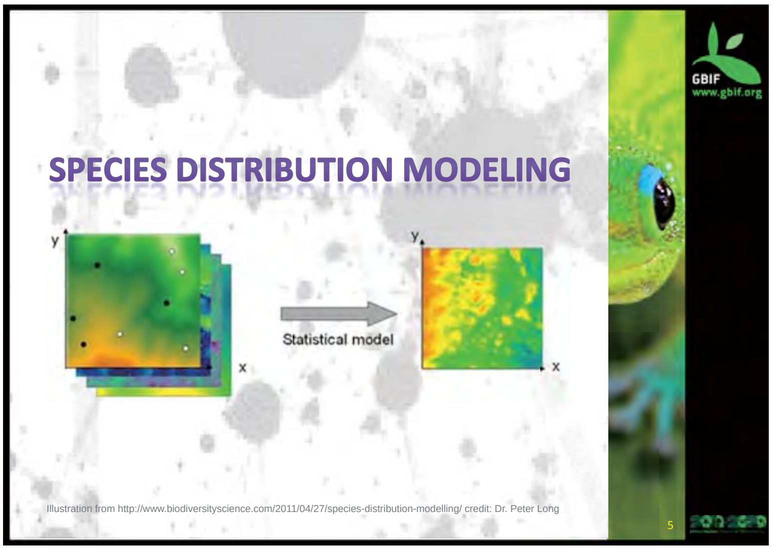


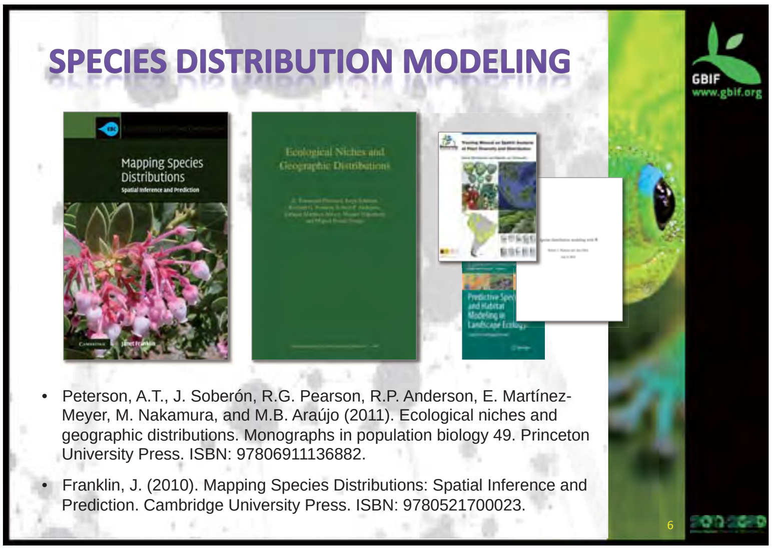

GBIF
www.gbif.org
GBIF Nodes training - Promoting data use IV
Governing Board (GB20) meeting in Berlin
5B
GB20 Nodes Training Course
5B: Latest trends in data analysis
Dag Endresen
GBIF Norway, Natural History Museum, University in Oslo (NHM-UiO)
Global Biodiversity Information Facility (GBIF)
5th October 2013
2011-2020


• Introduction to data analysis and species distribution modeling.
• The R language including the dismo package (& biomod).
• Workflows using Kepler and Taverna.
• Review some available resources.
GBIF
www.gbif.org
5B
2


INTRODUCTION TO DATA ANALYSIS
Illustration from the R home page: http://www.r-project.org/, credit: V. De Genève
GBIF
www.gbif.org
3


DATA ANALYSIS: "BIG DATA"
• Data analysis is today often the process of finding the right data in a massive flow of information.
• And then understanding the process underlying the data and to discover the important patterns in the data.
GBIF
www.gbif.org
4


SPECIES DISTRIBUTION MODELING
y
x
Statistical model
y
x
Illustration from http://www.biodiversityscience.com/2011/04/27/species-distribution-modelling/ credit: Dr. Peter Long
GBIF
www.gbif.org
5


SPECIES DISTRIBUTION MODELING
Mapping Species Distributions
Ecological Niches and Geographic Distributions
• Peterson, A.T., J. Soberón, R.G. Pearson, R.P. Anderson, E. Martínez-Meyer, M. Nakamura, and M.B. Araújo (2011). Ecological niches and geographic distributions. Monographs in population biology 49. Princeton University Press. ISBN: 97806911136882.
• Franklin, J. (2010). Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press. ISBN: 9780521700023.
GBIF
www.gbif.org
6

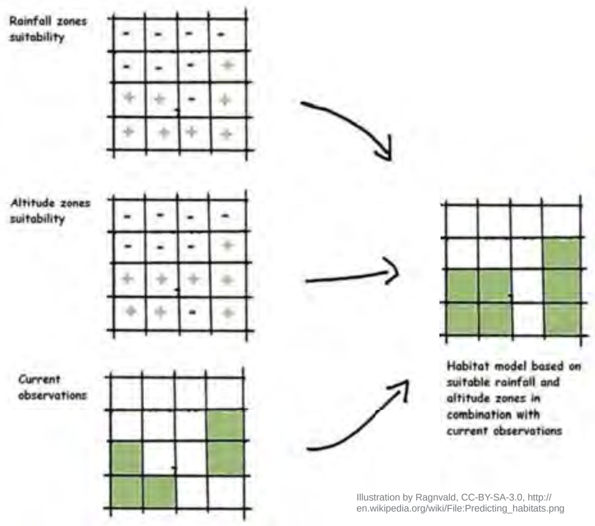


Illustration by Ragnvald, CC-BY-SA-3.0, http://en.wikipedia.org/wiki/File:Predicting_habitats.png

7

PRESENCE/ABSENCE

- So-called "**presence/absence**" or "**abundance**" data is collected in formal biological surveys of a set of randomized sites where both the presence and absence of target species are recorded.
- Presence/absence data is well suited for regression methods and is well explored in ecology (e.g. generalized linear or additive, GLM, GAM, or regression trees such as RF, BRT).

8

PRESENCE-ONLY

- However, much of the occurrence data from herbaria and museum specimen collections that are made available in GBIF are of the so-called "**presence-only**" category.
- The success of mobilizing presence-only data in GBIF has lead to numerous new methods specialized in analyzing such presence-only data (BIOCLIM, GARP, Maxent, ...).

9

PSEUDO-ABSENCE

- Many of the new methods developed to analyze presence-only data address the lack of absence data by:
  - Creating **pseudo-absences**, e.g. randomly sample an equal number "absence" points by different strategies.
  - Analyzing (all) **background points** as representatives of unsuitable environments.
- Other methods can model what is characteristic of the sites of recorded species occurrence without looking at sites where the species is assumed absent.
  - E.g. rule-based, principal component, factorial, clustering or machine-learning methods can be used.

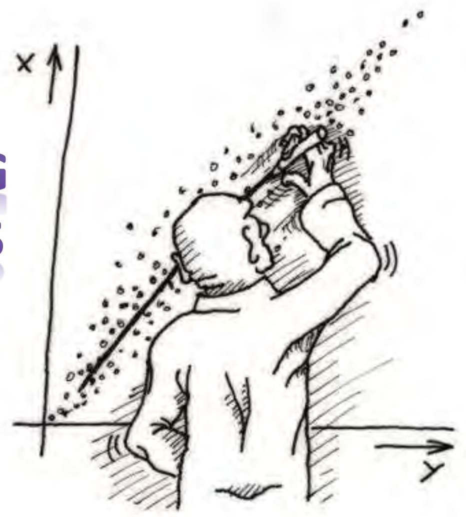
10

DATA QUALITY ISSUES

- Notice also ***disturbed areas*** that can be environmentally suitable for the species even if no species occurrences are found here.
- Low or variable ***detectability*** of the species can provide similar problems.
- ***Sample bias*** often provide another fundamental problem of species occurrence data where some areas in the landscape are sampled more intensively than others (high density of ecologists and other biologists, accessibility by car etc...).

11

MODELING METHODS



CHOOSING MODELING METHOD

- Different modeling methods or algorithms tackle issues of data quality differently.
- Some modeling methods are more sensitive to some types of data quality issues and less sensitive to other issues.
- Choosing the appropriate data modeling method depend on the types of data quality issues you discover in your respective data set.
- Identifying the appropriate method is often the process of validating performance (on an independent test set).
- It is a bad strategy to simply choose the same method as performed well in your previous studies.

GBIF www.gbif.org

13

BIOCLIM (1984)

- BIOCLIM is the classic '*climate-envelope-model*' first released by Australian scientists around 1984.
- The basic BIOCLIM algorithm (Nix 1986, Busby 1991) finds the climatic range of the points for each climatic variable.
- To form a bounding box, or climate envelope.
- The so-called BIOCLIM variables are still widely used (originally 12, now in total 35 variables).

GBIF www.gbif.org

14

IN TOTAL 35 BIOCLIM VARIABLES

BIOCLIM uses monthly or weekly values of
- maximum temperature
- minimum temperature
- rainfall
- radiation
- evaporation

to derive the following bioclimatic parameters:
- P1. Annual Mean Temperature
- P2. Mean Diurnal Range(Mean(period max-min))
- P3. Isothermality (P2/P7)
- P4. Temperature Seasonality (Coefficient of Variation)
- P5. Max Temperature of Warmest Period
- P6. Min Temperature of Coldest Period
- P7. Temperature Annual Range (P5-P6)
- P8. Mean Temperature of Wettest Quarter
- P9. Mean Temperature of Driest Quarter
- P10. Mean Temperature of Warmest Quarter
- P11. Mean Temperature of Coldest Quarter
- P12. Annual Precipitation
- P13. Precipitation of Wettest Period
- P14. Precipitation of Driest Period
- P15. Precipitation Seasonality(Coefficient of Variation)
- P16. Precipitation of Wettest Quarter
- P17. Precipitation of Driest Quarter
- P18. Precipitation of Warmest Quarter
- P19. Precipitation of Coldest Quarter
- P20. Annual Mean Radiation
- P21. Highest Period Radiation
- P22. Lowest Period Radiation
- P23. Radiation Seasonality (Coefficient of Variation)
- P24. Radiation of Wettest Quarter
- P25. Radiation of Driest Quarter
- P26. Radiation of Warmest Quarter
- P27. Radiation of Coldest Quarter
- P28. Annual Mean Moisture Index
- P29. Highest Period Moisture Index
- P30. Lowest Period Moisture Index
- P31. Moisture Index Seasonality (Coefficient of Variation)
- P32. Mean Moisture Index of Highest Quarter MI
- P33. Mean Moisture Index of Lowest Quarter MI
- P34. Mean Moisture Index of Warmest Quarter
- P35. Mean Moisture Index of Coldest Quarter

GBIF www.gbif.org

15

WORLDCLIM PROVIDES 19 BIOCLIM VARIABLES.

Maximum temperature, minimum temperature, rainfall --->

BIO1 = Annual Mean Temperature
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3 = Isothermality (BIO2/BIO7) (* 100)
BIO4 = Temperature Seasonality (standard deviation *100)
BIO5 = Max Temperature of Warmest Month
BIO6 = Min Temperature of Coldest Month
BIO7 = Temperature Annual Range (BIO5-BIO6)
BIO8 = Mean Temperature of Wettest Quarter
BIO9 = Mean Temperature of Driest Quarter
BIO10 = Mean Temperature of Warmest Quarter
BIO11 = Mean Temperature of Coldest Quarter
BIO12 = Annual Precipitation
BIO13 = Precipitation of Wettest Month
BIO14 = Precipitation of Driest Month
BIO15 = Precipitation Seasonality (Coefficient of Variation)
BIO16 = Precipitation of Wettest Quarter
BIO17 = Precipitation of Driest Quarter
BIO18 = Precipitation of Warmest Quarter
BIO19 = Precipitation of Coldest Quarter

GBIF www.gbif.org

16

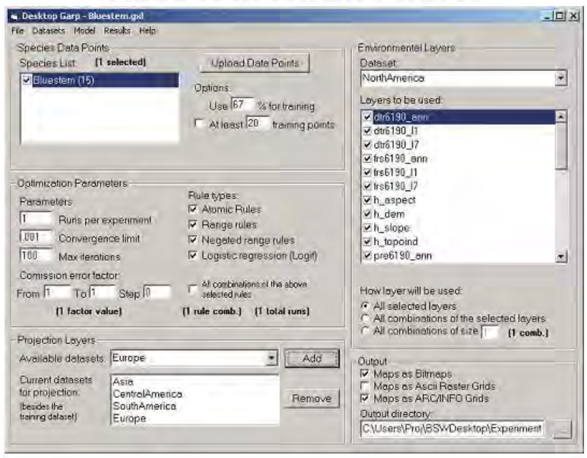
GARP (1999, 2002)

- Genetic Algorithm for Rule-set Production (GARP).
- Originally released as "**GARP algorithm**" around 1999.
- "**Desktop GARP**" software released around 2002 by the University of Kansas and CRIA in Brazil.
- http://www.nhm.ku.edu/desktopgarp/

GBIF www.gbif.org

17

DESKTOP GARP (2002)



GBIF www.gbif.org

18

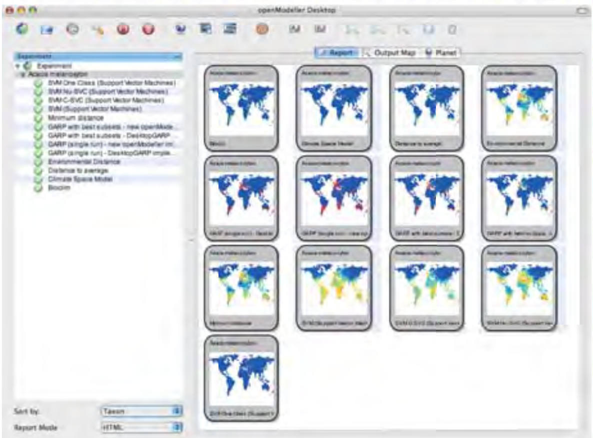
OPENMODELLER (2003, 2006)

- Originally developed by CRIA in 2003 as part of the speciesLink project.
- "**OM desktop**" released in 2006.
- Cross-platform, C++ framework, Python API, WS interface.
- Implements a number of different modeling algorithms (including Random Forest, ANN, SVM, Maxent, GARP, …).

