Curso de Mobilização de Dados de Biodiversidade

Secretariado do GBIF

Versão 12, May 2021

Índice

Descrição do curso . . . 1
- Audiência . . . 1
- Pré-requisitos . . . 1
- Objetivos de aprendizagem . . . 2
- Certificação . . . 2

Ficheiros para descarregamento . . . 2
- Vídeos . . . 3
- Dados do exercício . . . 3
- Folhas de exercício . . . 3

Fundamentos . . . 3
- Terminologia . . . 3
  - Definições . . . 3
  - Software . . . 4
  - Estructuras . . . 4
- Qualidade dos dados . . . 4
- Documentação . . . 9
- Fluxos de Trabalho de Digitalização . . . 9
- Ferramentas de software . . . 11
- Instalar OpenRefine . . . 12
  - Requisitos de Instalação . . . 12
  - Instalação no MS Windows . . . 12
  - Instalação no Mac . . . 14
  - Instalação no Linux . . . 20
- Revisão dos fundamentos . . . 23

Use Case I - Espécime de Herbário . . . 27
- Cenário . . . 27
  - Descrição . . . 28
  - Recolha de dados . . . 29
  - Descrição dos dados . . . 29
- Exercícios . . . 29
  - Planificação . . . 30
  - Captura de dados . . . 30
  - Gestão de dados . . . 30
  - Publicação de dados . . . 30
- Folhas de exercício . . . 30

Planificação . . . 30
- Recursos . . . 30
- Organizador . . . 30

Exercício 1a-c . . . 31
Exercício 1a . . . 32
Exercício 1b . . . 32
Exercício 1c . . . 33
Revisão . . . 33
Captura de dados . . . 34
Padrões e Darwin Core . . . 35
Tipos e origens de dados . . . 35
Captura de dados, processamento e qualidade . . . 35
Exercício 2 . . . 35
Revisão . . . 36
Gestão de dados . . . 44
Princípios da gestão de dados . . . 44
Ferramentas de gestão de dados . . . 44
Exercício 3a-c . . . 44
Exercício 3a . . . 45
Exercício 3b . . . 45
Exercício 3c . . . 46
Dicas do exercício . . . 46
Verificações da validação . . . 46
Ferramentas úteis . . . 47
Revisão . . . 47
Publicação de dados . . . 48
Conceitos de publicação de dados . . . 49
Introdução geral ao IPT . . . 49
Instalações IPT para treino . . . 49
Demonstração do IPT . . . 49
Exercício 4 . . . 49
Revisão . . . 50
Avaliação e certificação . . . 51
Rubrica de planeamento . . . 52
Rubrica de captura de dados . . . 53
Rubrica de gestão de dados . . . 55
Rubrica de publicação de dados . . . 60
Use o caso II - Espécie invasora . . . 62
Cenário . . . 63
Descrição . . . 63
Recolha de dados . . . 64
Descrição dos dados digitais . . . 66
Folha de exercício de invasoras . . . 66
Exercício 1 . . . 66

Exercício 1a . . . 66
Exercício 1b . . . 67
Exercício 2 . . . 67
Exercício 3 . . . 67
Exercício 4 . . . 68
Caso de Uso II - avistamentos de Lepidoptera . . . 68
Cenário . . . 68
Descrição . . . 69
Recolha de dados . . . 70
Exemplo de captura de dados analógicos . . . 70
Descrição dos dados digitais . . . 71
Folha de exercício Lepidoptera . . . 72
Exercício 1 . . . 72
Exercício 1a . . . 73
Exercício 1b . . . 74
Exercício 2 . . . 74
Exercício 3 . . . 74
Exercício 4 . . . 75
Caso de Uso III - Aves a partir de literatura . . . 75
Cenário . . . 75
Descrição . . . 76
Colheita de dados originais . . . 77
Descrição dos dados analógicos . . . 77
Descrição de dados digitalizados e traduzidos . . . 78
Descrição dos dados digitais . . . 79
Aves da literatura da folha de exercícios . . . 79
Exercício 1 . . . 79
Exercício 2 . . . 80
Exercício 3 . . . 81
Exercício 4 . . . 81
Tarefas finais . . . 82
CASO DE USO II . . . 82
CASO DE USO III . . . 82
Submissão do trabalho . . . 82
Avaliação do curso . . . 83
Documentação chave . . . 83
Darwin Core . . . 83
Publicação de dados . . . 83
Publicação de dados: IPT . . . 84
Digitalização . . . 84
GBIF . . . 84

Georreferenciação . . . 84
Espécies Invasoras . . . 85
Living Atlases . . . 85
Diversos . . . 85
OpenRefine . . . 85
Planeamento/Colaboração . . . 85
Qualidade . . . 86
Espécie sensível . . . 86
Taxonomia . . . 86
Glossário . . . 86
Apêndice: Artigos de dados . . . 89
Apêndice: Soluções . . . 90
Soluções da revisão dos fundamentos . . . 90
Planificar as soluções de revisão . . . 93
Soluções de revisão dos dados de captura . . . 94
Soluções de revisão da gestão de dados . . . 95
Soluções de revisão de dados . . . 97
Sugestão de solução do Use Case I . . . 98
Agradecimentos . . . 98
Design do curso e instrução . . . 98
Tradutores . . . 99
Francês . . . 99
Português . . . 99
Espanhol . . . 99
Recursos . . . 99
Suporte a recursos . . . 100
Cólofon . . . 100
Citação sugerida . . . 100
Colaboradores . . . 100
Licença . . . 100
URL Persistente . . . 100
Controlo de Documentos . . . 100

Descrição do curso

Este curso permite aos participantes planear e implementar os esforços de mobilização de dados de biodiversidade efetivamente, utilizando os padrões aceites da comunidade. O objectivo é aumentar o volume, a riqueza e a qualidade dos dados publicados através da rede GBIF. Este curso foi desenvolvido pela primeira vez no âmbito do programa de informação sobre biodiversidade para o desenvolvimento (BID), financiado pela União Europeia.

Os tópicos incluem:

- Gestão de projeto
- Captura de dados
- Gestão de dados
- Publicação de dados

Este curso é composto por instruções em vídeo, com questionários e exercícios práticos. Quando oferecido como oficina virtual, o trabalho de grupo e a interação social são encorajados.

Audiência

Este curso foi desenvolvido para indivíduos que trabalham como investigadores ou técnicos em investigação sobre a biodiversidade ou instituições políticas. A instrução fornecida é particularmente útil para aqueles que têm necessidade ou desejo de planear projectos de mobilização e/ou de mobilização da informação sobre biodiversidade para as respectivas instituições.

Pré-requisitos

1. Introdução ao curso do GBIF
2. Adicionalmente, para melhor utilização das actividades em torno deste curso, os participantes devem possuir as seguintes habilidades e conhecimentos:
    - Competências básicas no domínio da utilização de computadores e da Internet e, em particular, na utilização de tabelas.
    - Conhecimento básico sobre geografia e informática para a biodiversidade: conceitos geográficos e de mapeamento, taxonomia básica e regras de nomenclatura.
    - Disposição para divulgar o conhecimento adquirido no workshop com parceiros e colaboradores em seu projecto, adaptando os materiais de treinamento de dados da biodiversidade a contextos e idiomas específicos, mantendo o valor educativo.
    - Um bom comando de Inglês. Apesar dos esforços realizados para prover materiais em outras línguas, instruções/vídeos serão em inglês.

Objetivos de aprendizagem

- Aprenda conceitos-chave (fundamentos) de informação sobre biodiversidade, especialmente para a gestão de dados digitais de biodiversidade.
- Receba uma introdução ao padrão Darwin Core e seus componentes.
- Aprenda a compreender as diferentes etapas para planear um projeto de mobilização e como adaptar etapas a um projecto específico.
- Avalie uma estratégia de mobilização de dados para identificar potenciais lacunas, ineficiências e dificuldades.
- Desenvolva uma estratégia de mobilização de dados personalizada para um determinado quadro institucional.
- Aprenda a identificar os tipos de dados e a melhor maneira de capturar informações relevantes usando as melhores práticas, softwares existentes, ferramentas e técnicas.
- Utiliza ferramentas de software desenhadas para facilitar a captura de dados de biodiversidade para produzir dados de biodiversidade digitais a partir de fontes analógicas.
- Aprenda conceitos de qualidade de dados e receba uma introdução às ferramentas utilizadas para padronizar dados, validar dados e limpar dados.
- Utilize ferramentas de software para avaliar a aptidão para a utilização de um conjunto de dados de biodiversidade
- Utilize ferramentas de software desenhadas para limpeza de dados (biodiversidade).
- Aprenda o processo de tornar os dados de biodiversidade disponíveis gratuitamente online, também conhecido como publicação de dados, utilizando o Conjunto de Ferramentas Integradas para Publicação do GBIF (IPT).
- Defina os tipos e subtipos de dados publicáveis (se houver) para um conjunto de dados de biodiversidade.
- Use o GBIF IPT para publicar conjuntos de dados de biodiversidade utilizando as extensões apropriadas.
- Capacite outros no planeamento, captura, gestão e publicação dos dados da biodiversidade.

Certificação

Após a conclusão bem sucedida do curso e avaliação bem sucedida das recomendações (por treinadores e mentores), os participantes têm a oportunidade de receber uma certificação oficial sob a forma de Open Badge. Veja Assessment and Certification para mais detalhes.

Ficheiros para descarregamento

Todos os ficheiros do curso podem ser descarregados desta página. Ou se preferir, todos os ficheiros são vinculados individualmente ao longo do curso, a medida que ocorrem no currículo. Os ficheiros de vídeo também são incorporados ao longo do curso e são reproduzidos pelo Vimeo. Se tiver dificuldade em aceder ao Vimeo. com, faça o descarregamento dos ficheiros mp4 para

reproduzi-los localmente no seu computador.

Vídeos

Os vídeos são narrados em inglês. Há dobragem em francês, espanhol e português.

Foundations1.zip (73.7 MB)

Foundations2.zip (90.2 MB)

Planning.zip (51.3 MB)

Capture.zip (63.1 MB)

Management.zip (30.2 MB)

Publishing.zip (77.9 MB)

Appendix-Data-Papers.zip (97.5 MB)

Dados do exercício

Este ficheiro comprimido (ZIP 37.7 MB) contém os dados de exercício para todos os casos de uso.

Folhas de exercício

Este arquivo compactado (ZIP 1.4 MB) contém as folhas de exercício para todos os casos de uso. As folhas de exercícios estão escritas em inglês e devem ser completadas em inglês.

Fundamentos

Este módulo inclui instruções para garantir que todos os participantes estejam no mesmo nível antes de mergulhar nos tópicos de mobilização de dados. Receberá uma introdução a linguagem, terminologia e definições para alguns dos conceitos básicos, funções e processos que irá colocar em uso durante o resto do curso. Receberão também uma introdução à qualidade dos dados e aprenderão a importância da documentação. Por último, será solicitado a instalar o software OpenRefine como parte deste módulo.

Terminologia

Definições

Neste vídeo (12:02), poderá rever a terminologia usada neste curso. Se não puder assistir o vídeo incorporado, utilize o download localmente. (MP4 - 38.5 MB)

https://www.youtube.com/watch?v=FZAF5Sy8Nsc *(YouTube video)*

Software

Neste vídeo (05:58), revisará exemplos dos diferentes tipos de aplicações e softwares disponíveis no mundo da mobilização da informática para a biodiversidade. Se não puder assistir ao vídeo, use o download localmente. (MP4 - 18.9 MB)

▶ https://www.youtube.com/watch?v=vYfDIgBBKXY *(YouTube video)*

Estructuras

Neste vídeo (13:10), irá rever os tipos de campos e de dados que possuem dados, as estruturas que ajudam a organizar e proteger esses dados e o que eles significam para a integridade e a segurança de seus dados. Se não conseguir assistir ao vídeo embutido, use o download localmente. (MP4 - 38.8 MB)

▶ https://www.youtube.com/watch?v=msnVbZvIy2E *(YouTube video)*

Qualidade dos dados

Neste vídeo (12:26), poderá rever a terminologia usada neste curso. Se não puder assistir o vídeo incorporado, utilize o download localmente. (MP4 - 44.5 MB)

▶ https://www.youtube.com/watch?v=5o7TcS2K7Cw *(YouTube video)*

Abaixo encontrará uma leitura selecionada do guia do Arthur Chapman: “Princípios da qualidade de dados”. Documento completo, referências e traduções podem ser encontradas em GBIF.org.

Antes que uma discussão detalhada sobre a qualidade dos dados e a sua aplicação a dados de ocorrência de espécies possa ser iniciada, há vários conceitos que precisam ser definidos e descritos. Estes, incluem o próprio termo "qualidade de dados", os termos "exatidão" e "precisão", que são comumente mal aplicados, e o que queremos dizer com dados de espécies primárias e dados de ocorrência de espécies.

Dados de ocorrência-espécie

Os dados de ocorrência de espécies são utilizados aqui para incluir dados de etiquetas de espécimes anexados a espécimes ou coleções alojados em museus e herbaria, dados de observações e dados de pesquisas ambientais. Em geral, os dados são o que chamamos de “point-based”, embora transecto (dados de transectos de pesquisas ambientais, colheitas ao longo de um rio), polígono (observações de uma área definida como um parque nacional) e dados de grid (observações ou registos de pesquisa de uma grid regular) também são incluídos. Em geral, estamos a falar de dados georeferenciados – i.e. registros com referências geográficas que os amarram a um local em particular no espaço - seja com uma coordenada georeferenciada (e.g. latitude e longitude, UTM) ou não (descrição textual de uma localidade, altitude, profundidade) – e hora (data, hora do dia).

Em geral, os dados também são vinculados a um nome taxonômico, mas as coleções não identificadas também podem ser incluídas. O termo tem sido ocasionalmente utilizado indistintamente com o termo "dados primários de espécies".

Dados primários de espécies

"Dados de espécies primárias" é usado para descrever dados e dados da coleção bruta sem quaisquer atributos espaciais. Inclui dados taxonómicos e nomenclaturais sem atributos espaciais, como nomes, taxa e conceitos taxonómicos sem referências geográficas associadas.

Exatidão e precisão

A exatidão e a precisão são regularmente confundidas e as diferenças não são geralmente compreendidas.

A exatidão refere-se à proximidade de valores medidos, observações ou

estimativas para o valor real ou verdadeiro (ou para um valor que é aceito como sendo verdadeiro – por exemplo, as coordenadas de um ponto de controle da pesquisa).

A precisão (ou Resolução) pode ser dividida em dois tipos principais. A precisão estatística é a proximidade com que repetidas observações se relacionam a si mesmas. Eles não têm nada a ver com a sua relação com o verdadeiro valor e podem ter uma grande precisão, mas uma baixa exatidão. Precisão numérica é o número de algarismos significativos em que uma observação é registada e tornou-se muito mais evidente com o advento dos computadores. Por exemplo, uma base de dados pode produzir um registro decimal de latitude/longitude em 10 casas decimais – i.e. ca. 1 mm quando na realidade, o registo tem uma resolução não superior a 10-100 m (3-4 casas decimais). Isto conduz frequentemente a uma falsa impressão tanto da precisão como da exatidão.

Estes termos – exatidão e precisão – também podem ser aplicados aos dados não espaciais, bem como aos dados espaciais. Por exemplo, uma coleção pode ter uma identificação para o nível de subespécie (i.e. tem alta precisão), mas seja o táxon errado (ou seja, tenha baixa exatidão), ou ser identificado apenas para o nível de Família (alta exatidão, mas baixa precisão).

Qualidade dos dados

A qualidade dos dados é multidimensional, e envolve gestão de dados, modelação e análise, controlo de qualidade e garantia, armazenamento e apresentação. Como dito de forma independente por Chrisman (1991) e Strong et al. (1997), a qualidade dos dados está relacionada à utilização e não pode ser avaliada independentemente do utilizador. Numa base de dados, os dados não têm real qualidade ou valor (Dalcin 2004); só têm valor potencial que é reconhecido apenas quando alguém usa os dados para fazer algo útil. A qualidade da informação está relacionada com a sua capacidade de satisfazer os seus clientes e de atender as necessidades dos clientes (English 1999).

Redman (2001) sugeriu que para que os dados estejam em condições de serem utilizados, devem ser acessíveis, exactos, oportunos, completos, consistente com outras fontes, relevantes, abrangentes,

fornecerem um adequado nível de detalhe, serem fáceis de ler e fáceis de interpretar.

Uma questão que um detentor de dados pode precisar considerar é o que pode precisar ser feito com a base de dados, para aumentar a sua utilização para um público mais amplo (i.e. aumentar seu potencial uso ou relevância) e torná-la assim adequada para um leque mais amplo de propósitos. Haverá uma troca entre o aumento da usabilidade e o volume de esforço necessário para adicionar funcionalidade e usabilidade extra. Isso pode exigir coisas como atomizar campos de dados, adicionar informações de georeferenciamento, etc.

Garantia de Qualidade/ Controle de Qualidade

A diferença entre controlo de qualidade e garantia de qualidade nem sempre é clara. Taulbee (1996) faz a distinção entre controlo de Qualidade e Garantia de Qualidade e sublinha que uma não pode existir sem a outra, se quisermos que os objectivos de qualidade sejam cumpridos. Ela define o controlo de qualidade como um juízo de qualidade com base em normas internas, processos e procedimentos estabelecidos para controlar e monitorizar a qualidade; e a garantia da qualidade como uma avaliação da qualidade com base em normas externas ao processo e é a revisão das actividades e dos processos de controlo de qualidade para garantir que os produtos finais cumprem padrões de qualidade pré-determinados.

Numa abordagem mais orientada para as empresas, Redman (2001) define a Garantia de Qualidade como “as atividades que são concebidas para produção de produtos de informação isentos de defeito para atender às necessidades mais importantes dos clientes, pelo menor custo possível”.

Como estes termos devem ser aplicados na prática não é claro, e na maioria dos casos, os termos parecem ser em grande medida utilizados como sinónimo para descrever a prática global da gestão da qualidade dos dados.

Incerteza

A incerteza pode ser considerada como uma "medida da ineficácia do conhecimento ou informação sobre uma quantidade desconhecida cujo

verdadeiro valor poderia ser estabelecido se um dispositivo de medição perfeito estivesse disponível" (Cullen e Frey 1999). A incerteza é uma propriedade da compreensão dos dados por parte do observador e tem mais a ver com o observador do que com os dados em si. Há sempre incerteza nos dados; a dificuldade é registar, compreender e visualizar essa incerteza para que os outros também a possam entender. A incerteza é um termo fundamental na compreensão dos riscos e na avaliação de riscos.

Erro

Erro engloba tanto a imprecisão dos dados quanto sua inexatidão. Há muitos fatores que contribuem para o erro. O erro é geralmente visto como aleatório ou sistemático. Um erro aleatório tende a se referir ao desvio do verdadeiro estado de forma aleatória. O erro sistemático ou enviezamento advém de uma mudança uniforme nos valores e, por vezes, é descrito como tendo uma "exatidão relativa" no mundo cartográfico (Chrisman 1991). Ao determinar "aptidão para o uso", o erro sistemático pode ser aceitável para algumas aplicações, e inaptas para outras.

Um exemplo pode ser a utilização de um datum1 geodésico diferente – onde, se utilizado ao longo da análise, não pode causar quaisquer maiores problemas. Os problemas surgirão no entanto onde uma análise utiliza dados de diferentes fontes e com diferentes tendências – por exemplo, fontes de dados que usam diferentes datum geodésico, ou onde as identificações podem ter sido realizadas usando uma versão anterior de um código nomenclatural.

"Porque o erro é inevitável, ele deve ser reconhecido como uma dimensão fundamental dos dados" (Chrisman 1991). Somente quando um erro é incluído em uma representação dos dados é possível responder a perguntas sobre limitações dos dados, e até mesmo limitações do conhecimento atual. Os erros conhecidos nas três dimensões do espaço, atributo e tempo têm de ser medidos, calculados, registados e documentados.

Validação e Limpeza

Validação é um processo usado para determinar se os dados são incompletos, não razoáveis ou não acurados. O processo pode incluir

verificações de formato, verificações de integridade, verificações de razoabilidade, verificações de limite, revisão dos dados para identificar "outliers" (geográficos, estatísticos, temporais ou ambientais) ou outros erros e avaliação de dados por especialistas em áreas temáticas (e.g. especialistas taxonómicos). Estes processos resultam geralmente em sinalizadores, documentação e subsequente verificação de registos suspeitos. As validações podem também envolver o controlo da conformidade com padrões, regras e convenções aplicáveis. Uma etapa chave na validação e limpeza de dados é identificar as causas dos erros detectados e concentrar-se em impedir que esses erros ocorram novamente (Redman 2001).

Limpeza de dados refere-se ao processo de “reparação” de erros nos dados que foram identificados durante o processo de validação. O termo é sinónimo de "limpeza de dados", embora alguns usem a limpeza de dados para englobar a validação de dados e a limpeza de dados. No processo de limpeza de dados, é importante que os dados não se percam inadvertidamente e que as alterações à informação existente sejam efectuadas com muito cuidado. Muitas vezes é melhor manter ambos os antigos (dados originais) e os novos (dados corrigidos) lado a lado na base de dados, para que, caso sejam cometidos erros no processo de limpeza, as informações originais possam ser recuperadas.

Documentação

Neste vídeo (09:47), daremos uma visão geral da importância da documentação no que se refere à gestão e publicação de dados. Você aprenderá sobre mapeamento de dados, relações e metadados. Se não puder assistir o vídeo embutido, poderá utilizar o download localmente. (MP4 - 29.2 MB)

https://www.youtube.com/watch?v=Z5-SYImGRGc *(YouTube video)*

Fluxos de Trabalho de Digitalização

Este vídeo (07:20) sobre Fluxo de Trabalho de Digitalização identifica cinco clusters (ou etapas) no processo de digitalização de objectos de colecções de história natural usando imagens digitais, e estas fases podem ser facilmente adaptadas a outras fontes de dados de biodiversidade. Se não conseguir assistir ao vídeo embutido do Vimeo, poderá utilizar o download localmente. (MP4 - 26,8 MB)

▶ https://vimeo.com/120369455 *(Vimeo video)*

> Como salienta o vídeo, os protocolos de digitalização variam de instituição para instituição, mas é essencial que o protocolo escolhido seja acordado, documentado e respeitado.

Não ensinamos a digitalização por si só durante o workshop, uma vez que pode ser um curso de uma semana. Em vez disso, centramo-nos na introdução básica à captura de dados sobre biodiversidade. No entanto, queremos fornecer-lhe recursos em matéria de digitalização, como sabemos que há interessados nisso.

Há muitas maneiras de organizar os esforços de digitalização e por isso, a digitalização pode parecer assustadora no início. É importante lembrar que, na maioria dos casos, outra pessoa já tentou digitalizar os mesmos tipos de espécimes e objectos que você planeia. Neste exercício nós apresentamos alguns recursos práticos de fluxo de trabalho de digitalização para ajudá-lo a começar. Estas constituirão também a base de trabalho que iremos desenvolver no workshop sobre a selecção, modificação e avaliação dos fluxos de trabalho.

Algumas etapas do processo podem incluir:

- **Curadoria de pré-digitalização e "staging"**: Isso inclui a preparação da fonte de dados para o processo de digitalização, incluindo a atribuição de identificadores únicos que ajudarão a referir à fonte sem erro e a manter todas as informações derivadas juntas.
- **Captura de imagem**: Isso inclui um planeamento razoável, não apenas na própria imagem capturada (e.g. definição da sequência de trabalho, seleção de hardware adequado), mas também sobre como e onde as imagens serão armazenadas e tratadas.
- **Processamento de imagens**: Isso inclui controle de qualidade, conversão de ficheiro, etc.
- **Captura de dados eletrônicos**: O núcleo do processo de digitalização inclui a captura de informações-chave em um banco de dados. O vídeo destaca que o método mais comum para inserir a informação é através de um teclado, mas cada vez mais as instituições estão a voltar-se para tecnologias avançadas de entrada de dados.
- **Georeferenciação**: informações geográficas são muito importantes para a análise da biodiversidade, então os projectos de digitalização devem procurar extrair as informações geográficas mais acuradas possíveis.

Biocoleções Digitalizadas Integradas (iDigBio) é o centro de coordenação do Recurso Nacional dos Estados Unidos para Avanço da Digitalização da Coleção de Biodiversidade (ADBC). Estes, conduzem um esforço a nível nacional para tornar disponíveis para a comunidade científica dados e imagens de milhões de espécimes biológicos num formato electrónico normalizado, para agências governamentais, estudantes, educadores e o público em geral. Eles produziram vários vídeos que discutem o processo de digitalização.

Há outros vídeos na série iDigBio em que você pode estar interessado, se quiser aprender mais sobre fluxos de trabalho específicos para diferentes tipos de espécime:

- "Digitalização de coleções não-secas" (4:34 minutos) https://vimeo.com/120369690
- “Fluxos de Trabalho de Imagens para a Digitalização de Espécimes de Vertebrados (Dry-

preserve)" (7:25 mins) https://vimeo.com/160615629

- "Digitalização de espécimes de Herbarium" (7:34 mins) https://vimeo.com/120369768

Ferramentas de software

Revisar ferramentas de software utilizadas em informática para a biodiversidade

Durante as actividades do curso, demonstraremos e trabalharemos com diversas ferramentas de software relacionadas à digitalização de dados, qualidade de dados e transformação. Você provavelmente já usa vários deles em seu trabalho diário.

Os formadores, mentores e os participantes de cursos anteriores compilaram uma lista com informações sobre ferramentas de software sobre informática para a biodiversidade. A lista fornece ligações para os seus principais portais, fatos-chave e uma síntese de pontos fortes e fracos.

