

Стандарти за дигитализация на колекциите на Института по биоразнообразие и екосистемни изследвания при БАН и Националния природонаучен музей при БАН в системата Specify 7

Преамбюл

Основна ценност на Института по биоразнообразие и екосистемни изследвания при БАН (ИБЕИ-БАН) и Националния природонаучен музей при БАН (НПМ-БАН) са колекциите от биологични и геологични образци заедно с прилежащата към тях информация и метаданни. За да се осигурят запазването на колекциите и информацията в бъдещето, интегритета на данните, както и възможността да бъдат използвани, е необходимо да се разработят, прилагат и актуализират набор от съответните стандарти.

Документът и стандартите в него са изработени под лиценз за ползване Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Съдържание

Стандарт за транслитериране	2
Стандарт за мета-данни	5
Стандарт за уникален идентификатор на колекционните единици	9
Стандарт за дигитални изображения	12
Стандарт за етикети	16

Стандарт за транслитериране

Обща част

Резюме: Описва системата от правила за предаване на буквите от българската кирилска азбука чрез латински букви или съчетания от латински букви. Осигурява еднозначно предаване на български географски имена, имена на исторически личности, културни реалии, както и термини с български произход.

Отнася се за: всички колекции.

Целеви групи: вътрешно-институционално ползване, външни потребители.

Базирано на: Закон за транслитерацията, РБ, последно изменение ДВ. бр. 98 от 13 Декември 2019 г. Достъпно на <https://www.lex.bg/laws/ldoc/2135623667>.

Необходимост от прилагане на стандарта: Правилата за транслитерация, определени в Закона за транслитерацията, се прилагат от всички юридически и физически лица, които извършват и ползват транслитерация на географски имена, имена на исторически личности и културни реалии, както и термини с български произход в съответната научна област. ИБЕИ-БАН и НПМ-БАН като организации, които съставят, издават или публикуват [...] справочни материали (колекциите), следва да прилагат правилата за транслитерация, установени със Закона за транслитерацията.

Стандарт

Системата за транслитерацията на българските букви с латински букви е, както следва:

Кирилица	Латиница	Кирилица	Латиница	Кирилица	Латиница
А, а	A, a	К, к	K, k	Ф, ф	F, f
Б, б	B, b	Л, л	L, l	Х, х	H, h
В, в	V, v	М, м	M, m	Ц, ц	Ts, ts
Г, г	G, g	Н, н	N, n	Ч, ч	Ch, ch
Д, д	D, d	О, о	O, o	Ш, ш	Sh, sh
Е, е	E, e	П, п	P, p	Щ, щ	Sht, sht
Ж, ж	Zh, zh	Р, р	R, r	Ъ, ъ	A, a
З, з	Z, z	С, с	S, s	Ь, ь	Y, y
И, и	I, i	Т, т	T, t	Ю, ю	Yu, yu
Й, й	Y, y	У, у	U, u	Я, я	Ya, ya

Буквени съчетания:

Кирилица	Латиница
дж	dzh
дз	dz
ьо	yo
йо	yo
"ия", в края на думата	ia

Задължително изписване:

България Bulgaria

Транслитерират се:

Географски термини част от географско име (планина, равнина, низина, плато, град, село, река, езеро, залив и други):

Стара планина	Stara planina
Атанасовско езеро	Atanasovsko ezero

Прилагателни в състава на географското име ("северен", "южен", "източен", "западен", "централен" и други подобни):

Централен Балкан	Tsentralen Balkan
София-юг	Sofia-yug
Перник-север	Pernik-sever

Съставни имена, съобразно с правилата на българския правопис:

Златни пясъци	Zlatni pyasatsi
Горна Оряховица	Gorna Oryahovitsa

Превеждат се:

Географските термини, които не са част от географското име, се превеждат според правилата на езика, избран за превод:

Нос Емине Саре Емине

Имената на личности и географските имена от съвременни чужди книжовни езици:

- които използват версия на латинската азбука, се изписват в оригиналния им вид;
- които не използват версия на латинската азбука, се изписват съгласно приетата системата за транслитерация.