GBIF www.gbif.org

19

OPENMODELLER DESKTOP (2006)



GBIF www.gbif.org

20

MAXENT (2004) $f_1(\mathbf{z}) = f(\mathbf{z})e^{\eta(\mathbf{z})}$

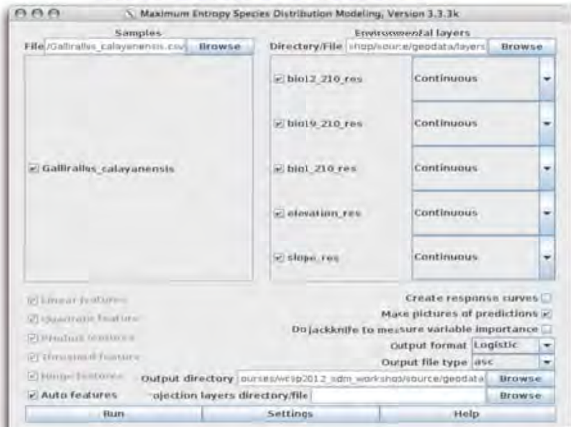
- Maxent Java SDM software released in 2004.
- Well suited and with high performance for presence-only data.
- By default Maxent randomly samples 10,000 background points.
- Maxent currently has six feature classes: linear, product, quadratic, hinge, threshold and categorical.
- It is common to mask the study area – i.e. setting no-data values outside the area of interest.
- Assumption: Maxent relies on an unbiased sample.
  - One fix is to provide background data of similar bias.
- Assumption: environment layers have grid cells of equal area.
  - In un-projected latitude-longitude-degree data, grids cells to the north and south of the equator have smaller area.
  - On fix could be to re-project to an equal-area-projection.
- Available at: http://www.cs.princeton.edu/~schapire/maxent/

GBIF www.gbif.org

PRINCETON UNIVERSITY

21

MAXENT (2004)



GBIF www.gbif.org

PRINCETON UNIVERSITY

22

MODEL PERFORMANCE

- The relatively large number of species distribution (SDM) modeling methods and software implementations lead to studies comparing SDM method performances.
- The high score of **Maxent** observed by Elith *et al.* (2006) contributed to the increased popularity of this method.

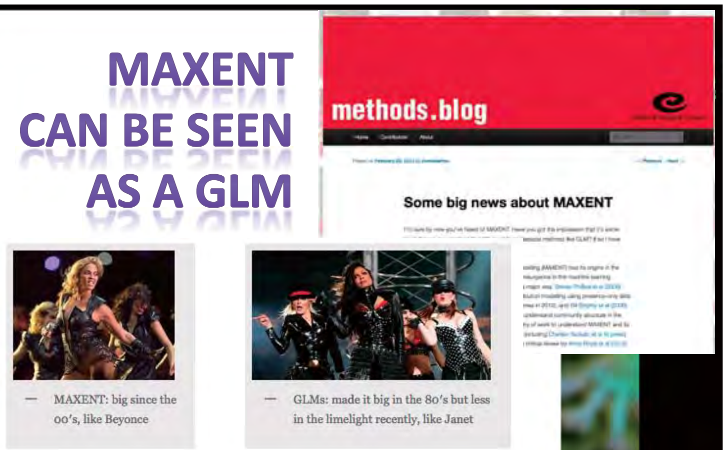
Elith*, J., H. Graham*, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M. and E. Zimmermann, N. (2006), Novel methods improve prediction of species' distributions from occurrence data. Ecography, 29: 129–151. doi: 10.1111/j.2006.0906-7590.04596.x

GBIF www.gbif.org

PRINCETON UNIVERSITY

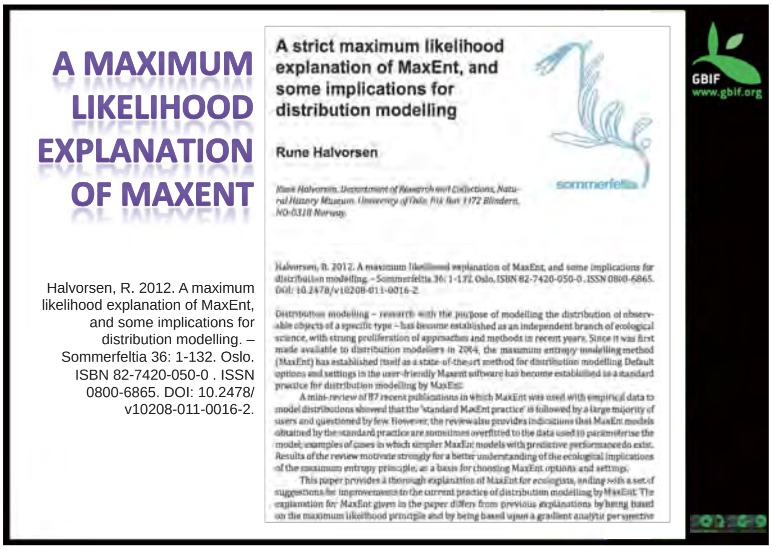
23

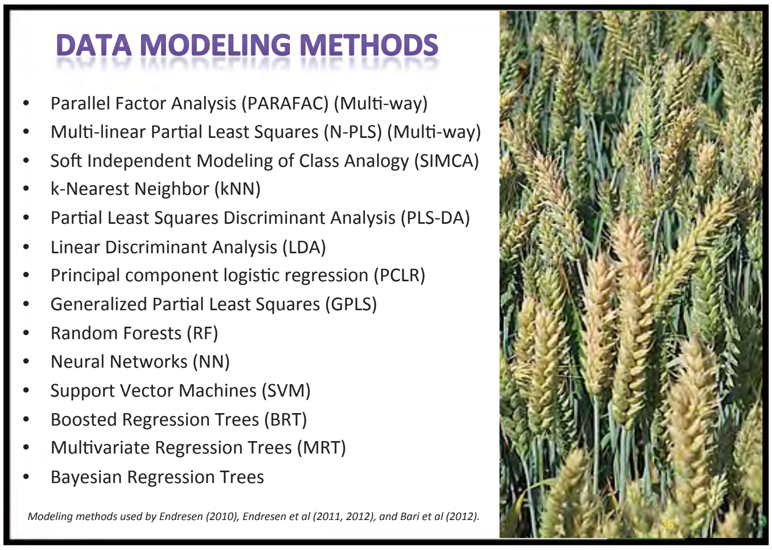
MAXENT CAN BE SEEN AS A GLM

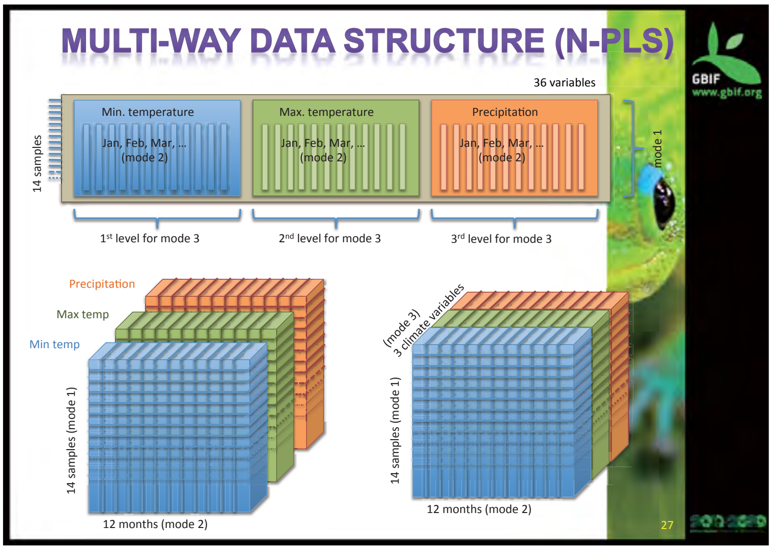


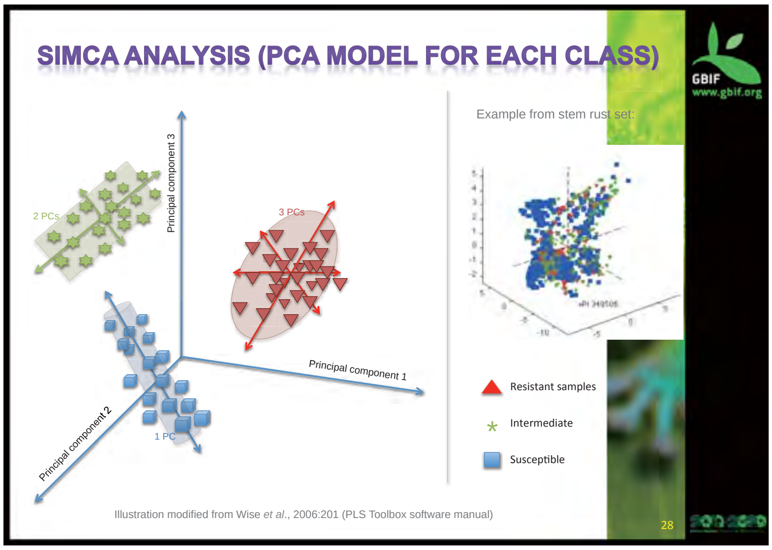
http://methodsblog.wordpress.com/2013/02/20/some-big-news-about-maxent/

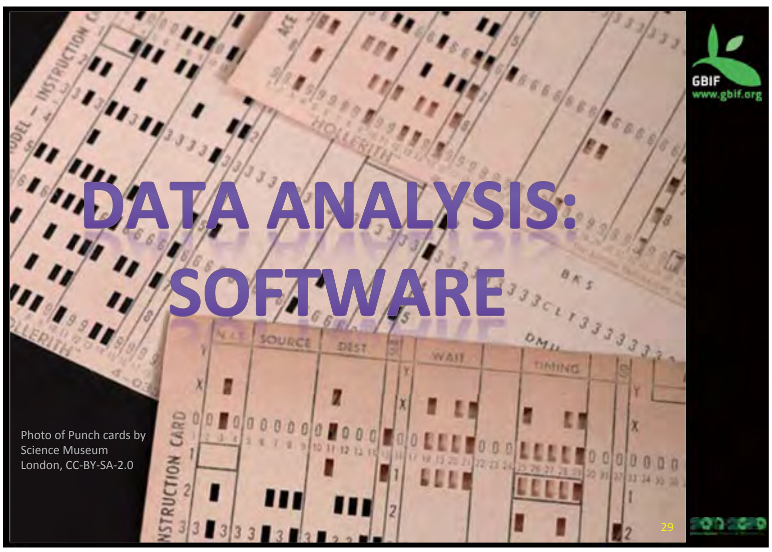
24

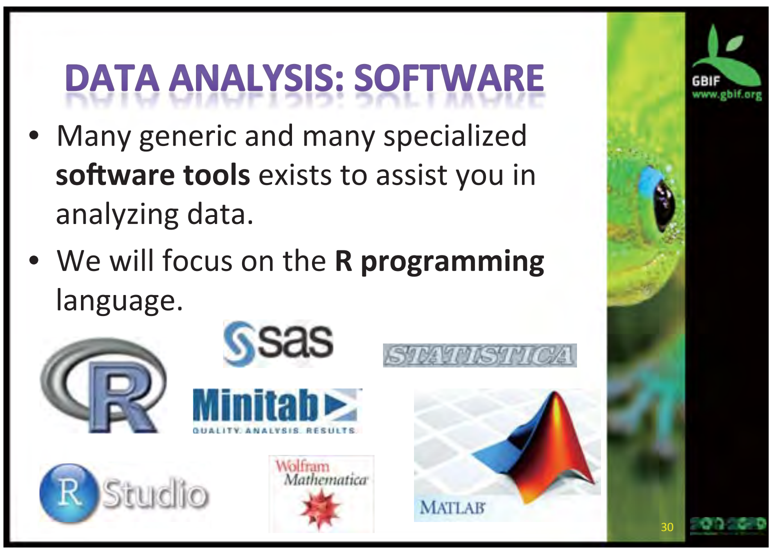

A MAXIMUM LIKELIHOOD EXPLANATION OF MAXENT
Halvorsen, R. 2012. A maximum likelihood explanation of MaxEnt, and some implications for distribution modelling. – Sommerfeltia 36: 1-132. Oslo. ISBN 82-7420-050-0 . ISSN 0800-6865. DOI: 10.2478/ v10208-011-0016-2.
A strict maximum likelihood explanation of MaxEnt, and some implications for distribution modelling
Rune Halvorsen
sommerfeltia
GBIF
www.gbif.org


DATA MODELING METHODS
• Parallel Factor Analysis (PARAFAC) (Multi-way)
• Multi-linear Partial Least Squares (N-PLS) (Multi-way)
• Soft Independent Modeling of Class Analogy (SIMCA)
• k-Nearest Neighbor (kNN)
• Partial Least Squares Discriminant Analysis (PLS-DA)
• Linear Discriminant Analysis (LDA)
• Principal component logistic regression (PCLR)
• Generalized Partial Least Squares (GPLS)
• Random Forests (RF)
• Neural Networks (NN)
• Support Vector Machines (SVM)
• Boosted Regression Trees (BRT)
• Multivariate Regression Trees (MRT)
• Bayesian Regression Trees
Modeling methods used by Endresen (2010), Endresen et al (2011, 2012), and Bari et al (2012).


MULTI-WAY DATA STRUCTURE (N-PLS)
36 variables
14 samples
Min. temperature
Max. temperature
Precipitation
Jan, Feb, Mar, ... (mode 2)
Jan, Feb, Mar, ... (mode 2)
Jan, Feb, Mar, ... (mode 2)
mode 1
1st level for mode 3
2nd level for mode 3
3rd level for mode 3
Precipitation
Max temp
Min temp
14 samples (mode 1)
12 months (mode 2)
(mode 3) 3 climate variables
14 samples (mode 1)
12 months (mode 2)
GBIF
www.gbif.org
27


SIMCA ANALYSIS (PCA MODEL FOR EACH CLASS)
Example from stem rust set:
2 PCs
3 PCs
1 PC
Principal component 3
Principal component 1
Principal component 2
Resistant samples
Intermediate
Susceptible
Illustration modified from Wise et al., 2006:201 (PLS Toolbox software manual)
GBIF
www.gbif.org
28


DATA ANALYSIS: SOFTWARE
Photo of Punch cards by Science Museum London, CC-BY-SA-2.0
GBIF
www.gbif.org
29


DATA ANALYSIS: SOFTWARE
• Many generic and many specialized software tools exists to assist you in analyzing data.
• We will focus on the R programming language.
sas
STATISTICA
R
Minitab
QUALITY. ANALYSIS. RESULTS.
R Studio
Wolfram Mathematica
MATLAB
GBIF
www.gbif.org
30

INSTALLING "R" AND "R STUDIO"

- Download the latest version of R at http://cran.r-project.org/
- Download the latest R Studio at http://www.rstudio.com/
- Available for Windows, Linux and Mac.

