

Projekt DC07P02OUK002

Národní autority v prostředí muzeí a galerií - interoperabilita s NK ČR

Taxonomické databáze

Pracovní materiál

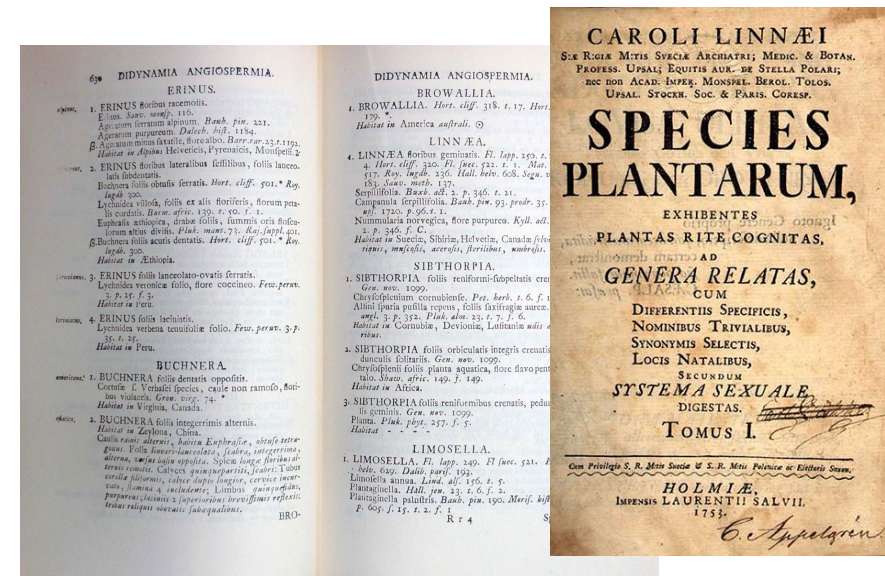
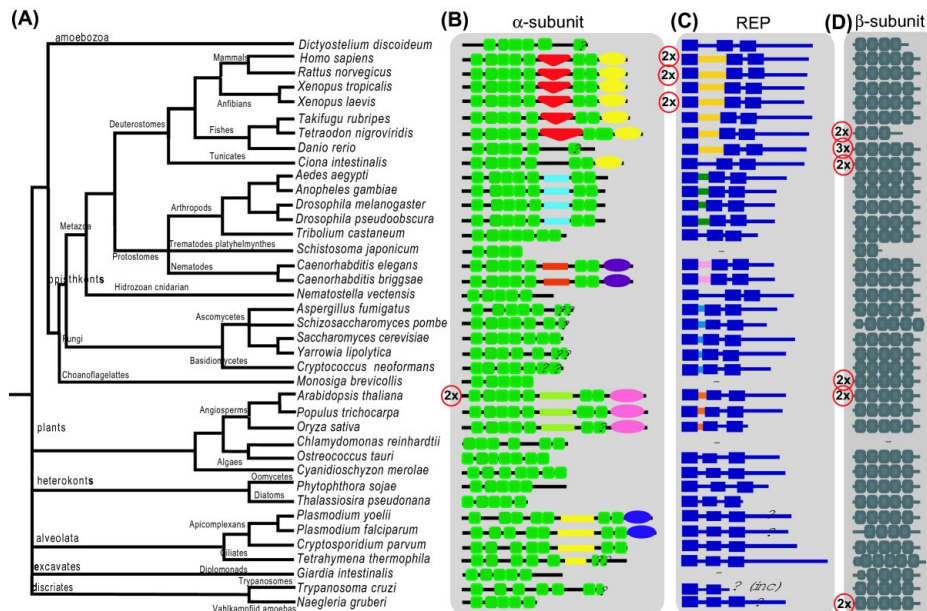
Mgr. David Cigánek, 8.3.2010

taxonomie

- Vytváří a studuje systém
- Zabývá se různými formami příbuznosti a následnosti taxonomických jednotek
- Jména pouze používá, ale nevytváří
- Taxonomická pravidla jsou zcela formální a musejí se vždy za každých okolností dodržet
- Kdyby se prokázalo, že smrk patří mezi papoušky, tak tam bude zařazen!