Стандарт за мета-данни

Обща част

Резюме: Стандартът за метаданни е изискване, което има за цел да установи общо разбиране за значението или семантиката на данните, за да осигури правилното използване и интерпретация на данните от техните собственици и потребители. За постигнето на това общо разбиране, трябва да се дефинират редица характеристики или атрибути на данните, известни също като метаданни.

Метаданните са „данни, които предоставят информация за други данни“, но не са самите те данни, като текст на съобщение или самото изображение. Видове метаданни са например:

- **Описателни метаданни** — описателната информация за ресурс. Използва се за откриване и идентификация. Той включва елементи като заглавие, резюме, автор и ключови думи.
- **Структурни метаданни** — метаданни за контейнери с данни. Указват как са събрани сложните обекти, например как страниците са подредени, за да образуват глави. Той описва видовете, версиите, връзките и други характеристики на цифровите материали.
- **Административни метаданни** — информацията за подпомагане на управлението на ресурс, като тип ресурс, разрешения и кога и как е създаден.
- **Правни метаданни** — предоставят информация за създателя, притежателя на авторски права и публичното лицензиране, ако са предоставени.

Метаданните не са строго обвързани с една от тези категории, тъй като могат да опишат част от данни по много други начини.

Darwin Core

Данните за биоразнообразието се съдържат в разнообразни източници, съхранявани в различни формати на много различни хардуерни и софтуерни платформи.

Важна стъпка към разбирането на глобалните модели на биоразнообразието е, да се осигури стандартизиран поглед върху тях и хетерогенните източници на данни да имат оперативна съвместимост. Основно изискване за постигане на подобна съгласуваност е дефинирането на общи термини.

Това се постига чрез създаването на стандарти, един от които е Darwin Core, стандарт за данни за публикуване и интегриране на информация за биоразнообразието. Darwin Core (често съкратено на DwC) е разширение на Dublin Core за информатика на биоразнообразието. Той има за цел да осигури стабилни идентификатори и дефиниции за споделяне на информация за биологично разнообразие (биоразнообразие) и да улесни споделянето на тази информация. Термините, описани в този стандарт, са част от по-голям набор от речници и технически спецификации, които се разработват и се поддържат от информационните стандарти за биоразнообразието (TDWG). DwC се основава предимно на таксони, тяхното присъствие в природата, както е документирано от наблюдения, екземпляри, проби и свързана информация. Стандартът DwC включва речник на термини (в друг контекст те могат да се наричат свойства, елементи, полета, колони, атрибути или концепции) и документация, предоставяща референтни дефиниции, примери и коментари. В стандарта са включени документи, описващи как се управляват тези термини, как наборът от термини може да бъде разширен за нови цели и как термините могат да се използват. Всеки термин на DwC включва определение и дискусии, предназначени да насърчат последователното използване на термините в приложения и дисциплини.

Въпреки че типовете данни и ограниченията не са предоставени в дефинициите на термините, се дават препоръки как да се ограничат стойностите, където е уместно, например чрез предлагане на използване на контролирани речници.

DwC е базиран на версии и непрекъснато се развива, а работните групи често добавят към документацията практически примери, които обсъждат, прецизират и разширяват нормативните дефиниции на всеки термин. Този подход към документацията позволява на стандарта да се адаптира към нови цели, без да нарушава съществуващите приложения.

На практика DwC отделя дефиницията и семантиката на отделните термини от приложението на тези термини в различни технологии. Darwin Core предоставя отделни насоки как да кодирате термините като RDF, XML или текстови файлове.

Отнася се за: всички колекции.

Целеви групи: вътрешно-институционално ползване, външни потребители.

Базирано на: <http://www.tdwg.org/standards/450> , Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, et al. (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLOS ONE 7(1): e29715. <https://doi.org/10.1371/journal.pone.0029715>.

Необходимост от прилагане на стандарта: Specify е адаптиран да експортира данни в DwC, съдържа полетата, които предвижда DwC.

Стандарт

Преглед на приетите понастоящем термини и концепции може да се намери в краткото справочно ръководство на DwC, поддържано от TDWG: <https://dwc.tdwg.org/terms/>.

В DwC (последна актуализация при разписване на стандартите: 2021-09-18): <http://rs.tdwg.org/dwc/terms/> са дефинирани над 215 термина. Основните типове категории са извадени във фигура 1.