31

R STUDIO GUI

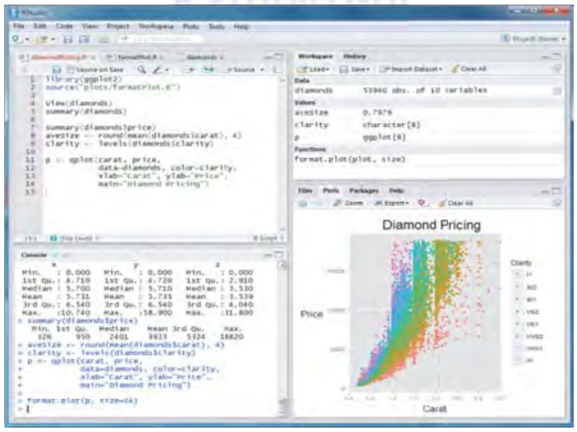


Illustration from the R Studio home page: http://www.rstudio.com/ide/

32

DISMO (R)

- Hijmans, R.J., S. Philips, J. Leathwick, and J. Elith (2013). Dismo: distribution modeling [Dismo R package]. Available at http://cran.r-project.org/web/packages/dismo/index.html
- Hijmans, R.J. and J. Elith (2013). Species distribution modeling with R. [Dismo vignette manual]. Available at http://cran.r-project.org/web/packages/dismo/vignettes/sdm.pdf
- Hijmans, R.J. (2013). Introduction to the 'raster' package (version 2.1-49). Available at http://cran.r-project.org/web/packages/raster/vignettes/Raster.pdf
- Elith, J. and J. Leathwick (2013). Boosted regression trees for ecological modeling. [Dismo vignette manual]. Available at http://cran.r-project.org/web/packages/dismo/vignettes/brt.pdf
- Dismo was released around 2009.

33

BIOMOD (R)

- Thullier W., Georges D., and Engler R. (2013). Biomod2: Ensemble platform for species distribution modeling [biomod2 R package]. Available at http://cran.r-project.org/web/packages/biomod2/index.html
- Thullier W., Georges D., and Engler R. (2013). [BIOMOD R Package]. Available at https://r-forge.r-project.org/projects/biomod/
- Thuiller W., Lafourcade B., Engler R. & Araujo M.B. (2009). BIOMOD – A platform for ensemble forecasting of species distributions. Ecography, 32, 369-373.
- BIOMOD released around 2008, biomod2 released in 2012. See also: http://www.will.chez-alice.fr/Software.html

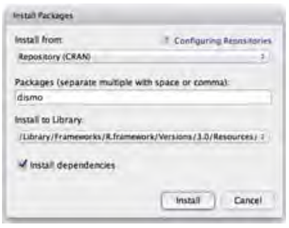
34

INSTALL DISMO

- From the R command line window:

```
install.packages(c('raster', 'rgdal', 'dismo', 'rJava'))
```

- … or use the R Studio GUI:

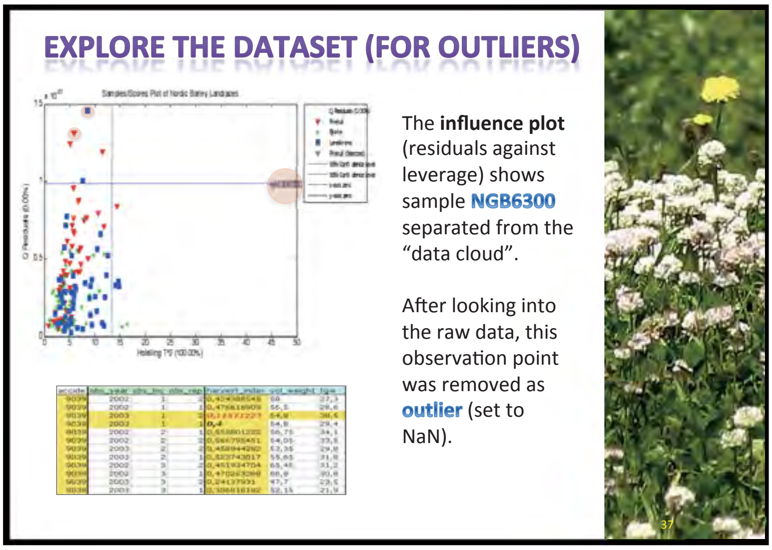


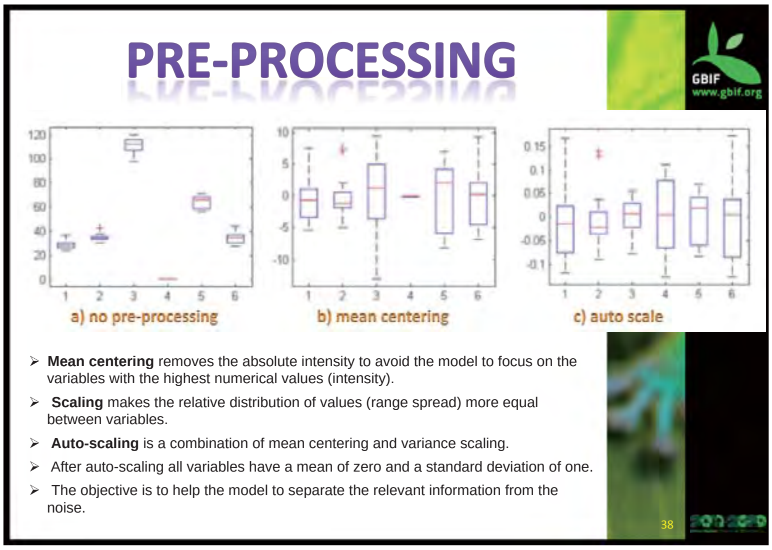
35

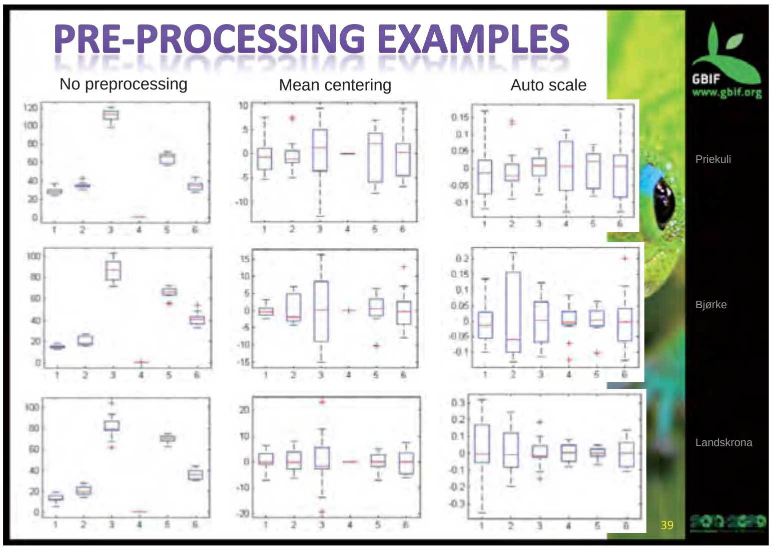
STEPS TO FOLLOW:

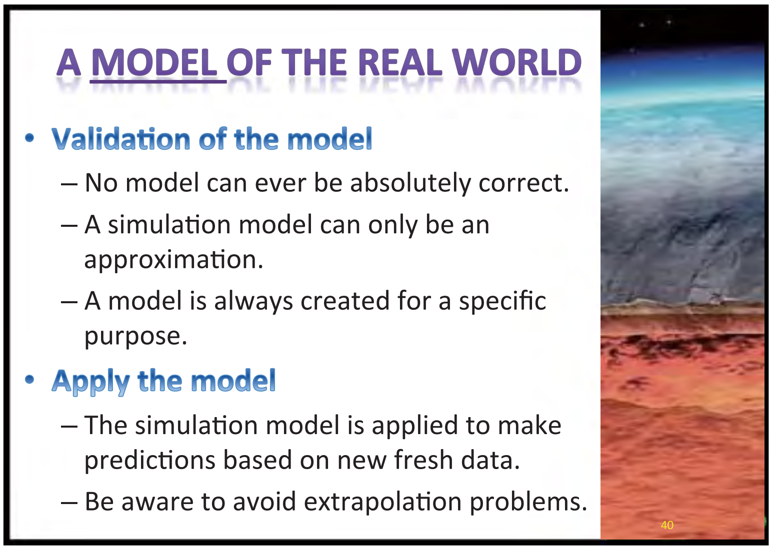
- Data collection and preparation.
- Geo-referencing site locations.
- Initial data exploration.
- Pre-processing of dataset.
- Choose modeling method.
- Calibration of model.
- Validation of model.
- Validation of prediction results.

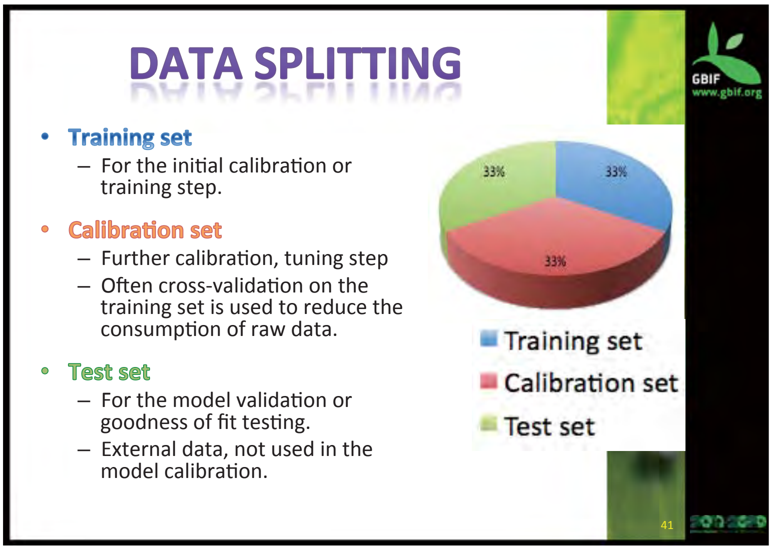
36


EXPLORE THE DATASET (FOR OUTLIERS)
The influence plot (residuals against leverage) shows sample NGB6300 separated from the "data cloud".
After looking into the raw data, this observation point was removed as outlier (set to NaN).
37

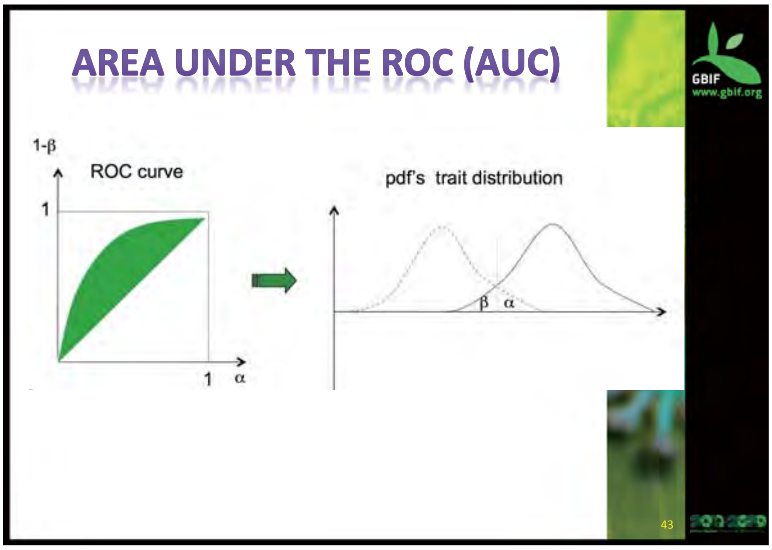

PRE-PROCESSING
GBIF
www.gbif.org
a) no pre-processing
b) mean centering
c) auto scale
Mean centering removes the absolute intensity to avoid the model to focus on the variables with the highest numerical values (intensity).
Scaling makes the relative distribution of values (range spread) more equal between variables.
Auto-scaling is a combination of mean centering and variance scaling.
After auto-scaling all variables have a mean of zero and a standard deviation of one.
The objective is to help the model to separate the relevant information from the noise.
38

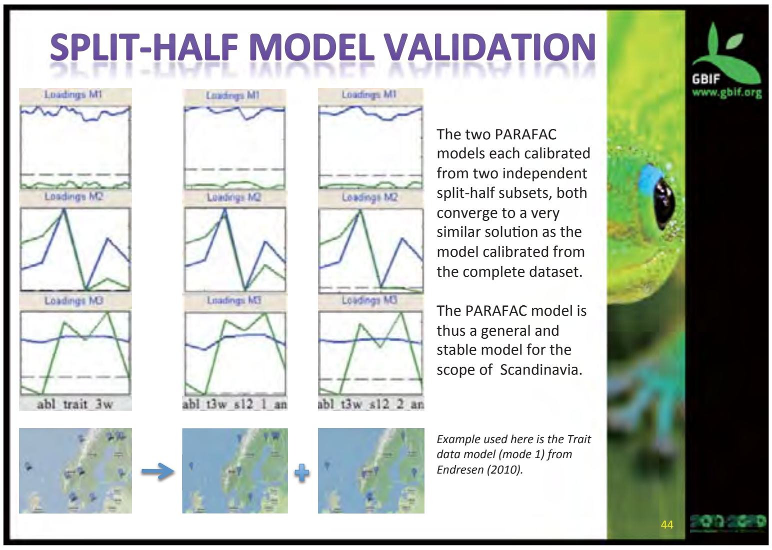

PRE-PROCESSING EXAMPLES
No preprocessing
Mean centering
Auto scale
GBIF
www.gbif.org
Priekuli
Bjørke
Landskrona
39

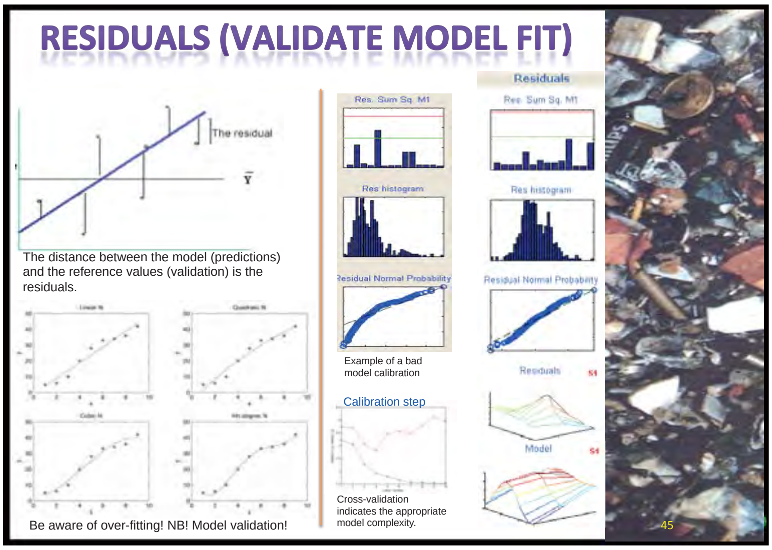

A MODEL OF THE REAL WORLD
Validation of the model
– No model can ever be absolutely correct.
– A simulation model can only be an approximation.
– A model is always created for a specific purpose.
Apply the model
– The simulation model is applied to make predictions based on new fresh data.
– Be aware to avoid extrapolation problems.
40

