

Data standards and protocols for biological collections data

Renato De Giovanni <renato@cria.org.br>

Centro de Referência em Informação Ambiental (CRIA)

Access to biodiversity data has never been more important. Global climate change, disease spread and habitat loss are some of the issues that require access to valid and comprehensive biodiversity data to be dealt with. Significant part of this data is stored and maintained by numerous biological collections across the world, which constitute an invaluable source of information about the historic distribution and diversity of life. Providing uniform and simultaneous access to this data has always been a great challenge given its distributed and heterogeneous nature. The first biological collections network appeared in the end of the 90s, but only during the last years was there considerable progress in the amount of data available online. Since the first initiatives, different strategies involving software architecture, data standards and protocols were created and adopted by regional, thematic and global networks. Architectures ranged from centralized databases to fully distributed networks, including hybrid approaches. Distributed online queries and data harvesting techniques were used. With the growing need to access, integrate and analyze all biodiversity data, standards are becoming increasingly important. The adoption of common data standards by providers and networks leads to reduced effort, cost, time, and risk to develop and maintain software applications. In some cases, however, the diversity of standards and protocols caused interoperability issues, leading to integrating efforts in the biodiversity informatics community. The presentation will provide an overview of the main networks dealing with biological collections' data, with particular emphasis on the architecture, standards and protocols being used. New trends and developments will also be discussed.