Descarregue o Link:../course-docs/Software-database-EN.xlsx[Software-database-EN.xlsx,opts=download]. (23 KB)

Ao analisar um software de biodiversidade que você não utilizou antes, é necessário considerar como o adaptaria para os seus fins. Você encontrará abaixo uma lista com a qual poderá começar a sua avaliação. É inspirada nas "características de uma boa solução de banco de dados" do manual GBIF "Iniciando um Projecto de Digitalização":

- **Preço**: Um dos fatores mais determinantes. Esteja ciente de outros custos além do preço da licença do software, tal como o hardware necessário para executá-lo, manutenção, actualização e o conhecimento para executá-lo.
- **Funcionalidade**: Você precisa ter clareza sobre o que espera que o software alcance, e certifique-se de que ele o faça eficientemente. Não se distraia com funcionalidades adicionais que podem tornar o software mais complexo desnecessariamente.
- **Estabilidade**: Algumas soluções estão há muito tempo no mercado e são apoiadas por instituições ou empresas sólidas e têm maior probabilidade de não apresentarem "bugs" e de terem sistemas bons para resolver quaisquer problemas que surjam. Isto também tornará mais provável que seja actualizado e portabilizado para sistemas operacionais mais modernos.
- **Escalabilidade**: Alguns softwares tem boa performance quando trabalham fora da caixa, mas o desempenho diminui após algum tempo ou ao utilizá-los com maiores quantidades de dados ou quando vários utilizadores acedem-nos, simultaneamente. Verifique as opiniões de outros utilizadores on-line.
- **Integração**: Certifique-se de que o software aceita e produz os formatos de dados que você utiliza e precisa. Transformação de dados é uma tarefa demorada.
- **Suporte a idioma**: É essencial que todos que usam o software possam compreender a sua interface, e a documentação que tornará possível o seu uso.
- **Documentação e suporte técnico**: Certifique-se de explorar a documentação existente e os mecanismos de suporte. Você pode ter certeza que em algum momento vai precisar.
- **Curva de aprendizagem**: Alguns softwares podem necessitar de treinamento específico para aprender a usá-los, enquanto outros são mais intuitivos e podem ser aprendidos enquanto os

usam, apoiados por sistemas de ajuda in-line.

Instalar OpenRefine

Instalar software necessário para actividades mais tarde no curso

OpenRefine é uma ferramenta com um conjunto de recursos para trabalhar com dados tabulares que melhora a qualidade geral de um conjunto de dados. É uma aplicação que é executada no seu próprio computador, como um pequeno servidor web, e para usá-lo seu navegador web deve apontar para esse servidor web. Portanto, pense no OpenRefine como uma aplicação web pessoal e privada.

Utilizaremos o OpenRefine durante a parte de mobilização de dados do curso, especialmente durante os exercícios práticos. Será necessário instalar o OpenRefine no seu portátil. Se é um utilizador qualificado de computador, pode seguir estas etapas para instalar o software no seu computador. Se não é confiante, por favor peça ajuda. Consulte OpenRefine download page para obter mais detalhes.

Palavras-passe administrativas podem ser necessárias para instalar o software.

Requisitos de Instalação

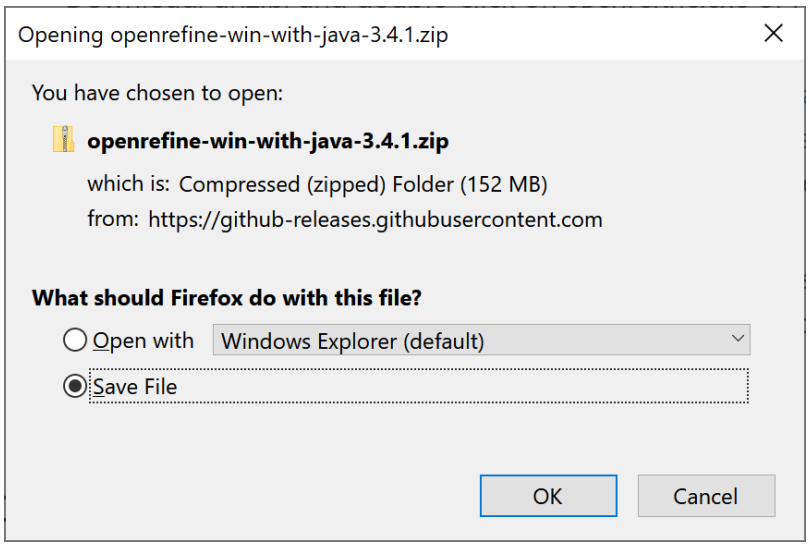
1. Apenas para utilizadores do Linux: Java JRE instalado.
2. Google Chrome, Microsoft Edge ou Mozilla Firefox instalado. Internet Explorer não é suportado.

A última versão estável é o OpenRefine 3.4.1, lançado em 24 de setembro de 2020. Instruções detalhadas de instalação estão disponíveis em https://docs.openrefine.org/manual/installing/.

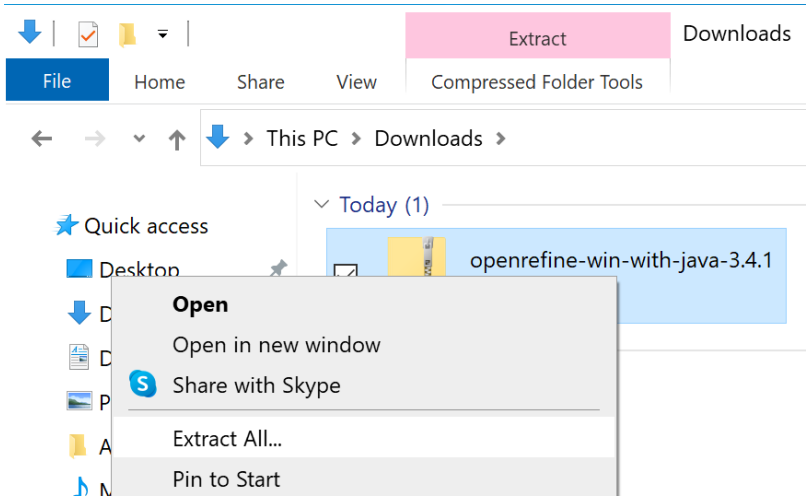
Instalação no MS Windows

1. Transfira o kit Windows **com Java embutido**. Escolha gravar o ficheiro em vez de abri-lo.
2. Encontre o ficheiro transferido. Clique com o botão direito e escolha "Extrair todos...". Descompactação e um duplo clique no openrefine.exe ou refine.bat, se o primeiro não funcionar.
3. Uma janela de comando aparecerá (não a feche) e de seguida uma nova janela do navegador web irá mostrar a aplicação.

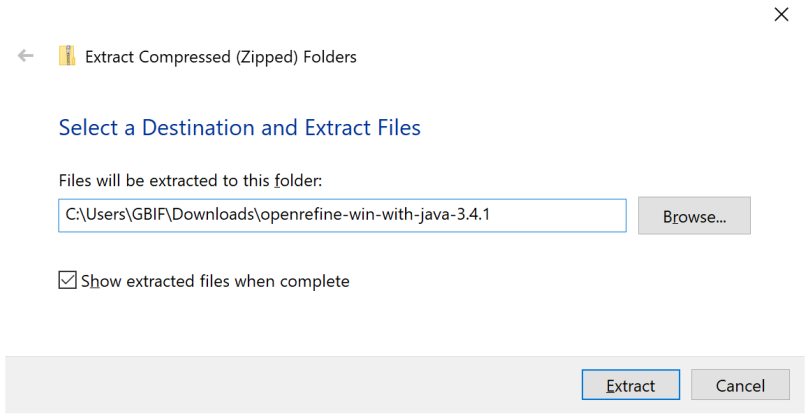
Instruções detalhadas para o MS Windows (clique para expandir)

1. 

Transfira o kit Windows com Java embutido. Escolha gravar o ficheiro em vez de abri-lo.

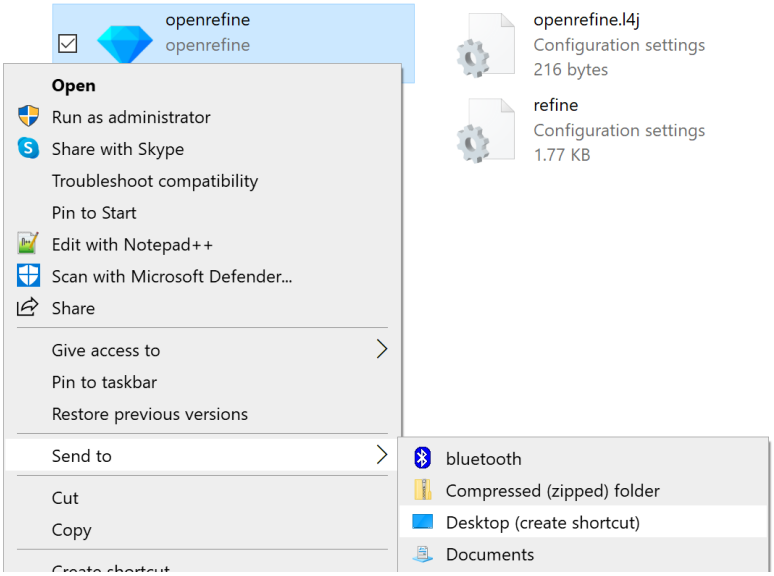
2. 

Encontre o ficheiro que transferiu. Clique com o botão direito do rato e escolha "Extrair Todos..."

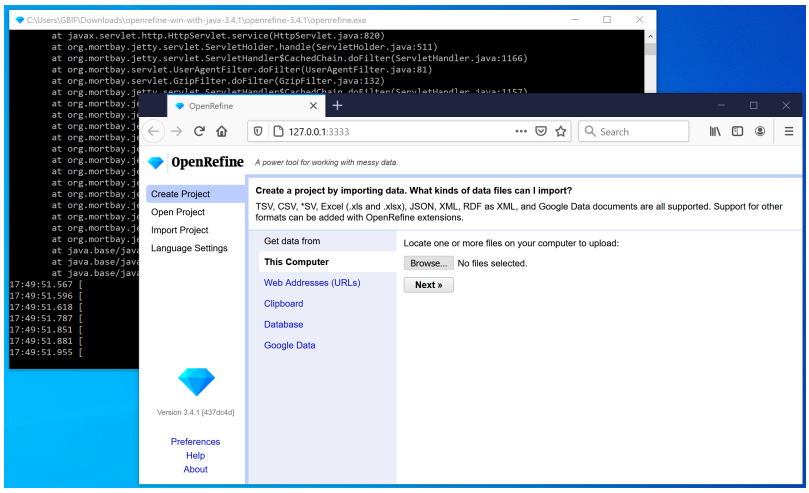
3. 

Clique "Extrair"

4.



Encontre os ficheiros extraídos. Opcionalmente, clique com o botão direito em "openrefine" e escolha "Enviar para → Desktop (criar atalho)" para criar um atalho na sua área de trabalho. Em seguida, clique duas vezes em "openrefine"

5. 

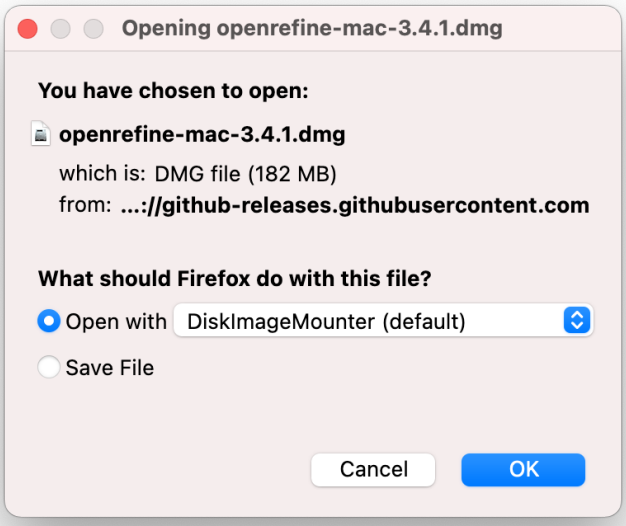
Uma janela preta da consola irá abrir, e de seguida também o navegador. O OpenRefine está pronto para ser usado.

Instalação no Mac

1. Transfira o Mac kit.
2. Descarregue, abra, arraste o ícone para a pasta Aplicações. Não precisa instalar o Java separadamente.
3. Clique duas vezes nele e uma nova janela do navegador da Web exibirá a aplicação.

Instruções detalhadas para Mac (clique para expandir)

1.



Transfira o Mac kit, e escolha abrir.

2.

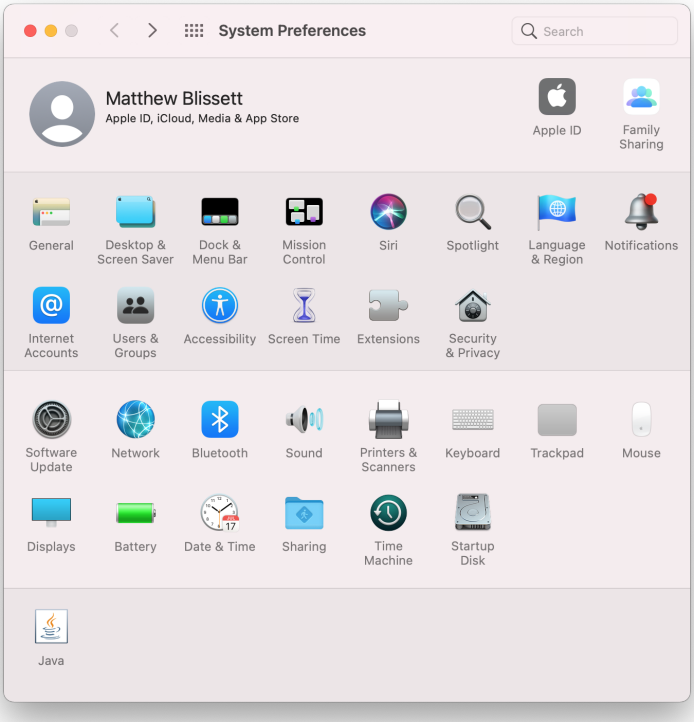
Um aviso é mostrado. Clique em "OK".

3.

System Preferences

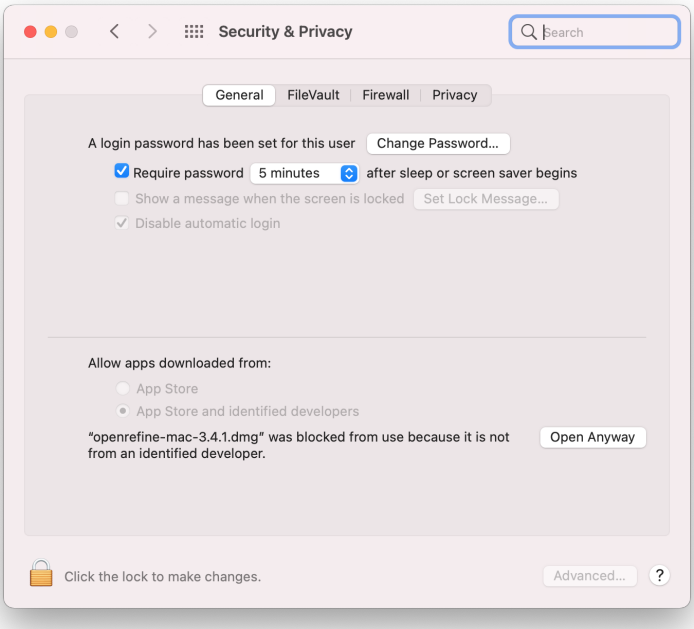
Abrir Preferências do Sistema.

4.



Abrir Segurança & Privacidade

5.

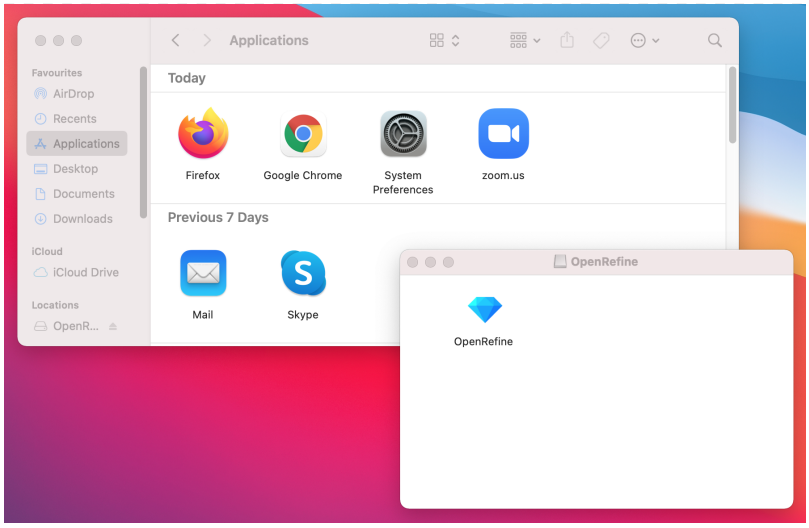


Escolha "Abrir de qualquer forma" em baixo.

6.

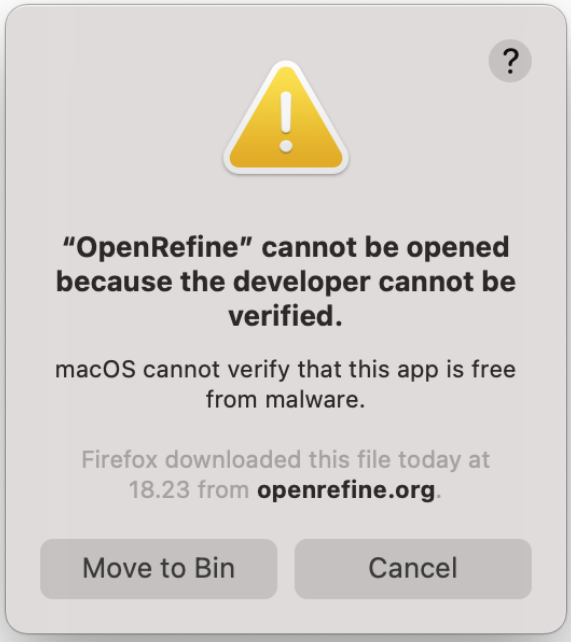
Escolha "Abrir"

7.



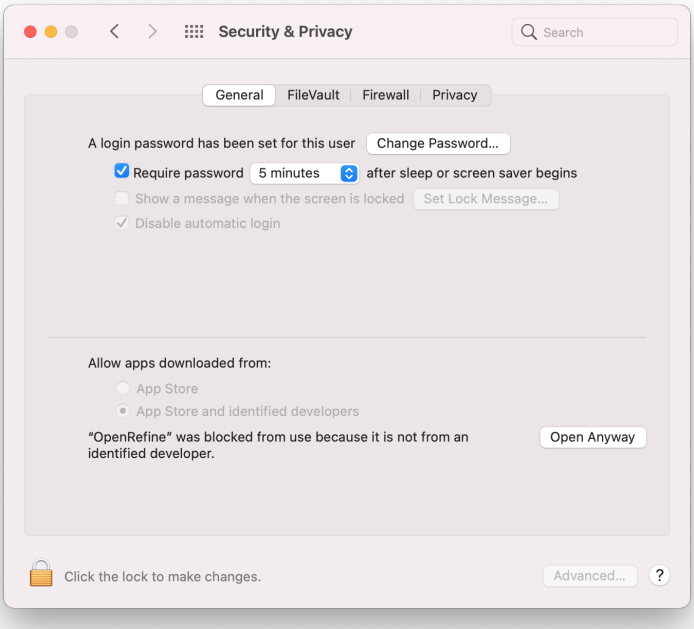
Finalmente, o arquivo da aplicação está aberto! Arraste-o para a pasta Aplicativos.

8.



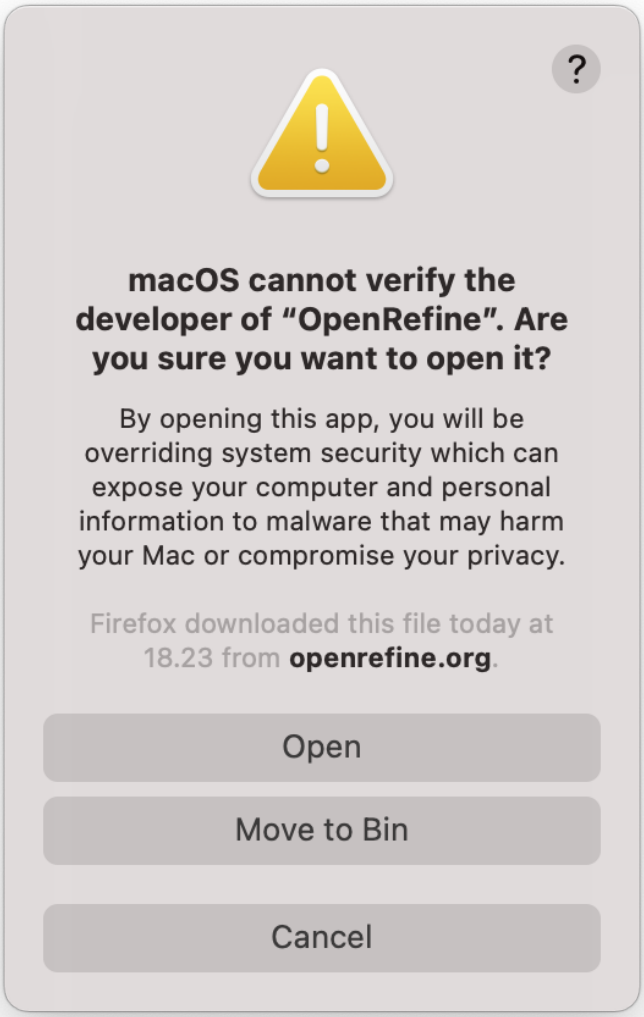
Faça duplo click no ícone OpenRefine. Aparece outro aviso de segurança!

9.



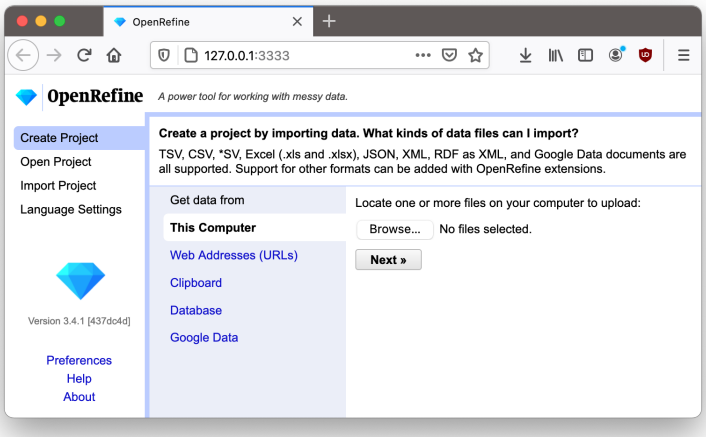
Volte para "Segurança e Privacidade" e clique em "Abrir de qualquer forma" — novamente.

10.



(Para evitar esses avisos, os programadores OpenRefine teriam de pagar à Apple.)

Clique em "Abrir".

11. 

Finalmente! A aplicaação está em execução.

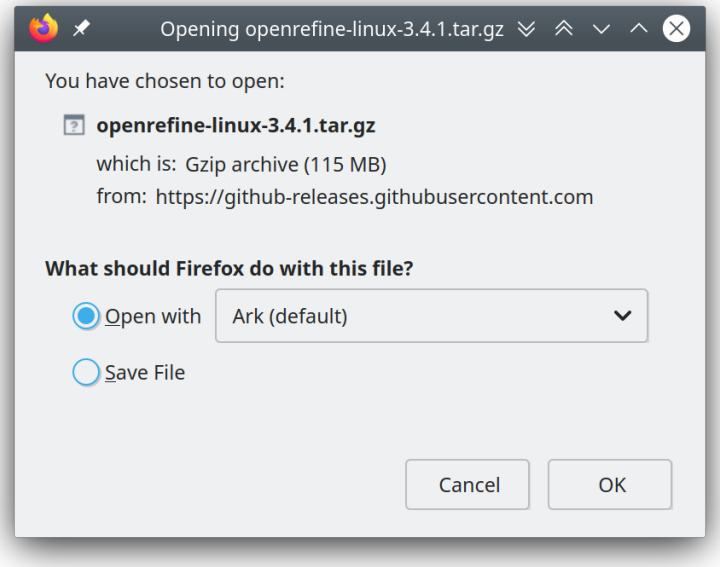
Instalação no Linux

1. Transfira o Linux kit.
2. Descarregue, extraia, em seguida digite ./refine para iniciar. Este requer que o Java esteja instalado no seu computador.

Instruções detalhadas para Linux (clique para expandir)

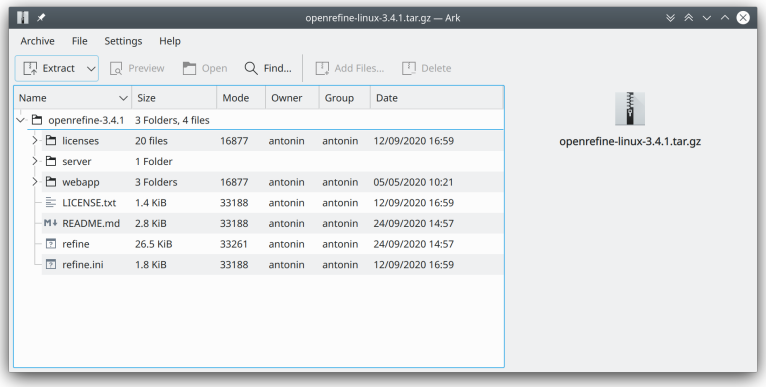
Estas instruções são para o KDE (por exemplo, Kubuntu, SuSE), mas o processo é similar para Gnome (por exemplo, Ubuntu, Red Hat, CentOS).

1.



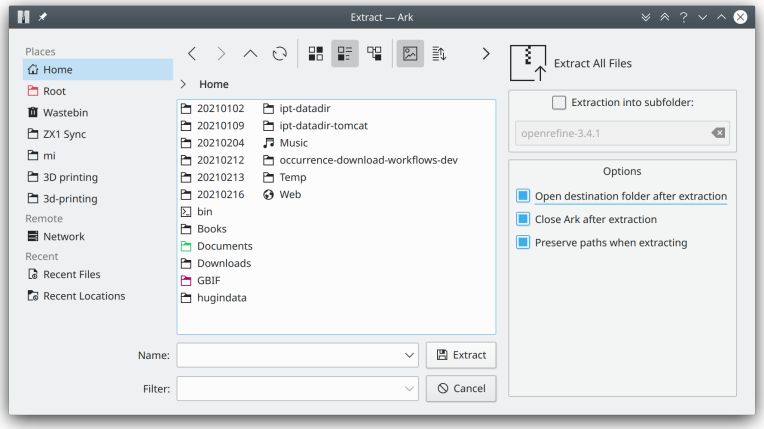
Transfira o Linux kit. Abra-o.