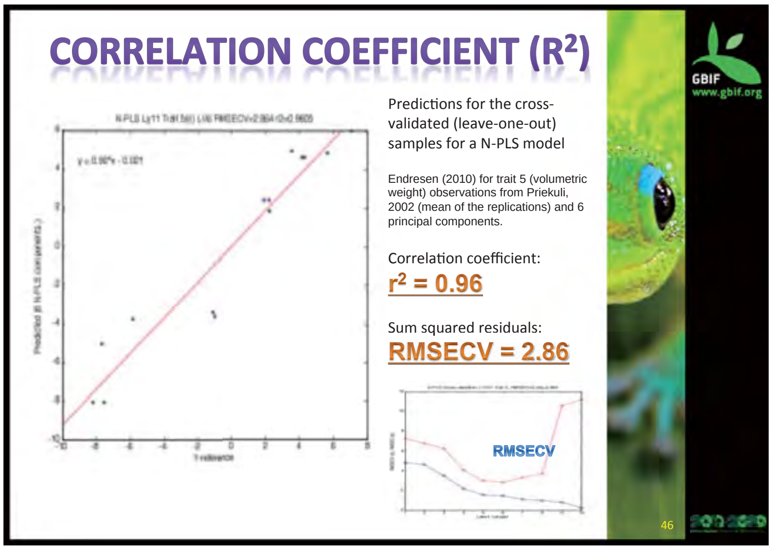

DATA SPLITTING
GBIF
www.gbif.org
Training set
– For the initial calibration or training step.
Calibration set
– Further calibration, tuning step
– Often cross-validation on the training set is used to reduce the consumption of raw data.
Test set
– For the model validation or goodness of fit testing.
– External data, not used in the model calibration.
33%
33%
33%
Training set
Calibration set
Test set
41

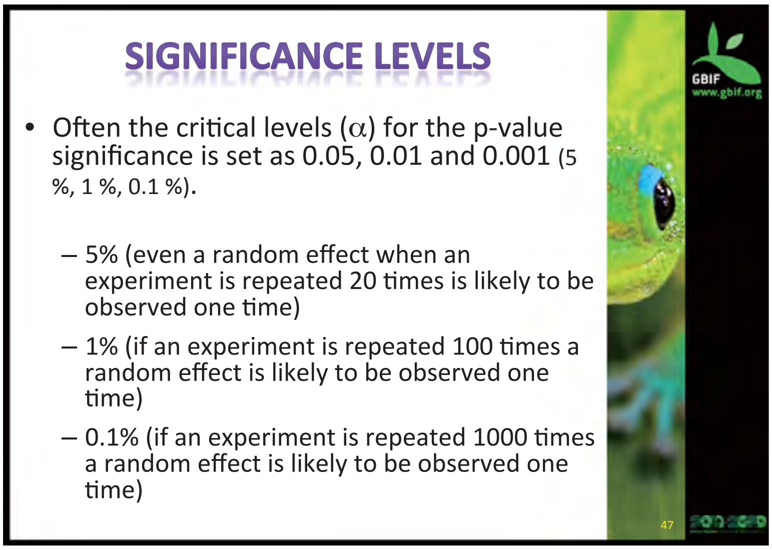

VALIDATE:
- MODEL PERFORMANCE
- MODEL RESULTS
GBIF
www.gbif.org
VALIDATE
APPROVED
APPROVED
VALIDATED
DISPUTED
42

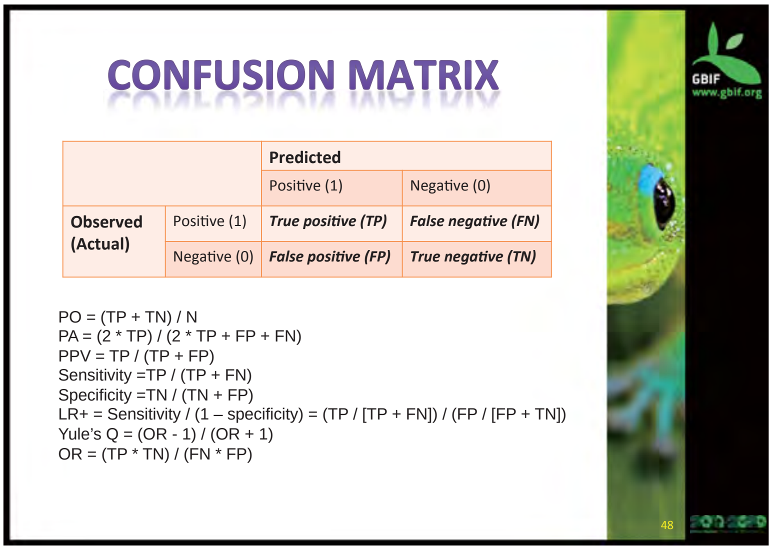

AREA UNDER THE ROC (AUC)
GBIF
www.gbif.org
1-β
ROC curve
1
1
α
pdf's trait distribution
β α
43

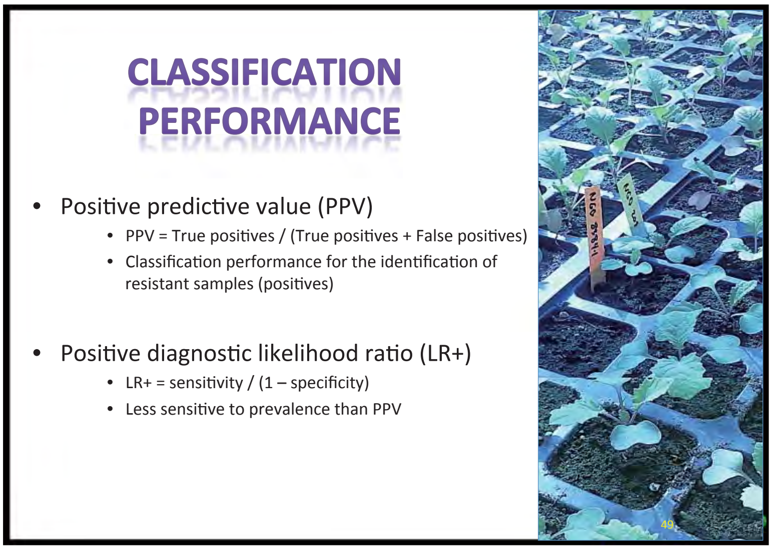

SPLIT-HALF MODEL VALIDATION
GBIF
www.gbif.org
Loadings M1
Loadings M2
Loadings M3
abl_trait_3w
abl_t3w_s12_1_an
abl_t3w_s12_2_an
The two PARAFAC models each calibrated from two independent split-half subsets, both converge to a very similar solution as the model calibrated from the complete dataset.
The PARAFAC model is thus a general and stable model for the scope of Scandinavia.
Example used here is the Trait data model (mode 1) from Endresen (2010).
44

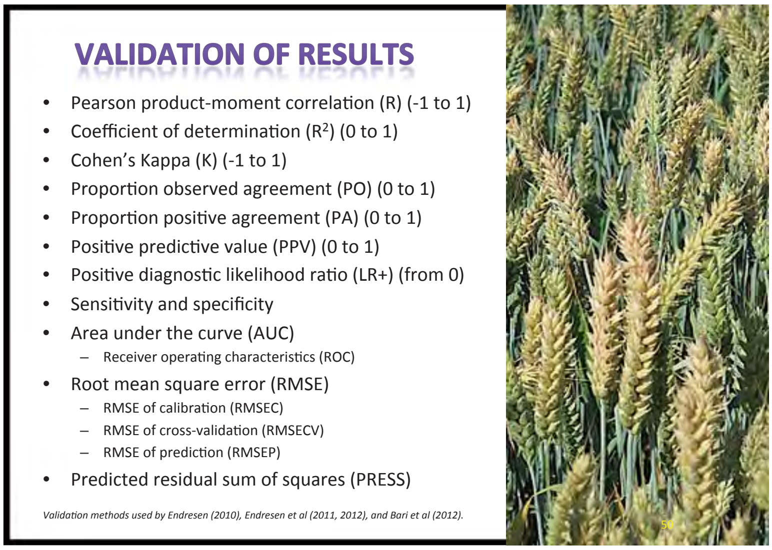

RESIDUALS (VALIDATE MODEL FIT)
The residual
Ȳ
The distance between the model (predictions) and the reference values (validation) is the residuals.
Be aware of over-fitting! NB! Model validation!
Res. Sum Sq. M1
Res histogram
Residual Normal Probability
Example of a bad model calibration
Calibration step
Cross-validation indicates the appropriate model complexity.
Residuals
Model
45


CORRELATION COEFFICIENT (R²)
GBIF
www.gbif.org
Predictions for the cross-validated (leave-one-out) samples for a N-PLS model
Endresen (2010) for trait 5 (volumetric weight) observations from Priekuli, 2002 (mean of the replications) and 6 principal components.
Correlation coefficient:
r² = 0.96
Sum squared residuals:
RMSECV = 2.86
RMSECV
46

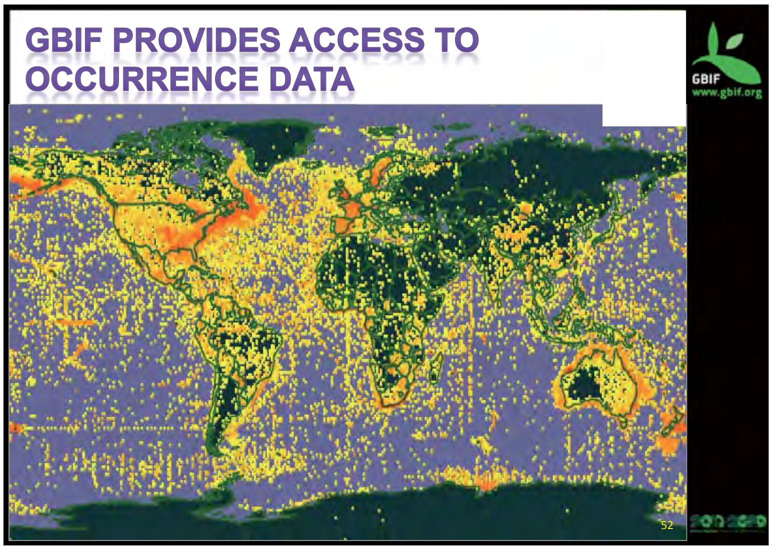

SIGNIFICANCE LEVELS
GBIF
www.gbif.org
• Often the critical levels (α) for the p-value significance is set as 0.05, 0.01 and 0.001 (5 %, 1 %, 0.1 %).
– 5% (even a random effect when an experiment is repeated 20 times is likely to be observed one time)
– 1% (if an experiment is repeated 100 times a random effect is likely to be observed one time)
– 0.1% (if an experiment is repeated 1000 times a random effect is likely to be observed one time)
47

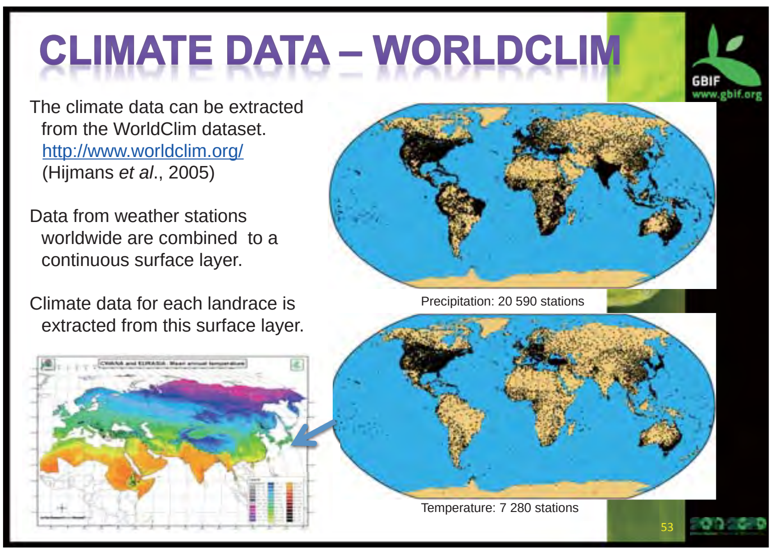

CONFUSION MATRIX
GBIF
www.gbif.org
Predicted
Positive (1)
Negative (0)
Observed (Actual)
Positive (1)
True positive (TP)
False negative (FN)
Negative (0)
False positive (FP)
True negative (TN)
PO = (TP + TN) / N
PA = (2 * TP) / (2 * TP + FP + FN)
PPV = TP / (TP + FP)
Sensitivity =TP / (TP + FN)
Specificity =TN / (TN + FP)
LR+ = Sensitivity / (1 – specificity) = (TP / [TP + FN]) / (FP / [FP + TN])
Yule's Q = (OR - 1) / (OR + 1)
OR = (TP * TN) / (FN * FP)
48

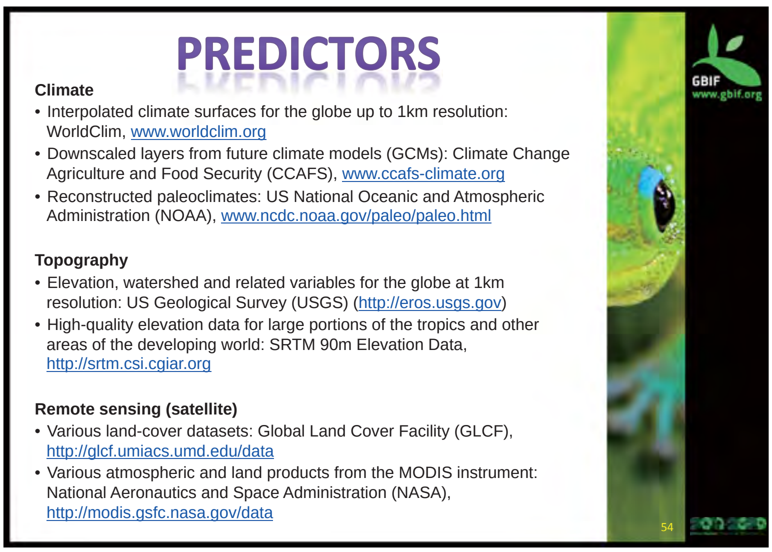

CLASSIFICATION PERFORMANCE
• Positive predictive value (PPV)
• PPV = True positives / (True positives + False positives)
• Classification performance for the identification of resistant samples (positives)
• Positive diagnostic likelihood ratio (LR+)
• LR+ = sensitivity / (1 – specificity)
• Less sensitive to prevalence than PPV
49


VALIDATION OF RESULTS
• Pearson product-moment correlation (R) (-1 to 1)
• Coefficient of determination (R^2) (0 to 1)
• Cohen's Kappa (K) (-1 to 1)
• Proportion observed agreement (PO) (0 to 1)
• Proportion positive agreement (PA) (0 to 1)
• Positive predictive value (PPV) (0 to 1)
• Positive diagnostic likelihood ratio (LR+) (from 0)
• Sensitivity and specificity
• Area under the curve (AUC)
– Receiver operating characteristics (ROC)
• Root mean square error (RMSE)
– RMSE of calibration (RMSEC)
– RMSE of cross-validation (RMSECV)
– RMSE of prediction (RMSEP)
• Predicted residual sum of squares (PRESS)
Validation methods used by Endresen (2010), Endresen et al (2011, 2012), and Bari et al (2012).