nomenklatura

- Zabývá se jmény taxonů,
- Vytváří pravidla pro definici a užívání jmen,
- Systém řeší jen do té míry, pokud ovlivňuje tvorbu jmen,
- Nomenklatorická pravidla jsou formalizovaná, platná jen pro danou skupinu organismů a poskytují určitou volnost pro arbitrární rozhodování nomenklatorických komisí



taxonomická jednotka

- skupina definovaná podřízenou jednotkou kterou obsahuje (ovšem podle ICBN je to vždy rod!)
- formální definice skupiny nestanovuje které další rody do ní patří/nepatří
- skutečný obsah skupiny (příslušnost těch či oněch rodů do ní) je na rozhodnutí konkrétního taxonoma, resp. jeho vědeckého názoru

člověk

Australopithecus

šimpanz

gorila

orangutan

gibon



Hominidae

sensu stricto

>

>

>

>

sensu lato

Tyrannosaurus vs. *Manospondylus*

V roce 1905 byl popsán rod *Tyrannosaurus* (na základě druhu *Tyrannosaurus rex*).

Podle dalších výzkumů se ukázalo, že rod *Tyrannosaurus* (Osborn, 1905) je totožný s už dříve popsaným rodem *Manospondylus* (Cope, 1892) a současně popsaným rodem *Dynamosaurus* (Osborn, 1905).

Platným rodovým jménem by měl být *Manospondylus* - z důvodu časové priority. Jméno *Tyrannosaurus* se ale postupem času stalo vžitým a populárním.

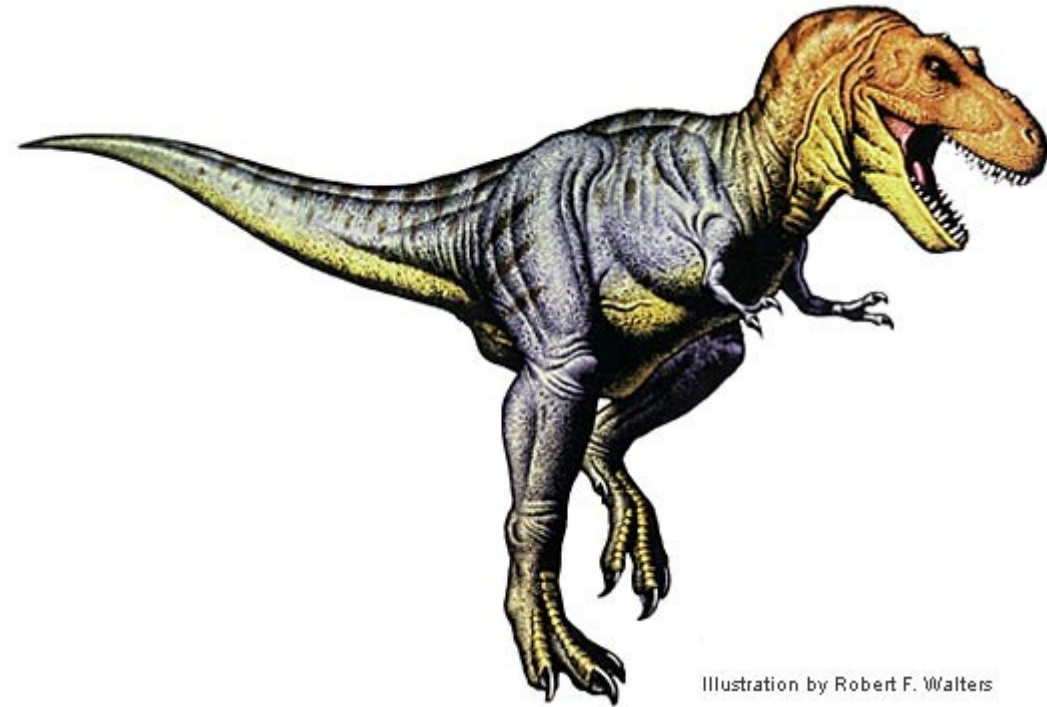
„dosavadní jméno musí být zachováno“
pokud

"starší synonymum nebo homonymum
nebylo po roce 1899 užíváno jako validní
jméno"

a zároveň

"mladší synonymum nebo homonymum bylo
pro některý taxon užíváno jako domnělé
validní jméno v alespoň 25 pracích
publikovaných nejméně 10 autory v
předchozích 50 letech..."