2.



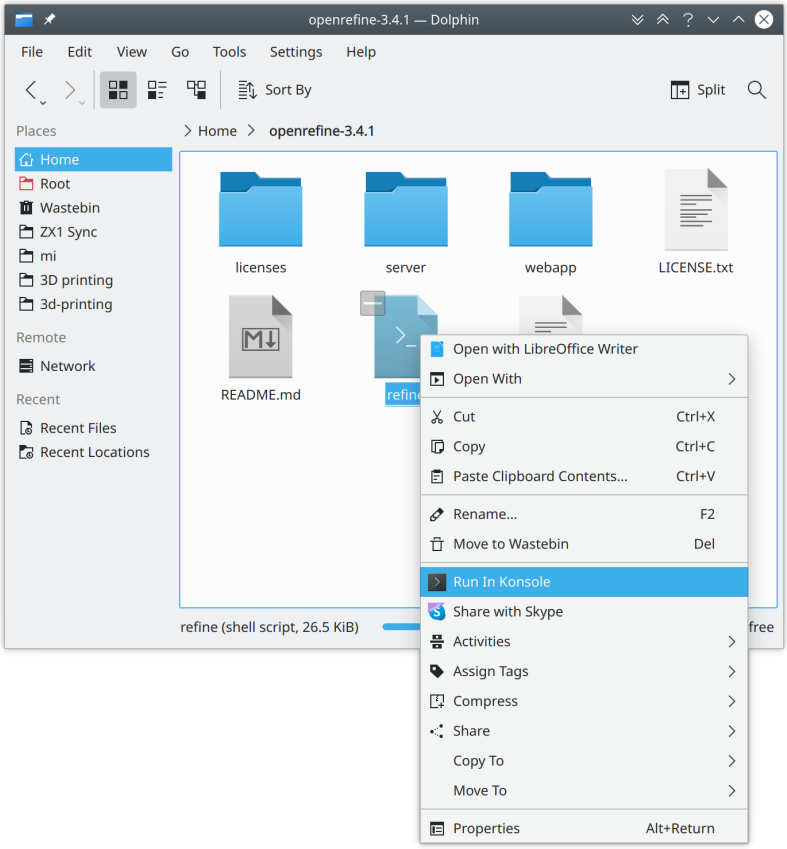
Clique em "Extrair" para descompactar o aplicativo transferido.

3.

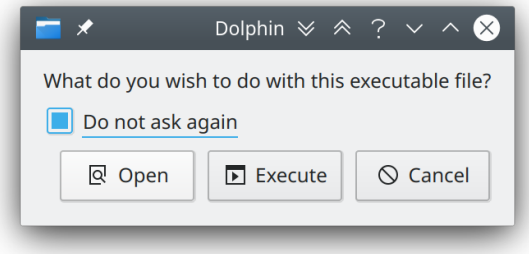


Escolha o local adequado. Eu também selecionei "Abrir pasta de destino após a extração" e

"Fechar Ark após a extração"

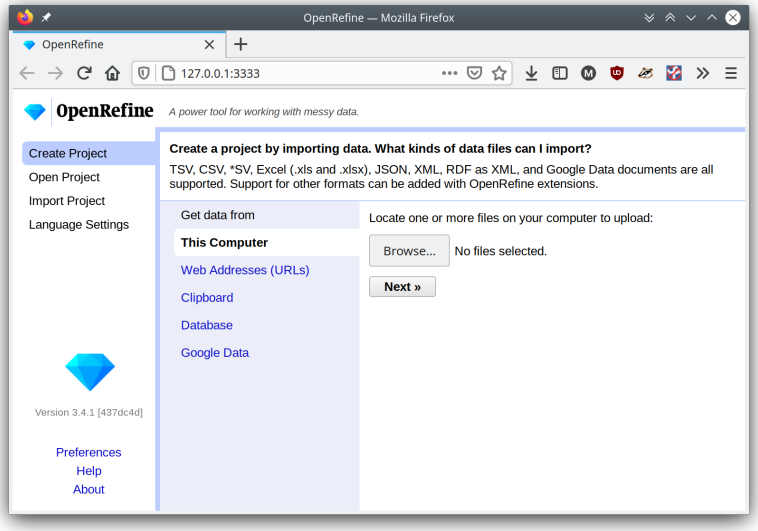
4. 

Clique com o botão direito em "refine" e escolha "Executar em Konsole". Isso é necessário para possa sair do OpenRefine mais tarde, fechando a janela do Konsole.

5. 

Confirme que deseja executar a aplicação transferida.

6.



O OpenRefine está a correr agora.

Revisão dos fundamentos

Faça um "quiz" sobre os conceitos aprendidos nesta secção.

1. Para a afirmação fornecida, insira o termo correto (banco de dados, língua da base de dados, programa da base de dados)
    - Combina e apresenta funções e recursos para manipular dados, em conjunto em uma interface unificada
      _____
    - coleção estruturada e organizada de dados e/ou informações seguradas em um computador
      _____
    - a forma pela qual um humano se comunica com um computador
      _____
2. Se você abrir um ficheiro de dados e ver o seguinte, de qual problema suspeitaria?

   `□tre, ou ne pas □tre, c□est l□ la question.`

   - ☐ Nenhuma das opções
   - ☐ Está corrompido
   - ☐ A codificação errada foi utilizada para abrir o ficheiro
   - ☐ O remetente usou uma fonte estranha
3. Para o software fornecido, insira o tipo de software (captura de dados, gestão de dados, limpeza de dados, publicação de dados).
    - Integrado Publishing Toolkit (IPT)
      _____
    - Especificar
      _____ E _____
    - iNaturalist
      _____
    - OpenRefine
      _____
4. Para o exemplo dado, insira o tipo de dados correcto (binário, booleano, float, inteiro, inteiro longo, texto, texto não estruturado)
    - 1236975
      _____
    - 01101111
      _____
    - Caminhamos 5 milhas pela estrada a oeste do correio, no centro da cidade. Depois seguimos 2 milhas a norte em um caminho de terra até o rio. Então continuámos a oeste ao longo do rio por mais 5 milhas.
      _____
    - 1024

_____

- 29.0

_____

- Sim/Não

_____

- 6 coelhos foram observados

_____

5. Qual destes termos descreve um nome de "campo/coluna"?

   - ☐ Atribuído
   - ☐ Descritivo
   - ☐ Identificador
   - ☐ Legível
   - ☐ Único
   - ☐ Interface de utilizador

6. Quais destes termos descreve uma "etiqueta de campo"?

   - ☐ Atribuído
   - ☐ Descritivo
   - ☐ Identificador
   - ☐ Legível
   - ☐ Único
   - ☐ Interface de utilizador

7. Para cada afirmação, insira a estrutura correta (linha, coluna, tabela)

   - Todos os dados se referem a um conceito SINGLE.

     _____

   - Um atributo tem o tipo de campo/dado de cada registo.

     _____

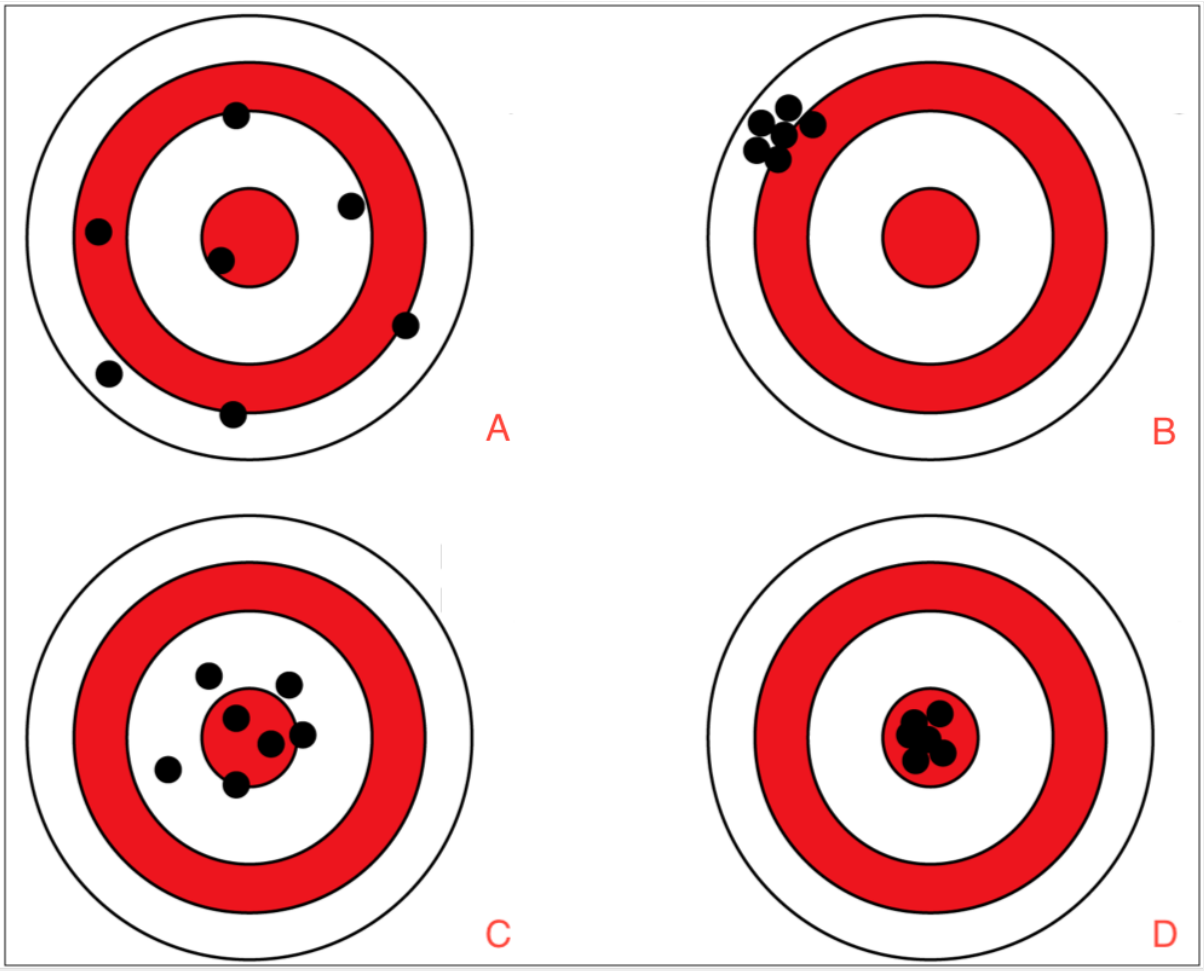
   - Atributos de um registo permanentemente.

     _____

8. Quem determina a aptidão para o uso de seus dados?

   - ☐ O director do museu ou do departamento
   - ☐ Os utilizadores dos dados para investigação ou educação
   - ☐ O coletor de dados no campo
   - ☐ A pessoa inserindo os dados na base de dados

9. Para as afirmações fornecidas, insira a imagem correspondente. (A, B, C, D)



- Alta exatidão, baixa precisão

  _____

- Baixa exatidão, alta precisão

  _____

- Alta exatidão, alta precisão

  _____

- Baixa exatidão, baixa precisão

  _____

10. Identifique as relações de dados em que o Dataset B precisa ser mesclado ao conjunto de dados A (0:1, 1:0, 1:1, 1:∞, ∞:1, ∞:∞). Nem todas as relações são utilizadas.

    - O campo de coletor de dados existe em ambos os conjuntos de dados A e B

      _____

    - O campo País só existe no conjunto de dados B

      _____

    - O campo Nome existe no conjunto de dados A, mas o dataset B contém campos de Primeiro Nome e Sobrenome

      _____

    - campo de ID existe nos conjuntos de dados A e B

      _____

    - Elevação existe no conjunto de dados A, mas não no conjunto de dados B

_____

- Data existe no conjunto de dados A, mas Dia, Mês e Ano são campos separados no conjunto de dados B

_____

11. Metadados são importantes porque (selecione as declarações VERDADEIRAS):

    - ☐ permite que os utilizadores determinem se um conjunto de dados é adequado para o seu uso.
    - ☐ permite que você compartilhe coordenadas exatas para cada ocorrência.
    - ☐ permite que você saiba sob quais termos legais a reutilização de dados é permitida.
    - ☐ isto também se aplica a todos os materiais suplementares e associados, incluindo imagens, vídeos e outras mídias.
    - ☐ permite que você saiba sobre a próxima exposição/horário de funcionamento da instituição.

Use Case I - Espécime de Herbário

Familiarize-se com o cenário do caso.

O caso I, é um caso prático para o planeamento, captura de dados, gestão de dados e módulos de publicação de dados. É recomendável que você desacarregue o folha de exercícios (MS Word 345 KB) para que possa fazer anotações conforme trabalha nos exercícios. Uma solução sugerida será fornecida no apêndice da solução. O caso I não é avaliado.

Cenário

Um Projecto de Mobilização de Dados em um Herbário Regional

Eriocaulon bilobatum Morong *coletado na Guatemala por Coleção de Referência Rápida (RRC) Express Museum of Natural History - Keller Action Science Center (licenciado sob* CC-BY-NC 4.0*)*

Esta narrativa foi desenvolvida como base para exercícios práticos no curso de mobilização de dados para a biodiversidade e o conceito e conteúdo do exercício foi desenvolvido por Alberto González-Talaván, Néstor Beltrán, Nicolas Noé, Sharon Grant. Os dados são de um conjunto de dados real, mas foram modificados para os fins desses exercícios. É um cenário fictício e destina-se apenas a fins instrutivos.

Descrição

A Universidade de Planícies Brancas é uma instituição de ensino superior bem reconhecida na Guatemala e uma referência nacional à investigação sobre biodiversidade. O Departamento de Biologia de Plantas mantém herbário de médio tamanho contendo aproximadamente 80,000 espécimes colhidos dentro e em torno da região local e datam de meados do século Xx até o presente. A coleção inclui espécimes importantes como tipos e endémicos.

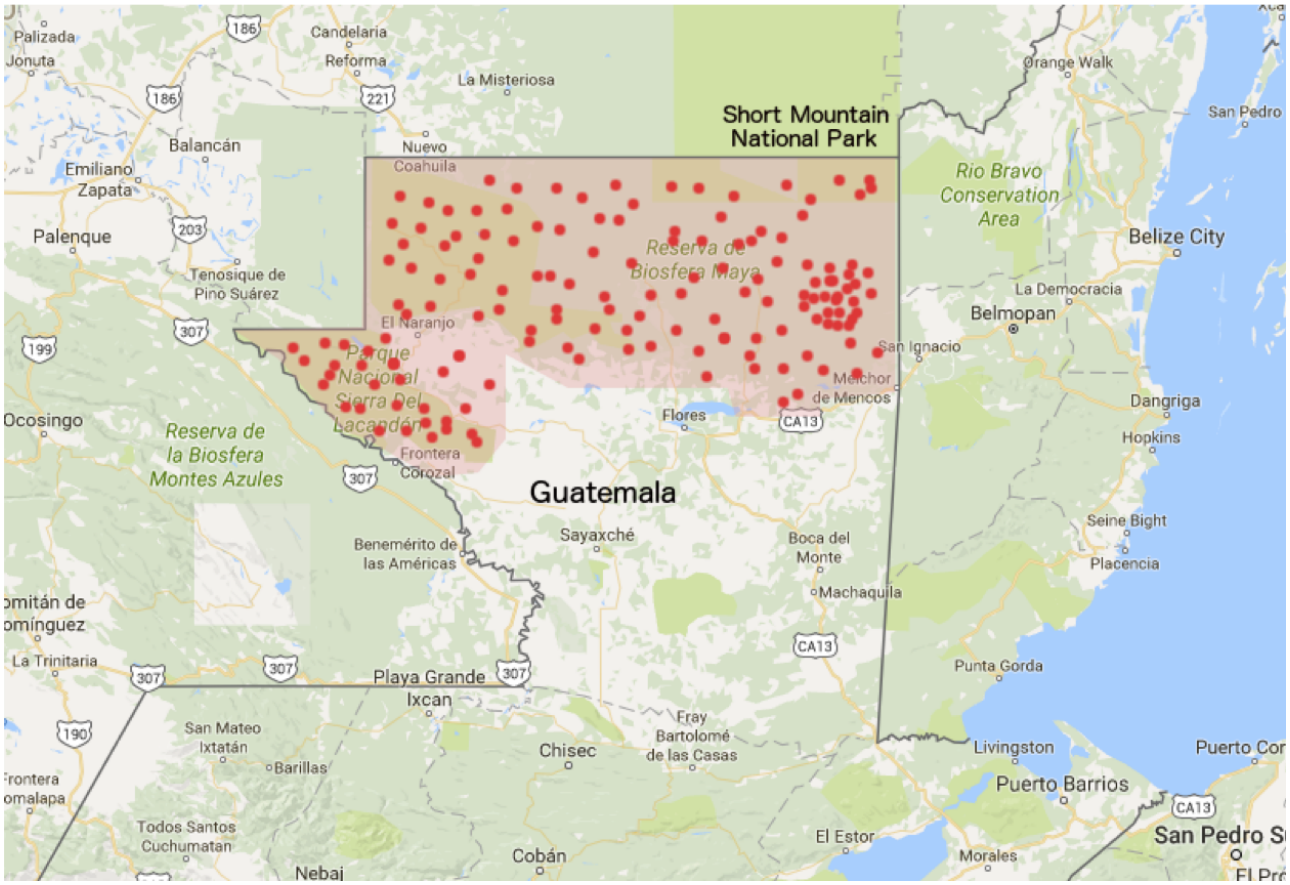
Actualmente, os cuidados da coleção são atribuídos ao professor de Sistemática de Plantas, que realiza tarefas de curadoria, bem como seus trabalhos regulares de investigação e ensino. O Departamento Administrativo é responsável pelas tarefas administrativas do dia-a-dia do herbário, tais como a compra de materiais, incluindo papel e etiquetas. Os docentes e os alunos que colaboram no departamento trabalham ocasionalmente e actualizam as identificações sobre os espécimes e há dois botânicos aposentados que regularmente se voluntariam na coleção ajudando o professor a preparar os empréstimos.

A Universidade já tem uma busca on-line em suas coleções de bibliotecas mantidas pela equipa central da Universidade, em servidores hospedados externamente. Nenhum espécime de história

natural é actualmente integrado a esta plataforma.

O Director do Departamento de Biologia de Plantas garantiu recentemente 50.000USD, dois anos de financiamento à captura de imagem e publicação online das informações de coleções botânicas. A equipa quer aproveitar esta oportunidade para estabelecer um protocolo de digitalização e publicação permanentes, que dará maior visibilidade ao herbário e continuará a atrair financiamento.

Recolha de dados



O herbário é composto por aproximadamente 80 mil espécimes coletados de 1960 a 2015, concentrados principalmente no hotspot de biodiversidade — o Parque Nacional de Montanhas Curtas. A coleção continua a aumentar em resultado de intercâmbios, doações e vários projectos de investigação activos. Para cada espécime, informações sobre coletor, hora, data, local e taxonomia estão documentadas. Até à data, não foram tomadas quaisquer medidas de controlo de qualidade sobre os dados e não houve nenhuma captura de imagens de espécimes sistemática.

Descrição dos dados

O professor tem uma base de dados simples e não-relacional em seu computador, que serve como um índice de muitos, mas não de todos os espécimes. Quaisquer imagens dos espécimes existentes são mantidas localmente pelos investigadores individuais que as tiraram.

Exercícios

Os exercícios individuais para este caso de uso estão localizados nos módulos correspondentes.

Planificação

Exercício 1a-c

Captura de dados

Exercício 2

Gestão de dados

Exercício 3a-c

Publicação de dados

Exercício 4

Folhas de exercício

UC1-Herbarium-Exercise_EN.docx (MS Word 345 KB)

Planificação

No módulo de planificação, você analisará as principais etapas de planejamento do projecto e aprenderá a criar um fluxo de trabalho viável. Além disso, você criará um plano/fluxo de trabalho de projecto idealizado com base no CASO de ESTUDO I. Você identificará objectivos, tarefas, principais stakeholders, funções e atribuirá tarefas específicas às etapas.

Recursos

Neste vídeo (15:09), você aprenderá como definir os elementos que terão impacto em um projecto e analisará as interações entre os elementos. Se não conseguir assistir o vídeo incorporado, pode utilizar o download localmente. (MP4 - 25.1 MB)

https://www.youtube.com/watch?v=VRvUdMjd93c *(YouTube video)*

Organizador

Este vídeo (15:52) concentra-se em tarefas e nas formas como elas podem ser organizadas a fim de preparar planos de trabalho práticos e documentação clara. Se não puder assistir o vídeo embutido, utilize o download localmente. (MP4 - 26,2 MB)

https://www.youtube.com/watch?v=uhhK6B2VwIs *(YouTube video)*

Exercício 1a-c

Para esta actividade, você usará uma plataforma de tabletop virtual. Confira as instruções do exercício abaixo.

Referências

- Definição de papeis
- Definições de estágios

Para ajudá-lo a trabalhar nos exercícios de planeamento para todos os casos, disponibilizamos cinco mesas virtuais de cartas de jogos.

Porque estes são espaços partilhados, nós pedimos que siga essas regras para que todos que queiram utilizá-las possam:

1. Limpar quando terminar. Se você utilizar uma tabela de cartão, certifique-se de devolvê-la para seu estado limpo e original quando terminar. Isso significa deitar todas as cartas fora e remover/excluir quaisquer novas cartas que tenha adicionado ao baralho. Pode fazer isso clicando nos botões de retorno abaixo de cada baralho de cartas.
2. Se estiver ocupada, tente outra mesa. Caso veja que uma mesa de cartas está em uso, utilize um dos outros links de mesas de cartas. Se não for possível encontrar uma vazia, avise-nos em training@gbif.org para que possamos verificar as mesas para ter certeza de que quaisquer mesas não foram deixadas sem participação (ou se precisarmos criar uma mesa adicional para você).
3. Não deixe uma mesa sem vigilância. Por favor, não deixe seu trabalho em uma mesa por longos períodos de tempo. Se tiver de se afastar durante uma ou duas horas, isso é aceitável, mas por favor não deixe as cartas sem vigilância por mais de 4 horas. Se precisar sair da sua configuração por um longo período de tempo, por favor use a plataforma eLearning para notificar o grupo e os treinadores do link que está usando, por que você precisa deixar suas cartas e quando voltará para concluir seu trabalho.
4. Tire capturas de tela com frequência. Não podemos garantir que as cartas não serão redefinidas acidentalmente por alguém do workshop ou por alguém do seu próprio projecto. Você não pode redefinir as cartas uma vez que elas tenham sido redefinidas. Se quiser enviar uma captura de tela com seu exercício, por favor faça.
5. Os treinadores reservam-se o direito de redefinir as mesas de cartas. Se observarmos que uma ou mais mesas contém cartas que não foram guardadas e que não foram utilizados por mais de quatro horas sem quaisquer notificações no Fórum da plataforma eLearning, vamos limpar e redefinir a mesa para outros usarem.

Links para Sala de Jogos (cada mesa será aberta em uma nova janela/aba):

BLUE: https://playingcards.io/ndqppx

GREEN: https://playingcards.io/zpwe9j

PURPLE: https://playingcards.io/y4h6ky

RED: https://playingcards.io/4x4e2j

YELLOW: https://playingcards.io/m6j9sk

ORANGE: https://playingcards.io/868jgh

Muito apreço e agradecimento à Jwalant Patel e a Eric Ma por terem encontrado e ajudado a criar as mesas de jogos on-line para este exercício e à Kate Webbink pela competência artística. A plataforma online é fornecida por https://playingcards.io/.

Exercício 1a

Leia USE CASE I (se ainda não tiver lido).

Utilize os cartões e selecione os objectivos que mais se assemelham aos esboços do projecto e em seguida, escolha as tarefas que precisariam ser executadas para os completar. Depois, identifique as pessoas/recursos que o projecto tem disponível e atribua cartas para elas conforme apropriado. Por último, atribua-os aos stakeholders e afiliações.

1. Revise as cartas GOALS com o grupo, selecione e exponha os que atendem ao caso de uso.
2. Revise as cartas de TAREFA com o grupo.
3. Atribua as cartas de TAREFA a cada uma das cartas GOAL selecionados a partir da descrição do caso.
4. Identifique as instituições e as pessoas mencionadas no texto de caso e tome nota das mesmas.
5. Exponha as cartas de AFFILIATIONS na mesa.
6. Revise as cartas STAKEHOLDER, identifique qualquer mencionado no caso de uso e depois decida a que AFFILIATION pertencem.
7. Revise as cartas ROLE, identifique qualquer uma mencionada no caso de uso e decida a que grupo STAKEHOLDER pertencem.
8. Crie notas as cartas, conforme apropriado.
9. Uma vez que as cartas estejam atribuídas, tire fotos/capturas de tela.
10. Use o exercise sheet para escrever as suas respostas.

Questão

Há recursos ou objectivos faltando nas cartas que você considera essenciais para a conclusão bem sucedida do projecto? Aponte-os na folha de resposta.

Exercício 1b

Utilize a análise de stakeholders e goals do exercício 1a e desenvolva um fluxo de trabalho usando as cartas STAGE.

1. Re-leia a introdução do caso de uso, se necessário.

2. Decida quais TASKS pelas quais cada ROLE será responsável.
3. REveja os cartões STAGE e ordene as TASKS de modo apropriado.
4. Use a folha de exercício descarregada anteriormente para fornecer as suas respostas.

Questões

- Há estrangulamentos óbvios no fluxo de trabalho? Por exemplo: há muitas tarefas para um determinado papel/recurso?
- Quais são as questões que considera importantes para o éxito da mobilização de dados, do ponto de vista de cada um dos papéis/partes interessadas? Por exemplo: o que são os resultados entregues? Pensando no calendário geral do projeto, elas são realistas?
- Consolidar notas e priorizar por ordem de importância.
- Se tiver tempo, pode explorar diferentes combinações, já que diferentes cenários são possíveis em diferentes contextos, ou até mesmo tentar documentar a situação para o seu próprio projeto.

Exercício 1c

Este exercício deve ser utilizado quando o curso for ensinado virtualmente ou presencial, como uma atividade em grupo.

Depois dos exercícios o relator de cada grupo irá:

1. Apresentar as partes interessadas e/ou tarefas em falta, e explicar as razões pelas quais foram adicionadas.
2. Destacar as duas questões/tópicos mais críticas que o grupo identificou.