DATA SOURCES
GBIF
www.gbif.org
51

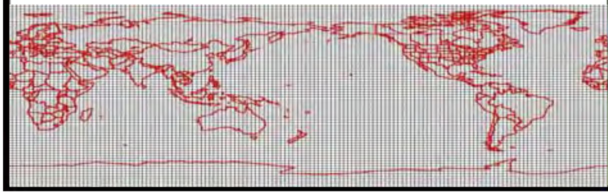

GBIF PROVIDES ACCESS TO OCCURRENCE DATA
GBIF
www.gbif.org
52


CLIMATE DATA – WORLDCLIM
The climate data can be extracted from the WorldClim dataset.
http://www.worldclim.org/
(Hijmans et al., 2005)
Data from weather stations worldwide are combined to a continuous surface layer.
Climate data for each landrace is extracted from this surface layer.
Precipitation: 20 590 stations
Temperature: 7 280 stations
GBIF
www.gbif.org
53


PREDICTORS
Climate
• Interpolated climate surfaces for the globe up to 1km resolution: WorldClim, www.worldclim.org
• Downscaled layers from future climate models (GCMs): Climate Change Agriculture and Food Security (CCAFS), www.ccafs-climate.org
• Reconstructed paleoclimates: US National Oceanic and Atmospheric Administration (NOAA), www.ncdc.noaa.gov/paleo/paleo.html
Topography
• Elevation, watershed and related variables for the globe at 1km resolution: US Geological Survey (USGS) (http://eros.usgs.gov)
• High-quality elevation data for large portions of the tropics and other areas of the developing world: SRTM 90m Elevation Data, http://srtm.csi.cgiar.org
Remote sensing (satellite)
• Various land-cover datasets: Global Land Cover Facility (GLCF), http://glcf.umiacs.umd.edu/data
• Various atmospheric and land products from the MODIS instrument: National Aeronautics and Space Administration (NASA), http://modis.gsfc.nasa.gov/data
GBIF
www.gbif.org
54

OTHER SPATIAL DATA

- Harmonized World Soil Database, www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML
- Abundance of other species (forage, prey, predators, sympatric, or interacting in other ways…) [e.g. SDM results]
- Relevant links and data at DIVA-GIS website (country level, global level, global climate, species occurrence); near global 90-meter resolution elevation data, high-resolution satellite images (LandSat), www.diva-gis.org/Data



55

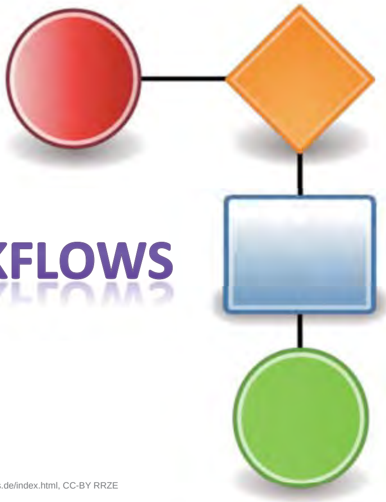
WORLD-MAP

A visual background map can often be useful for plotting your locations.

- True Marble is a free raster freely available down to 250 meter resolution, www.unearthedoutdoors.net/global_data/true_marble/download
- The DIVA-GIS project provides a useful collection of country based vector data on borders, roads, water bodies, place names, etc., www.diva-gis.org/Data
- Global Administrative Areas (GADM), www.gadm.org
- Database with eight million place names with geographical coordinates: GeoNames, www.geonames.org

56

WORKFLOWS



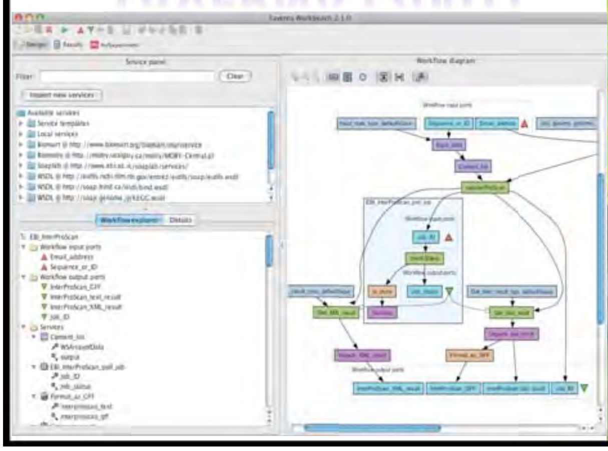
57

SCIENTIFIC WORKFLOW

- A scientific workflow is a ***string of distributed data analysis steps***.
- Data and computation steps can be online as web-services.
- A **workflow graph** describes a series of computational steps that can be **reused** for different input data or modified for exploring variation in a computational pathway.
- Specialized workflow management systems such as **Kepler** or **Taverna** provide a visual front-end to build workflow graphs.
- Buzzwords: e-Science and Grid computing.

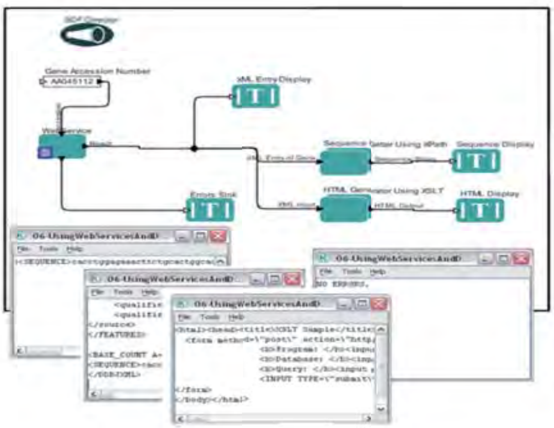
58

TAVERNA (2001)



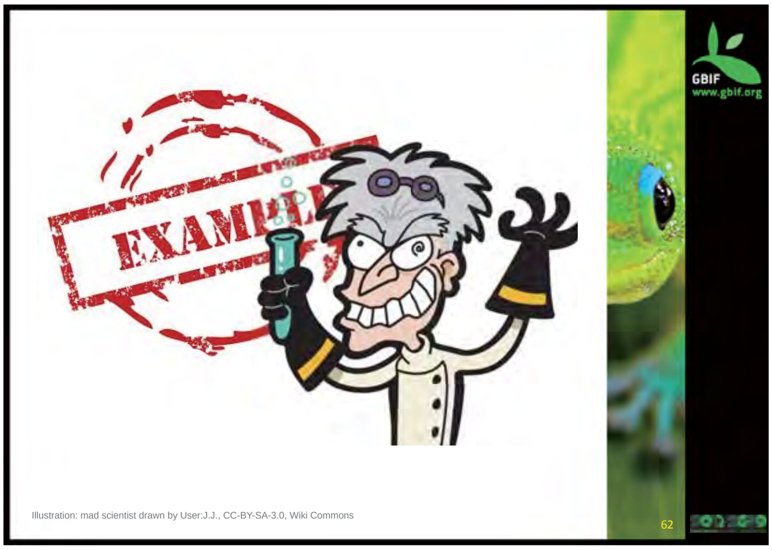
59

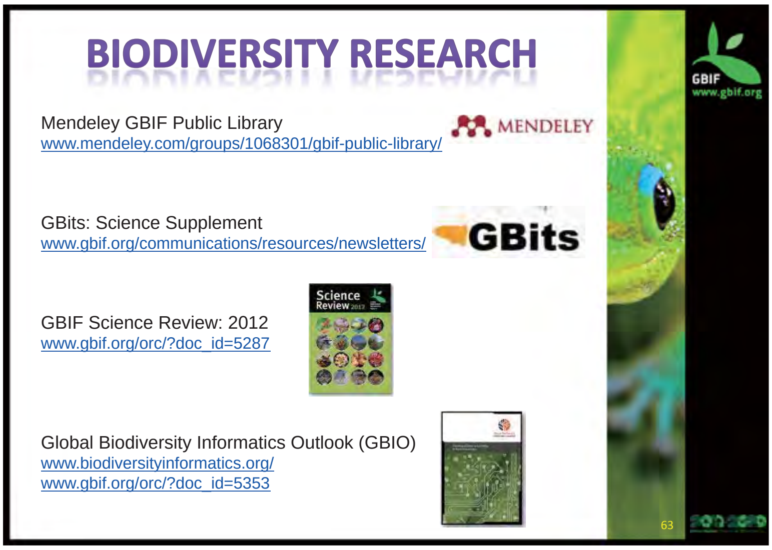
KEPLER (2002)

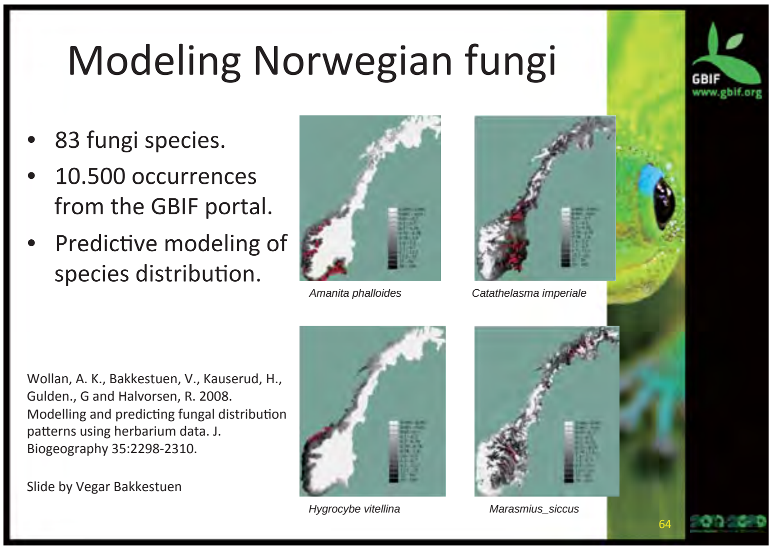


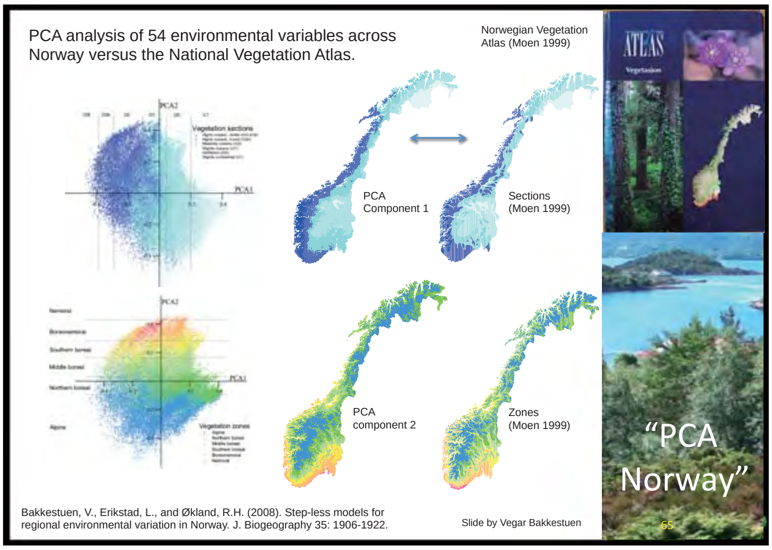
60

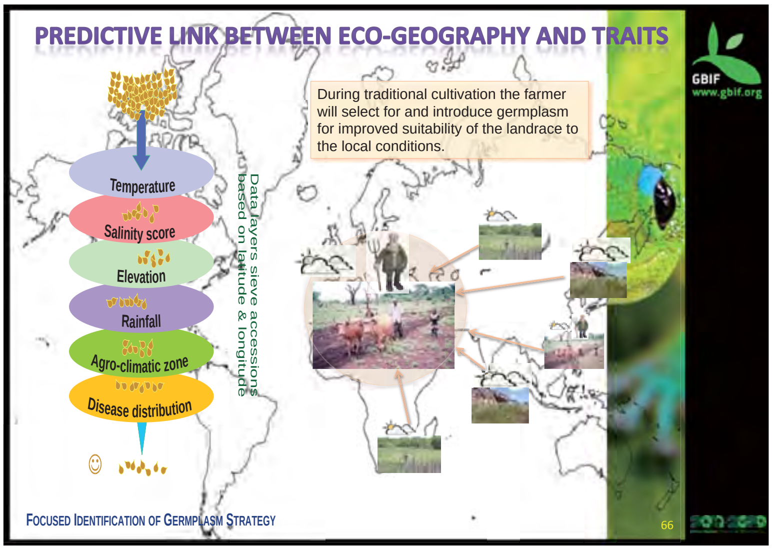

BioVeL
Biodiversity Virtual e-Laboratory
Forum
eNews Subscription
BioVeL Portal
Home
About BioVeL
Workflows
Resources
Support
News and Updates
Workflows
Introduction
Workflows
Getting started
Data Refinement Workflow
Ecological Niche Modeling Workflow
Population Modeling Workflow
myexperiment
BiodiversityCatalogue
http://www.biovel.eu/index.php/workflows
GBIF
www.gbif.org
61

