

巨大な海洋生物多様性情報群から”正しい”データを引き出すために To extract “right” information from a huge marine biodiversity information pool

*細野 隆史¹、齋藤 秀亮¹、坪井 誠司¹

*Hosono Takashi¹, Hideaki Saito¹, Seiji Tsuboi¹

1. 海洋研究開発機構 国際海洋環境情報センター

1. Global Oceanographic Data Center (GODAC), Japan Agency for Marine-Earth Science and Technology (JAMSTEC)

近年、海洋生態学分野では国際的な観測ネットワークの構築により、類似の情報を扱うデータベースが非常に多く稼働し始めている。これらのデータベースに集約された情報は、第三者に利用され、国際的にインパクトのある研究成果を生み出している。例えば、2017年のNature誌で出版された海洋生物多様性に関する論文のほとんどはデータベースを基にしたメタ解析論文となっている。情報技術のさらなる発達とともに、生物多様性情報の増加がさらに加速し、それらを用いたデータ駆動型研究が発展していくことは間違いない。情報ソースの拡大はメタ解析の可能性を広げるが、その一方で、メタ解析に必須となるリソース間のデータ統合において複雑化をもたらし、データ抽出の作業規模を増大させる。そして、データ統合の複雑化と大規模化は必然的に、重複データ（本質的に同一の情報を持つが、表記のみ異なるデータ）が生成される危険性も増大させる。これらは、長年データ利用者である研究者が個々人の努力で解決してきた問題ではある。しかし、飛躍的な情報量の増大を考えた場合、生物多様性分野に特化したデータの統融合技術が不可欠であり、この技術の開発がこの分野の裾野を広げ情報活用を最大化する鍵となる。本発表では、データ処理に煩わされることなく、全ての利用可能なデータを使用するためにどのような技術開発が必要であるのか議論したい。

キーワード：海洋生物多様性、データベース、データ駆動科学

Keywords: marine biodiversity, database, data-driven science