Pontos de discussão potenciais entre grupos:

- Que semelhanças e diferenças se podem ver nos fluxos?
- Houve questões comuns que surgiram entre os grupos?

Revisão

Faça um "quiz" sobre os conceitos aprendidos nesta secção.

1. Qual é a ordem dos cinco agrupamentos do Processo PMBoK?

   - ☐ Planeamento, Inicialização, Acompanhamento e Controle, Execução, Conclusão
   - ☐ Inicialização, Planeamento, Execução, Acompanhamento e Controle, Conclusão
   - ☐ Inicialização, Planeamento, Execução, Conclusão, Acompanhamento e Controle
   - ☐ Inicialização, Planeamento, Acompanhamento e Controle, Execução, Conclusão

2. Quais são os tipos de resultados a entregar? (múltiplas respostas corretas)

   - ☐ Declarado
   - ☐ Implicado
   - ☐ Estimado
   - ☐ Direto
   - ☐ Indireto
   - ☐ Tentativa

3. O que é um estrangulamento?

   - ☐ um bloqueio que atrasa o desenvolvimento ou progresso
   - ☐ um espaço onde algo ou alguém está em falta
   - ☐ um problema, ou situação que impede alguém de fazer algo, ou torna algo impossível.

4. Quais é que são exemplos de tarefas de mobilização? (múltiplas respostas corretas)

   - ☐ Afiliação
   - ☐ Publicação
   - ☐ Imagem
   - ☐ Georreferênciação
   - ☐ Maior atenção pública

Captura de dados

Neste módulo, irá aprender sobre o conceito de padrões, em particular, o Padrão Darwin Core e seus componentes. Aprenderá também os tipos de dados primários sobre biodiversidade e a melhor forma de partilhar essas informações através do GBIF. Por último, irá rever os princípios da qualidade de dados no contexto da captura de dados e aprenderá sobre a qualidade e a coerência dos dados (especialmente sobre temas como a georeferenciação, datas, nomes e verificações cruzadas de taxa).

Padrões e Darwin Core

Neste vídeo (15:37), irá aprender como interagir com os padrões no dia-a-dia. Seguidamente, ser-lhe-ão apresentados os Padrões de Informações de Biodiversidade, incluindo Padrão Darwin Core que irá usar durante deste curso. Se não puder assistir ao vídeo integrado, poderá utilizar o download localmente. (MP4 - 27 MB)

https://www.youtube.com/watch?v=S02PJHPsRAs *(YouTube video)*

Tipos e origens de dados

Neste vídeo (10:45), poderá rever os dados primários de biodiversidade que podem ser partilhados através do GBIF. Se não conseguir ver o vídeo integrado, pode clicar descarregar e ver localmente. (MP4 - 19 MB)

https://www.youtube.com/watch?v=wKeOveydjsw *(YouTube video)*

Questões

- O seus dados são de um tipo de dados diferente do que pensou originalmente?
- Com que tipo de dados trabalha?
- Como publicaria os seus dados através do GBIF (qual o Core e/ou extensão)?

Captura de dados, processamento e qualidade

Neste vídeo (09:11), irá explorar os princípios da qualidade de dados aplicada à captura de dados, em particular a transcrição de dados das etiquetas de colecção, notas de campo, folhas de cálculo, etc. Se não puder assistir o vídeo incorporado, poderá clicar descarregar e ver localmente. (MP4 - 19 MB)

https://www.youtube.com/watch?v=QkDJIkmwBMA *(YouTube video)*

Exercício 2

Nesta atividade, irá concluir um exercício simulando a captura de dados. Começará a trabalhar com os termos Darwin Core e tomar decisões sobre dados que são necessários para a sua organização/projeto, e irá considerar quais os dados que serão partilhados mais tarde durante a publicação.

Leia USE CASE I (se ainda não tiver lido).

Imagine que é a pessoa designada para transcrever os dados encontrados nas folhas de herbário.

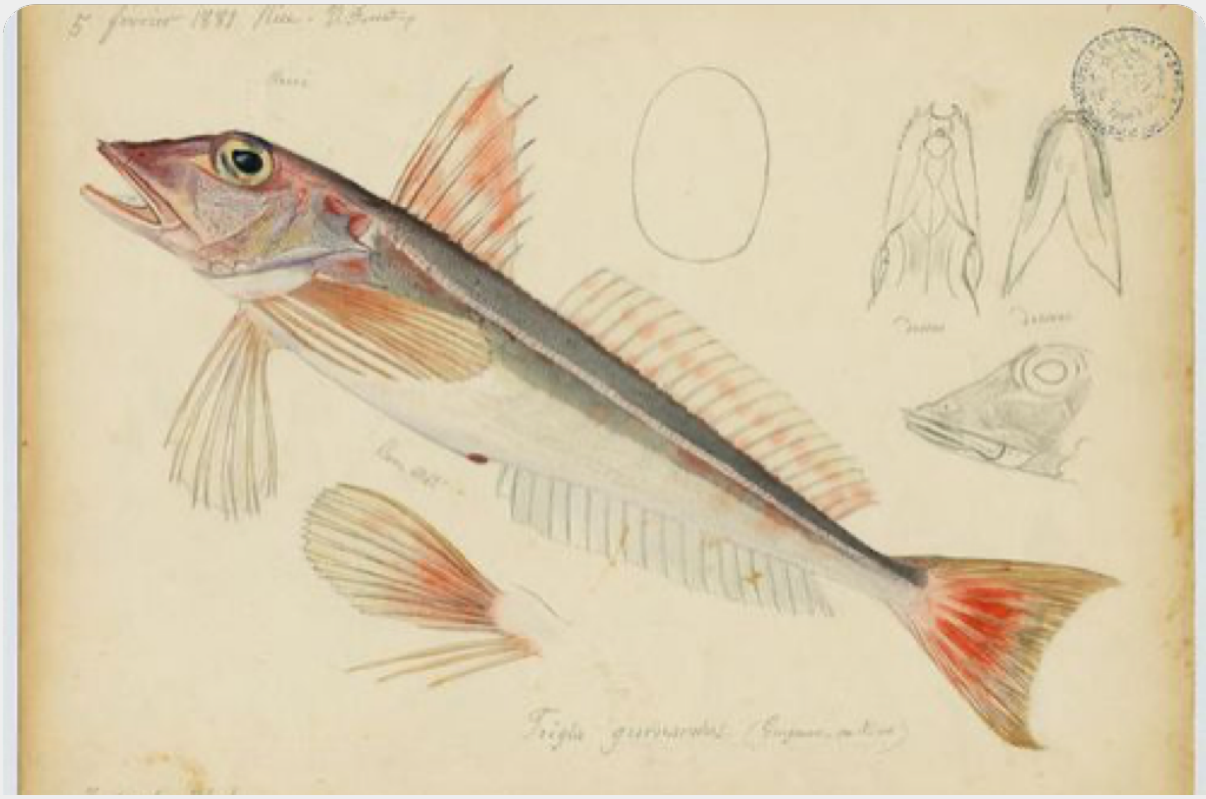
1. Link de download UC1-2-base-material.zip. (34.4 MB). Existem 10 imagens. Duas imagens por espécime para um total de cinco espécimes As folhas de herbário estão em espanhol (os dados podem chegar-lhe por várias vias e noutras línguas que não a sua), mas deve poder reconhecer os dados contidos nos campos das etiquetas. Lembre-se de usar ambas as imagens por registro para compilar as informações.
2. Descarregue o modelo de tabela: UC1-2-occurrence-template.xlsx (57.3 KB) para transcrever as informações encontradas em cada uma das imagens dos cinco espécimes.
3. Use a folha de exercício descarregada anteriormente para fornecer as suas respostas.

pode precisar de adicionar campos à tabela pois poderá capturar mais informações das etiquetas do que foram planeadas no modelo.

Revisão

Faça um "quiz" sobre os conceitos aprendidos nesta secção.

1. Qual(ais) o(s) tipo(s) de conjunto de dados que escolheria para uma coleção de ictiologia?



Eutrigla gurnardus (Linnaeus, 1758) | *Muséum d'histoire naturelle de Nice*

- ☐ ocorrência
- ☐ checklist
- ☐ evento de amostragem

2. Qual(ais) o(s) tipo(s) de conjunto de dados você escolheria para uma lista de espécies invasoras?

Jacinto de água (Eichhornia crassipes) observeada no Bourail, Nova Caledónia, onde é uma espécie introduzida e invasora segundo o GRIIS. Fotografia por gérard (2016) com a licença CC BY-SA 2.0

- ☐ ocorrência
- ☐ checklist
- ☐ evento de amostragem

3. Que tipo(s) de conjunto de dados escolheria para a flora e fauna de um estudo de impacto ambiental?

   Os estudos de avaliação do impacto ambiental são realizados por peritos para avaliar a biodiversidade e os biótopos de uma determinada zona. antes, durante e depois de ser afectada por actividades humanas (obras rodoviárias, turbinas eólicas, mineração, construção civil, etc.).

Entomologista colhendo borboletas *by Matthieu Gauvain (CC-BY-SA)*

- ☐ ocorrência
- ☐ checklist
- ☐ evento de amostragem

4. Qual(is) tipo(s) de conjunto de dados você escolheria para dados de rastreamento de pássaros?

   Dados de rastreamento de pássaros são obtidos com dispositivos específicos, como GPS inseridos em aves vivas, permitindo aos cientistas poderem rastrear rotas migratórias ou os locais de reprodução.

Abutre-grifo *observado na Gamla Nature Reserve by □□□□□□□ - MinoZig (CC0)*

- ☐ ocorrência
- ☐ checklist
- ☐ evento de amostragem

5. Qual(is) tipo(s) de conjunto de dados você escolheria para dados de armadilhas de insectos?

Armadilha de insecto *por miheco (CC-BY-SA)*

- ☐ ocorrência
- ☐ checklist
- ☐ evento de amostragem

6. Que tipo(s) de conjunto de dados você escolheria para gestão dados um de parque nacional?

   Os dados adquiridos no contexto da gestão de áreas protegidas (como parques nacionais, mas também reservas naturais mais pequenas) podem ser diversos e ter diferentes origens: investigação botânica, rastreio de animais marcados, observações de guardas, e até dados de ciência-cidadã ou dados inferidos de imagens partilhadas nas redes sociais.

Elefantes do Sri Lankan *observado por pen_ash.*

- ☐ ocorrência
- ☐ checklist
- ☐ evento de amostragem

7. Qual(ais) o(s) tipo(s) de conjunto de dados que você escolheria para um bioblitz de ciência cidadã?

   Os dados de ciência cidadã são frequentemente recolhidos no âmbito de trabalhos de campo temáticos conhecidos como "bioblitz. Os voluntários normalmente reúnem-se numa determinada área e passam o dia tentando observar e identificar tantas espécies quantas as que puderem nessa área.

   Os dados de cada participante são capturados e adicionados à ferramenta de captura ou gestão de dados do programa de ciência cidadã.

À procur de pássaros com a equipa do parque *por US National Park Service (reutilização autorizada na pesquisa de imagens do google)*

- ☐ ocorrência
- ☐ checklist
- ☐ evento de amostragem

8. Qual(ais) o(s) tipo(s) de conjunto de dados que escolheria para lista regional de espécies?

Rinoceronte negro *observado no Magdeburg Zoo, Alemanha por Mani300*

- ☐ ocorrência
- ☐ checklist
- ☐ evento de amostragem

Gestão de dados

Neste módulo, irá analisar os principais conceitos, ferramentas relacionadas e melhores práticas para a gestão de dados, particularmente, a limpeza e normalização de dados.

Princípios da gestão de dados

Neste vídeo (09:49), irá analisar um conjunto importante de princípios necessários para melhorar os dados através dos processos de limpeza de dados. Se não conseguir assistir ao vídeo incorporado, pode carregar o descarregar localmente. (MP4 - 16.6 MB)

▶ https://www.youtube.com/watch?v=4ijm1cJeVHE *(YouTube video)*

Ferramentas de gestão de dados

Neste vídeo (06:42), irá analisar um conjunto importante de princípios necessários para melhorar os dados através dos processos de limpeza de dados. Se não conseguir assistir ao vídeo incorporado, pode carregar o descarregar localmente. (MP4 - 10.3 MB)

▶ https://www.youtube.com/watch?v=Ru3vWiYU3gw *(YouTube video)*

Exercício 3a-c

Para esses exercícios, irá executar verificações de validação técnicas e consistentes, melhorará os dados com diferentes ferramentas e aprenderá como usar OpenRefine.

Leia USE CASE I (se ainda não tiver lido).

A sua instituição faz parte da “Associação Global de Poales (GPA)”. Esta associação garantiu financiamento para a publicação de uma flora atualizada para o grupo e solicitou ao seu herbário que participasse e fornecesse todos os registos de alta qualidade que possa ter para esta ordem de plantas. A ordem está bem representada na sua coleção, pelo que poderá contribuir substancialmente para este esforço.

Exercício 3a

Verificações de validação

Neste exercício, vamo-nos concentrar em erros técnicos e realizar uma verificação de validação básica para identificar erros técnicos. Consulte Verificações da validação para obter informações sobre os tipos de erros.

1. Link de download UC1-3ab-data-cleaning.csv. (207.5 KB)
2. Importe o arquivo CSV para o Excel usando o assistente do Excel. Consulte o Excel-tips-EN.pdf (PDF, 7 MB) para obter instruções de importação para o seu sistema operativo (Windows, Mac, Linux).
3. Localizar e corrigir os erros manualmente.
4. Use a folha de exercício descarregada anteriormente para fornecer as suas respostas.

Exercício 3b

Outras ferramentas de gestão de dados

A associação GPA forneceu-lhe uma lista de elementos de qualidade de dados para verificar:

- Todos os nomes das plantas (nome completo) estão escritos corretamente
- Todos os nomes das plantas pertencem à ordem
- Todos os registos têm coordenadas
- Todas as coordenadas estão dentro do país declarado e convertidas em formato decimal
- Todas as datas estão na coluna correta e no formato YYYY-MM-DD

As três categorias de erros são:

- Erros nomenclaturais
- Erros de formatação
- Erros geográficos / outliers
    1. Consulte Ferramentas úteis para completar o exercício. Não está limitado a essas ferramentas, pode usar qualquer ferramenta que quiser.
    2. Use o mesmo arquivo do exercício anterior.
    3. Faça a correção APENAS para a família Eriocaulaceae (possivelmente, quererá filtrar os dados)
    4. Corrija os erros encontrados no conjunto de dados usado no exercício 3a (exercício anterior), usando as ferramentas à sua escolha e documente as alterações que realizar na folha de exercícios.
    5. Corrija todo o arquivo se tiver tempo.
    6. Use a folha de exercício descarregada anteriormente para fornecer as suas respostas.

Exercício 3c

Neste vídeo (03:27), você aprenderá sobre OpenRefine. Você pode usar o OpenRefine para padronizar e melhorar a qualidade de seus dados. Se você não conseguir assistir ao vídeo incorporado, você pode descarregar localmente. (MP4 - 3.8 MB)

https://www.youtube.com/watch?v=_YFw_bfwc3Y *(YouTube video)*

OpenRefine

Neste exercício, utilizamos o OpenRefine para melhorar a qualidade de um conjunto de dados usando os recursos padrão, serviços web existentes e expressões regulares.

1. Link de download UC1-3c-open-refine.csv. (207.5 KB)
2. Descarregue e complete os exercícios pelo OpenRefine-Exercise3c-EN.pdf. (PDF, 1.1 MB) Também disponível no Francês e Espanhol.
3. Use a folha de exercício descarregada anteriormente para fornecer as suas respostas.

Dicas do exercício

Verificações da validação

Erros técnicos Relativamente simples, muitas vezes capaz de ser automatizado, **verifica a integridade dos dados**. Isso pode indicar exportações incorretas, mapeamento de dados, deslizamento de campo (e.g. moveu 1 coluna para a direita) ou dados em falta na fonte.

- **Integridade**: Se todos os dados e metadados estão disponíveis - todos os campos estão presentes, todos os campos estão preenchidos?
- **Limites**: Por exemplo, os dias foram dados no intervalo de 1-31 (dependendo do mês)
- **Tipo de dados**: Por exemplo, o campo Data contém uma data ou um número?
- **Formato de dados**: Por exemplo, as datas são fornecidas como 01/01/2010 ou 01/Jan/10?

Erros de consistência

Aplicação de regras do mundo real aos dados. Isso pode indicar informações incorretas de registos antigos, erros de transcrição ou de pós-processamento. Alguns são complexos para implementar e **requerem conjuntos de dados de referência para nova verificação**. E.g. uma lista de coletores e hábitos de coleta conhecidos. Estas regras podem ser recolhidas através de utilizadores e analistas de dados.

- **Taxonomia**: Por exemplo, se identificado ao nível de espécies, foram fornecidos o nome científico binómio e entradas nos campos de gênero e espécie?
- **Moeda**: As datas de coleção, identificação, actualização e digitalização são consistentes?
- **Outliers**: Detectar outliers, mas lembre-se de que nem todos os outliers são necessariamente erros. Por exemplo, compare contra uma extensão de espécies conhecidas ou um intervalo

ambiental conhecido (mas lembre-se que outliers podem ser identificações erradas, ao invés de coordenadas incorretas).

- **Geografia**: As coordenadas estão dentro da localidade ou região identificadas? Por exemplo, há alguma ocorrência terrestre no mar ou ocorrências marinhas em terra?
- **Padrões de amostragem**: O detalhe da ocorrência corresponde aos padrões de colheita conhecidos da organização ou coletor? Algum registo parece ter sido criado após a morte de um coletor (poderia ser um coletor diferente com um nome similar)? Por exemplo, algum registo de mamíferos está atribuído a um grupo de observação de pássaros?
- **Exatidão e precisão**: Por exemplo, quaisquer registos geo-referenciados indicando uma precisão ou exatidão muito alta de um período de colheita pré-GPS (ou pré-acurácia do GPS)?
- **Métodos de colheita**: Diferentes métodos de pesquisa (por exemplo, transects e buscas de área) têm características específicas. Os registos são consistentes com o método fornecido?

Ferramentas úteis

- **GBIF Name Parser**: https://www.gbif.org/tools/name-parser
- **Global Names Resolver**: http://resolver.globalnames.org
- **Correspondência de nome para o catálogo da vida**: https://data.catalogueoflife.org/tools/name-match
- **TNRS**: https://tnrs.biendata.org
- **WoRMS**: https://www.marinespecies.org/aphia.php?p=match
- **InfoXY**: http://splink.cria.org.br/infoxy?criaLANG=en
- **Georeferencing Calculator**: http://georeferencing.org/georefcalculator/gc.html
- **Conversão de coordenadas Canadensys**: http://data.canadensys.net/tools/coordinates
- **Análise de datas do Canadensys**: http://data.canadensys.net/tools/dates
- **Google Maps**: https://maps.google.com/

Revisão

Faça um "quiz" sobre os conceitos aprendidos nesta secção.

1. Por que é melhor limpar seus dados?
   - ☐ para torná-los o mais adequados para uso possível
   - ☐ para atingir seus objetivos de qualidade de dados
   - ☐ os dados devem ser limpos pelos utilizadores, não pelos detentores
2. Como você deve organizar seu fluxo de trabalho de limpeza de dados?
   - ☐ trabalhe sozinho, você conhece seus dados melhor
   - ☐ peça a seus colegas por experiência
   - ☐ trabalhe a nível institucional para harmonizar fluxos de trabalho de qualidade de dados
3. Qual é melhor:
   - ☐ previnir erros de ocorrer
   - ☐ corrigir os erros assim que você os encontrar em seu banco de dados ou folha de cálculo
   - ☐ não limpar erros, mas documentá-los durante o processo, para que as pessoas que reutilizam seus dados saibam onde estão
4. A responsabilidade da qualidade dos dados é de quem?
   - ☐ Da(s) pessoa(s) que registam os dados no campo
   - ☐ De quem transcreve os dados
   - ☐ Do gestor da base de dados
   - ☐ De todos os envolvidos na gestão de dados
   - ☐ Das pessoas que usam seus dados
   - ☐ GBIF
5. Quais ferramentas podem ser utilizadas para limpar seus dados?
   - ☐ Excel e outras ferramentas de gestão de ficheiros
   - ☐ OpenRefine
   - ☐ Seu software de base de dados
   - ☐ Ferramentas online, como o Scientific Names Resolver ou o Google Maps

Publicação de dados

Neste módulo, irá aprender sobre conceitos de publicação de dados, incluindo o IPT, núcleos e extensões, e a importância das licenças, metadados, campos obrigatórios e alojamentos de conjuntos de dados.

Conceitos de publicação de dados

Neste vídeo (11:45), irá aprender sobre os conceitos de publicação de dados e receberá uma introdução ao Integrated Publishing Toolkit (IPT). Se não puder assistir o vídeo integrado, poderá usar download localmente. (MP4 - 20 MB)

https://www.youtube.com/watch?v=b9O0d9ukjSQ *(YouTube video)*

Introdução geral ao IPT

Neste vídeo (06:56), você receberá uma visão geral da interface de publicação de dados do IPT. Se você não conseguir assistir ao vídeo incorporado, pode descarregar localmente. (MP4 - 8.7 MB)

https://www.youtube.com/watch?v=gHXsaN_JWeI *(YouTube video)*

Instalações IPT para treino

Se ainda não recebeu um login, entre em contato com training@gbif.org e ser-lhe-á fornecido um login e senha para o IPT do curso.

https://training-ipt-a.gbif.org/

https://training-ipt-b.gbif.org/

https://training-ipt-c.gbif.org/

Demonstração do IPT

Neste vídeo (24:16), irá aprender como publicar um conjunto de dados de ocorrência usando um IPT. Se não conseguir assistir ao vídeo integrado, você pode usar o download localmente. (MP4 - 52.6 MB)

https://www.youtube.com/watch?v=eDH9IoTrMVE *(YouTube video)*

Exercício 4

Neste exercício, irá publicar um conjunto de dados de ocorrência usando o IPT.

Leia USE CASE I (se ainda não tiver lido).

Publicação de dados

Depois de limpar os dados no conjunto de dados de Poales, a equipa considera que a publicação dos dados on-line através da rede GBIF pode ser uma boa maneira de tornar este esforço visível. Foi solicitado que você liderasse esse trabalho de publicação, com base no conjunto de dados.

1. Neste momento, você precisa de uma conta em uma das course IPTs. Se você ainda não recebeu um login, entre em contato com training@gbif.org e será fornecido um login e senha em um dos IPT do curso.
2. Link para descarregar UC1-4-poales-publishing.csv. (233.5 KB)
3. Utilize a instalação do IPT atribuído e publique este ficheiro.
4. Use a folha de exercício descarregada anteriormente para fornecer as suas respostas.

Revisão

Faça um "quiz" sobre os conceitos aprendidos nesta secção.

1. O que significa publicação de dados no contexto do GBIF?
   - ☐ Significa exportar um ficheiro csv dos seus dados limpos que que podem ser partilhados com os seus colegas
   - ☐ Escrever um artigo com a descriçãodos seus dados, e o(s) protocolo(s) que usou para obtê-los, capturá-los e limpá-los
   - ☐ Tornar os seus conjuntos de dados de biodiversidade publicamente acessíveis e detectáveis num formato padronizado
2. O que é um IPT?
   - ☐ uma ferramenta que o ajuda a gerir e corrigir os seus dados
   - ☐ uma ferramenta que o ajuda a publicar os seus dados através do GBIF
   - ☐ uma ferramenta que o ajuda a produzir um Artigo de Dados
3. Quais licenças e dispensas Creative Commons que são recomendadas pelo GBIF para publicação de dados?
   - ☐ CC-BY, CC-BY-SA e CC-BY-ND
   - ☐ CC0, CC-BY e CC-BY-NC
   - ☐ CC0, CC-BY e CC-BY-SA
4. Quais são os três Núcleos (Cores) que pode escolher para um recurso IPT?
   - ☐ Metadata Core, Occurrence Core, Multimedia Core
   - ☐ Taxon Core, Collection Core, MeasurementOrFact Core
   - ☐ Occurrence Core, Taxon Core, Event Core
5. Quantas ficheiros de Extensão pode um conjunto de dados ter?
   - ☐ zero
   - ☐ um
   - ☐ tantos quanto necessário

Avaliação e certificação

Neste módulo, você irá rever os critérios a serem usados para avaliar as atribuições e para realizar a certificação.

Após a conclusão bem sucedida do curso e avaliação bem sucedida das recomendações (por treinadores e mentores), os participantes têm a oportunidade de receber uma certificação oficial sob a forma de Open Badge.

An overall score of 2.5-2.9 earns a BASIC Biodiversity Data Mobilization badge

An overall score of 3.0-4.0 earns an ADVANCED Biodiversity Data Mobilization badge

Os participantes devem submeter (em inglês) o Use Case II (escolha entre duas opções) e Use Case III, e cada caso será classificado conforme as rubricas educacionais do curso. As rubricas definem os níveis de habilidade e desempenho em que os objectivos de aprendizagem do curso são construídos.

Revise as rubricas para garantir a compreensão das habilidades que serão avaliadas para a certificação.