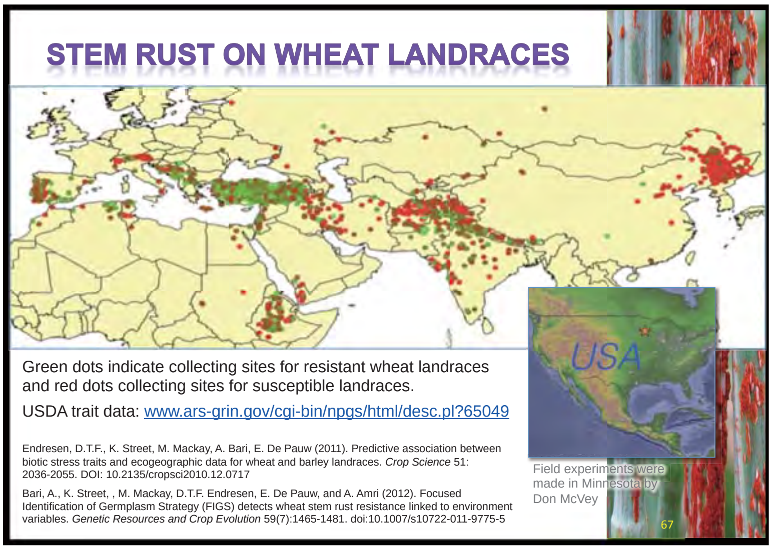

EXAMPLE
Illustration: mad scientist drawn by User:J.J., CC-BY-SA-3.0, Wiki Commons
GBIF
www.gbif.org
62

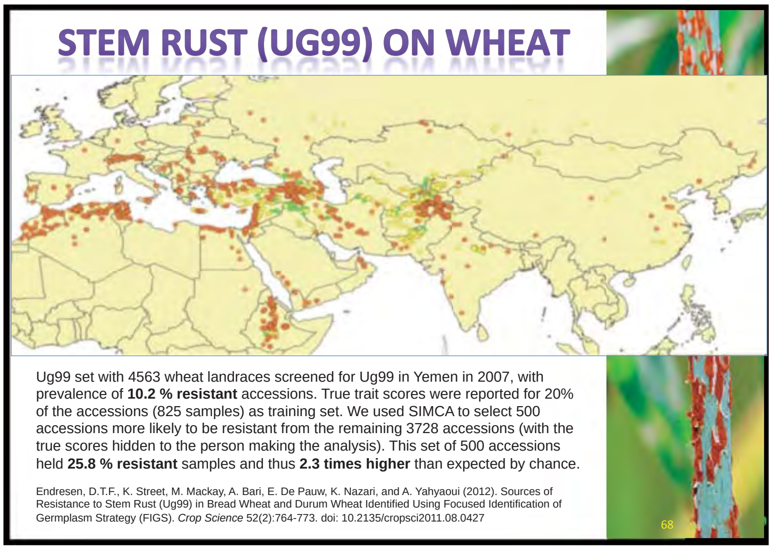

BIODIVERSITY RESEARCH
Mendeley GBIF Public Library
www.mendeley.com/groups/1068301/gbif-public-library/
MENDELEY
GBits: Science Supplement
www.gbif.org/communications/resources/newsletters/
GBits
GBIF Science Review: 2012
www.gbif.org/orc/?doc_id=5287
Science Review
Global Biodiversity Informatics Outlook (GBIO)
www.biodiversityinformatics.org/
www.gbif.org/orc/?doc_id=5353
GBIF
www.gbif.org
63

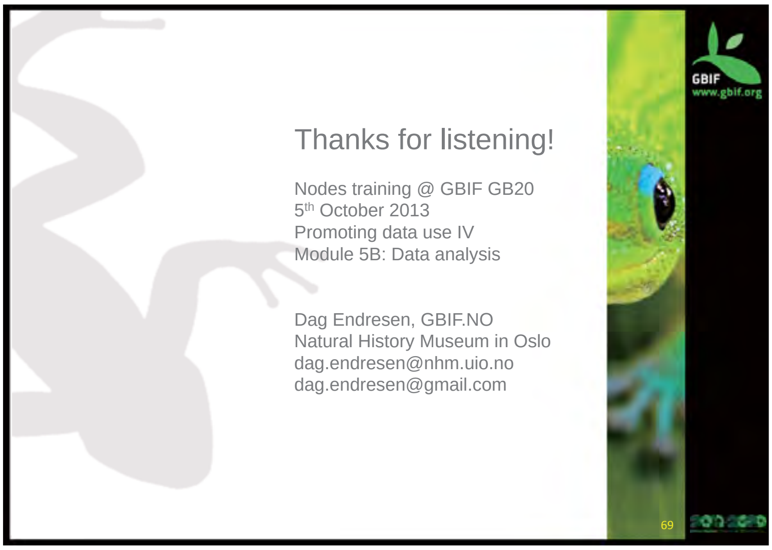

Modeling Norwegian fungi
83 fungi species.
10.500 occurrences from the GBIF portal.
Predictive modeling of species distribution.
Wollan, A. K., Bakkestuen, V., Kauserud, H., Gulden., G and Halvorsen, R. 2008. Modelling and predicting fungal distribution patterns using herbarium data. J. Biogeography 35:2298-2310.
Slide by Vegar Bakkestuen
Amanita phalloides
Catathelasma imperiale
Hygrocybe vitellina
Marasmius_siccus
GBIF
www.gbif.org
64


PCA analysis of 54 environmental variables across Norway versus the National Vegetation Atlas.
Norwegian Vegetation Atlas (Moen 1999)
PCA Component 1
Sections (Moen 1999)
PCA component 2
Zones (Moen 1999)
ATLAS
"PCA Norway"
Bakkestuen, V., Erikstad, L., and Økland, R.H. (2008). Step-less models for regional environmental variation in Norway. J. Biogeography 35: 1906-1922.
Slide by Vegar Bakkestuen
65


PREDICTIVE LINK BETWEEN ECO-GEOGRAPHY AND TRAITS
During traditional cultivation the farmer will select for and introduce germplasm for improved suitability of the landrace to the local conditions.
Temperature
Salinity score
Elevation
Rainfall
Agro-climatic zone
Disease distribution
Data layers sieve accessions based on latitude & longitude
FOCUSED IDENTIFICATION OF GERMPLASM STRATEGY
GBIF
www.gbif.org
66


STEM RUST ON WHEAT LANDRACES
Green dots indicate collecting sites for resistant wheat landraces and red dots collecting sites for susceptible landraces.
USDA trait data: www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?65049
Endresen, D.T.F., K. Street, M. Mackay, A. Bari, E. De Pauw (2011). Predictive association between biotic stress traits and ecogeographic data for wheat and barley landraces. *Crop Science* 51: 2036-2055. DOI: 10.2135/cropsci2010.12.0717
Bari, A., K. Street, , M. Mackay, D.T.F. Endresen, E. De Pauw, and A. Amri (2012). Focused Identification of Germplasm Strategy (FIGS) detects wheat stem rust resistance linked to environment variables. *Genetic Resources and Crop Evolution* 59(7):1465-1481. doi:10.1007/s10722-011-9775-5
USA
Field experiments were made in Minnesota by Don McVey
67


STEM RUST (UG99) ON WHEAT
Ug99 set with 4563 wheat landraces screened for Ug99 in Yemen in 2007, with prevalence of **10.2 % resistant** accessions. True trait scores were reported for 20% of the accessions (825 samples) as training set. We used SIMCA to select 500 accessions more likely to be resistant from the remaining 3728 accessions (with the true scores hidden to the person making the analysis). This set of 500 accessions held **25.8 % resistant** samples and thus **2.3 times higher** than expected by chance.
Endresen, D.T.F., K. Street, M. Mackay, A. Bari, E. De Pauw, K. Nazari, and A. Yahyaoui (2012). Sources of Resistance to Stem Rust (Ug99) in Bread Wheat and Durum Wheat Identified Using Focused Identification of Germplasm Strategy (FIGS). *Crop Science* 52(2):764-773. doi: 10.2135/cropsci2011.08.0427
68


GBIF
www.gbif.org
Thanks for listening!
Nodes training @ GBIF GB20
5th October 2013
Promoting data use IV
Module 5B: Data analysis
Dag Endresen, GBIF.NO
Natural History Museum in Oslo
dag.endresen@nhm.uio.no
dag.endresen@gmail.com
69

GBIF Governing Board 20

Module 6B: New GBIF Tools II
2013 Portal and NPT Startup

Daniel Amariles
IT Leader, National Biodiversity Information System of Colombia
Global Biodiversity Information Facility (GBIF)

Saturday, October 5th, 2013

SUMMARY

The development of the new GBIF portal has included the complete redesign of its web services, which are now offered in a complete API that is the base for all the data related pages in the portal.

This is an introduction to the new system, for those who are considering building (web) applications using the API.

This presentation corresponds to module 6B of the GB20 training course for Nodes held in October 2013 in Berlin, Germany.

INDEX

INTRODUCTION TO NPT STARTUP

"NPT Startup is a Drupal site that sets up a web presence for GBIF Participant Nodes."

INTRODUCTION TO NPT STARTUP

Demo site
http://nptstartup.gbif.org/

Presentation video
http://vimeo.com/63640790

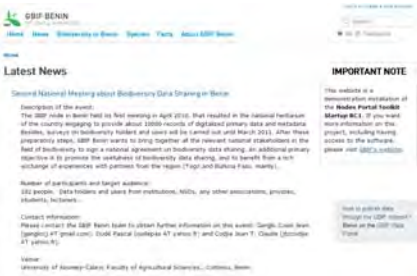
GitHub Repository
https://github.com/gbif/gbif-npt-startup/releases

NPT STARTUP FEATURES

- Biodiversity information coming from the GBIF Data Portal and from the Encyclopedia of Life.

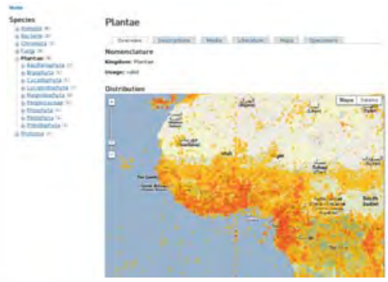
NPT STARTUP FEATURES

- Common website features like news, forums and blogs so the website can be used by a GBIF Participant Node as a communication website.



NPT STARTUP FEATURES

- Includes an initial country checklist, generated by assembling scientific names that are extracted from species occurrence data within a country available in the GBIF network.
- Biodiversity data of a species are shown as static species distribution maps.



NPT STARTUP FEATURES

- Further information available about a species includes a species description and media from the Encyclopedia of Life (EOL).

NPT STARTUP FEATURES

- At the country level, NPT Startup provides some basic facts on data publishing activities.

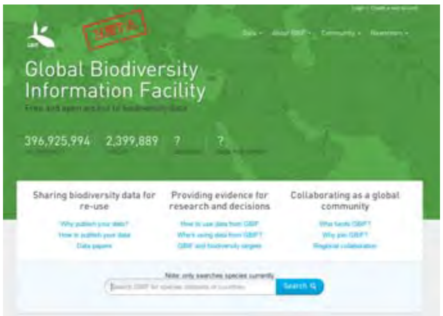
NPT STARTUP FEATURES

- Shows scientific articles and reports using GBIF-mediated data, published by authors from the country and/or about the biodiversity of the country.

INDEX

1. Introduction to NPT Startup
2. **Introduction to the 2013 Data Portal**
3. The Data Portal API
4. From the Portal to the API
5. The mapping API
6. Use the API data
7. Practical exercise

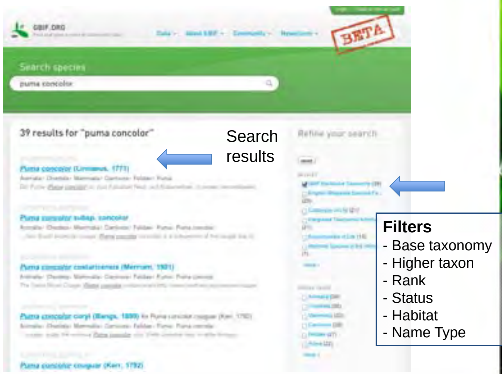
INTRODUCTION TO THE 2013 DATA PORTAL



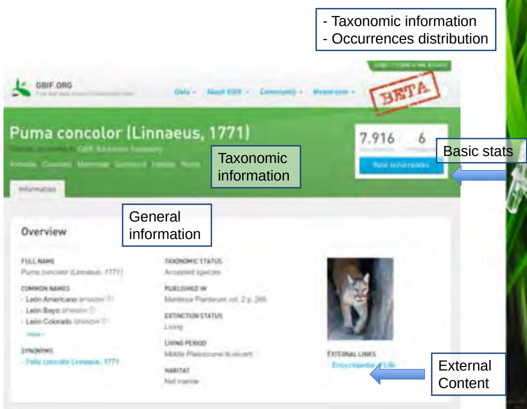
MAIN SECTIONS

- Explore occurrences
- Explore datasets
- Explore species
- Explore by country

SPECIES SEARCH



SPECIE OVERVIEW



INDEX

1. Introduction to NPT Startup
2. Introduction to the 2013 Data Portal
3. **The Data Portal API**
4. From the Portal to the API
5. The mapping API
6. Use the API data
7. Practical exercise

DATA PORTAL API

An interface to access the data published through the GBIF network using Web Services.

