

Oficina de Trabalho

Estratégia Nacional de Monitoramento da Biodiversidade

Ministério do Meio Ambiente
Instituto Chico Mendes de Conservação da Biodiversidade
Ministério da Ciência e Tecnologia

Brasília, DF - 23 a 25 de agosto de 2010

Palestra **Estratégias para armazenamento de dados biológicos e ambientais: experiências do PELD/LTER e PPBio do Brasil** (Flávia Costa)

1. Bancos de dados são específicos para cada tipo de dado
2. Pode ser relativamente simples montar e manejar um banco de dados quando o tipo de dado é simples e não muda com o tempo – p. Ex, GENBANK ou bancos de dados de coleções de espécimes
3. Existem muitas iniciativas para armazenar dados de coletas biológicas, quando estas são centradas no indivíduo coletado, e não faz sentido inventar novos sistemas para isso. Várias coleções (Herbários, Museus) tem feito esforços para colocar suas informações em bancos padronizados e que foram adotados por consenso (BRHAMS e Specify p.ex). Os esforços atuais devem ser para capacitar os gerentes de coleções a usar estes sistemas e em interligar as coleções on line, como já tem sido feito em alguns lugares (dar exemplo do BRHAMS on line e dos trabalhos do CRIA)
4. Outros tipos de dados (p.ex os dados de inventários biológicos) são mais variáveis e precisam de abordagens mais específicas. É muito difícil prever todos os possíveis tipos de dados e estrutura das amostragens a priori, e portanto, definir uma estrutura de banco de dados que funcione para tudo (pode citar e mostrar trechos do trabalho “como manejar dados mal”)
5. Dada a complexidade dos dados de inventários biológicos e dados ambientais associados, muitos esforços de desenvolver sistemas de armazenagem para eles tem sido frustrados – na maior parte das vezes o sistema é criado mas fica vazio, pois os usuários tem dificuldade de usá-lo ou ele não se adapta às mudanças
6. Reconhecendo a limitação de soluções do tipo “omnibus” (que serve pra tudo) vários profissionais da área de manejo de dados tem proposto que a criação de repositórios de dados é uma solução melhor para o problema de armazenar dados com qualidade. A solução pode ser vista como definitiva ou como um passo intermediário até a migração para bancos de dados.
7. Um repositório não é um “saco” ou “armário” onde os dados são jogados e guardados sem nenhuma regra. O que faz dos repositórios locais que guardam dados com qualidade é que eles tem uma grande preocupação com os metadados. Os metadados são informações que descrevem os dados e sem eles, os dados não tem nenhum valor ou utilidade, seja lá qual for o sistema de armazenagem.
8. Já foram desenvolvidas linguagens e modelos para metadados de determinadas disciplinas, e no caso de dados de inventários biológicos e dados ambientais, um bom modelo é o Ecological Metadata Language, desenvolvido por xxx e usado pelo International Long Term Ecological Research Program

9. Além de metadados bem escritos, outro requisito importante para que os dados armazenados em repositórios sejam úteis é a definição de chaves primárias suficientemente detalhadas, para permitir que os dados de diferentes estudos possam ser integrados (dar exemplos) ou transferidos para bancos de dados no futuro.
10. Um repositório de dados precisa ter um elemento fundamental para funcionar: o “filtro humano”. Embora a checagem de alguns tipos de dados possa ser automatizada, a checagem de metadados não pode, e se os metadados não tiverem sua qualidade verificada, nenhuma checagem de qualidade dos dados vai fazê-los mais úteis. Além disso, existem checagens de erros em planilhas de dados que não podem ser automatizadas e dependem de uma revisão cuidadosa de pessoas experientes, e do diálogo entre os manejadores do repositório (os filtros humanos) e os pesquisadores que geraram os dados.
11. Em suma, para um repositório de dados funcionar, é necessário:

I. Equipamentos

- a. Pessoas dedicadas a receber e checar os dados e metadados, dialogar com os geradores dos dados para corrigir os erros e finalmente disponibilizar os dados e metadados on line
- b. Servidores para armazenar os dados e metadados

II. Decisões

- a. Arquitetura da rede de recebimento e disponibilização dos dados (recomenda-se uma rede com um ponto central, mas vários pontos regionais, interligados)
- b. Variáveis mínimas para as chaves primárias das tabelas de dados (descritores de tempo, espaço, e responsáveis pela amostragem, p.ex)
- c. Adoção de algum(s) modelo de metadados

III. Capacitação

- a. Das pessoas dedicadas a gerir os dados e metadados, para serem capazes de detectar os erros e dialogar com os geradores dos dados
- b. Dos fornecedores de dados (pesquisadores e afins), para entender o que são e como escrever metadados, e para construir as tabelas de dados de forma que as informações originais não sejam sumarizadas e que as chaves primárias importantes estejam presentes