Current status and challenges of open data in biodiveristy field

*Utsugi Jinbo[1]

1.National Museum of Nature and Science

Biodiversity informatics is a research field aiming to share species-level biodiversity data, including taxon names and occurrence records based on specimens or observations. Such data is indispensable for many kind of activities relevant to biodiversity, researches, plannings for biodiversity conservations, and cultural activities. Until now, global projects, such as the Global Biodiversity Information Facility (GBIF), provide frameworks to accumulate and share biodiversity data. Recently, people realized that biodiversity data should be open to facilitate and promote the data reuse. GBIF decided to change dataset licenses to the Creative Commons License, and many project have been launched to publish biodiversity data openly available (e.g. iDigBio, pro-iBiosphere, and Atlas of Living Australia). Open-access journals directly linked to biodiversity databases are also good solutions to make the data open.
In Japan, only small amounts of resources on biodiversity have been hitherto openly published. Most of them depend on individual efforts. The author has been published taxon name databases as open data by the assistance of GBIF Japan Node. In addition, some projects are also currently in progress. Linked Open Data for ACademia (LODAC) is a project to publish academic data in Linked Open Data (LOD) format. Both of taxon and occurrence data are converted and stored in LOD format, and linked to other LOD data such as DBPedia. The openly published records are used as a core component of the biodiversity LOD.
Many people still have difficulty to openly publish their data. One issue is there are no cultures and incentives to encourage open data publications. A data citation system like that for scientific papers using Digital Object Identifier (DOI), and a data paper publication framework could be solutions for this issue, though there have not been any evaluation systems in research communities. Another issue is a risk of the publication of sensitive data, that is, distribution records of endangered or commercially valuable species. The improvement of data interoperability between biodiversity and other fields is also a challenge for the future.
In this presentation, the author will show the current progress of open data activities on biodiversity information in Japan, and discuss on future challenges and collaborations.

Keywords: biodiversity informatics, open data, databases, taxonomy, linked data