

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE  
FAKULTA ŽIVOTNÍHO PROSTŘEDÍ

KATEDRA APLIKOVANÉ GEOINFORMATIKY A ÚZEMNÍHO PLÁNOVÁNÍ



Zhodnocení dat výskytu mořských savců  
publikovaných prostřednictvím sítí GBIF a OBIS

BAKALÁŘSKÁ PRÁCE

Vedoucí práce: Ing. Vítězslav Moudrý, PhD.

Bakalant: Sára Marcinková

2019

# ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Fakulta životního prostředí

## ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Sára Marcinková

Aplikovaná ekologie

### Název práce

**Zhodnocení dat výskytu mořských savců publikovaných prostřednictvím sítě GBIF a OBIS**

### Název anglicky

**Assessment of marine mammals data published through GBIF and OBIS networks**

---

### Cíle práce

Cílem práce je zhodnotit data výskytu mořských savců publikovaných prostřednictvím globálních sítí GBIF (Global Biodiversity Information Facility) a OBIS (Ocean Biogeographic Information System). Formulace dílčích cílů je úkolem autorky.

### Metodika

Veřejně přístupná data o výskytu druhů publikovaná skrze globální agregátory dat jsou stále více využívána v ekologických studiích. Mezi nejznámější agregátory patří Global Biodiversity Information Facility (GBIF) a Ocean Biogeographic Information System (OBIS). Ačkoliv tímto způsobem dostupná data dosahují závratných počtů (např. více než 500 milionů záznamů v případě GBIF), mnohé z těchto záznamů jsou prostorově nepřesné, neposkytují informaci o době pozorování, nebo jsou dokonce duplikovány. Úkolem řešitelky je zpracovat literární rešerši na téma sdílení ekologických dat, přičemž se zaměří na doposud identifikované problémy související s kvalitou poskytovaných dat. V praktické části zhodnotí kvalitu záznamů dostupných pro mořské savce.

### Doporučený rozsah práce

30 – 40 stran

## Klíčová slova

Formulace klíčových slov je úkolem autora.

---

## Doporučené zdroje informací

- Amano, T., Lamming, J. D., & Sutherland, W. J. (2016). Spatial gaps in global biodiversity information and the role of citizen science. *Bioscience*, 66(5), 393-400.
- Costello, M. J., Vanhoorne, B., & Appeltans, W. (2015). Conservation of biodiversity through taxonomy, data publication, and collaborative infrastructures. *Conservation Biology*, 29(4), 1094-1099.
- Ficetola, G. F., Rondinini, C., Bonardi, A., Katariya, V., Padoa-Schioppa, E., & Angulo, A. (2014). An evaluation of the robustness of global amphibian range maps. *Journal of Biogeography*, 41(2), 211-221.
- Franz, N. M., & Sterner, B. W. (2018). To increase trust, change the social design behind aggregated biodiversity data. *Database*, 2018.
- Mesibov, R. (2013). A specialist's audit of aggregated occurrence records. *ZooKeys*, (293), 1.
- Otegui, J., Ariño, A. H., Encinas, M. A., & Pando, F. (2013). Assessing the primary data hosted by the Spanish node of the Global Biodiversity Information Facility (GBIF). *PLoS One*, 8(1), e55144.
- 

## Předběžný termín obhajoby

2018/19 LS – FŽP

## Vedoucí práce

Ing. Vítězslav Moudrý, Ph.D.

## Garantující pracoviště

Katedra aplikované geoinformatiky a územního plánování

Elektronicky schváleno dne 17. 3. 2018

**doc. Ing. Petra Šímová, Ph.D.**

Vedoucí katedry

Elektronicky schváleno dne 18. 3. 2018

**prof. RNDr. Vladimír Bejček, CSc.**

Děkan

## **Prohlášení**

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně pod vedením Ing. Vítězslava Moudrého, PhD. a že jsem uvedla veškeré literární prameny a publikace, ze kterých jsem čerpala. Dále prohlašuji, že se tištěná verze shoduje s verzí odevzdanou v informačním systému.

V Praze dne 25.4 2019

Podpis:.....

## **Poděkování**

Touto cestou bych ráda poděkovala Ing. Vítězslavu Moudrému, PhD. za odborné rady a konzultace, které mi v průběhu vedení práce poskytl. Ing. Kateřině Gdulové za pohotovou pomoc. A v neposlední řadě také své rodině a nejbližším přátelům, kteří mě po celou dobu studia podporovali.