Rubrica de planeamento

Planeamento

Aptidões	**Desempenho iniciado 1 Desempenho em desenvolvimento 2**	**Desempenho pleno 3**	**Desempenho de excelência 4**	**A. Conhecimento sobre os diferentes elementos que fazem parte de um plano de mobilização de dados de biodiversidade**
Mostra compreensão sobre algumas funções e tarefas exigidas num plano de mobilização sólido, mas não é capaz de diferenciar quem executa cada tarefa no plano ou confunde as funções.	Compreende muitas das funções e tarefas necessárias num plano de mobilização, mas ainda perde as ligações e as interações entre elas.	Compreende as principais tarefas e funções necessárias num plano de mobilização e como elas interagem.	Pode identificar funções e tarefas adicionais que podem ser necessárias para situações específicas	B. Capacidade de aplicar os diferentes elementos dum plano de mobilização de dados de biodiversidade a um determinado contexto institucional (por exemplo, o seu próprio)

Aptidões	Desempenho iniciado 1 Desempenho em desenvolvimento 2	Desempenho pleno 3	Desempenho de excelência 4	A. Conhecimento sobre os diferentes elementos que fazem parte de um plano de mobilização de dados de biodiversidade
Pode ver como algumas das tarefas e papéis se relacionam com um determinado contexto institucional, mas tem dificuldades quando o contexto muda, mesmo que ligeiramente, em relação às referências genéricas usadas.	Pode ver como a maioria das tarefas e funções podem ser aplicadas a um determinado contexto institucional, mas ainda tem lacunas quando a situação é diferente das referências genéricas usadas.	É capaz de identificar todas as tarefas e funções relevantes e como podem ser traduzidas para um determinado contexto institucional, mesmo que a situação não seja idêntica à das referências genéricas utilizadas.	Pode aplicar os diferentes elementos de um plano de mobilização de forma criativa, incorporando novos elementos não encontrados nas referências genéricas utilizadas.	C. Capacidade de escrever / redigir um documento de planeamento claro
Consegue reunir matéria-prima que seria a base de um plano (por exemplo, como itens), mas tem dificuldade em traduzir isso numa narrativa.	Consegue construir um plano básico de trabalho, mas carece de coerência entre os seus elementos.	Consegue escrever um documento coerente e bem articulado que inclua todos os elementos relevantes dum plano de trabalho.	Consegue escrever planos concisos, com resumos eficazes e níveis crescentes de detalhe	D. Capacidade para avaliar um plano de mobilização de dados de biodiversidade

Rubrica de captura de dados

Captura de dados

Aptidões	Desempenho iniciado 1 Desempenho em desenvolvimento 2	Desempenho pleno 3	Desempenho de excelência 4	A. Capacidade de identificar o tipo de dados digitais que podem ser extraídos de uma fonte de dados de biodiversidade (ou seja, que podem ser publicados usando a rede GBIF)
Pode identificar apenas os tipos de dados mais evidentes de fontes comuns de dados de biodiversidade (por exemplo, ocorrências de espécimes de coleção de história natural). Mostra pouca compreensão do potencial de publicação online usando o GBIF.	Frequentemente, pode identificar corretamente, pelo menos um tipo de dados digital que pode ser extraído de fontes de dados comuns. Tem dificuldade em identificar quais podem ser publicados atualmente usando o GBIF.	Consegue sempre identificar um (ou mais) tipos de dados digitais que podem ser extraídos de fontes de dados comuns. Pode identificar qual desses tipos pode ser publicado atualmente usando o GBIF.	Consegue sempre identificar um ou mais tipos de dados digitais que podem ser extraídos de fontes de dados comuns e incomuns. Pode identificar qual desses tipos pode ser publicado atualmente usando o GBIF e quais estão em avaliação. Pode identificar núcleos e extensões de dados usados para publicar esses tipos de dados.	B. Capacidade de extrair informações relevantes de uma fonte de dados de biodiversidade em estruturas de dados simples (por exemplo, folhas de cálculo) que seguem os padrões internacionais

Aptidões	Desempenho iniciado 1 Desempenho em desenvolvimento 2	Desempenho pleno 3	Desempenho de excelência 4	A. Capacidade de identificar o tipo de dados digitais que podem ser extraídos de uma fonte de dados de biodiversidade (ou seja, que podem ser publicados usando a rede GBIF)
Só pode extrair grandes blocos de informações óbvias (por exemplo, todas as informações geográficas como uma única unidade) que são evidentes na fonte de dados. Mostra pouco conhecimento dos padrões atuais para registro de dados de biodiversidade.	Pode recuperar vários itens de informação da fonte de dados (mas não todos) e pode desagregá-los em partes significativas. Mostra algum conhecimento básico dos padrões mais comuns (por exemplo, DwC) e os campos de dados mais usados nesses padrões.	Pode identificar todas as informações valiosas numa fonte de dados e extrair os elementos obrigatórios numa estrutura de dados padrão (por exemplo, uma tabela baseada em DwC Simples). Pode identificar informações ausentes e inferir a partir de informações existentes (por exemplo, derivar o nome de um país a partir de uma província).	Pode identificar todas as informações valiosas em uma fonte de dados complexa e dividi-las em partes significativas que se traduzem diretamente em padrões internacionais. Pode identificar informações críticas ausentes na fonte e inferir a partir dos dados existentes ou de informações adicionais sobre a fonte (metadados).	C. Capacidade de compreender e aplicar princípios básicos de qualidade de dados ao processo de captura de dados

Rubrica de gestão de dados

Gestão de dados

Aptidões	Desempenho iniciado 1 Desempenho em desenvolvimento 2	Desempenho pleno 3	Desempenho de excelência 4	A. Capacidade de avaliar a qualidade (ou seja, identificar problemas e seus tipos) de um conjunto de dados de biodiversidade.
Só usa verificações visuais para analisar a qualidade. Não consegue diferenciar entre os tipos de erros. Consegue detectar valores ausentes em campos obrigatórios e inconsistências de dados graves.	Só consegue usar técnicas muito básicas (por exemplo, classificação) para analisar a qualidade dos dados. Consegue detectar incompatibilidades entre nomes de campo e conteúdo. Consegue identificar erros técnicos de forma consistente, mas apenas os erros de consistência mais comuns num conjunto de dados.	Consegue usar ferramentas e técnicas específicas para avaliar a qualidade. Reconhece o nível mínimo de desagregação / normalização necessário para uso comum e publicação. Consegue identificar consistentemente erros técnicos e a maioria dos erros de consistência num conjunto de dados.	Usa uma abordagem sistemática para análise de conjunto de dados cobrindo todos os principais domínios de dados. Consegue identificar consistentemente erros técnicos e de consistência num conjunto de dados. Consegue usar outras fontes de dados (por exemplo, metadados ou outros conjuntos de dados) para identificar ou inferir erros de consistência num conjunto de dados.	B. Capacidade de realizar correção de formato de dados.

Aptidões	Desempenho iniciado 1 Desempenho em desenvolvimento 2	Desempenho pleno 3	Desempenho de excelência 4	A. Capacidade de avaliar a qualidade (ou seja, identificar problemas e seus tipos) de um conjunto de dados de biodiversidade.
Só consegue fazer correções manualmente nas tabelas. Mostra conhecimento genérico sobre o uso de tipos de formato em dados digitais (por exemplo, datas, strings, números)	Consegue identificar pelo menos uma ferramenta específica para corrigir automaticamente os erros de formato, mas só consegue usá-la em casos específicos. Em alternativa, usa mecanismos simples (por exemplo, 'localizar e substituir') para resolver problemas.	Consegue usar pelo menos uma ferramenta para corrigir automaticamente os erros de formato.	Consegue usar recursos avançados de mais de uma ferramenta para corrigir erros de formato.	C. Capacidade de realizar correção de dados nomenclaturais.

Aptidões	Desempenho iniciado 1 Desempenho em desenvolvimento 2	Desempenho pleno 3	Desempenho de excelência 4	A. Capacidade de avaliar a qualidade (ou seja, identificar problemas e seus tipos) de um conjunto de dados de biodiversidade.
Só consegue fazer correções manualmente nas tabelas. Só usa conhecimento pessoal de grupos taxonómicos conhecidos.	Consegue identificar pelo menos uma ferramenta específica para corrigir automaticamente erros nomenclaturais, mas só consegue usá-la em casos específicos. Em alternativa, usa mecanismos simples (por exemplo, 'localizar e substituir') para resolver problemas.	Consegue usar pelo menos uma ferramenta para corrigir automaticamente os erros nomenclaturais. É capaz de encontrar e usar informações nomenclaturais de referência adequadas para os grupos taxonómicos com os quais costuma trabalhar.	Consegue usar mais de uma ferramenta para corrigir erros nomenclaturais. É capaz de encontrar e usar informações nomenclaturais de referência adequadas para grupos taxonómicos fora das suas áreas de especialização.	D. Capacidade de realizar correção de dados geográficos.

Aptidões	Desempenho iniciado 1 Desempenho em desenvolvimento 2	Desempenho pleno 3	Desempenho de excelência 4	A. Capacidade de avaliar a qualidade (ou seja, identificar problemas e seus tipos) de um conjunto de dados de biodiversidade.
Só consegue fazer correções manualmente nas tabelas. Só usa conhecimento pessoal de áreas geográficas conhecidas.	Consegue identificar pelo menos uma ferramenta específica para mapear e / ou corrigir automaticamente erros nas informações geográficas, mas só consegue usá-la em casos específicos. Em alternativa, usa mecanismos simples (por exemplo, 'localizar e substituir3') para resolver problemas.	Consegue usar pelo menos uma ferramenta para mapear e / ou corrigir automaticamente erros nas informações geográficas. É capaz de encontrar e usar informações geográficas de referência adequadas num formato adequado para as áreas com as quais trabalha normalmente.	Consegue usar mais de uma ferramenta para mapear e / ou corrigir automaticamente os erros nas informações geográficas. É capaz de encontrar e usar informações geográficas de referência num formato adequado para áreas fora da sua área de especialização.	E. Capacidade de usar softwares específicos (por exemplo, OpenRefine) como ferramentas para limpeza de dados.

Aptidões	Desempenho iniciado 1 Desempenho em desenvolvimento 2	Desempenho pleno 3	Desempenho de excelência 4	A. Capacidade de avaliar a qualidade (ou seja, identificar problemas e seus tipos) de um conjunto de dados de biodiversidade.
Consegue identificar pelo menos uma ferramenta de limpeza de dados. Consegue identificar os principais recursos de uma ferramenta de limpeza de dados (por exemplo, OpenRefine).	Consegue identificar várias ferramentas de limpeza de dados. Consegue usar um ou alguns dos recursos básicos do software de limpeza de dados para limpar um conjunto de dados (por exemplo, criar um projeto OpenRefine, usar facetas, filtragem, clustering ou reconciliação).	Consegue usar todos os recursos básicos de um software de limpeza de dados para limpar um conjunto de dados (por exemplo, em OpenRefine: facetas, filtragem, agrupamento, reconciliação).	Consegue usar os recursos avançados de um ou mais pacotes de software de limpeza de dados para limpar conjuntos de dados (por exemplo, em OpenRefine: usar API, expressões regulares, Google Refine Expression Language).	F. Capacidade de documentar procedimentos de transformação de dados.

Rubrica de publicação de dados

Publicação de dados

Aptidões	Desempenho iniciado 1 Desempenho em desenvolvimento 2	Desempenho pleno 3	Desempenho de excelência 4	A. Conhecimento sobre padrões de dados de informações sobre biodiversidade (BDI).
Mostra conhecimento limitado ou nenhum conhecimento sobre os padrões de dados BDI e quais desses padrões de dados são aceitos pelo GBIF.	Consegue identificar padrões BDI e sabe quais são aceites pelo GBIF, mas não sabe onde localizar informações sobre como usá-los. Não consegue identificar quais termos são obrigatórios.	Conhece os padrões BDI aceitos pelo GBIF. Consegue encontrar uma lista de extensões e núcleos de dados aceitos. Publica conjuntos de dados de acordo com os padrões GBIF exigidos e / ou recomendados para termos de dados e metadados e sabe como encontrar as definições dos termos.	Mostra compreensão sobre as características e limitações dos vários padrões de BDI.	B. Capacidade de analisar a adequação de um conjunto de dados de biodiversidade para publicação por meio do GBIF.
Mostra conhecimento limitado ou nenhum conhecimento sobre os critérios formais que um conjunto de dados precisa cumprir para ser publicável através do GBIF.	Conhece os critérios formais que um conjunto de dados precisa cumprir para ser publicável através do GBIF, mas não consegue avaliar se um determinado conjunto de dados os cumpre.	Consegue avaliar corretamente se um conjunto de dados pode ser publicado atualmente através do GBIF. Consegue atribuir pelo menos um tipo de dados válido (= núcleo) a um conjunto de dados com base na descrição fornecida pelo portador de dados e após ter analisado o conjunto de dados.	Consegue identificar mais de uma opção de publicação para um conjunto de dados (quando possível).	C. Uso do IPT: capacidade de produzir / analisar metadados de alta qualidade.

Aptidões	Desempenho iniciado 1 Desempenho em desenvolvimento 2	Desempenho pleno 3	Desempenho de excelência 4	A. Conhecimento sobre padrões de dados de informações sobre biodiversidade (BDI).
Mostra conhecimento limitado ou nenhum conhecimento sobre as características de bons metadados.	Conhece as características de bons metadados, mas tem dificuldade em reconhecê-los.	Conhece as características de bons metadados e como reconhecê-los. Consegue produzir recomendações sobre como melhorar os metadados existentes.	Conhece as características dos metadados de alta qualidade e como produzi-los.	D. Uso de IPT: capacidade de fazer upload / conectar dados e mapeá-los para núcleos e extensões existentes.
Consegue carregar conjuntos de dados de arquivo único no IPT, mas não consegue mapeá-los para nenhum núcleo.	Só consegue carregar conjuntos de dados de arquivo único no IPT e mapear para um único tipo de núcleo sem extensões.	Consegue carregar vários arquivos em um IPT como parte de um único conjunto de dados e mapeá-los corretamente para um núcleo e pelo menos uma extensão. Consegue usar o recurso de valor constante IPT.	Consegue carregar vários arquivos em um IPT como parte de um único conjunto de dados e mapeá-los corretamente para um núcleo e várias extensões. Consegue usar o recurso de tradução de dados IPT.	E. Uso do IPT: capacidade de usar a ferramenta para publicar e registar conjuntos de dados.

Use o caso II - Espécie invasora

Familiarize-se com o cenário do caso.

Você tem uma escolha para o Use Case II entre dois cenários:

- Checklist de espécies invasoras
- Amostragem de Lepidoptera

Sua escolha para Use Case II será avaliada.

Cenário

Rastreamento de espécies invasoras

Leucaena leucocephala (Lam.) de Wit *observed in Hawaii by Sharon Grant (licensed under* CC-BY-NC 4.0*)*

Esta narrativa foi desenvolvida como base para exercícios práticos no curso de mobilização de dados para a biodiversidade, e o conceito e o conteúdo do exercício foi desenvolvido por Sharon Grant, John Wieczorek, David Bloom e Laura Anne Russell.

É um cenário fictício baseado em um conjunto de dados real e destinado apenas para fins instrutivos. O conjunto de dados original é atribuído a Simpson A (2016). Comitê de Espécies Invasoras da Ilha Big Island - Pest Reports- 2005-2010. Versão 4.1. United States Geological Survey. Occurrence Dataset acessada via GBIF.org em 2017-07-13.

Descrição

O Conselho de Espécies Invasoras (HISC) recebeu um subsídio federal para colaborar com escolas de ensino médio como parte do currículo de toda a zona do Havaí em Espécies Invasoras, para aumentar o conhecimento local sobre essas espécies, aumentar a colheita de dados e produzir checklists para áreas com lacunas de dados. Um Gestor de Projecto em tempo integral é empregado para supervisionar o projecto. Todos os fundos e alocações são geridos pelo Tesoureiro do HISC.

O Gestor do Comité de Espécies Invasoras de cada ilha (ISC) recebeu uma sub-prémio para criar programas locais de educação e colher dados. Os programas treinaram estudantes do liceu para tornarem-se mentores de alunos e para facilitar a colheita de imagens e de dados, por membros da comunidade local. O Comité de Espécies Invasoras da Grande Ilha (BIISC) recebeu mais um sub-prémio para estender sua base de dados central, para receber os dados de cada ISC, fornecer as

escolas participantes os seus próprios portais, e manter um portal de dados único e com ferramenta de busca, para servir esforços de investigação governamentais, públicas e acadêmicas.

Foram seleccionadas duas escolas em cada ilha, porque estavam localizadas em zonas onde os conhecimentos e a documentação de avaliação de espécies invasoras eram escassos ou inexistentes. Professores trabalharam com seu Conselho de Espécies Invasoras (ISC) local para a criação de materiais de ensino detalhando 21 espécies importantes de plantas invasoras, incluindo a forma de identificar as fases de vida de cada espécie e os métodos de controlo mais eficazes.

A Universidade do Havaí em Maui (UHM) tem um curso sobre engajamento da comunidade. Como parte da avaliação final do curso, quatro alunos da Universidade estão validando as identificações das imagens e descrições enviadas por cada escola para seu ISC local.

Recolha de dados

Os alunos de cada escola organizaram uma série de pesquisas comunitárias diárias em suas freguesias. Participantes, guiados por Técnicos de Detecção Antecipada do ISC e Mentores de Estudantes, visitaram vários locais onde receberam guias de fotos e atribuíram uma rota para acompanhar durante os eventos de amostragem. Ao longo de cada rota, eles tinham a tarefa de identificar as espécies-alvo e de tirar 1 a 3 fotos delas usando telemóveis com GPS.

Detalhes, descrevendo cada observação das 21 espécies de interesse, foram gravados usando um formulário de colheita de dados digital, durante cada evento de amostragem de dados da comunidade. Os participantes carregaram as imagens capturadas em telemóveis e foram encorajados a clicar seus locais usando um mapa do Google, embutido no formulário, para atribuir latitude e longitude a cada observação. O formulário foi baseado no formulário de relatório de pragas HISC.

REPORTER INFORMATION

Report Number: T07222017172139

First Name:

Last Name:

Email: e.g. pestreport@643pest.com

Phone: e.g. 808-123-4567 or 8081234567

PEST SIGHTING INFORMATION

NOTE: Asterisk [*] and red label color indicate a required field.

*Date of Pest Sighting:

*Pest Name:

*Pest Description:
(Plant: size; flower, foliage, or fruit color / scent / orientation; habitat)
(Insect / animal: size; color; plant / host found on or nearby; habitat)

*Island of Pest Sighting:
(Please choose an island before entering information into the Location field).

--Select Island--

Location of Pest Sighting:
(Street address, cross streets, city, mile marker, place name or general area)

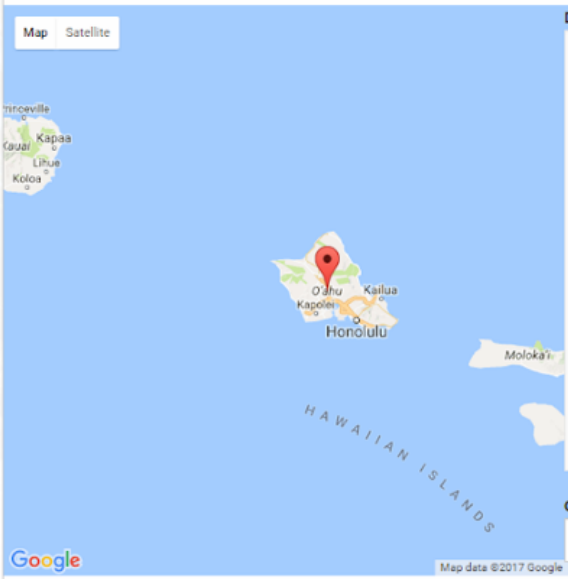
Additional Comments:

Image Upload (*Allowed File Types: .jpg, .png, .gif | Upload up to 3 images*):

Drag & drop files here ...

Browse ...

PEST SIGHTING LOCATION DETAILS



Directions:

1. The map will have automatically moved to the island chosen from the **Island** drop-down list above. If an island was not chosen or the incorrect island was chosen, please go back and make a new selection. Choosing an island is a **mandatory** step.

2. Use the **map tools** at left (Zoom In, Zoom Out, Pan) for help finding and zooming in to a desired area or location. Use the **Geo-location Search** box below to search for and zoom in to a specific geo-location (address, city, place name) in Hawaii.

3. To pinpoint the **exact location of the pest sighting**: a) click on and drag the red map marker to a particular location; or b) click on any particular location on the map to move the red map marker to that location. The **coordinates (Latitide, Longitude)** submitted with the pest report are shown in the boxes below and reflect the final position of the red map marker.

Geo-location Search:

Enter an address, city, place name...

Marker Coordinates:

Latitude

Longitude

Descrição dos dados digitais

Um banco de dados criado e hospedado pelo Departamento de Computação da UHM, contém imagens e dados do formulário on-line, mas esses dados não estão acessíveis ao público. Os dados foram exportados como ficheiros de valor separado por vírgula (.csv) e dados aos quatro alunos de graduação da UHM para validação taxonómica, usando as imagens e descrições enviadas. O analista BIISC GIS utilizou as coordenadas do Google Maps e imagem EXIF para verificar a qualidade das observações e adicionar quaisquer georeferências em falta. Os mentores dos alunos renomearam todos os ficheiros de imagem para coincidir com o número da observação para cruzar as referências posteriormente no BIISC.

Folha de exercício de invasoras

Baixar o folha de exercício. (MS Word, 342 KB)

Exercício 1

Planeamento

Você é o Gestor ISC local e devido ao sucesso dos planos de aula e pesquisas da comunidade, mais 10 escolas da sua ilha gostariam de criar os seus próprios projectos para o ano seguinte. Você gostaria de os acolher, mas o seu financiamento ISC expirará ao final deste ano. O HISC indicou que irá olhar favoravelmente para a submissão de um pequeno subsídio para expandir seus programas no ano seguinte e a BIISC ofereceu apoio.

Exercício 1a

Analise as implicações financeiras em aumentar o número de escolas

1. Avalie as seguintes opções para expandir o número de escolas participantes. Só poderá selecionar DUAS dessas opções, então você precisa escolher sabiamente.
2. Use a folha de exercícios para fornecer suas respostas.

Opções

1. Pague estagiários de verão adicionais para trabalhar no ISC local para coordenar pesquisas.
2. Oferecer apoio financeiro à BIISC para preparar portais para cada escola nova.
3. Oferecer compensação financeira aos estudantes de nível superior. Você não poderá pagar aos quatro deles o equivalente a um salário normal, mas poderia cobrir os custos dos horários parciais para dois deles.
4. Contratar uma empresa de software para construir uma base de dados faça a ingestão automática dos dados diretamente do formulário online. O sistema incluirá uma interface de administração para permitir a manipulação de dados e exportações csv.
5. Financie quatro atividades de divulgação (por exemplo, um BioBlitz) para promover a sensibilização nas comunidades e aumentar a participação voluntária.
6. Prepare e realize uma formação reutilizável para que os professores das escolas ensinem a

preparar os dados para submissão ao BIISC.

Exercício 1b

Atribuir funções

O novo projeto tem as seguintes pessoas disponíveis para processamento e mobilização de dados.

1. Atribuir funções para maximizar a eficiência do processamento e transformação de dados para produzir dados com a melhor qualidade possível?
2. Use a folha de exercícios para fornecer suas respostas.

Funções

- Análise BIISC GIS: Ferramentas avançadas para uso do computador, GIS e análise de dados.
- ISC Manager: Boas aptidões no uso do computador.
- ISC Outreach Associado: Boa aptidão na identificação de campo; uso básico de computador. Especialista em redes sociais.
- Mentor do aluno 1: conhecimento taxonómico básico. Uso básico de computador.
- Mentor do aluno 2: conhecimento taxonómico básico. Uso básico de computador.
- Aluno de Botânica 1: Conhecimento taxonómico avançado. Aptidões de programação.
- Aluno de Botânica 2: Conhecimento taxonómico avançado.
- Aluno de Botânica 3: Conhecimento taxonómico avançado.
- Aluno de Botânica 4: Conhecimento taxonómico avançado.

Exercício 2

Captura de dados

A BIISC planeja agora disponibilizar todos os dados do projeto publicamente através da publicação de conjuntos de dados através do GBIF. Como Assistente de Outreach da BIISC, deve identificar os campos relevantes do Darwin Core para acomodar os dados do formulário online. Notou que os dados adicionais que descrevem espécies e locais foram adicionados ao formulário de dados pelos alunos que realizam validações. Para acomodar esses dados, precisa estender a estrutura de dados para agregar os dados do formulário on-line com a taxonomia e georeferências adicionadas.

1. Link de download UC2-IS-2-ForCapture.csv. (7.1 MB)
2. Usando o conjunto de dados descarregado, produza uma folha de cálculo como exemplo da estrutura de dados estendida e dos campos que identificou como relevantes para o Darwin Core.
3. Use a tabela para fornecer as suas respostas e enviar a folha de cálculo.

Exercício 3

Gestão de dados

Ao longo do verão, os estagiários do gabinete principal do HISC criaram checklists a partir dos dados de ocorrência originais que foram colhidos e ampliados a partir do formulário online. Assumindo o papel do Líder do Projeto HSC, deve agora realizar verificações finais de qualidade antes da publicação.

1. Link de download UC2-IS-3-ForCleaning.xlsx. (156 KB)
2. Avalie o conjunto de dados e identifique que tipos de erros estão presentes.
3. Identifique as formas possíveis de corrigir esses problemas e realizar essas correções para todos os erros que for possível.
4. Use a tabela para fornecer as suas respostas e enviar a folha de cálculo.

Exercício 4

Publicação de dados

O HISC está pronto para publicar os dados da checklist e ocorrências associadas através do GBIF. Para este exercício, irá desempenhar o papel do Líder do Projeto. A sua responsabilidade é publicar online os dados da checklist limpa e ocorrências associadas, através da rede GBIF.

1. Link de download UC2-IS-4-ForPublication.xlsx. (99 KB)
2. Utilize a instalação do IPT fornecida anteriormente para publicar o conjunto de dados fornecido.
3. Use a folha de exercícios para fornecer as suas respostas e ligar ao conjunto de dados publicado.

Caso de Uso II - avistamentos de Lepidoptera

Familiarize-se com o cenário do caso.

Você tem uma escolha para o Use Case II entre dois cenários:

- Checklist de espécies invasoras
- Amostragem de Lepidoptera

Sua escolha para Use Case II será avaliada.

Cenário

Amostragem de Lepidoptera ao nível dos países

Papilio machaon Linnaeus, 1758 *observed in Israel by* □□□□ □□□□ *(licensed under* CC-BY-NC 4.0*)*

Esta narrativa foi desenvolvida como base para exercícios práticos do curso de mobilização de dados para a biodiversidade e o conceito e conteúdo do exercício foram desenvolvido por Alberto González-Talaván, com base no trabalho de Alberto González-Talaván, Danny Vélez, Larissa Smirnova, Laura Russell, Mélianie Raymond e Nicolas Noé. É um cenário fictício e destina-se apenas a fins instrutivos.