Record-level Terms	Dublin Core terms, institutions, collections, nature of data record	Simple Darwin Core (flat)
Occurrence	evidence of species in nature, observers, behavior, associated media, references.	
Event	sampling protocols and methods, date, time, field notes	
Location	geography, locality descriptions, spatial data	
Identification	linkage between Taxon and Occurrence	
Taxon	scientific names, vernacular names, names usages, taxon concepts, and the relationships between them	
GeologicalContext	geologic time, chrono-stratigraphy, biostratigraphy, lithostratigraphy	
ResourceRelationship	explicit relationships between identified resources (e.g., one organism to another, taxon to location, etc.)	Generic Darwin Core (relational)
MeasurementOrFact	measurements, facts, characteristics, assertions, references	

Фигура 1. Darwin Core категории: Simple Darwin Core се състои от седем категории термини (в зелено) (от Wicczorek et al. 2012).

Стандарт за уникален идентификатор на колекционните единици

Обща част

Резюме: Описва изискванията за създаване на уникален идентификатор на колекционните единици от колекциите на ИБЕИ-БАН и НПМ-БАН. Придаването на уникален идентификатор на всяка колекционна единица (обект) позволява оптимизирано управление на колекциите (напр. проследяване на местоположението и движението на колекционните единици).

Отнася се за: всички колекции.

Съответен DarwinCode термин: CatalogNumber
(<http://rs.tdwg.org/dwc/terms/catalogNumber>)

Целеви групи: вътрешно-институционално ползване: създаване и администрация на колекции.

Базирано на: вътрешни решения на Изпълнителен съвет на Консорциум „Надграждане на разпределена научна инфраструктура „Разпределена система от научни колекции – България (DiSSCo–BG): Етап 1“.

Необходимост от прилагане на стандарта: задължително.

Стандарт

Всеки уникален идентификатор (*catalogNumber*) следва да е универсален и уникален по отношение на проследимост.

При въвеждането в съответната колекция, всяка колекционна единица получава буквено-цифров уникален идентификатор в информационната система. Генерираният идентификатор на всяка колекционна единица в информационната система следва да бъде наличен и свързан физически със колекционната единица (например, на етикет или маркиран върху самия физически обект).

Идентификаторът (*catalogNumber*) се състои от следните компоненти:

- България се съкращава BG според стандарта ISO 3166-1 alpha-2
- Префиксите за ИБЕИ-БАН и НПМ-БАН: BG-IBER и BG-NMNHS

- Код на колекция (*CollectionCode*), която е комбинацията от код на институцията и следните тубуквени съкращения на колекциите, разделени с тире:

имена на колекции в Specify	ИБЕИ-БАН	НПМ-БАН
Минералогия	-	MIN
Безгръбначни животни	INV	INV
Ентомология	-	ENT
Ихтиология	-	ICH
Земноводни и влечуги	-	HER
Орнитология	-	ORN
Бозайници	-	MAM
Гръбначни животни	VER	-
Гъби	SOMF	MYC
Бриология	SOMB	-
Ботаника	-	BOT
Хербариум	SOM	-
Палеоботаника	-	PBT
Палеоботаника и палинология	PB	-
Палеозоология	-	PZO
Фосилни бозайници	FMC	-

Подколекции няма.

Потенциални други колекции се именуваат посредством уникален трибуквен код на латиница.

Пример за код на колекция: Гъби в НПМ-БАН – BG-NMNHS-MYC

- Пореден автоматично зададен от системата цифров номер:

000000000001



Състои се от 12 цифри. При изписване е допустимо да не се изписват 0-те от лявата част.

Компонентите се разделят с тире. Така, крайният идентификатор е например *BG-NMNHS-MYC-000000123456*.

За колекционна единица със съществуващ идентификатор (създаден преди въвеждането ѝ в Specify) следва съществуващият идентификатор да се въведе в съответното поле за колекцията (напр. Alt Cat Number).

Номерацията на документи (напр. за приемане или отдаване на колекционни единици) е с формат: NMNHS-трибуквен идентификатор на колекцията-автоматичен шест цифров пореден номер)/година.

Например, за колекция Ботаника на НППМ-БАН: *NMNHS-BOT-000123/2021*.