IV. vydání ICZN, 2000



***Tyrannosaurus*, Osborn 1905
(*nomen protectum*)**

Syn: *Manospondylus*, Cope 1892 (*nomen oblitum*);
Dynamosaurus, Osborn 1905; ?*Nanotyrannus*,
Bakker, Williams & Currie 1988; *Stygivenator*,
Olshevsky 1995; *Dinotyrannus*, Olshevsky 1995

Authority

Nejpodstatnější částí jsou jména a dále vazby vysvětlující synonymii („viz, viz též“), šíří termínu a jeho příslušnost do různých pojmových systémů. Skupiny jsou pojmenovány a dále se s nimi nakládá víceméně jako se jmény.

Taxonomie

Systém tvoří skupiny (taxony), jména identifikující tyto skupiny a vazby, které definují:

- vztahy mezi jmény (platná jména, synonyma)
- vztahy mezi skupinami (nadřízenost / podřízenost, inkluzivitu / exkluzivitu)

Každá skupina, jméno nebo vazba jsou vždy přiřazeny k jedné nebo více taxonomiím, což se udává autorskou citací. (atribut? funkční podmínka? omezení?)

„Biological classification (sometimes known as "Linnaean taxonomy") is still generally the best known form of taxonomy.“ (<http://en.wikipedia.org/wiki/Taxonomy>)

Authority	Taxonomie	Ontologie
jména	jména	instance, objekty
vazby	skupiny (taxony)	třídy
	vazby (vertikální, horizontální)	vazby, vztahy
		atributy
		funkční podmínky
		omezení
		pravidla
		axiomy
		události

Taxonomie se zdá být obecnější obdobou autoritního systému jak jej známe z knihovního prostředí. Zcela univerzální pak je ontologické schéma, pomocí něhož lze popsat jakýkoli obecný znalostní systém.

Možné strategie vytváření taxonomického systému v projektu „Národní autority...“ CITE M:

- A) obohatit stávající autoritní systém NK ČR tak, aby byl schopen pracovat i se skupinami, vazbami a alternativními taxonomiemi,
- B) s využitím specializovaných nástrojů pro počítačové zpracování dat v taxonomii vytvořit prostředí umožňující využívat volně dostupných mezinárodně spravovaných taxonomií,
- C) formou namapování na CIDOC CRM navrhnout obecný spojovací prvek mezi instancemi jmen taxonomického systému a objekty v paměťových institucích.

Authority	Taxonomie	Ontologie
jména	jména	instance, objekty
vazby	skupiny (taxony)	třídy
	vazby (vertikální, horizontální)	vazby, vztahy
		atributy
		funkční podmínky
		omezení
		pravidla
		axiomy
		události

A) obohatit stávající autoritní systém NK ČR tak, aby byl schopen pracovat i se skupinami, vazbami a alternativními taxonomiemi

- + zřejmý výchozí stav,
- + dosažitelný výsledek i bez rozsáhlého výzkumu,
- + maximální využití stávajícího schématu NA,
- ještě pevnější přimknutí ke 40. let starým standardům skupiny MARC,
- nutnost řešit řadu problémů nestandardními nouzovými postupy,
- relativně nepatrná využitelnost současných trendů (sémantický web, ontologie, CIDOC CRM...) > obstarožně vyhlížející projekt, který nebude chtít nikdo podporovat

CITeM navrhne způsob jak do s použitím standardních autoritních formátů realizovat taxonomickou autoritu a připraví pravidla pro práci s ní. Implementace se ujme Cosmotron nebo jiný dodavatel v návazném projektu.

Autority	Taxonomie	Ontologie
jména	jména	instance, objekty
vazby	skupiny (taxony)	třídy
	vazby (vertikální, horizontální)	vazby, vztahy
		atributy
		funkční podmínky
		omezení
		pravidla
		axiomy
		události

B) s využitím volně dostupných taxonomických nástrojů vytvořit prostředí pro využívání dostupných mezinárodně spravovaných taxonomií

- + bez nutnosti vytvářet a především spravovat používané taxonomie
- + široká mezinárodní kompatibilita
- + ověřené a vědecky podložené postupy
- v zahraničí mnoho alternativních přístupů, žádný všeobecně přijímaný
- omezená možnost redakčně ovlivňovat vytvářené taxonomie

CITeM vyhledá mezinárodně spravované taxonomie a k jedné nebo části z nich navrhne konektory pro využívání v CMS. Vlastní realizace je otázkou návazného projektu řešeného nejlépe ve spolupráci s NK ČR. Naplňování i redakce systému zůstane nadále v kompetenci správce příslušné služby, která se bude pouze vzdáleně využívat.