Descrição

A International Butterfly Amateur Network (IBAN) tem proporcionado um enquadramento para os grupos de observação amadores nacionais capturarem dados sobre a ocorrência de borboletas (Lepidoptera) desde 2009. Uma vasta rede de observadores amadores usa um protocolo padrão baseado em percursos de Pollard para capturar esta informação em folhas de papel que enviam para a sua sede nacional. Alguns destas sedes digitalizam essas informações em folhas de cálculo, mas outros não dispõem de recursos humanos para o fazer e enviam os registos em papel ao IBAN para processamento. O IBAN elabora um relatório anual baseado nos avistamentos fornecidos por estes membros nacionais. com mapas de distribuição atualizados e análise de tendências demográficas para algumas espécies chave.

A sede do IBAN é principalmente composta por voluntários. Com a crescente popularidade da ciência cidadã e o interesse geral das borboletas enquanto grupo notável de organismos, são recebidos cada vez mais dados todos os anos e as folhas de dados em papel acumulam rapidamente dados não digitalizados. O comité diretivo do IBAN está a tentar identificar um fluxo de trabalho mais eficiente e ágil para a criação de dados digitais, porque gostariam de começar a publicar esses

dados online regularmente. Também gostariam de começar a processar imagens digitais que os seus voluntários já estão a capturar com dispositivos móveis, como telefones e tablets. O seu objectivo último é aumentar o perfil da rede e reforçar as colaborações com os governos locais e regionais para influenciar as políticas de conservação de Lepidoptera nos países envolvidos.

Não existe atualmente qualquer acordo formal entre o IBAN e os amadores que capturam dados, por exemplo, para definir as formas como os dados podem ser utilizados. O comité diretivo tem algumas preocupações de que, quando começarem a publicar os dados online, terão de formalizar esta acordo.

Recolha de dados

O protocolo recomendado - percursos de Pollard - é baseado em transetos que variam entre 300 e 600 m de comprimento, divididos em 50 m secções. Cada transeto deve abranger um único tipo de habitat.

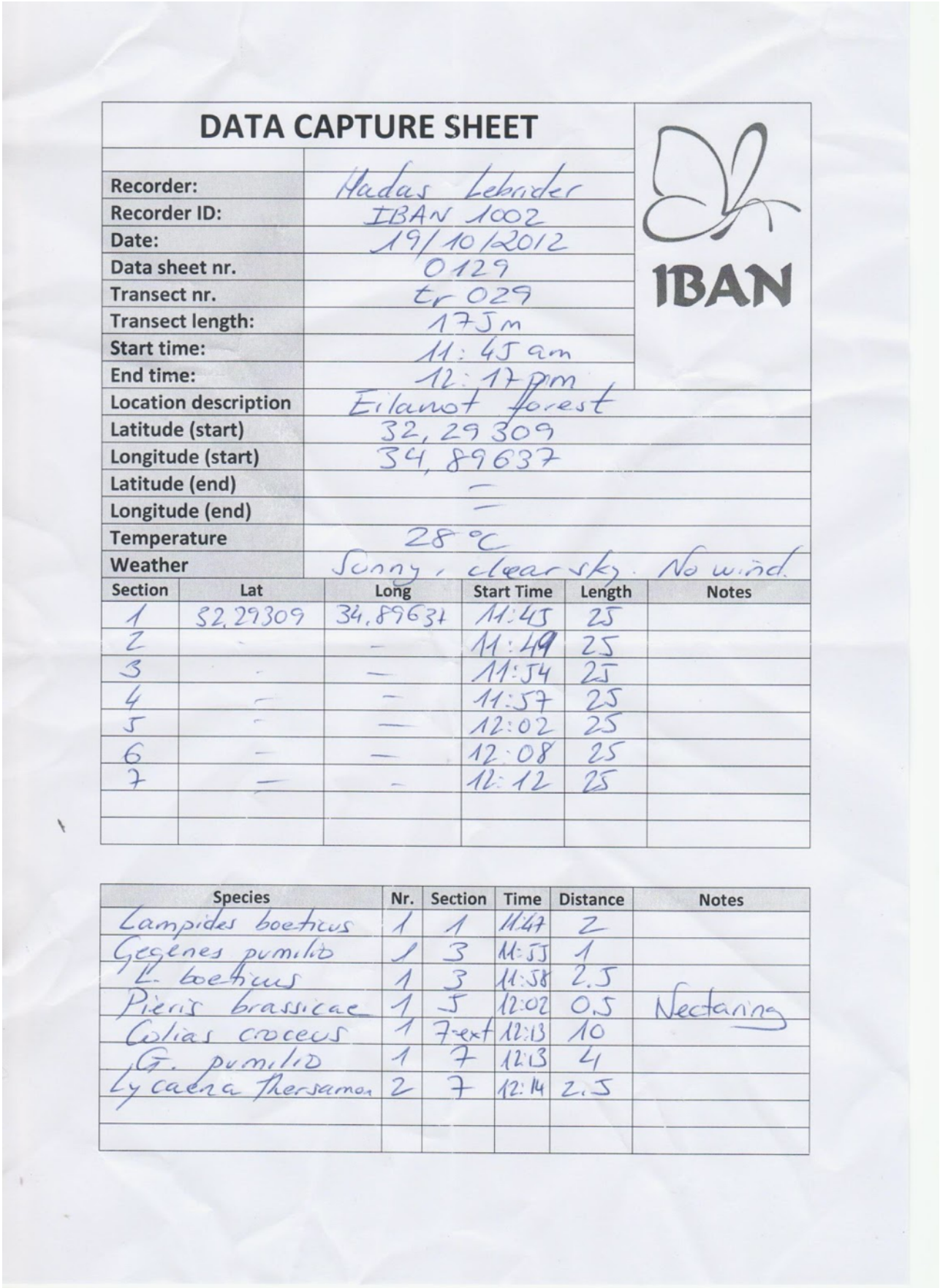
Em cada visita, os participantes que percorrem os transetos têm de contar todas as espécies de Lepidoptera que podem ser vistas dentro de 5 m da linha de transeto. Os comportamentos especiais (posturas de ovos ou néctar), bem como níveis de desenvolvimento (por exemplo, larva ou ovos), também devem ser registrados.

Para a maioria dos países, estes esforços de amostragem ocorrem uma vez a cada duas semanas, desde o início de Outubro até ao final de Junho.

Existem medidas de controle de qualidade em curso: cada registo reportado é sinalizado "Aguardando aprovação". O estado do registo é alterado apenas para "Aprovado" após a verificação por um especialista taxonómico designado. As espécies avistadas fora da temporada regular ou área de distribuição são sinalizadas para verificação adicional.

A hora do dia e as condições meteorológicas são registadas no início do transecto. Ao longo da v, o número de indivíduos para cada espécie observada é contado. As espécies não identificadas são contadas e registadas ao nível da família ou como um complexo predefinido de duas ou três espécies semelhantes. As borboletas observadas fora do intervalo de 5 metros são registadas como "Extra+o número da seção mais próxima" (e.g. 5-extra). A hora final do transecto também é registada.

Exemplo de captura de dados analógicos



DATA CAPTURE SHEET

IBAN

Recorder:	Hadas Lebrider
Recorder ID:	IBAN 1002
Date:	19/10/2012
Data sheet nr.	0129
Transect nr.	tr 029
Transect length:	175m
Start time:	11:45 am
End time:	12:17 pm
Location description	Eilanot forest
Latitude (start)	32,29309
Longitude (start)	34,89637
Latitude (end)	–
Longitude (end)	–
Temperature	28 °C
Weather	Sunny, clear sky. No wind.

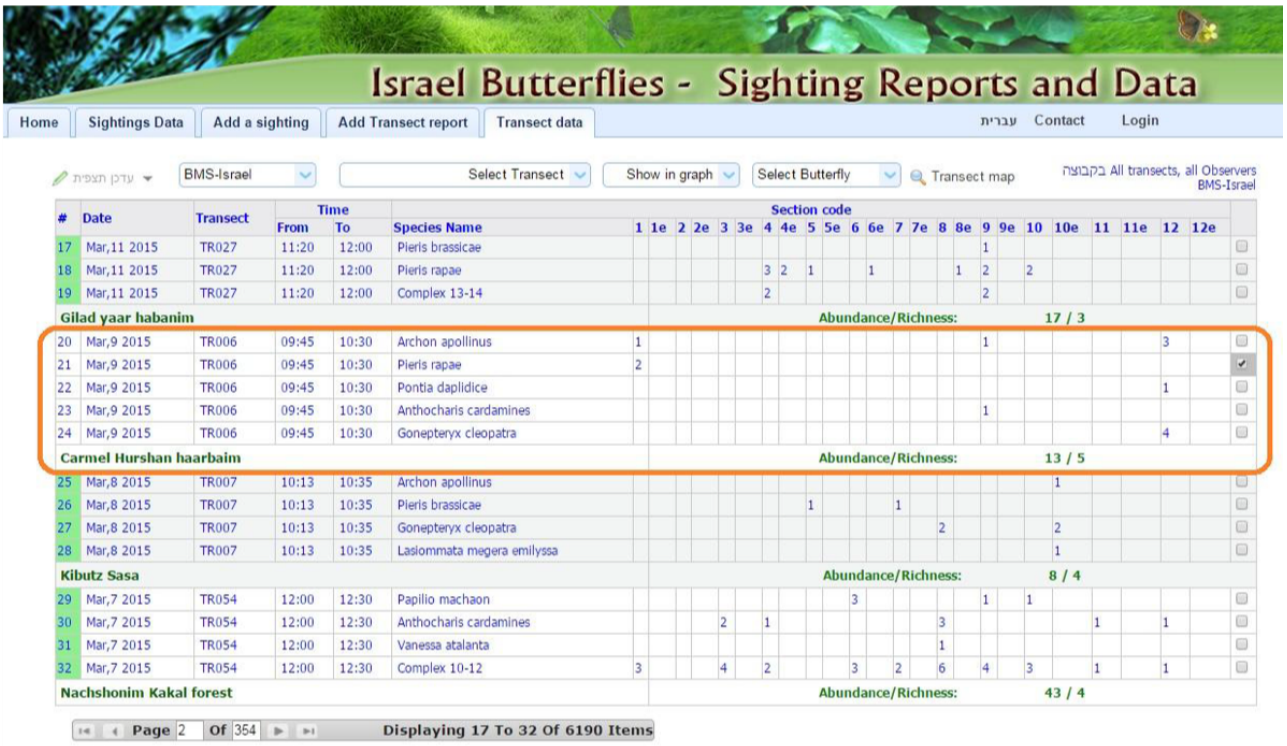
Section	Lat	Long	Start Time	Length	Notes
1	32,29309	34,89637	11:45	25	
2	–	–	11:49	25	
3	–	–	11:54	25	
4	–	–	11:57	25	
5	–	–	12:02	25	
6	–	–	12:08	25	
7	–	–	12:12	25	

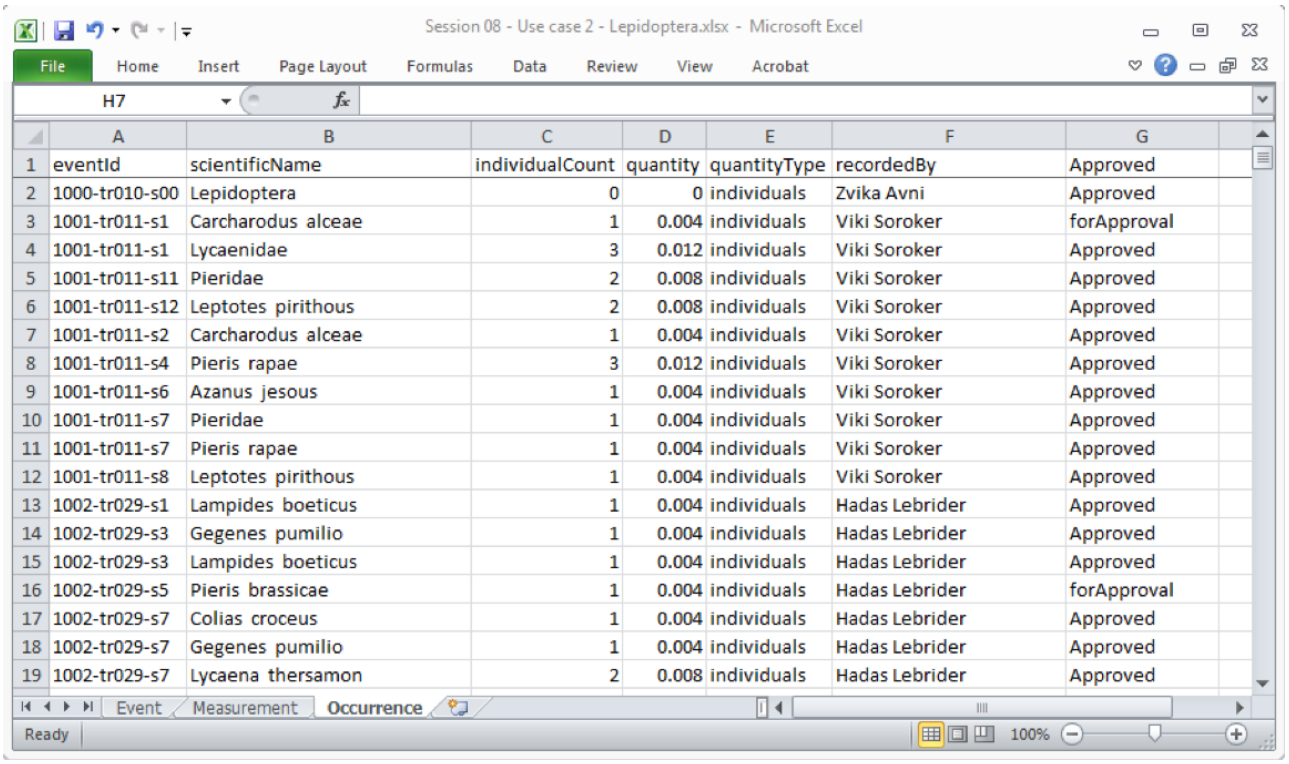
Species	Nr.	Section	Time	Distance	Notes
Lampides boeticus	1	1	11:47	2	
Gegenes pumilio	1	3	11:55	1	
L. boeticus	1	3	11:58	2.5	
Pieris brassicae	1	5	12:02	0.5	Nectaring
Colias croceus	1	7-ext	12:13	10	
G. pumilio	1	7	12:13	4	
Lycaena thersamon	2	7	12:14	2.5	

Descrição dos dados digitais

Alguns gabinetes nacionais usam grupos de voluntários para digitalizar os registos em papel e produzir tabelas em formato digital. As tabelas são muito simples e incluem três fichas técnicas. Uma contém as informações ligadas aos esforços de amostragem, a segunda das condições

meteorológicas e a terceira as espécies encontradas e do número de indivíduos observados pelo observador-amador.





	A	B	C	D	E	F	G
1	eventId	scientificName	individualCount	quantity	quantityType	recordedBy	Approved
2	1000-tr010-s00	Lepidoptera	0	0	individuals	Zvika Avni	Approved
3	1001-tr011-s1	Carcharodus alceae	1	0.004	individuals	Viki Soroker	forApproval
4	1001-tr011-s1	Lycaenidae	3	0.012	individuals	Viki Soroker	Approved
5	1001-tr011-s11	Pieridae	2	0.008	individuals	Viki Soroker	Approved
6	1001-tr011-s12	Leptotes pirithous	2	0.008	individuals	Viki Soroker	Approved
7	1001-tr011-s2	Carcharodus alceae	1	0.004	individuals	Viki Soroker	Approved
8	1001-tr011-s4	Pieris rapae	3	0.012	individuals	Viki Soroker	Approved
9	1001-tr011-s6	Azanus jesous	1	0.004	individuals	Viki Soroker	Approved
10	1001-tr011-s7	Pieridae	1	0.004	individuals	Viki Soroker	Approved
11	1001-tr011-s7	Pieris rapae	1	0.004	individuals	Viki Soroker	Approved
12	1001-tr011-s8	Leptotes pirithous	1	0.004	individuals	Viki Soroker	Approved
13	1002-tr029-s1	Lampides boeticus	1	0.004	individuals	Hadas Lebrider	Approved
14	1002-tr029-s3	Gegenes pumilio	1	0.004	individuals	Hadas Lebrider	Approved
15	1002-tr029-s3	Lampides boeticus	1	0.004	individuals	Hadas Lebrider	Approved
16	1002-tr029-s5	Pieris brassicae	1	0.004	individuals	Hadas Lebrider	forApproval
17	1002-tr029-s7	Colias croceus	1	0.004	individuals	Hadas Lebrider	Approved
18	1002-tr029-s7	Gegenes pumilio	1	0.004	individuals	Hadas Lebrider	Approved
19	1002-tr029-s7	Lycaena thersamon	2	0.008	individuals	Hadas Lebrider	Approved

Folha de exercício Lepidoptera

Link de download exercise sheet. (MS Word, 342 KB)

Exercício 1

Planeamento

O volume de dados analógicos (registos em papel) que chegam na sede IBAN logo excederá a sua capacidade de digitalizar. Assim, a direcção decidiu reconsiderar a actual abordagem desta área de trabalho que tem vindo a crescer sem gestão nos últimos anos. Até à data, é assim que o trabalho tem sido organizado:

- Os registos de papel chegam através dos correios. A secretária abre os pacotes e recolhe os registos.
- Há cinco voluntários com capacidades básicas em informática usando dois computadores partilhados para digitalizar os registos de papel. Esses voluntários também são cientistas cidadãos e portanto estão familiarizados com a taxonomia da ordem Lepidoptera, e com as espécies que ocorrem no país onde se situa a sede do IBAN.
- Os voluntários que fazem a digitalização aparecem sempre que têm tempo, para que normalmente verifiquem por telefone a disponibilidade de computadores. Por vezes, há incompatibilidades de tempo e alguns têm de ir para casa, uma vez que os dois computadores estão ocupados. Outras vezes, ambos os dois computadores estão desocupados.
- Quando eles digitalizam, geralmente escolhem um papel de registo de cada vez a partir do monte de registos, e digitalizam-no (se puderem). Os problemas comuns que ocorrem são:
  - o voluntário não sabe se as espécies observadas (ocorrem erros de escrita),
  - o voluntário não sabe onde ocorreu a amostragem
  - o voluntário não consegue ler a caligrafia ou a linguagem em que alguns dos comentários são escritos.
- Uma única especialista taxonómico recebe todas as tabelas digitalizadas e produz os mapas de relatório e distribuição com base nelas. Normalmente ela tem de descartar cerca de 15% dos dados digitalizados por causa de inconsistências, erros ou outros erros que não tem tempo para verificar.

Exercício 1a

Analisar a componente financeira de seu novo plano de digitalização

A direcção está a analisar as seguintes opções para o seu novo plano de digitalização, que tem todas as implicações financeiras no seu já reduzido orçamento. Sabem que só podem implementar DUAS destas opções, pelo que têm de escolher com sensatez. Use a folha de exercícios para fornecer uma recomendação sobre quais duas opções eles devem seleccionar e explicar por que as escolheu.

1. Opção 1: Compre mais três computadores para que todos os voluntários possam digitalizar simultaneamente.
2. Opção 2: Oferecer suporte financeiro aos escritórios nacionais para comprar scanners de mesa e enviar/partilhar os registos eletronicamente, ao invés de por correio.
3. Opção 3: Oferecer compensação financeira aos voluntários. Eles não poderão pagar a todos os cinco o equivalente a um salário normal, mas poderão cobrir os custos de posições a tempo parcial para três dos voluntários.
4. Opção 4: Adquira um software de digitalização de biodiversidade existente em inglês, que inclua a verificação de entrada taxonómica e ferramentas auxiliares para corrigir a informação geográfica.

5. Opção 5: Contrate de uma empresa de software para desenvolver um programa de digitalização personalizado. Pelo mesmo preço do software comercial, os programadores fornecerão uma solução na linguagem local, que corresponderá perfeitamente ao esquema de dados original e também fornecerá um portal de dados web para expor os resultados do esforço de digitalização.
6. Opção 6: Organize um curso para que os cinco voluntários melhorem suas capacidades em taxonomia, uso de computadores e padrões de informática para a biodiversidade.

Exercício 1b

Atribuir funções

Estes são os recursos humanos disponíveis para este esforço de digitalização. Como designar os papéis para maximizar a eficiência do processo de digitalização e produzir dados da maior qualidade possível? Use a folha de exercícios para fornecer as suas respostas.

1. Assistente administrativo. Nenhum conhecimento taxonómico. Uso básico de computador. Pode ler 3 idiomas.
2. Voluntário 1. Conhecimento taxonómico básico. Uso básico de computador.
3. Voluntário 2. Conhecimento taxonómico básico. Uso básico de computador.
4. Voluntário 3. Conhecimento taxonómico básico. Uso básico de computador. Pode ler 3 idiomas.
5. Voluntário 4. Conhecimento taxonómico básico. Uso básico de computador. Pode ler 3 idiomas.
6. Voluntário 5. Conhecimento taxonómico básico. Uso avançado do computador (incluindo GIS e ferramentas de análise de dados).
7. Especialista taxonómico. Conhecimento taxonómico avançado. Uso avançado do computador (incluindo GIS e ferramentas de análise de dados).

Exercício 2

Captura de dados

Imagine que é um dos voluntários que digitaliza os registos em papel recebidos na sede do IBAN. Você recebeu dois registos em papel.

1. Descarregue os registos 1 e 2 UC2-LS-2-ForCapture.zip. (943 KB)
2. Que estrutura de dados usaria que seja adequada aos dados nesses registos?
3. Criar uma tabela usando esta estrutura e os dados dos registos.
4. Use a tabela para fornecer as suas respostas e enviar a folha de cálculo.

Exercício 3

Gestão de dados

Assumindo o papel de um dos voluntários com competências informáticas avançadas, imagine que lhe foi atribuída a responsabilidade pelas questões de qualidade dos dados. A sua tarefa principal é

reduzir a quantidade de dados que atualmente são descartados (cerca de 15%) antes do processamento devido a erros e inconsistências. Recebeu um conjunto de dados como o produto bruto do esforço de digitalização.

1. Link para download UC2-LS-3-ForCleaning.xlsx. (44 KB)
2. Avalie o conjunto de dados e identifique que tipos de erros estão presentes.
3. Identifique as formas possíveis de corrigir esses problemas e faça essas correções para o maior número possível de erros.
4. Use a tabela para fornecer as suas respostas e enviar a folha de cálculo.

Exercício 4

Publicação de dados

Para este exercício, você assumirá o papel do especialista taxonómico colaborando com a sede do IBAN. Algumas das suas responsabilidades anteriores (escrever o relatório anual e produzir os mapas de distribuição base) foram entregues aos voluntários, e agora recebeu uma nova responsabilidade: publicar online os dados limpos através da rede GBIF. O voluntário responsável pela qualidade dos dados forneceu um conjunto de dados para ser publicado.

1. Link para download UC2-LS-4-ForPublication.xlsx. (58 KB)
2. Utilize a instalação do IPT fornecida anteriormente para publicar o conjunto de dados fornecido.
3. Use a folha de exercícios para fornecer as suas respostas e ligar ao conjunto de dados publicado.

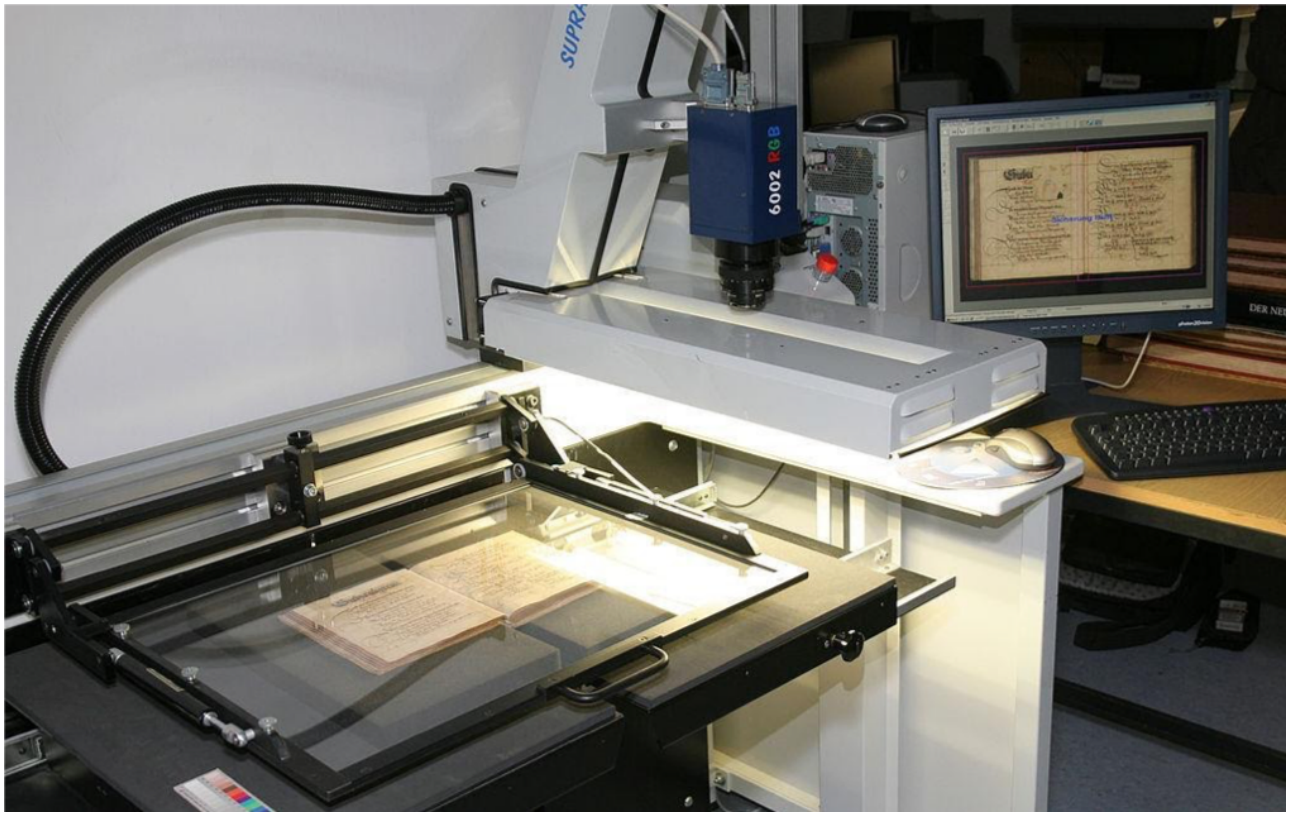
Caso de Uso III - Aves a partir de literatura

Familiarize-se com o cenário do caso.