Стандарт за дигитални изображения

Обща част

Резюме: Описва общи насоки за изготвяне и съхраняване на снимков документационен материал на колекционни единици (биологични и геологични). Резултатът от прилагането на стандарта е изготвянето на референтен дигитален снимков материал, който да осигури на потребителите на цифровизирани музейни бази данни количеството и качеството на визуална информация, необходими за първично запознаване със състоянието на колекционните единици.

Отнася се за: всички колекции.

Целеви групи: вътрешно-институционално ползване, външни потребители/изследователи.

Необходимост от прилагане на стандарта: по усмотрение на кураторите, в зависимост от спецификите на съответната колекция

Стандарт

Целта на създаването и поддържането на дигитален снимков материал за колекционните единици е да бъде предоставена допълнителна информация към съответните записи, която да послужи като първична информация за учени и куратори, интересувачи се от съответния обект и като база за сравнение на състоянието на материала в бъдеще. Съответно качеството на снимковият материал следва да отговаря на тези нужди. Важно е да се има предвид, че снимките не е необходимо да отговарят на изисквания, според които да бъдат използвани за описване/идентифициране на вид/таксономична единица, или научна публикация.

Поради разнородните колекции и техните специфики не е възможно създаването на единен и работещ детайлен стандарт, който да бъде стриктно спазван от всички.

Въпреки това, съществуват основни общи минимални изисквания. Те представляват баланс между нуждите на кураторите (напр. улесняване на работата с

колекцията, популяризирането ѝ), външните потребители на данните от колекцията (напр. учени с интерес за работа с колекционните материали) и техническото време и ресурси, необходими за изработването на снимковия материал и поддържането му.

Поради гореизброените причини, основни препоръки са:

- Добро изображение е това, в което осветлението и позиционирането са стандартизирани и по този начин обектът може да бъде сравнен с други изображения със сходно качество. Последователният контрол на ключови фактори ни позволява да създадем снимка, в която: (а) яркостта и контрастът са равномерни, без области с "нулеви данни", (б) фокусът е рязък и (в) позата и ракурсът на обекта са сходна с тези, използвана за други сравними обекти от същия тип колекции;
- Силно препоръчително е всяка снимка да съдържа съответен мащаб (напр. физическа мащабна линия), позициониран така, че да позволява приблизителното определяне на размера на обекта (не е нужно да позволява 100% точно морфометрично измерване);
- За обекти, при които цветовете носят важна информация се препоръчва използването на стандартна цетова референция (напр. Pantone), която да бъде заснета и описана заедно с обекта, за да могат потребителите да коригират цветовете.

Получената от снимката информация е важна за всеки учен и куратор като улеснява работата със съответният образец, затова постарайте се да следвате тези четири стъпки.

1. Заснемане на ясен и четим образ – етикетът е видим и се разчита, поставена е скала, образът е чист и показва състоянието и по възможност максимално количество полезна информация за съответният обект. Максимално голяма част от обекта е на фокус, няма отблясъци и тъмни участъци;
2. Позиция (ъгъл на заснемане) – в зависимост от значимостта на обекта (напр. при типов екземпляр, рядък вид, интересен обект) е препоръчително да се направят изображения от различни информативни позиции;

3. Корекции след заснемането – чрез използване на специализиран софтуер за работа със снимки, проверете и коригирайте при нужда нивата на цветовете (RGB), осветеността, контраста;
4. Коректно и информативно именуване на файловете – ако използвате вече готови снимки или първо снимате, а след това качвате снимката в Specify, именувайте ги така, че да е ясно към кой Collection Object се отнася съответната снимка.

Оптимални характеристики на снимката:

- резолюция – 300 dpi;
- размер – <20 Mb;
- формат – JPEG или TIFF.

Основни разлики между JPEG/JPG и TIFF формати:

- JPEG/JPG използва компресия със загуби на оригинална информация, което означава, че се намалява качеството на картината (загуби в резки преходи в интензивността и цветовия оттенък), за да се постигне по-малък и по-управляем размер на файла;
- TIFF обикновено пази цялата оригинална информация или използва компресия без загуби (напр. LZW, препоръчваме ползването ѝ), което ги прави по-добър избор за извършване на всякакви работи по тяхното редактиране. Това означава, че TIFF файловете обикновено са по-големи и заемат повече дисково/сървърно пространство.