Autority	Taxonomie	Ontologie
jména	jména	instance, objekty
vazby	skupiny (taxony)	třídy
	vazby (vertikální, horizontální)	vazby, vztahy
		atributy
		funkční podmínky
		omezení
		pravidla
		axiomy
		události

C) formou namapování na CIDOC CRM navrhnout obecný spojovací prvek mezi instancemi jmen taxonomického systému a objekty v paměťových institucích

- + dostatečně obecný základ pro solidní řešení
- + široká mezinárodní kompatibilita
- + potenciál pro velký mezinárodní projekt
- potenciál pro velký mezinárodní průser
- mimořádná metodická a organizační náročnost
- nerealizovatelné bez úzké spolupráce s autory CIDOC CRM a dalšími experty
- bezpodmínečně nutný důraz na vyváženost projektu po teoretické a praktické stránce

CITeM se transformuje v progresivní metodickou instituci evropského významu, najme specialisty a navrhne teoretické řešení. Pro jeho implementaci se bude muset počkat na speciálně pro tento účel vytvořený dotační titul z Bruselu, nebo lépe na novou EU.

Authority	Taxonomie	Ontologie
jména	jména	instance, objekty
vazby	skupiny (taxony)	třídy
	vazby (vertikální, horizontální)	vazby, vztahy
		atributy
		funkční podmínky
		omezení
		pravidla
		axiomy
		události

B) s využitím volně dostupných taxonomických nástrojů vytvořit prostředí pro využívání dostupných mezinárodně spravovaných taxonomií

CITeM vyhledá mezinárodně spravované taxonomie...

Seznam jmen cévnatých rostlin květeny ČR

Vzniká postupně od 90. let, aktuální verze (2007) zahrnuje všechna jména použitá v Klíči ke květeně České republiky (*Kubát et al. 2002*) + navíc přidáno několik jmen v r. *Taraxacum* a *Rubus*, opraveny chyby.

Rozsah: 4,8k jmen (jen cévnaté rostliny)

Licence: ?



<http://www.ibot.cas.cz/index.php?p=seznam&site=default>

- + kompletní květena ČR
- + důvěryhodný zdroj
- + včetně autorských citací, synonym, zařazení do čeledí
- + obsahuje i oficiální česká jména

- podoba (zazipované XLS, DBF, MDB soubory)
- nelze využívat online
- omezené aktualizace
- málo použitelné pro exotické taxony

Wikispecies

Za podpory Nadace Wikimedia začíná v srpnu 2004. Na základě linnéovské taxonomie je postupně vytvářena ve wiki-formátu a linkována k článkům o jednotlivých taxonech na Wikipedii.

Rozsah: 210k druhů (všechny formy života)

Licence: Creative Commons

„Wikispecies is free, because life is in the public domain!“

<http://species.wikimedia.org>



- + návaznost na další wiki-projekty
- + svobodná licence, nulová cena
- + obecná popularita
- + značný rozsah, růstový potenciál
- + včetně cizojazyčných jmen taxonů

- neřeší alternativní taxonomie
- volně splývá s Wikipedií
- místy mezernatý systém
- výrazně nadšenecký projekt

Encyclopedia of Life

Stránka byla s prvními 30k záznamy spuštěna 26. 2. 2008 a brzy se stala velmi populární. Hodnota projektu se odhaduje na 110M dolarů. Cílem je katalogizace všech vědě známých druhů – 1.8M. Informace kompiluje z několika vědeckých databází, v dozorčí radě projektu zasedají renomovaní vědci světového formátu.

Rozsah: 1,5M druhů (všechny formy života)

Licence: Creative Commons

<http://www.eol.org>



- + výrazně odborný charakter
- + rozsáhlé informace k taxonům
- + vyčerpávající rozsah

- webové služby v beta-verzi
- zatím slabá podpora pro používání alternativních taxonomií

Global Names Index

Syntetický zdroj více než 30 velkých databází vědeckých jmen. Společná iniciativa GBIF a EOL.