O Caso de Uso III será classificado.

Cenário

Projeto de Mobilização de Dados a partir da Literatura “Aves caídas nos Faróis Dinamarqueses, 1883–1939”



Scanner de alta resolução *para o projeto de digitalização de livros de Heiko Hornig (licenciado sob* CC BY-SA 2.5*)*

Esta narrativa foi desenvolvida como base para os exercícios práticos no curso de mobilização de dados sobre biodiversidade, e o conceito e conteúdo do exercício foi desenvolvido por Alberto Gonzaz-Talavaėn, Andrea Hahn, Laura Russell e Sharon Grant. O exercício baseia-se numa adaptação anterior feita por Alberto González-Talaván, Danny Vélez, Larissa Smirnova, Laura Russell, Mélianie Raymond e Nicolas Noé.

Trata-se de um cenário fictício baseado num projecto real e num conjunto de dados e destina-se apenas a fins instrutivos. projeto original e dataset original são atribuídos ao nó dinamarquês GBIF, DanBIF.

Descrição

O Museu da História Natural da Dinamarca (NHM-DK) é um centro de investigação associado à Universidade de Copenhaga. A sua biblioteca é um membro da associação de bibliotecas nacionais que recebeu recentemente financiamento estatal para disponibilizar online os recursos mantidos pelos seus membros. O NHM-DK gostaria de começar a digitalização dos cadernos de campo, publicações em revistas científicas e livros detidos na sua biblioteca, alguns dos quais têm um valor histórico significativo.

Após uma breve consulta com os seus parceiros regulares, a NHM-DK recebeu uma sugestão do chefe do gabinete de gestão do Parque Nacional de Nordjylland. Estes gostariam que o conteúdo de uma literatura clássica fosse digitalizado para um projecto em execução: "Aves nos Faróis Dinamarquesa, 1883-1939" (Em dinamarquês, “Fuglene ved de danske Fyr, 1883–1939”). Querem usar quaisquer dados de ocorrência registrados nesses livros de dois faróis (Lodbjerg Fyr e Hanstholm Fyr) para um projeto de uma exposição no local.

O NHM-DK iniciou discussões com o seu nó nacional GBIF, DanBIF, sobre a mobilização das informações contidas nestes volumes, ou seja, preservar seu conteúdo para o futuro e proporcionar acesso online para todos. Com a participação do DanBIF, existe a intenção de publicar e registar os dados extraídos com o GBIF. Uma vez que o GBIF requer que uma licença seja aplicada com todos os dados publicados. o museu decidiu publicar os dados com uma licença Creative Commons que permite o uso de dados com atribuição (CC-BY).

Os serviços de TI necessários são fornecidos pela Unidade Tecnológica da Universidade de Copenhaga, assim como por todos os projectos digitais de museus.

The NHM-DK deputy director, who is coordinating this piece of work has developed a general outline for the work:

1. O museu executará a digitalização da literatura usando dois funcionários da biblioteca treinados no uso do scanner de bibliotecas para digitalizar volumes delicados. Eles também extrairão o texto com software de digitalização através de OCR (Reconhecimento de Caracteres).
2. Três voluntários da Sociedade Ornitológica de Copenhaga (COS) que colaboram regularmente com o museu e estão familiarizados com as aves da região foram associados para ajudar e completarão a transferência de dados dos PDFs digitalizados para formato de folha de cálculo. Eles terão de ir ao museu e usar os computadores disponíveis na biblioteca para ter acesso aos arquivos armazenados na intranet do museu (rede privada).
3. The Ornithology Curator in the NHM-DK Bird Department will lead the team responsible for taxonomic checking, data curation, cleaning, format and transformation, and will oversee the entry of metadata for the published dataset. The team includes a collaborating researcher from Sweden, and two postdoctoral students. They have been selected for this task because they are used to working with digital biodiversity data. They will all use their own work computers.
4. O Coodenador do Nó DanBIF garantirá que a instituição está adequadamente registada no GBIF como um publicador de dados e que o director adjunto e o curador da ornitologia têm credenciais e acesso adequados à instância do IPT do DanBIF para carregar e publicar os dados.

Colheita de dados originais

No período de 1883-1939, estiveram em funcionamento na Dinamarca 45 faróis e faróis flutuantes. Estes faróis foram utilizados por várias espécies de aves durante as noites do período de migração de aves entre os anos de 1886 e 1939. A presença e as actividades dessas aves foram registadas, especialmente pelos responsáveis desses faróis que recolheram também espécimes enviados para o museu em Copenhaga. Estes pássaros foram cuidadosamente preservados e catalogados por gestores de colecção no museu e ainda hoje é possível encontrar ali espécimes. Foram também documentadas observações sobre as condições meteorológicas durante a noite em que as aves eram observadas pelos faroleiros.

Descrição dos dados analógicos

Este é um exemplo da descrição de uma série de observações de espécies de um dos livros (em alemão, exceto o nome comum para a espécie que é fornecido em Danish).

2. *Cerchneis tinnunculus*, Linn. (»Taarnfalk«).

nistet in mehreren Kirchthürmen in der Gegend von **Viborg**. An den Kirchen von **Mönsted** und **Daubjerg** gab es im Jahre 1880 mehrere Sippen, jede von 7 Stücken, was auf einen besonders reichlichen Futtervorrath auf den umliegenden grossen Haiden (Eidechsen) schliessen lässt. In der Gegend von **Thisted**, wo der Thurmfalke sehr häufig brütet, waren die Jungen stets zu 6 vorhanden, nur ein einziges Mal fand ich auf **Egebäksande** eine Sippschaft von 7. (H.)

Bei **Horsens** habe ich den Thurmfalken an den Kirchen von **Vähr**, **Hansted**, **Hundslund**, **Thyrsted** und **Oelsted** brütend gefunden. In dem Kirchthurme von **Vähr** fand ich ein Nest mit 6 Eiern 6. Mai 1875, 6 Eier 30. April 1876, 4 Eier 28. April 1878; 1877 und 1879 waren keine da, 1880 hatten sie 5 Eier am 2. Mai, 1882 ebenfalls 5 Eier und 1883 sieben.

In **Oelsted** 5 Eier am 3. Mai 1879, 1882 ebenfalls 5 und 1883 6 Eier.

Auf den Kirchen (Frauen- und Petri-) von **Kopenhagen** habe ich ihn oft im Herbste und Winter 1879 gesehen; im Frühjahr 1880 brütete er auf der Frauenkirche, dem »runden Thurm« und dem Kanzleigebäude (in einem Ventil) und gewiss auch auf der Petrikirche. (F.)

Descrição de dados digitalizados e traduzidos

Este é um exemplo da saída digitalizada e traduzida do exemplo analógico acima.

Output of the OCR software	Translation into English
2. *Cerchneis tinnunculus*, Linn. (»Taarnfalk«) nistet in mehreren Kirchthürmen in der Gegend von **Viborg**. An den Kirchen von **Mönsted** und **Daubjerg** gab es im Jahre 1880 mehrere Sippen, jede von 7 Stücken, was auf einen besonders reichlichen Futtervorrath auf den umliegenden grossen Haiden (Eidechsen) schliessen lässt. In der Gegend von **Thisted**, wo der Thurmfalke sehr häufig brütet, waren die Jungen stets zu 6 vorhanden, nur ein einziges Mal fand ich auf **Egebäksande** eine Sippschaft von 7. (H.) Bei **Horsens** habe ich den Thurmfalken an den Kirchen von **Vähr**, **Hansted**, **Hundslund**, **Thyrsted** und **Oelsted** brütend gefunden. In dem Kirchthurme von **Vähr** fand ich ein Nest mit 6 Eiern 6. Mai 1875, 6 Eier 30. April 1876, 4 Eier 28. April 1878; 1877 und 1879 waren keine da, 1880 hatten sie 5 Eier am 2. Mai, 1882 ebenfalls 5 Eier und 1883 sieben. In **Oelsted** 5 Eier am 3. Mai 1879, 1882 ebenfalls 5 und 1883 6 Eier. Auf den Kirchen (Frauen- und Petri-) von **Kopenhagen** habe ich ihn oft im Herbste und Winter 1879 gesehen; im Frühjahr 1880 brütete er auf der Frauenkirche, dem »runden Thurm« und dem Kanzleigebäude (in einem Ventil) und gewiss auch auf der Petrikirche. (F.)	2. *Cerchneis tinnunculus*, Linn. ("Taarnfalk") nests in several steeples around **Viborg**. At the churches of **Mönsted** and **Daubjerg** there were several family groups in 1880, each of 7 individuals, suggesting a particularly abundant source of food on the surrounding heather (lizards). In the area of **Thisted**, where the tower falcon broods very often, young were always present in broods of 6, only once did I find a group of 7 on **Egebäksande** (H.) In **Horsens** I found kestrels brooding on the churches of **Vähr**, **Hansted**, **Hundslund**, **Thyrsted** and **Oelsted**. In the steeple of **Vähr** I found a nest with 6 eggs on 6 May 1875, 6 eggs on 30 April 1876, 4 eggs on 28 April 1878; in 1877 and 1879 there were none, on 2 May 1880 they had 5 eggs, in 1882 also 5 eggs, and in 1883 seven. In **Oelsted** 5 eggs on 3 May 1879, in 1882 also 5, and in 1883 6 eggs. On the churches (Our Lady's and St. Peter's) of **Copenhagen**, I have often seen it in the autumn and winter of 1879; in spring 1880, it brooded on Our Lady's church, the "round tower" and the law firm building (in a valve) and certainly also on St. Peter's Church. (F.)

Descrição dos dados digitais

Estudando o extrato do livro, os voluntários da Sociedade de Ornitologia de Copenhaga sugerem extrair os seguintes dados do texto digitalizado e traduzido:

- O nome científico como aparece no livro
- Nome(s) comum(ns) em dinamarquês como aparece no livro
- Localidade
- Ano/mês/dia
- Número de indivíduos observados
- Sexo
- Estágio de vida
- Observações
- URL da página de livro digitalizado em que a ocorrência é fornecida

Aves da literatura da folha de exercícios

Descarregar o folha de exercício. (MS Word, 342 KB)

Exercício 1

Planeamento

A equipa precisa desenvolver um fluxo de trabalho sustentável para digitalizar os recursos de literatura, extrair destes qualquer informação valiosa sobre biodiversidade e publicá-lo on-line via GBIF. Precisam de desenvolver um plano que possa ser sustentado no futuro depois de terminado o financiamento da associação das bibliotecas nacionais.

A secção scenario deste caso de uso inclui uma breve descrição do fluxo de trabalho concebido pelo vice diretor. Baseando-se no fluxo de trabalho e no texto que o acompanha, complete o seguinte:

1. Identifique as diferentes partes interessadas que participam neste projeto
2. Identifique a afiliação e atribua cada uma delas a um grupo de partes interessadas
3. Identifique os papéis associados a eles e atribua as tarefas pelas quais eles são actualmente responsáveis
4. Faça uma análise crítica do fluxo de trabalho, identifique potenciais riscos e lacunas e sugira maneiras de melhorar o fluxo de trabalho, de maximizar a eficiência do projecto de digitalização e de produzir dados de maior qualidade possível.
5. Use a folha de exercícios para fornecer suas respostas.

Exercício 2

Captura de dados

As verificações e o reconhecimento de caracteres (OCR) dos livros foram concluídos. Os dados de ocorrência devem agora ser extraídos dessas fontes e compilados em um formato de folha de cálculo.

Os dados originais estavam em alemão e, para torná-los mais amplamente utilizáveis quando publicados on-line, o gestor do projeto gostaria de torná-los disponíveis em inglês.

> 2. Cerchneis tinnunculus, Linn. ("Taarnfalk") nests in several steeples around Viborg.
>
> At the churches of Mönsted and Daubjerg there were several family groups in 1880, each of 7 individuals, suggesting a particularly abundant source of food on the surrounding heather (lizards). In the area of Thisted, where the tower falcon broods very often, young were always present in broods of 6, only once did I find a group of 7 on Egebäksande (H.)
>
> In Horsens I found kestrels brooding on the churches of Vähr, Hansted, Hundslund, Thyrsted and Oelsted. In the steeple of Vähr I found a nest with 6 eggs on 6 May 1875, 6 eggs on 30 April 1876, 4 eggs on 28 April 1878; in 1877 and 1879 there were none, on 2 May 1880 they had 5 eggs, in 1882 also 5 eggs, and in 1883 seven.
>
> In Oelsted 5 eggs on 3 May 1879, in 1882 also 5, and in 1883 6 eggs.
>
> On the churches (Our Lady’s and St. Peter’s) of Copenhagen, I have often seen it in the autumn and winter of 1879; in spring 1880, it brooded on Our Lady’s church, the "round tower" and the law firm building (in a valve) and certainly also on St. Peter’s Church. (F.)

1. Assuma o papel de um voluntário encarregado de transformar o texto traduzido numa tabela

com ocorrências individuais. As ocorrências precisarão de números únicos atribuídos a cada uma elas.

2. Crie uma tabela usando os campos de dados listados no Descrição dos dados digitais usando os dados encontrados no exemplo acima registado por: Fr. Lütken.
3. Use a tabela para fornecer as suas respostas e enviar a folha de cálculo.

Nos exemplos utilizados, as ocorrências individuais nem sempre contêm dados para completar todas as colunas da tabela.

Exercício 3

Gestão de dados

Os dados foram agora compilados num formato de folha de cálculo pelos voluntários da Sociedade Ornitológica de Copenhaga. Assumindo o papel do Curador Ornitológico no Departamento de Aves, foi-lhe atribuída a responsabilidade dos problemas de qualidade dos dados no conjunto de dados.

Através da georreferenciação retrospectiva, as coordenadas foram adicionadas ao conjunto de dados juntamente com a localidade, mas nenhuma outra geografia ao nível superior. Uma vez que todas as observações foram feitas na Dinamarca, o continente e o país podem facilmente ser acrescentados. Além disso, somente o nome científico foi fornecido. Os níveis superiores da taxonomia podem ser obtidos usando software como o OpenRefine. Também está consciente de que existem erros tipográficos que foram cometidos pelos digitalizadores.

1. Link de download UC3-DL-3-ForCleaning.zip. (45 KB)
2. Identificar e corrigir quaisquer anos inválidos.
3. Verificar e corrigir a taxonomia.
4. Verificar se as coordenadas estão corretas para as duas localizações fornecidas. Corrija as que não estão. As coordenadas devem estar no formato decimal.
5. Adicione quaisquer dados para elementos ausentes que possam ser derivados usando os dados disponíveis
6. Lembre-se de manter as informações originais fornecidas e documentar suas alterações e assunções como parte dos registos individuais e metadados.
7. Use a tabela para fornecer as suas respostas e enviar a folha de cálculo.

o conjunto de dados deve conter apenas os anos no intervalo 1883-1939

Exercício 4

Publicação de dados

Para este exercício, irá desempenhar o papel da pessoa responsável pela publicação online dos dados limpos através da rede GBIF. Foi fornecido um ficheiro multimédia e um ficheiro de histórico de identificação que deverá ser publicado juntamente com as observações. O membro da equipa

responsável pela qualidade dos dados forneceu conjuntos de dados limpos para serem publicados.

1. Link de download UC3-DL-4-ForPublication.zip. (65 KB)
2. Utilize a instalação do IPT fornecida anteriormente para publicar o conjunto de dados fornecido.
3. Use a folha de exercícios para fornecer as suas respostas e ligar ao conjunto de dados publicado.

Tarefas finais

Como tarefas finais, irá completare submeter o Caso de Uso II e o Caso de Uso III para avaliação.

CASO DE USO II

Existem duas opções para o CASO DE USO II (Espécie invasora OU avistamentos de Lepidoptera). Só precisa de selecionar um deles para a sua atividade.

Ficheiros necessários para submissão:

- folha de exercícios concluída (é aceitável MS Word Doc ou similar)
- tabela de captura de dados (é aceitável MS Excel, csv, txt ou similar)
- conjunto de dados limpo/padronizado (é aceitável MS Excel, csv, txt ou similar)

CASO DE USO III

Há apenas uma opção para USE CASE III (Aves a partir de literatura).

Ficheiros necessários para submissão:

- folha de exercícios concluída (é aceitável MS Word Doc ou similar)
- tabela de captura de dados (é aceitável MS Excel, csv, txt ou similar)
- conjunto de dados limpo/padronizado (é aceitável MS Excel, csv, txt ou similar)

Inclua o número de caso de uso e do exercício, juntamente com seu nome em todos os ficheiros enviados. Por exemplo, Russell-UC2-IS-exercise-sheet.docx, Russell-UC2-IS-2.xlsx, Russell-UC2-IS-3.xlsx. **Todos os ficheiro devem ser submetidos em Inglês**. Entre em contato com training@gbif.org se tiver alguma dúvida.

Submissão do trabalho

Os trabalhos podem ser enviados a partir da versão online (HTML) do curso.

Avaliação do curso

Conclua a avaliação do curso

Documentação chave

As seguintes referências fornecem mais detalhes sobre os tópicos abordados neste curso. Todos os links abrem em uma nova janela/guia.

Darwin Core

- Darwin Core Terms: A quick reference guide
- Simple DarwinCore
- Darwin Core Questions & Answers
- Darwin Core Extensões registadas no GBIF

Publicação de dados

- Guia rápido de publicação de dados através do GBIF
- Como publicar dados de biodiversidade através de GBIF.org
- Torne-se um publicador de dados com GBIF
- Melhores Práticas para a Publicação de Dados de Biodiversidade de Avaliações de Impacto Ambiental + Secretariado GBIF & IAIA: Associação Internacional para Avaliação de Impacto (2020).
- https://doi.org/10.35035/doc-b8hq-me03[Guia para empresas privadas tornarem-se publicadores de dados através de GBIF: Documento modelo para apoiar o processo de autorização interno para se tornar um publicador do GBIF + Rui Figueira, Pedro Beja, Cristina Villaverde, Miguel Vega, Katia Cezoish, Tainan Messina, Anne-Sophie Archambeau, Rukaya Johaadien, Dag Endresen & Dairo Escobar (2020).
- Publishing DNA-derived data through biodiversity data platforms + Anders F. Andersson, Andrew Bissett, Anders G. Finstad, Frode Fossøy, Marie Grosjean, Michael Hope, Thomas S. Jeppesen, Urmas Kõljalg, Daniel Lundin, R. Henrik Nilsson, Maria Prager, Cecilie Svenningsen & Dmitry Schigel (2020).
- Classes de conjuntos de dados suportados pelo GBIF
- GBIF data quality requirements for publishing
- licenciamento de dados do GBIF
- Checklist core templates
- Occurrence core templates
- Sampling event core templates

- Melhores práticas para amostragem de dados de eventos
- Partilhando imagens, sons e vídeos no GBIF
- Data papers
- Documentos de dados publicados

Publicação de dados: IPT

- The GBIF Integrated Publishing Toolkit: Facilitando a Publicação de Dados de Biodiversidade Eficiente na Internet + Robertson et al. (2014)
- Instalar o IPT ou não instalar o IPT
- IPT data hosting centres
- Instalação IPT / Configurar webinar
- Instalando o vídeo IPT
- IPT in Practice demo video

Digitalização

- iDigBio Digitization Resources
- iDigBio Collections Digitization Workflows
- iDigBio Digitization Workflows e Protocols
- iDigBio specimen image capture guide
- Guia de 10 etapas para gerir imagens com seus dados de biodiversidade

GBIF

- O que é o GBIF
- Plano Estratégico
- Torne-se um membro
- Science Review
- Estabelecendo um efetivo Nó participante do GBIF: Conceitos e considerações gerais + Secretariado do GBIF (2019).

Georreferenciação

- Melhores Práticas de georeferenciação + Arthur D. Capítulo e John R. Wieczorek (2020).
- Guia rápido de georeferenciação + Paula F. Zermoglio, Arthur D. Chapman, John R. Wieczorek, Maria Celeste Luna e David A. Bloom (2020).
- Manual de cálculo de georeferenciação + David A. Bloom, John R. Wieczorek & Paula F. Zermoglio (2020).

- Recursos georeferenciados

Espécies Invasoras

- GRISS - Registo Global de Espécies Introduzidas e Invasoras
- TrIAS - Rastreio de espécies alienígenas invasoras

Living Atlases

- Living Atlases
- Documentação técnica chave ALA

Diversos

- Guia VertNet para abrir ficheiros de texto em Excel
- Guia de licenciamento de dados Vertnet

OpenRefine

- OpenRefine documentation
- OpenRefine regular expressions
- Guía para la limpieza de datos sobre biodiversidad con OpenRefine + Paula F. Zermoglio, Camila A. Plata Corredor, John R. Wieczorek, Ricardo Ortiz Gallego & Leonardo Buitrago (2021).
- Usando o Google Refine e bancos de dados taxonómicos (EOL, NCBI, uBio, WORMS) para limpar dados desorganizados + + iPhylo blog post. Rod Page 2012.
- Reconciling author names using Open Refine and VIAF + iPhylo blog post. Rod Page 2013.
- Validating scientific names with the GBIF Portal web service API + Guest post was written by Gaurav Vaidya, Victoria Tersigni and Robert Guralnick 2013.
- iDigBio Cleaning data with OpenRefine + iDigBio 2013.
- Have We Got the Names "Right"? + Canadensys 2014.
- Cleaning data with OpenRefine + Desmet and Brosens 2016 TDWG.
- EasyOpen Redlist + Querying the IUCN Red List, using a species list, OpenRefine, and some pre-written code. Olly Griffin July 2019.

Planeamento/Colaboração

- Agile
- O que é SCRUM
- SCRUM Framework
- metodologia Kanban.

- Guia de Scrum
- GitHub

Qualidade

- Princípios da Qualidade dos Dados + Capítulo Arthur 2005.
- Princípios e Métodos de Limpeza de Dados: Espécies Primárias e Dados de Ocorrência de espécies + Arthur Chapman 2005.
- Tenha cuidado com as datas no Excel + DataOne 2014.
- Codificação de caracteres para iniciantes
- Guia MVZ para Localidades no Campo Notas

Espécie sensível

- Melhores Práticas Actuais para Generalização de Ocorrências de Espécies Sensíveis + Arthur D. Capítulo 2020.

Taxonomia

- conjuntos de dados de checklists do GBIF e lacunas de dados
- GBIF Labs - Nomes Parser
- GBIF Labs - Species Matching
- Global Names Resolver
- Marine Name Matching Strategy for taxonomic quality control
- Nomenmatch

Glossário

ALA

Atlas of Living Australia. O nó australiano do GBIF, que desenvolveu um portal de dados de código aberto, agora amplamente utilizado dentro da comunidade GBIF & parceiros, para portais nacionais de biodiversidade.

API

Interface de Programação de Aplicação. Um conjunto de métodos de comunicação claramente definidos entre vários componentes de software.

BID

Informações de biodiversidade para o desenvolvimento. Um projecto financiado pela UE coordenado pelo GBIF com o objectivo de aumentar a capacidade de mobilização de dados nas regiões de África, das Caraíbas e do Pacífico.

BIFA

Fundo de Biodiversidade para a Ásia.

Licenças CC

Creative Commons. Esta é uma série de licenças criadas pela organização Creative Commons que permitem a partilha e reutilização da criatividade e do conhecimento através da disponibilização de ferramentas legais gratuitas. Três delas podem ser atribuídas a conjuntos de dados partilhados no GBIF: CC0, CC BY e CC BY-NC.

Vocabulário controlado

Este é um conjunto restrito de termos que são usados como valores possíveis para um determinado campo. Podemos considerá-la como uma lista de pesquisas ou uma lista para um determinado campo. Por exemplo, o campo basisOfRecord DwC deve conter apenas um desses valores: "PreservedSpecimen", "FossilSpecimen", "LivingSpecimen", "HumanObservation", "MachineObservation". Podemos dizer que a lista de valores é um vocabulário controlado para esse domínio.

DwC

Darwin Core é um padrão de dados sobre biodiversidade, mantido por TDWG & amplamente utilizado na comunidade e parceiros do GBIF. Trata-se de um conjunto de termos normalizados (ou nomes de campos) e das suas definições, que são utilizados para partilhar a informação sobre biodiversidade.

DOI

Identificador Digital de Objecto. Um identificador persistente ou utilizado para identificar unicamente objectos. Os DOIs são amplamente usados principalmente para identificar informações acadêmicas, profissionais e governamentais, como artigos de periódicos, relatórios de pesquisa e conjuntos de dados e publicações oficiais.

DwC-A

Darwin Core Archive. Um ficheiro comprimido (compactado) contendo todas as informações necessárias para partilhar com o GBIF, para um determinado recurso. Cada ficheiro contém três tipos de ficheiros:

1. os dados actuais, em um ou mais fucheiros de texto: occurrence.txt/event.txt/measurmentoffact.txt etc
2. um ficheiro de mapeamento: rtf.xml
3. um ficheiro de metadados (EML): eml.xml Quando você publica usando o IPT, cria um ficheiro Darwin Core, que é partilhado com GBIF. Além disso, quando descarregar dados do portal GBIF, também poderá escolher um formato DwC-A.

GUID

IDentificador globalmente único

IPT

Integrated Publishing Toolkit. É uma aplicação web de código aberto e livre (software) para publicação de dados de biodiversidade. O próprio software vive em um servidor (seja na sua

instituição ou em qualquer outro lugar) que deve ter acesso à internet 24/7. É usado para criar e lidar com ficheiros do Darwin Core Archive que podem ser partilhados e usados por qualquer pessoa, incluindo o GBIF.

Empréstimo

No contexto das colecções de história natural, este é o procedimento de concessão de empréstimos de espécimes entre instituições.

LSID

Identificador de Ciências da Vida. São identificadores persistentes e exclusivos para objetos biológicos.

Publicação de Dados

No que se refere ao GBIF, dispomos de uma definição muito específica de publicação de dados. Refere-se a tornar os conjuntos de dados de biodiversidade acessíveis ao público e detectáveis, de forma padronizada, através de um ponto de acesso, tipicamente um endereço da Web (uma URL).

Recurso

Um Recurso é o termo coletivo usado para referir a um conjunto de dados específico e seus metadados uma vez que tenha sido carregado para uma instância do IPT.

TDWG

Grupo de Trabalho de Bases de Dados Taxonómicos, agora renomeado Padrões de Informações de Biodiversidade.