Списък с подбрана литература за изготвяне на дигитални изображения

Brecko J. & Mathys A. (2020) Handbook of best practice and standards for 2D+ and 3D imaging of natural history collections. *European Journal of Taxonomy* 623: 1–115. doi: 10.5852/ejt.2020.623

Brecko J., Mathys A., Dekoninck W., Laponce M., VandenSpiegel D. & Semal P. (2014) Focus stacking: Comparing commercial top-end set-ups with a semi-automatic low

budget approach. A possible solution for mass digitization of type specimens. *ZooKeys* 464: 1–23. doi: 10.3897/zookeys.464.8615

Callomon P. (2021) A basic guide to specimen photography in museum collections. Department of Malacology, Academy of Natural Sciences of Drexel University, 27 pp.

Erolin C., Jarron M. & Csetenyi L.J. (2017) Zoology 3D: Creating a digital collection of specimens from the D'Arcy Thompson Zoology Museum. *Digital Applications in Archaeology and Cultural Heritage* 7: 51–55. doi: 10.1016/j.daach.2017.11.002

Flemons P. & Berents P. (2012) Image based digitisation of entomology collections: leveraging volunteers to increase digitization capacity. *ZooKeys* 209: 203–17. doi: 10.3897/zookeys.209.3146

Merckx J., Van Roie M., Gómez-Zurita J. & Dekoninck W. (2018) From theory to practice: a photographic inventory of museum collections to optimize collection management. *Biodiversity Informatics* 13: 38–48. doi: 10.17161/bi.v13i0.7036

Moeseneder C. (2017) Focus-stacking macro photography system for around \$1,200. *Scarabs* 83: 1–7.

ICEDIG Project Outcomes, DiSSCo Knowledgebase:
<https://know.dissco.eu/handle/item/164/simple-search?filterquery=Digitisation&filtername=subject&filtertype>equals>

Стандарт за етикети

Обща част

Резюме: Описва общи насоки за изготвяне и поставяне на етикети на колекционни единици (биологични и геологични). Резултатът от прилагането на стандарта е базисна информация за колекционната единица, лесно достъпна за прочит от хора и дигитални технологии, включваща уникалното определяне на единицата.

Отнася се за: всички колекции.

Целеви групи: вътрешно-институционално ползване, външни потребители/изследователи.

Необходимост от прилагане на стандарта: Поставянето на етикет е задължително за всяка колекционна единица.

Стандарт

Изработването и поставянето на информативен етикет е задължителен компонент от създаването и съхраняването на всяка колекционна единица! Информацията на етикета определя колекционната единица като уникална (чрез вписания идентификатор) и осъществява връзката на физическия обект с дигиталната информацията за него съхранявана в системата Specify.

Задължителни елементи при етикетиране на обект, въведен в системата Specify, са включването на уникалния идентификатор и машинно генериран код на колекционната единица. Останалата информация е по усмотрение на кураторите, в зависимост от спецификите на съответната колекция.

По изключение, етикет може да не се поставя ако уникалният идентификатор на колекционната единица се поставя директно върху самия обект (напр. ако се изписва върху костен или геологичен материал). Въпреки това, препоръчва се това да е изключение, а не практика. В такива случаи, етикет се добавя върху контейнера за съхранение.

Машинно генерираните QR кодове позволяват съхранението на относително големи буквено-цифрово-символни данни и могат да бъдат генерирани и разчитани посредством безплатни приложения. По този начин, освен дублиране на информацията от етикета която може да бъде четена от хора директно, може да се генерира и хипервръзка към конкретния запис за колекционната единица в системата Specify.

Информацията на етикета трябва да позволява възможността за бързо запознаване с основната информация за колекционната единица. За биологични единици, това оптимално включва:

- Уникален буквено-цифров идентификатор на колекционната единица към системата Specify, ако колекционната единица е вече въведена;
- Научно име (ако е налична таксономична идентификация на обекта);
- Локация на намиране (по възможност с географски координати);
- Година на намиране;
- Колектор;
- Пол, възраст, местообитание и друга първична информация;
- Машинно генериран код (QR код).

Размерите на етикета, големината на текста и разположението на информацията следва да бъдат изготвени според типа на колекцията и изискванията на куратора.