Parser umožňuje stahování informací k hledaným jménům (až 5k najednou!) ve 4 různých formátech vč. XML.

Rozsah: 18,4M jmen (všechny formy života)

Licence: MIT, obsah jako public domain



<http://www.globalnames.org>

- + výrazně odborný charakter
- + vyčerpávající rozsah
- + syntetický charakter
- + parser, API

- nejasná kvalita deduplikace
- parser v době testu nefunkční
- neumožňuje alternativní taxonomie

Global Biodiversity Information Facility

Syntetický zdroj zprostředkovávající přístup k mnoha oborovým a národním databázím od více než 300 poskytovatelů. Checklist Bank (taxonomický thesaurus) měl být podle informací na webu připraven k recenzi ve III. Čtvrtletí loňského

Rozsah: 196,5M záznamů o výskytu druhů,

Licence: zachovává práva poskytovatelů dat



<http://www.gbif.org>

- + výrazně odborný charakter
- + rozsah

- pouze informace o výskytu druhů
- neumožňuje alternativní taxonomie

Catalogue of Life

Společná iniciativa Species 2000 a ITIS zahájená v červnu 2001. Cílem je sestavení katalogu všech druhů organismů do roku 2011. Práce probíhají kompilací 66 taxonomických databází ve spolupráci více než třemi tisíci specialisty z celého světa. Vydává se roční checklist a dynamický checklist (vydáváný častěji, méně rozsáhlý, obohacen o mezinárodní checklisty).

Rozsah: 1 160 711 (aktuální Annual Checklist 2009)

Licence: zdarma pro nekomerční využití, vždy nutná citace zdroje

Catalogue of Life: Dynamic Checklist
indexing the world's known species

<http://www.catalogueoflife.org>



- + výrazně odborný charakter
- + expertní řízení
- + zřejmě velmi kvalitní deduplikace

- API ?
- neumožňuje alternativní taxonomie

BioLib

Původní český projekt, nyní mezinárodního charakteru. Jako zdroj dat využíván i některými většími systémy. Zaměřen jak na odbornou, tak i laickou veřejnost.

Spolupracuje (především na mapovacích projektech) s orgány ochrany přírody, Národním muzeem a vědeckou komunitou.

Deklarován jako nekomerční. Provozovatelem je Český Podnikatelský Servis s.r.o.

Rozsah: 713k taxonů, 991k vědeckých názvů, 257k národních názvů, 110k obrázků

Licence: nositelem autorských práv jsou autoři textů a fotografií

BioLib.cz

<http://www.biolib.cz>

- + soukromý nadšenecký projekt s odborným přesahem
- + wiki-like řízení vstupu dat
- + národní názvy taxonů

- API ?
- udržitelnost projektu ?
- neumožňuje alternativní taxonomie
- kolísající aktivita webu

NCBI Taxonomy

Zdroj zaměřený na využití pro studie v genetice, molekulární biologii, biotechnologii a medicíně. Spravuje National Center for Biotechnology Information.

Rozsah: 160k (jen druhy s dostupnými DNA daty)

Licence: ?



<http://www.ncbi.nlm.nih.gov/Taxonomy>

- + profesionální projekt
- + vládní (U.S.) financování

- taxonomie mimo primární zájem
- neumožňuje alternativní taxonomie
- jako autoritní zdroj zjevně nepoužitelné

Tree of Life Web Project

Edukativně laděný projekt Arizonské univerzity. Obsah rozdělený na samostatné stránky věnované konkrétním taxonům, zejména z vyšších hierarchických úrovní. Od roku 2007 spolupracuje s EOL. S možností vytváření alternativních taxonomií se počítá teprve do budoucna.