DATA PORTAL API

GBIF Portal Registry API:
http://portaldev.gbif.org/developer/registry

GBIF All Webservices API:
http://dev.gbif.org/wiki/display/POR/Webservice+API

"All services take parameters as query string encoded GET key=value parameters and respond with an application/**json** content type"

AVAILABLE API

- Checklist Bank Services: Name usage
- Checklist Bank Services: GBIF Backbone Taxon Lookup
- Checklist Bank Services: Search Service
- Checklist Bank Services: Metrics
- Registry Services: Node
- Registry Services: Organization
- Registry Services: Network
- Registry Services: Technical Installation
- Registry Services: Dataset
- Registry Services: Dataset Search Service
- Occurrence Services: Occurrence Record
- Occurrence Services: Search Service
- Occurrence Services: Download Service
- Cube Metrics Services: Occurrence Service
- Crawler Services: Dataset Crawl Service

API CALL EXAMPLES

Names API

http://api.gbif.org/lookup/name_usage/?name=oenanthe&kingdom=plants

Dataset API

http://api.gbif.org/dataset/52a423d2-0486-4e77-bcee-6350d708d6ff

Organization API

http://api.gbif.org/organization/10980920-6dad-11da-ad13-b8a03c50a862

Names Usage API

http://api.gbif.org/name_usage/1

Occurrence search API

http://api.gbif.org/occurrence/search?TAXON_KEY=6746

INDEX

1. Introduction to NPT Startup
2. Introduction to the 2013 Data Portal
3. The Data Portal API
4. **From the Portal to the API**
5. The mapping API
6. Use the API data
7. Practical exercise

QUERYING ORGANIZATION

Get organization UUID

10980920-6dad-11da-ad13-b8a03c50a862

QUERYING AN ORGANIZATION

To build the API URL:

http://api.gbif.org/organization/ 10980920-6dad-11da-ad13-b8a03c50a862

API Base URL + Endpoint URL + Organization UUID

Use REST client to EXPLORE THE CODE

USE THE API DATA

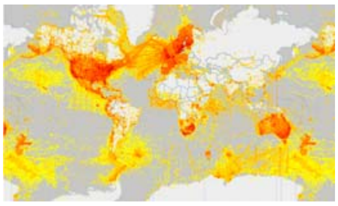
INDEX

1. Introduction to NPT Startup
2. Introduction to the 2013 Data Portal
3. The Data Portal API
4. From the Portal to the API
5. **The mapping API**
6. Use the API data
7. Practical exercise

MAPPING API v1.0

“Looking to overlay GBIF content on your maps? Try this simple to use, flexible tile rendering service”

http://uat.gbif.org/developer/maps



MAPPING API v1.0

This service is intended for use wth commonly used clients such as the google maps api, leaflet JS library or the modest maps JS library.

Google maps

Leaflet

http://leafletjs.com/

These libraries allow the GBIF layers to be visualized with other content, such as those coming from web map service (WMS) providers. It should be noted that the mapping api is not a WMS service, nor does it support WFS capabilities.

CUSTOMIZING LAYER CONTENT

The format of the URL is as follows:

```
http://api.gbif.org/map/density/tile?x={x}&y={y}&z={z}
```

With the following required parameters:

type : TAXON, DATASET, COUNTRY or PUBLISHER

key : The appropriate key for the chosen type (a taxon key, dataset/publisher UUID or 2 letter ISO country code)

CUSTOMIZING LAYER CONTENT

The layer parameter is an optional multi-evaluated field in which it is declared the layers to be combined by the server for this tile.

- Basis of record
  - Fossil
  - Living specimen
- Year Range
  - Observations year range
  - Specimens’ year range
  - Other year ranges

Example:

```
&layer=OBS_2010_2020&layer=SP_2010_2020&layer=OTH_2010_2020&layer=LIVING&layer=FOSSIL
```

STYLING A LAYER

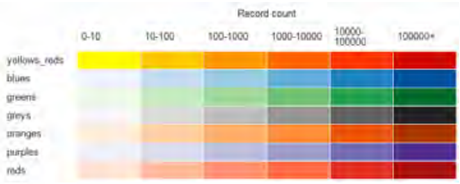
"Styling the configured layer is controlled through either the **colors** parameter or the **palette** parameter."

Using colors parameter:

```
&colors=%2C100%2C%23FF000033%7C100%2C10000%2C%2300FF0033%7C10000%2C%2C%230000FF33
```

Using palette parameter:

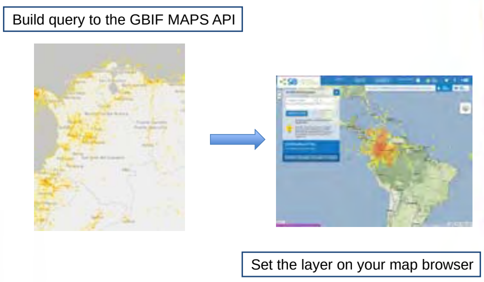
```
&palette=greens
```



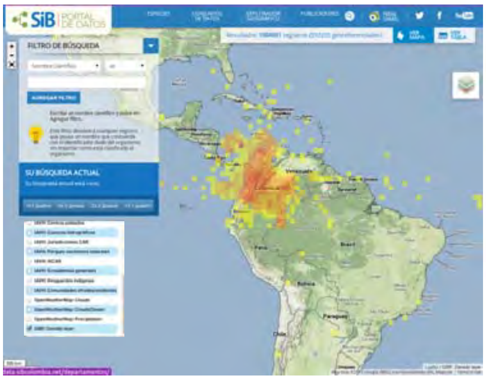
INDEX

1. Introduction to NPT Startup
2. Introduction to the 2013 Data Portal
3. The Data Portal API
4. From the Portal to the API
5. The mapping API
6. **Use the API data**
7. Practical exercise

USE THE API DATA



USE THE API DATA

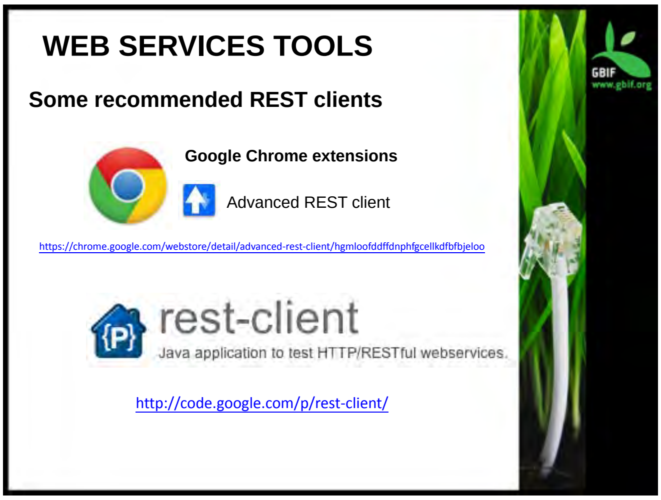


INDEX

1. Introduction to NPT Startup
2. Introduction to the 2013 Data Portal
3. The Data Portal API
4. From the Portal to the API
5. The mapping API
6. Use the API data
7. **Practical exercise**

PRACTICAL EXERCISE: PLAYING WITH PORTAL API

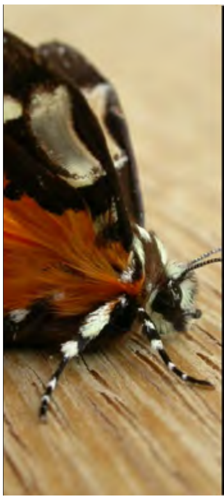
1. Explore data with a REST Client
2. Try to imagine how would you like to use those data in your portal or application
3. Make one or more wireframes about how to integrate the data
4. Homework: Implement a JSON client to get the data


WEB SERVICES TOOLS
Some recommended REST clients
Google Chrome extensions
Advanced REST client
https://chrome.google.com/webstore/detail/advanced-rest-client/hgmloofddffdnphfgcellkdfbfbjeloo
rest-client
Java application to test HTTP/RESTful webservices.
http://code.google.com/p/rest-client/
GBIF
www.gbif.org


WIREFRAMING
Recommended wireframing tool
Cacoo
https://cacoo.com/
GBIF
www.gbif.org


GBIF
www.gbif.org
GBIF Governing Board 20
Module 6B: New GBIF Tools II
2013 Portal and NPT Startup
Daniel Amariles
IT Leader, National Biodiversity Information System of Colombia
Global Biodiversity Information Facility (GBIF)
Saturday, October 5th, 2013

附錄 4. GBIF Secretariat update



GBIF
www.gbif.org

GBIF Secretariat update

Donald Hobern
GBIF Director
Global Biodiversity Information Facility (GBIF)

GBIF GB20, Berlin, 6-11 October 2013

GBIO Framework

Focus Area: Culture

- The context for sharing digital knowledge
  - Data must be **available for reuse**
  - Data must **follow standards** to support discovery and use
  - Data must be **preserved for future uses**
  - **Policies and practices** must reinforce open use
  - The whole community should collaborate to **curate data**
- Issues shared in common with all research domains
- Investments here will multiply value of other components

Focus Area: Data

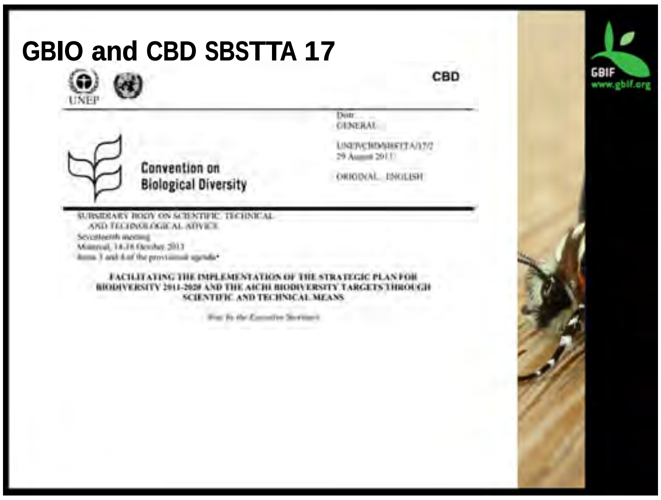
- The streams of primary biodiversity data
  - **Literature** and journals
  - Natural history **collections**
  - Professional and amateur **field observations** and surveys
  - Molecular **sequencing**
  - **Remote sensing** (including camera traps, acoustic monitoring, etc.)
- All deliver fundamental observations and measurements of biodiversity
- Foundations for analysis and understanding

Focus Area: Evidence

- Organised views of biodiversity data
  - Consistent assessment of **quality and fitness-for-use**
  - Comprehensive digital **nomenclature and taxonomy**
  - Access to all evidence for recorded **species occurrence**
  - Access to species **traits, measurements and interactions**
  - **Services and interfaces** to access data as needed
- Provide comprehensive organised views of all relevant data
- Act as a “lens” into primary data

Focus Area: Understanding

- The application of data to address questions
  - Integrate data into **spatial models**
  - Develop **temporal analyses**
  - Incorporate **biological reality** into models
  - Present **compelling representations** of biodiversity
  - Optimise **future investment** in biodiversity informatics
- Data-driven models for science and planning
- Integrate biodiversity with other research and data domains



GBIO and CBD SBSTTA 17

UNEP CBD

Convention on Biological Diversity

Distr. GENERAL

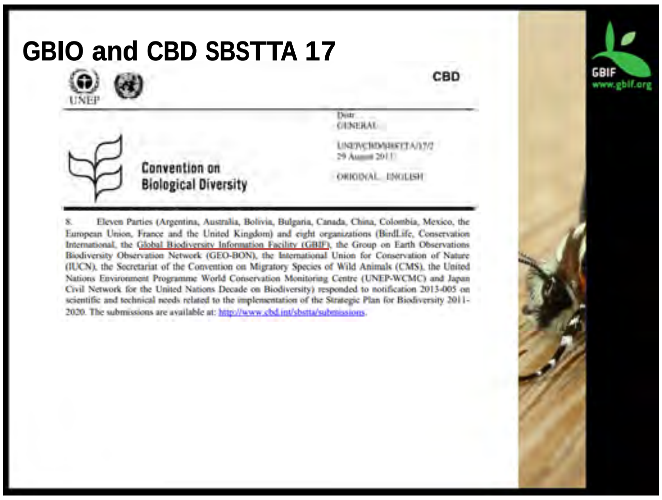
UNEP/CBD/SBSTTA/17/2
29 August 2013

ORIGINAL: ENGLISH

SUBSIDIARY BODY ON SCIENTIFIC, TECHNICAL AND TECHNOLOGICAL ADVICE
Seventeenth meeting
Montreal, 14-18 October 2013
Items 3 and 4 of the provisional agenda*

FACILITATING THE IMPLEMENTATION OF THE STRATEGIC PLAN FOR BIODIVERSITY 2011-2020 AND THE AICHI BIODIVERSITY TARGETS THROUGH SCIENTIFIC AND TECHNICAL MEANS

Note by the Executive Secretary



GBIO and CBD SBSTTA 17

UNEP CBD

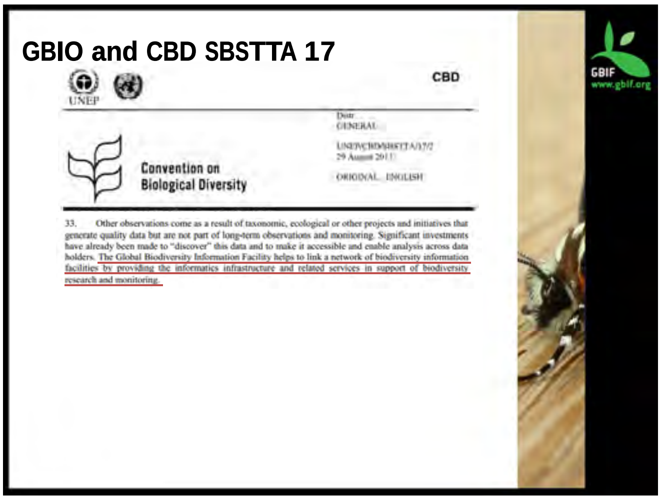
Convention on Biological Diversity

Distr. GENERAL

UNEP/CBD/SBSTTA/17/2
29 August 2013

ORIGINAL: ENGLISH

8. Eleven Parties (Argentina, Australia, Bolivia, Bulgaria, Canada, China, Colombia, Mexico, the European Union, France and the United Kingdom) and eight organizations (BirdLife, Conservation International, the Global Biodiversity Information Facility (GBIF), the Group on Earth Observations Biodiversity Observation Network (GEO-BON), the International Union for Conservation of Nature (IUCN), the Secretariat of the Convention on Migratory Species of Wild Animals (CMS), the United Nations Environment Programme World Conservation Monitoring Centre (UNEP-WCMC) and Japan Civil Network for the United Nations Decade on Biodiversity) responded to notification 2013-005 on scientific and technical needs related to the implementation of the Strategic Plan for Biodiversity 2011-2020. The submissions are available at: http://www.cbd.int/sbstta/submissions.



GBIO and CBD SBSTTA 17

UNEP CBD

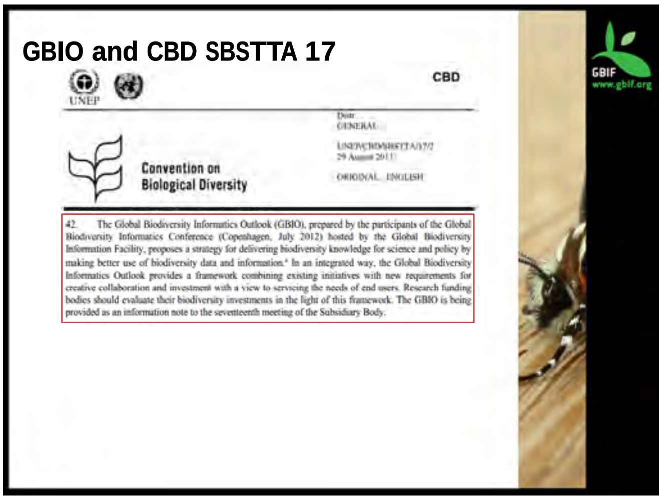
Convention on Biological Diversity

Distr. GENERAL

UNEP/CBD/SBSTTA/17/2
29 August 2013

ORIGINAL: ENGLISH

33. Other observations come as a result of taxonomic, ecological or other projects and initiatives that generate quality data but are not part of long-term observations and monitoring. Significant investments have already been made to "discover" this data and to make it accessible and enable analysis across data holders. The Global Biodiversity Information Facility helps to link a network of biodiversity information facilities by providing the informatics infrastructure and related services in support of biodiversity research and monitoring.



