

Міжнародні стандарти оцифрування природничих колекцій

В. Різун

Державний природознавчий музей НАН України (Львів)

International Standards for Digitization of Natural History Collections Data. — **Rizun, V.** — Creating of biodiversity databases according to the uniform standards developed by TDWG and cited in "Darwin Core" is recommended. According to these principles, database "Entomology" has been created in the State Museum of Natural History, NAS of Ukraine. Standard unification provides combining of different taxonomic data bases and creating the National biodiversity database in future; translation information into English provides database integrations to the International biodiversity database GBIF.

Практично у всіх музеях природничого спрямування проводилася і проводиться каталогізація наукових колекцій, а в останні десятиліття — переведення каталогів колекцій у електронний формат або ж створення баз даних. Зазвичай створюються окремі бази даних для певних систематичних груп живих організмів. Інформація, яка міститься у цих базах даних, на загал є для внутрішнього користування і для широкого наукового чи природоохоронного загалу є недоступною.

Натомість у світовій практиці заохочується «free and open access to biodiversity data» і створення єдиної бази даних з біорізноманіття. Це, по-перше, вимагає уніфікації усієї інформації, а, по-друге, — її англійськості. Прикладом такої бази є Global Biodiversity Information Facility (GBIF, <http://www.gbif.org>), яка на середину жовтня 2015 р. нараховує понад 577 млн. записів стосовно 1 млн. 611 тис. 321 виду (тварин, рослин, грибів, найпростіших, хромістів, бактерій, вірусів, архей), наданих 768 установами (переважно зберігачами різноманітних колекцій) з цілого світу.

У базі даних представлені колекційні дані, спостереження дослідників, автоматизовані спостереження, викопні матеріали, літературні дані і ін.). З України у цій базі даних представлена лише одна установа Інститут біології південних морів імені О. О. Ковалевського НАН України (припинив існування після анексії Криму Росією у 2014 р.). Така ситуація є неприйнятною якщо ми прагнемо інтеграції у світову наукову спільноту.

Об'єднання такої різноманітної інформації величезного об'єму забезпечується уніфікованим підходом до її систематизації завдяки переліку стандартів «Darwin Core» (<http://goo.gl/4Yno03>).

Фактично це є словник термінів (полів для вводу інформації) для покращення вводу інформації з біорізноманіття з визначеннями, прикладами і коментарями. Терміни описані у стандарті «Darwin Core» є частиною більших словників і технічних специфікацій які розробляються і підтримуються Biodiversity Information Standards (TDWG) (<http://www.tdwg.org/>). Свого часу GBIF опублікувало посібник з підготовки даних природничих колекцій для введення в бази даних «Digitisation of Natural History Collections Data» (2008) (<http://goo.gl/64xwbp>).

Наведемо для прикладу декілька уніфікованих назв полів для вводу інформації в оригінальному написанні:

- *LivingSpecimen* (живий зразок),
- *PreservedSpecimen* (законсервований/зафіксований зразок),
- *FossilSpecimen* (викопний законсервований/зафіксований зразок),
- *Taxon* (група організмів (sensu <http://goo.gl/5503C8>) яка вважається таксономістами гомогенною одиницею; наприклад: Рід *Truncorotaloides* як опубліковано в Brönnimann et al. 1953 в Journal of Paleontology Vol. 27(6) p. 817–820),
- *Event* (дія/подія що відбулася в певному місці протягом певного часу),
- *Habitat* (категорія чи опис оселища в якому відмічена *Event*),
- *SamplingEffort* (сума зусиль затрачених протягом *Event*; наприклад: «40 пастко-діб», «10 годин спостереження; 10 км пішохідний маршрут; 30 км автомобілем»),
- *Preparations* (перелік методів вжитих для препарування та фіксації зразка; наприклад «тушка», «шкіра», «скелет», «фіксований 70 % етанолом», «фіксований 4 % формаліном», «фото», «витяжка ДНК»)
- низка інших полів, передбачених для різних об'єктів та способів документації.

У Державному природознавчому музеї НАНУ за викладеними вище принципами формується україномовна база даних «Ентомологія», яка на кінець 2015 р. нараховує понад 25 тис. записів.

Отже, при створенні баз даних доцільно застосовувати розроблені TDWG принципи, зокрема «Darwin Core», що дасть змогу об'єднувати різноманітні таксономічні бази даних воедино. Переведення баз даних на англійську мову (паралельно з україномовним їх варіантом) дозволить вносити інформацію в міжнародну базу даних GBIF. При дотриманні принципів уніфікації можливе об'єднання баз даних як в межах окремих установ (музеїв, наукових інститутів) та між ними, так і, в перспективі, — створення загальнонаціональної бази даних з біорізноманіття.