Rozsah: >10k stránek (~ neznámý počet taxonů)

Licence: zachovává se copyright autorů obsahu,
XML data poskytována pod licencí Creative Commons

<http://tolweb.org/tree/>



- + grafika, ilustrativnost
- + XML formát, webové služby

- omezený rozsah
- primárně edukativní web
- neumožňuje alternativní taxonomie

Integrated Taxonomic Information System

Jeden z největších a nejrychleji rostoucích taxonomických informačních zdrojů. Partner GBIF, Species 2000 a vládních agentur v USA, Kanadě a Mexiku (ITIS-North America). Společně se Species 2000 (Catalogue of Life) zajišťují taxonomickou základnu mamutímu projektu Encyclopedia of Life (EOL).

Rozsah: necelých 500k jmen

Licence: public domain

<http://www.itis.gov>



- + profesionální projekt
- + vládní podpora (USA, Kanada...)
- + rychlý růst
- + zjevně kvalitní deduplikace
- + podrobně dokumentované webové služby, možnost stažení celé databáze

- zaměření především na severoamerické druhy (u rostlin výlučně)
- neumožňuje alternativní taxonomie

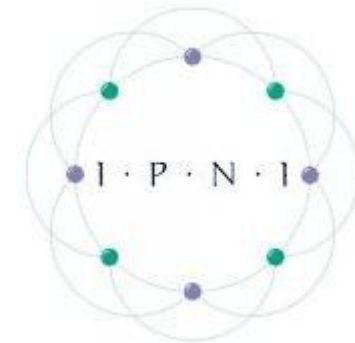
International Plant Names Index

Databáze jmen a bibliografie pro semenné rostliny a kaprad'orosty. Kompilováno na základě Index Kewensis a několika dalších digitálních/digitalizovaných zdrojů. Poskytuje pouze jmenný index, bez značení platnosti jména.

Rozsah: ? (z nepřímých zdrojů lze usuzovat na cca 1,3M)

Licence: free (copyright The Royal Botanic Gardens, The Harvard University Herbaria, The Australian National Herbarium)

<http://www.ipni.org>



- + profesionální projekt
- + udržitelnost, dlouhodobá podpora
- + kvalita dat srovnatelná s autoritami
- + URN identifikátory ke jménům

- zcela bez taxonomického systému
- chybějí starší supragenerická jména

uBio

„The uBio Taxonomic Name Server acts as a name thesaurus.“

Rozsah: NameBank - 11,1M jmen, 532k nomenklatorických skupin,
ClassificationBank – 90 klasifikací, 9,6M konceptů, 5,9M synonym

Licence:?



<http://www.ubio.org/>

- + vysoká míra konformity s autoritními systémy
- + profesionální projekt
- + řada webových služeb
- + variantní taxonomie

- ?

Srovnání jednotlivých služeb

	 Wsp. Wikispecies free species directory	 EoL Encyclopedia of Life	 GNI	 GBIF	 CoL Catalogue of Life	 BioLib	 NCBI	 ToL	 ITIS	 IPNI	 uBio
<i>Profi služba</i>	○	●	●	●	●	○	●	●	●	●	●
<i>Profi vstup & deduplikace</i>	○	●	●	●	●	○	●	●	●	●	●
<i>Variantní taxonomie</i>	○	●	○	○	○	○	○	○	○	○	●
<i>Webové služby, API</i>	○	●	●	●	○	○	○	○	●	●	●
<i>Dlouhodobá udržitelnost</i>	●	●	●	●	●	○	●	○	●	●	●
<i>Národní názvy taxonů</i>	●	○	○	○	○	●	○	○	○	○	●
<i>Významný růstový potenciál</i>	○	●	●	●	●	○	●	○	●	●	●
<i>Dostatečné geografické pokrytí</i>	○	●	●	●	○	●	○	○	○	●	●
<i>Dostatečné taxonomické pokrytí</i>	○	●	●	●	○	○	○	○	○	○	●
<i>Rozsah (mil. jmen, podle http://www.globalnames.org/data_sources)</i>	?	1,4	18,4	5,4	2,1	0,1	?	0,1	0,5	1,3	10,0

Návrh dalšího postupu v oblasti taxonomických autorit

- vybrat jeden až tři vhodné mezinárodní systémy,
- zvážit potřebnost samostatné autoritní mezivrstvy pro taxonomické authority (oproti přímé komunikaci s uživateli),
- navrhnout způsob propojení s NA ČR
- ve spolupráci s NK ČR připravit návazný projekt pro taxonomické authority (řešení návaznosti na NA, zapojení mimoknihovních uživatelů)