URN

Número Uniforme de Recursos

UUID

Identificador Universal Único

Apêndice: Artigos de dados

Um artigo de dados é um documento publicado em periódicos revisados por pares e que descreve um conjunto de dados. É preciso esforço para preparar, selecionar e descrever os dados. Os artigos de dados permitem o reconhecimento por esse esforço por meio de um artigo científico. Não cobrimos como criar artigos de dados neste curso, no entanto, como uma atividade opcional, pode assistir a este vídeo (51:51) apresentado por Lizanne Roxburgh. Neste vídeo, aprenderá mais sobre como publicar artigos de dados. Se não conseguir assistir ao vídeo incorporado, pode usar o download it locally. (MP4 - 99.2 MB)

https://vimeo.com/265350948 *(Vimeo video)*

Pode ler mais sobre artigos de dados no GBIF.org.

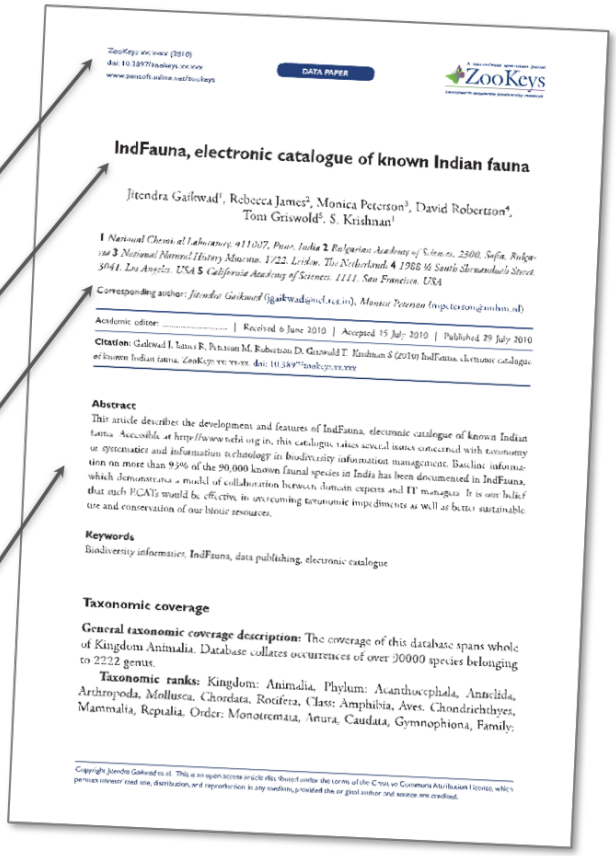
What it is: A scholarly publication of searchable metadata – a document describing a dataset, or a group of datasets

DOI: indexation and citation

Promote and publicize existence of data

Provide scholarly credit to data publishers through citable journal publications

Describe the data in a structured human-readable form



DATA PAPER

ZooKeys

IndFauna, electronic catalogue of known Indian fauna

Jitendra Gaikwad[1], Rebecca James[2], Monica Peterson[3], David Robertson[4], Tom Griswold[5], S. Krishnan[1]

Abstract

This article describes the development and features of IndFauna, electronic catalogue of known Indian fauna. Accessible at http://www.nbfi.org.in, this catalogue raises several issues concerned with taxonomy or systematics and information technology in biodiversity information management. Baseline information on more than 95% of the 90,000 known faunal species in India has been documented in IndFauna, which demonstrates a model of collaboration between domain experts and IT managers. It is our belief that such FCATs would be effective in overcoming taxonomic impediments as well as better sustainable use and conservation of our biotic resources.

Keywords

Biodiversity informatics, IndFauna, data publishing, electronic catalogue

Taxonomic coverage

General taxonomic coverage description: The coverage of this database spans whole of Kingdom Animalia. Database collates occurrences of over 90000 species belonging to 2222 genus.

Taxonomic ranks: Kingdom: Animalia, Phylum: Acanthocephala, Annelida, Arthropoda, Mollusca, Chordata, Rotifera, Class: Amphibia, Aves, Chondrichthyes, Mammalia, Reptilia, Order: Monotremata, Anura, Caudata, Gymnophiona, Family:

Veja os links abaixo para ver um artigo de dados tal como aparece no IPT, no GBIF.org e na revista cientifica. Qualquer deles está interligado.

- Revista: https://doi.org/10.3897/BDJ.2.e4244
- GBIF: https://www.gbif.org/dataset/3c9e2297-bf20-4827-928e-7c7eefd9432c
- IPT: http://130.120.204.55:8080/ipt/resource.do?r=harvestmen_of_french_guiana

Apêndice: Soluções

Este apêndice contém as respostas e informações adicionais para todos os questionários de revisão. Além disso, esta secção contém uma solução sugerida para o USE CASE I.

Soluções da revisão dos fundamentos

Para a afirmação fornecida, insira o termo correcto (base de dados, língua da base de dados, programa da base de dados)*

- combina e apresenta funções e recursos para manipular dados, juntos em uma interface unificada + **programa de base de dados**

- coleção de dados estruturada e organizada e/ou informações armazenadas em um computador + **base de dados**
- a maneira pela qual um humano se comunica com um computador + **linguagem de base de dados**

Se abrir um ficheiro de dados e vir o seguinte, de qual problema suspeitaria?

`□tre, ou ne pas □tre, c□est l□ la question.`

- A codificação incorreta foi utilizada para abrir o ficheiro

Para o software fornecido, insira o tipo de software (captura de dados, gestão de dados, limpeza de dados, publicação de dados).

- Integrated Publishing Toolkit (IPT) + **publicação de dados**
- Specify + **captura de dados** E **gestão de dados**
- iNaturalist + **captura de dados**
- OpenRefine + **limpeza de dados**

Para o exemplo dado, insira o tipo de dados correcto (binário, booleano, float, inteiro, inteiro longo, texto, texto não estruturado)

- 1236975 + **inteiro longo**
- 01101111 + **binário**
- Caminhamos 5 milhas pelo caminho a oeste desde a estação dos correios no centro da cidade. Depois, seguimos 2 milhas a norte num caminho de terra até o rio. Em seguida, continuamos a oeste ao longo do rio por mais 5 milhas. + **texto não estruturado**
- 1024 + **integer**
- 29,0 + **float**
- Sim/Não + **booleano**
- 6 coelhos foram observados + **texto**

Qual destes termos descreve um nome de "campo/coluna"

- Atribuído
- Identificando
- Único

Quais destes termos descreve uma "etiqueta de campo"

- Descrição
- Legível
- Interface do utilizador

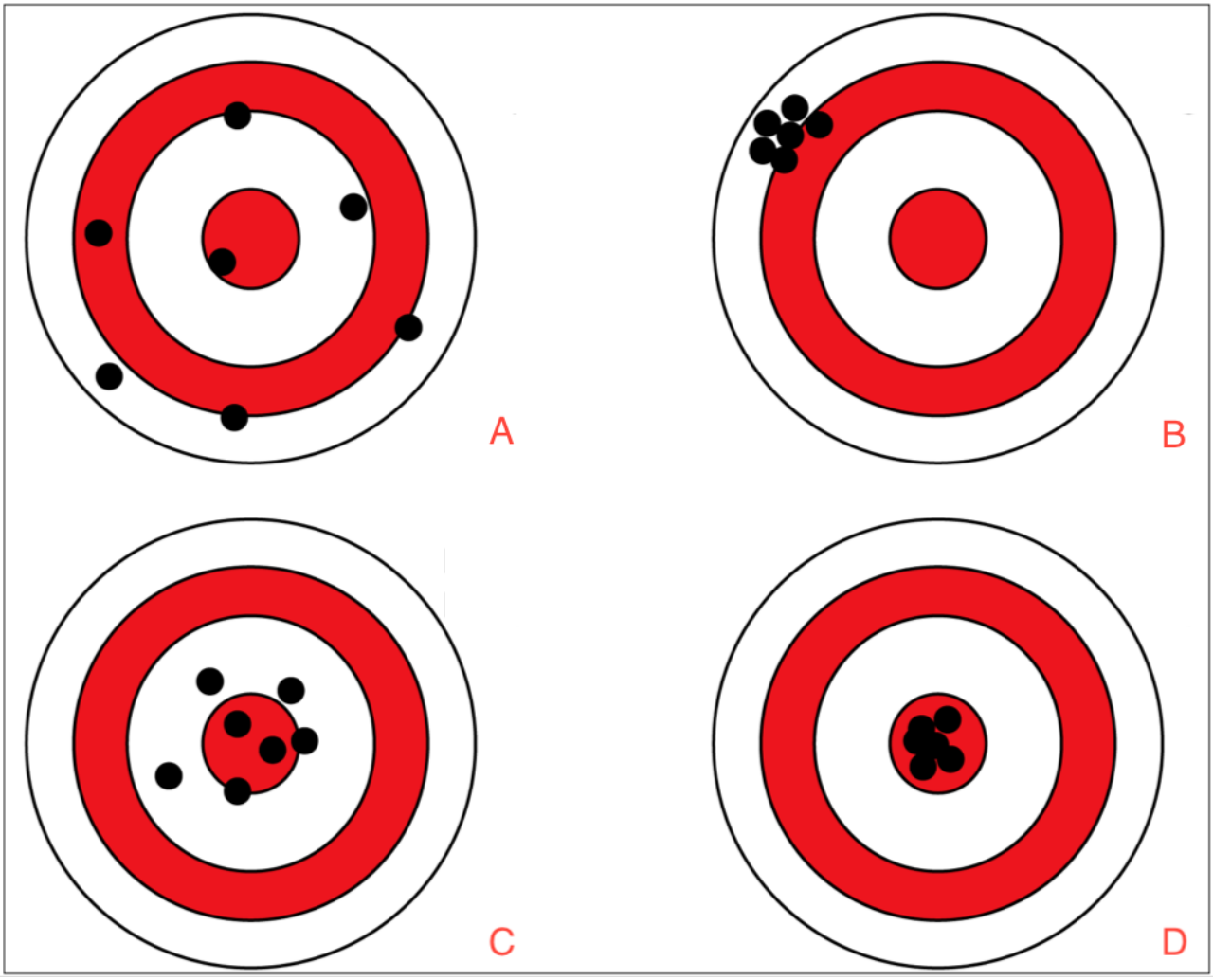
Para cada afirmação, insira a estrutura correta (linha, coluna, tabela)

- Todos os dados se referem a um conceito SINGLE. + **tabela**
- Um atributo tem o mesmo tipo de campo/dado para cada registo. + **coluna**
- Atributos de um registo SEMPRE permanecem juntos. + **linha**

Quem determina a aptidão para o uso de seus dados

- Os utilizadores dos dados para investigação ou educação

Para as instruções fornecidas, insira a imagem correspondente. (A, B, C, D)



- Alta exatidão, baixa precisão + **C**
- Baixa exatidão, alta precisão + **B**
- Alta exatidão, alta precisão + **D**
- Baixa exatidão, baixa precisão + **A**

Identifique as relações de dados em que o conjunto de dados B precisa ser mesclado ao conjunto de dados A (0:1, 1:0, 1:1, 1:∞, ∞:1, ∞:∞). Nem todas as relações são utilizadas

- O campo de coletor existe em ambos os conjuntos de dados A e B + **1:1**
- O campo país só existe no conjunto de dados B + **0:1**
- O campo nome existe no conjunto de dados A, mas o conjunto de dados B contém os campos Primeiro Nome e Sobrenome + **1:∞**

- Campo de ID existe em ambos os conjuntos de dados A e B + **1:1**
- Elevação existe no conjunto de dados A, mas não no conjunto de dados B + **1:0**
- A data existe no conjunto de dados A, mas o Dia, o Mês e o Ano são campos separados no conjunto de dados B + **1:∞**

Metadados são importantes porque (selecione as declarações VERDADEIRAS)

- permite que os utilizadores determinem se um conjunto de dados é adequado para o seu uso.
- permite que saiba sob que termos legais a reutilização de dados é permitida.

Planificar as soluções de revisão

Qual é a ordem dos cinco Grupos do Processo PMBoK?

- Iniciação, Planeamento, Execução, Monitorização e Controle, Encerramento

Saiba mais: https://quizlet.com/306742513/1-introduction-pmbok-guide-6th-edition-flash-cards/

Quais são os tipos de entrega?

- Especificado - SIM
- Implícito - SIM
- Estimado - NÃO
- Direto - SIM
- Indireto - SIM
- Sugestões - NÃO

O que é um estrangulamento?

- um bloqueio que atrasa o desenvolvimento ou o progresso - SIM um espaço onde algo ou alguém está desaparecido - NÃO, ISTO É UMA LACUNA
- um problema ou uma situação que impede alguém de fazer algo, ou que torna algo impossível. NÃO, ISTO É UMA BARREIRA

Quais são exemplos de tarefas de mobilização?

- Afiliação - NÃO, Este é um Tipo de Recurso
- Publicação - SIM
- Obtenção de imagens - SIM
- Georreferenciação - SIM
- Maior consciencialização pública - NÃO, esse é um objetivo implícito.

Soluções de revisão dos dados de captura

Qual(ais) o(s) tipo(s) de conjunto de dados que escolheria para uma coleção de ictiologia?

- ocorrência + A maioria das vezes, espécimes de bases de dados de colheita são partilhados como dados de ocorrência. Cada ocorrência (espécime ou grupo de espécimes) tem o seu próprio identificador único (às vezes derivado do seu número de catálogo na coleção fonte) e os campos Darwin Core usados para partilhá-los através do GBIF descrevem cada espécime: nome científico, a data em que foi recolhido no campo, quem o recolheu e/ou identificou, onde, etc. Cada coleção pode ter mais do que um espécime de uma mesma espécie, desde que cada espécime seja identificado por um identificador único.
- checklist + Também é possível criar e partilhar uma checklist taxonómica derivada de uma base de dados de coleção; neste caso, é recomendado partilhar a checklist como um conjunto de dados taxonómicos, com a lista de ocorrências (espécimes) associada, usando o Occurrence core como uma extensão do Taxon Core no GBIF IPT.

Qual(ais) o(s) tipo(s) de conjunto de dados que escolheria para uma lista de espécies invasovas?

- ocorrência + Alguns publicadores de dados partilharão conjuntos de dados de ocorrência provenientes de estudos ou de programas de acompanhamento de espécimes de alguma espécie invasora; quando os dados se concentram em indivíduos, em vez das espécies invasoras, em geral, podem ser partilhados como dados de ocorrência.
- checklist + Espécies invasovas podem ser rastreadas e monitorizadas a diferentes escalas (regional, nacional, temático...); uma vez que este tipo de conjunto de dados se concentra mais na espécie e sua distribuição num determinado âmbito geográfico, eles são partilhados principalmente como conjuntos de dados taxonómicos através do GBIF (ver resultados de pesquisa GRIIS).

Qual(ais) o(s) tipo(s) de conjuntos de dados que escolheria para a flora e fauna de um estudo de impacto ambiental?

- ocorrência + Os dados são registados pelos naturalistas no campo e podem ser partilhados como conjuntos de dados de ocorrência simples.
- evento de amostragem + Eles também podem ser partilhados como conjuntos de dados de eventos se houver protocolos padronizados (como pontos de vegetação, transectos, armadilhas...) para colher os dados.

Qual(ais) o(s) tipo(s) de conjuntos de dados que escolheria para os dados de rastreio de pássaros?

- ocorrência + Estes dados são partilhados como conjuntos de dados de ocorrência: idealmente, cada pássaro é identificado com o seu organismID, e cada ocorrência (GPS ping) tem a sua própria ocorrência, que é útil para rastrear os diferentes locais GPS do mesmo pássaro no âmbito do programa de rastreamento ou projeto. (Veja example)

Qual(ais) o(s) tipo(s) de conjuntos de dados que escolheria para dados de armadilhas de insetos?

- ocorrência + Embora tais dados possam ser partilhados como conjuntos de dados de ocorrências simples, é melhor se forem partilhados como conjuntos de dados de eventos, onde a localização, identificador e conteúdo de cada armadilha podem ser melhor detalhados.
- evento de amostragem + Armadilhas de insectos (bem como outras armadilhas, tais como armadilhas pitfall, armadilhas malaise...) são normalmente usadas em programas de monitorização para verificar a presença (ou ausência) de algumas espécies e/ou avaliar a sua abundância específica. Usar o campo "eventID" para identificar cada armadilha permite aos utilizadores obter todos os espécimes colhidos dentro de cada armadilha. A mesma lógica aplica-se a outros protocolos de campo, como transectos, pontos, câmaras remotas, etc. usando o núcleo do Evento (Event core) em vez do núcleo de Ocorrência (Occurrence core), poderá partilhar muito mais informações sobre o contexto da colheita de dados, e permitir que os utilizadores compreendam melhor (e possam até replicar) o seu trabalho.

Qual(ais) o(s) tipo(s) de base de dados você escolheria para a gestão de dados do parque nacional?

- ocorrência + registo de indivíduos de espécies
- checklist + É importante saber quantas espécies estão presentes no perímetro do parque/reserva e o seu estado de conservação.
- evento de amostragem + verifique e rastreie as populações

Qual(ais) o(s) tipo(s) de conjunto de dados que escolheria para um bioblitz de ciência cidadã?

- ocorrência + conjuntos de dados de Bioblitz são partilhados principalmente como conjuntos de dados de ocorrência.
- evento de amostragem + Dependendo do programa de ciência cidadã, podem ser usados protocolos de amostragem específicos por voluntários, e neste caso, os dados podem ser partilhados como um conjunto de dados do eventos.

Qual(ais) o(s) tipo(s) de conjunto de dados que escolheria para uma lista regional de espécies?

- checklist + Listas de espécies geográficas ou temáticas são frequentemente usadas para partilhar informações sobre as espécies presentes numa determinada área; na maior parte das vezes, essas listas também mencionam a distribuição de cada espécie, bem como o seu estado de conservação nesta área. As listas regionais de espécies podem proporcionar uma visão útil da biodiversidade e dos habitats de uma região, e precisa ser partilhado como conjuntos de dados taxonómicos, com ou sem ocorrências associadas.

Soluções de revisão da gestão de dados

Por que é melhor limpar os seus dados?

- para torná-los o mais possível adequados ao uso
- para atingir as suas metas de qualidade de dados

Deve sempre procurar gerir e publicar dados com a maior qualidade possível. Isto irá melhorar o seu trabalho diário (é mais fácil de trabalhar com dados organizados e limpos), assim como o trabalho de

potenciais reutilizadores dos seus dados, que precisam de compreendê-los e confiar na sua fonte antes de usá-los.

Como deve organizar o seu fluxo de trabalho de limpeza de dados?

- peça aos seus colegas o conhecimento especializado
- trabalhe ao nível institucional para harmonizar fluxos de trabalho de qualidade de dados

Ninguém sabe tudo sobre os dados da biodiversidade; deve procurar ajuda e aconselhamento com os seus colegas ou outras pessoas conhecidas. e certifique-se de aplicar as boas práticas recomendadas pela sua instituição enquanto limpa os seus dados.

Qual é o melhor:

- prevenir a ocorrência de erros
- corrija erros assim que os encontrar na sua base de dados ou tabela

A melhor maneira de evitar a propagação de erros nos seus dados é evitá-los no início do processo de colheita/registo de dados.

É claro que os erros são inevitáveis, portanto também deve limpá-los assim que os encontrar, e documentar o processo de limpeza.

Se não tem tempo ou recursos para limpar os seus dados corretamente, é melhor esperar conseguir fazê-lo em vez de publicar dados errados que podem confundir as pessoas.

De quem é a responsabilidade da qualidade dos dados?

- De todos os participantes na gestão dos dados

Cada pessoa envolvida no seu fluxo de trabalho de gestão de dados é pelo menos parcialmente responsável pela sua qualidade, desde os técnicos de campo até ao(s) gerente(s) da base de dados.

As pessoas que mais tarde possam utilizar os seus dados, podem informá-lo de qualquer erro que permaneca nos seus dados e devem utilizá-los de forma responsável na sua própria investigação, mas a qualidade inicial dos dados não é responsabilidade delas.

O GBIF pode realizar verificações automáticas nos seus dados (por exemplo, detecção de valores ausentes, outliers geográficos, nomes científicos desconhecidos), mas não deve ser responsabilizados pelos erros que ocorreram anteriormente no processo de gestão de dados.

Quais as ferramentas que podem ser usadas para limpar os seus dados?

- Excel e outras ferramentas de gestão de folhas de cálculo
- OpenRefine
- O seu software de base de dados
- Ferramentas on-line como o Scientific Names Resolver ou Google Maps

Todos os tipos de ferramentas podem ser usadas para limpar os seus dados, mas deve identificar

quais responderão às suas necessidades em termos de resolução taxonómica, georeferenciação, exclusão de duplicados e assim por diante. Pode encontrar Ferramentas úteis listados na seção de gestão de dados.

Soluções de revisão de dados

O que significa publicação de dados no contexto do GBIF?

- Tornar o(s) seu(s) dado(s) de biodiversidade acessível ao público e detectável em um formato padronizado

A publicação de dados através do GBIF significa tornar os seus conjuntos de dados de biodiversidade acessíveis ao público num formato padronizado (na maioria das vezes, Darwin Core), para que possa ser descoberto e reutilizado por outras pessoas.

O que é um IPT?

- uma ferramenta que o ajuda a publicar os seus dados através do GBIF
- uma ferramenta que ajuda a produzir um Artigo de Dados

O IPT (Integrated Publishing Toolkit) é um software escrito em Java que permite carregar e publicar dados através do GBIF. Não deve ser utilizado como instrumento de gestão de dados ou de limpeza de dados.

O IPT também pode ajudá-lo no processo de escrever e enviar um artigo de dados, graças ao ficheiro EML ele gera automaticamente quando preenche os metadados do seu recurso de dados.

Quais licenças e dispensas Creative Commons são recomendadas pelo GBIF para publicação de dados?

- CC0, CC-BY e CC-BY-NC

As licenças e dispensas Creative Commons recomendadas para publicar os seu(s) dados(s) no GBIF são CC0, CC-BY e CC-BY-NC. São licenças e/ou dispensas amplamente reconhecidas que se alinham com os requisitos internacionais de dados abertos para partilhamento e reutilização de dados.

Por favor, note que deve escolher apenas a dispensa/licença CC-BY ou CC-BY para o(s) seu(s) dados(s) relacionados com o BID.

Quais são os três núcleos (cores) que pode escolher para um recurso IPT?

- Núcleo de Ocorrência, Núcleo de Taxon, Núcleo de Eventos

Pode escolher um dos três Núcleos seguintes para cada um dos seus recursos IPT: Núcleo de Ocorrência, de Taxon ou de Evento.

O padrão Darwin Core também permite que associe extensões ao núcleo escolhido, como SimpleMultimedia ou MeasurementOrFact.

Os metadados são preenchidos numa secção separada do IPT sendo partilhados usando o padrão

EML, e não Darwin Core (que é usado apenas para dados).

Quantas extensões pode ter um conjunto de dados?

- tantas quantas necessárias

Uma vez que você escolheu um "core" para o seu recurso IPT, pode adicionar extensões de Darwin core a ele. Você pode adicionar apenas uma ou várias extensões, dependendo do tipo de "core" que escolheu e quais extensões são compatíveis com ele.

As extensões não são obrigatórias (você pode publicar um conjunto de dados sem qualquer extensão), mas pode ser útil se quiser partilhar informações adicionais que você não pode mapear com o "core" escolhido.

Sugestão de solução do Use Case I

suggested solution (PDF 144 KB)

Agradecimentos

Design do curso e instrução

O sucesso deste curso depende em grande parte do apoio prestado aos participantes pelos mentores da rede GBIF. Visite a página do GBIF em mentoria de melhoria de capacidade para ler mais sobre esses indivíduos e suas contribuições.

Os seguintes indivíduos são reconhecidos por suas contribuições significativas para o design do projecto, materiais e instruções do curso:

- Nestor Beltran*
- David Bloom
- Katia Cezón*
- Dag Endresen
- Alberto González-Talaván*
- Sharon Grant*
- Marie-Elise Lecoq
- Sophie Pamerlon*
- Nicolas Noé*
- Mélianie Raymond*
- Laura Anne Russell*
- John Wieczorek
- Paula Zermoglio

*Criadores do curriculum

Reconhecimento especial a Arthur Chapman pela reutilização de seus materiais sobre a qualidade dos dados.

Tradutores

Francês

- Maxime Coupremanne
- Jaures Gbètoho
- Marie Grosjean
- Patricia Mergen
- Sophie Pamerlon
- Andry Jean Marc Rakotomanjaka
- Y. Sabastian Wirsiy

Português

- Rui Figueira
- Clara Baringo Fonseca
- Keila Elizabeth Macfadem Juarez
- Tainan Messina

Espanhol

- Leonardo Buitrago
- Victor Chocho
- Camila Plata
- Anabela Plos
- William Ulate
- Paula Zermoglio

Recursos

- A plataforma online é fornecida por PlayingCards.io. Muito apreço e obrigado a Jwalant Patel e Eric Ma por terem encontrado e ajudado a criar as mesas de jogos on-line e a Kate Webbink pela competência artística.
- Os ícones utilizados neste curso são feitos por Freepik de www.flaticon.com
- OpenRefine
- Integrated Publishing Toolkit

Suporte a recursos

- Belgium Biodiversity Platform
- GBIF France
- GBIF Norway
- GBIF Spain
- SiB Colombia
- The Field Museum
- VertNet

Cólofon

Citação sugerida

GBIF Secretariat (2021) GBIF Biodiversity Data Mobilization Course. 12th edição. GBIF Secretariat: Copenhagen. https://doi.org/10.35035/ce-c6cr-6w42. [Data do curso.]

Colaboradores

O *Curso de Mobização de Dados do GBIF* foi originalmente desenvolvido como parte do Biodiversity Information Development, um programa financiado pela European Union. O currículo original foi criado por Nestor Beltran, Sharon Grant, Nicolas Nóe, Sophie Pamerlon, Alberto González-Talaván, Mélianie Raymond, Laura Anne Russell and Katia Cezón, com contribuições adicionais de treinadores, mentores e estudantes do GBIF.

Licença

GBIF Biodiversity Data Mobilization Course is licensed under Creative Commons Attribution 4.0 Unported License.

URL Persistente

https://doi.org/10.35035/ce-c6cr-6w42

Controlo de Documentos

12a edição, Maio 2021