GBIO and CBD SBSTTA 17

UNEP CBD

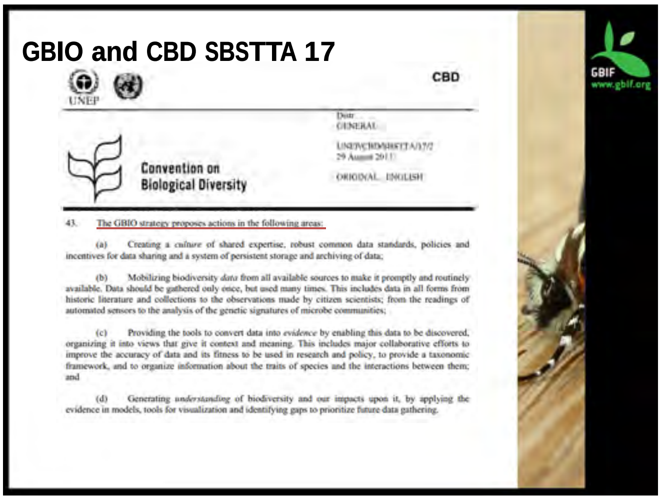
Convention on Biological Diversity

Distr. GENERAL

UNEP/CBD/SBSTTA/17/2
29 August 2013

ORIGINAL: ENGLISH

42. The Global Biodiversity Informatics Outlook (GBIO), prepared by the participants of the Global Biodiversity Informatics Conference (Copenhagen, July 2012) hosted by the Global Biodiversity Information Facility, proposes a strategy for delivering biodiversity knowledge for science and policy by making better use of biodiversity data and information.[a] In an integrated way, the Global Biodiversity Informatics Outlook provides a framework combining existing initiatives with new requirements for creative collaboration and investment with a view to servicing the needs of end users. Research funding bodies should evaluate their biodiversity investments in the light of this framework. The GBIO is being provided as an information note to the seventeenth meeting of the Subsidiary Body.



GBIO and CBD SBSTTA 17

UNEP CBD

Convention on Biological Diversity

Distr. GENERAL

UNEP/CBD/SBSTTA/17/2
29 August 2013

ORIGINAL: ENGLISH

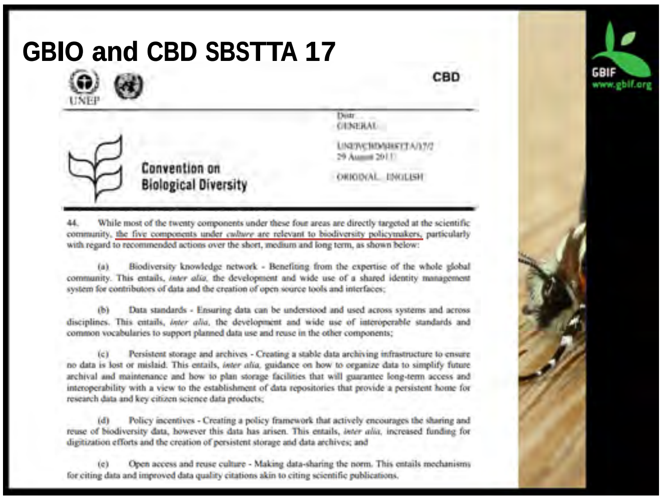
43. The GBIO strategy proposes actions in the following areas:

(a) Creating a *culture* of shared expertise, robust common data standards, policies and incentives for data sharing and a system of persistent storage and archiving of data;

(b) Mobilizing biodiversity *data* from all available sources to make it promptly and routinely available. Data should be gathered only once, but used many times. This includes data in all forms from historic literature and collections to the observations made by citizen scientists; from the readings of automated sensors to the analysis of the genetic signatures of microbe communities;

(c) Providing the tools to convert data into *evidence* by enabling this data to be discovered, organizing it into views that give it context and meaning. This includes major collaborative efforts to improve the accuracy of data and its fitness to be used in research and policy, to provide a taxonomic framework, and to organize information about the traits of species and the interactions between them; and

(d) Generating *understanding* of biodiversity and our impacts upon it, by applying the evidence in models, tools for visualization and identifying gaps to prioritize future data gathering.



GBIO and CBD SBSTTA 17

UNEP CBD

Convention on Biological Diversity

Distr. GENERAL

UNEP/CBD/SBSTTA/17/2
29 August 2013

ORIGINAL: ENGLISH

44. While most of the twenty components under these four areas are directly targeted at the scientific community, the five components under *culture* are relevant to biodiversity policymakers, particularly with regard to recommended actions over the short, medium and long term, as shown below:

(a) Biodiversity knowledge network - Benefiting from the expertise of the whole global community. This entails, *inter alia*, the development and wide use of a shared identity management system for contributors of data and the creation of open source tools and interfaces;

(b) Data standards - Ensuring data can be understood and used across systems and across disciplines. This entails, *inter alia*, the development and wide use of interoperable standards and common vocabularies to support planned data use and reuse in the other components;

(c) Persistent storage and archives - Creating a stable data archiving infrastructure to ensure no data is lost or mislaid. This entails, *inter alia*, guidance on how to organize data to simplify future archival and maintenance and how to plan storage facilities that will guarantee long-term access and interoperability with a view to the establishment of data repositories that provide a persistent home for research data and key citizen science data products;

(d) Policy incentives - Creating a policy framework that actively encourages the sharing and reuse of biodiversity data, however this data has arisen. This entails, *inter alia*, increased funding for digitization efforts and the creation of persistent storage and data archives; and

(e) Open access and reuse culture - Making data-sharing the norm. This entails mechanisms for citing data and improved data quality citations akin to citing scientific publications.

GBIO and CBD SBSTTA 17

UNEP

CBD

Distr.
GENERAL

Convention on Biological Diversity

ORIGINAL: ENGLISH

45. The Global Biodiversity Informatics Outlook provides a framework for managing, analysing, using and communicating biodiversity data. GBIO provides a framework around which funders, policymakers, researchers, information technology specialists, educators and the general public can unite to advance our ability to manage and analyze biodiversity information and data.

New portal

- Launch on Wednesday
- Key features
  - Integration of communications and data
  - Removal of arbitrary data limits
  - Improved country pages
  - Improved data publisher / metadata views
  - Improved maps
  - Real-time indexing of new / updated data
  - Built using public web services

Consultations

- Informatics consultations in 2013
  - Log-in processes for portal
  - Use of persistent identifiers
  - Licences for GBIF-mediated data
- Process works well – need more consistency
  - Clear life cycle with time lines
  - Visibility on portal
  - Mailbox and discussion area
- Consultations to prepare for 2014 WP
  - Email and website announcements late 2013
  - Refine WP activities early in 2014

Communication strategy

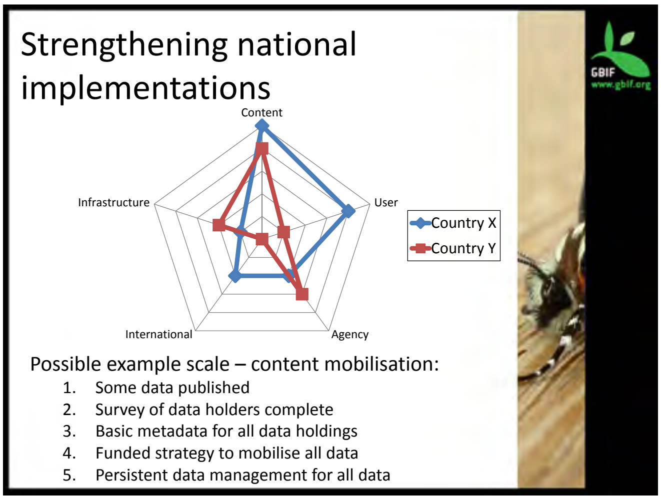
- Two fundamental needs for GBIF:
  - **Relevance** (for Participants and funders)
    - Why GBIF is important
    - What services GBIF deliver
    - How GBIF helps meet their goals and objectives
  - **Trust** (for research community)
    - What GBIF-mediated data offers
    - Why sharing data through GBIF is valuable
    - How GBIF can support continuous improvement in data
- Key audiences
  - National governments
  - MEAs and policy-oriented NGOs
  - Researchers and users of scientific data
  - Holders of biodiversity data
  - Research infrastructures and informaticians

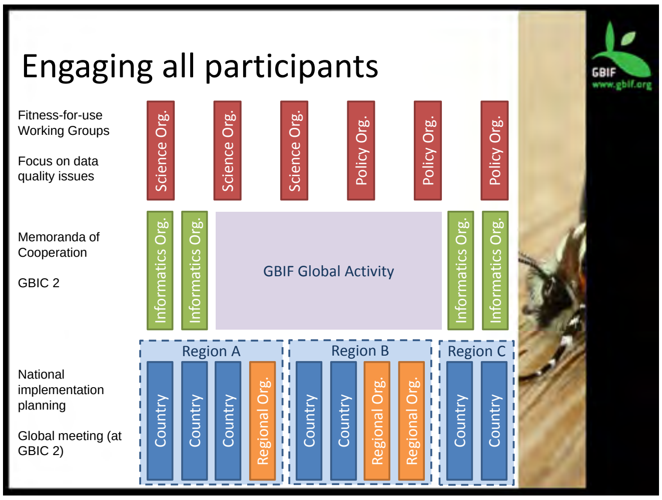
Information resources

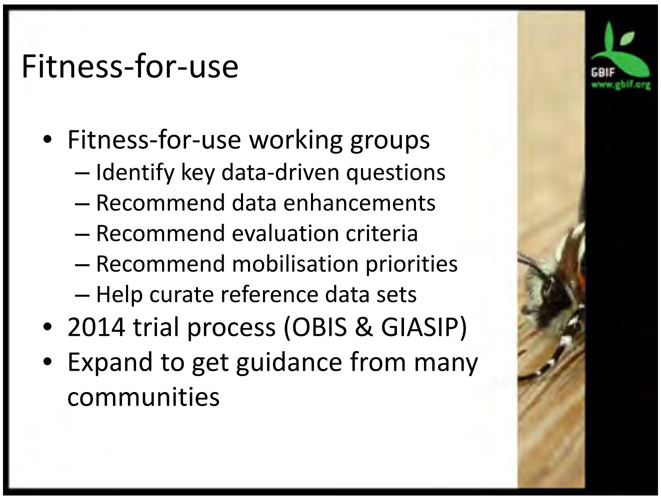
- Address key user groups, particularly:
  - Heads of Delegation
  - Node managers
  - Data publishers
  - Users
- Clarify roles/responsibilities/needs
- Deliver curriculum/reading list for each
  - Key documents
  - Getting started in role
  - Basic materials for GBIF training, etc.
- Formats and languages
  - Ensure relevance to whole community

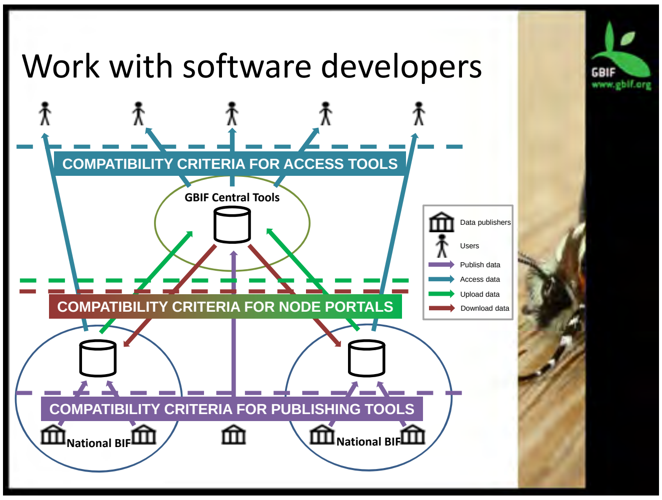
Strengthening national implementations

- Focus on key areas of relevance
  - Content mobilisation (2014)
  - Engagement with users (2014)
  - Support for national agencies
  - Other international frameworks
  - Research infrastructures
- Plans, support and reporting tools
  - To develop with Nodes
- Focus for training, mentoring, etc.
  - Curriculum components
  - Better foundations for new participants


Strengthening national implementations
Content
User
Agency
International
Infrastructure
Country X
Country Y
Possible example scale – content mobilisation:
1. Some data published
2. Survey of data holders complete
3. Basic metadata for all data holdings
4. Funded strategy to mobilise all data
5. Persistent data management for all data
GBIF
www.gbif.org


Engaging all participants
Fitness-for-use Working Groups
Focus on data quality issues
Memoranda of Cooperation
GBIC 2
National implementation planning
Global meeting (at GBIC 2)
Science Org.
Science Org.
Science Org.
Policy Org.
Policy Org.
Policy Org.
Informatics Org.
Informatics Org.
GBIF Global Activity
Informatics Org.
Informatics Org.
Region A
Region B
Region C
Country
Country
Country
Regional Org.
Country
Country
Regional Org.
Regional Org.
Country
Country
GBIF
www.gbif.org


Fitness-for-use
• Fitness-for-use working groups
– Identify key data-driven questions
– Recommend data enhancements
– Recommend evaluation criteria
– Recommend mobilisation priorities
– Help curate reference data sets
• 2014 trial process (OBIS & GIASIP)
• Expand to get guidance from many communities
GBIF
www.gbif.org


Work with software developers
COMPATIBILITY CRITERIA FOR ACCESS TOOLS
GBIF Central Tools
COMPATIBILITY CRITERIA FOR NODE PORTALS
COMPATIBILITY CRITERIA FOR PUBLISHING TOOLS
National BIF
National BIF
Data publishers
Users
Publish data
Access data
Upload data
Download data
GBIF
www.gbif.org


GBIF
www.gbif.org
Thank you
Donald Hobern, dhobern@gbif.org
GBIF Director
Global Biodiversity Information Facility (GBIF)
Berlin, 6-11 October 2013