## Abstrakt

Globální databáze nyní poskytují více než desítky milionů záznamů o výskytech druhů, které jsou velmi důležité při sledování změn v biologické rozmanitosti. Každý vydavatel přistupuje ke zpracování a šíření těchto záznamů jiným způsobem, čímž je zapříčiněna různá kvalita, rozsah a přesnost. Cílem této bakalářské práce je zhodnocení kvality dat mořských savců, v databázích Global Biodiversity Information Facility (GBIF) a Ocean Biogeographic Information System (OBIS). Z obou portálů byly datové sady staženy v roce 2015 a 2019 a poté podrobeny analýzám v prostředí ArcGIS. Hodnocení proběhlo na základě stanovených parametrů, mezi které byly zahrnuty: duplicitní záznamy, doba pořízení záznamu, výskyty druhů vůči pobřežní linii, přesnost souřadnic, změny chování dat a shodnost záznamů z obou databází. Výsledky ukazují, že poskytovaná data, která nejsou duplicitní, vyskytují se v oceánu a mají uvedenou dobu pořízení, zaujímají v roce 2015 53 % (GBIF), 56,4 % (OBIS) a v roce 2019, 60,7 % (GBIF) a 53,8 % (OBIS). Souřadnice, které nevykazovaly žádná desetinná místa, zaujímají v obou letech 20 % (GBIF) a 10 % (OBIS). Změny chování dat vykazují mezi lety 2015 až 2019 obrovské množství smazaných záznamů, které jsou v průběhu doplňovány novými. Uvedené nedostatky ukazují, že je důležité pokračovat v kontrole kvality záznamů a minimalizaci existujících chyb. Jedině tak se mohou globální databáze stát účinným nástrojem ochrany přírody.

**Klíčová slova:** záznamy výskytu, kvalita dat, GBIF, OBIS, mořští savci

..

## **Abstract**

Global database are providing more than 10 millions records about species occurrence, which are very important for monitoring of changes in biological diversity. Each publisher has different approach how to collect and publish their records, which makes difference in quality, range and accuracy. Task of this bachelor thesis is to have overview quality of records about marine mammals in database Global Biodiversity Information Facility (GBIF) a Ocean Biogeographic Information System (OBIS). From both databases records have been downloaded in 2015 and 2019 and than used by ArcGIS to have analytic results. Evaluation was based on defined criteria containing: duplicitous records, period of time when the record was taken, appearance of marine mammals by the coast lines, coordinates' accuracy, changes in records behavior and accordance of the records from both databases. Results proved that provided data, whose reports are not duplicated, situated in oceans and have stated time of obtaining record hold in the 2015 53 % (GBIF), 56,4 % (OBIS). Year 2019 has results 60,7 % GBIF and 53,8 % OBIS. Coordinates, which did not show any decimal place, reaches 20 % (GBIF) and 10 % (OBIS) in both years. Changes of records are providing huge differencies between years 2015 and 2019 due to lot of deleted records, which has been replaced by new records. Provides deficiencies prove that it is important to continue in verification of the record's quality and in minimization of existing errors. This is the only way to make global databases become an efficient tool for nature conservation.

**Keywords:** occurrence records, data quality. GBIF, OBIS, marine mammals

# Obsah

<b>1. Úvod</b> .....	<b>8</b>
<b>2. Cíle práce</b> .....	<b>9</b>
<b>3. Literární rešerše</b> .....	<b>10</b>
3.1 Primární data .....	10
3.2 Global Biodiversity Information Facility.....	11
3.3 Ocean Biogeographic Information System .....	11
3.4 Standardy globálních databází .....	12
3.4.1 Darwin Core a Integrated Publishing Toolkit .....	12
3.4.2 OBIS Schema .....	13
3.5 Chyby v datech .....	13
<b>4. Metodika</b> .....	<b>16</b>
4.1 Charakteristika vstupních dat .....	16
4.2 Identifikace duplicitních záznamů.....	17
4.3 Posouzení existence doby pořízení záznamu.....	19
4.4 Výskyty druhů vůči pobřežní linii .....	19
4.5 Evaluace přesnosti souřadnic.....	20
4.6 Změny chování dat.....	22
4.7 Shodné záznamy GBIF a OBIS.....	24
<b>5. Výsledky</b> .....	<b>26</b>
5.1 Identifikace duplicitních záznamů.....	26
5.2 Posouzení existence doby pořízení záznamu.....	26
5.3 Výskyty druhů vůči pobřežní linii .....	27
5.4 Celkové zhodnocení duplicitních záznamů, doby pořízení a výskytu druhů.....	27
5.5 Evaluace přesnosti souřadnic.....	29
5.6 Změny chování dat.....	31
5.7 Shodné záznamy GBIF a OBIS.....	33
<b>6. Diskuze</b> .....	<b>36</b>
<b>7. Závěr</b> .....	<b>38</b>
<b>Přehled literatury a použitých zdrojů</b> .....	<b>39</b>



# 1. Úvod

Druhová rozmanitost se v průběhu let neustále mění. Dochází k úbytku biotopů, mění se klimatické podmínky a v neposlední řadě jsou ohroženy jednotlivé druhy (Mittermeier et al., 2011). Příčinou těchto změn mohou být přírodní procesy jako orkány, povodně nebo přemnožení parazitů, ale také snaha lidí tyto procesy potlačit, čímž dochází k narušení základních principů fungování přírody. Pro hlubší pochopení těchto principů, vedoucích k aktivnímu zapojení do ochrany přírody, je potřeba integrovat a zkoumat velké množství dat (Otegui et al., 2013). V současné době jsou konkrétní data rozptýlena v mnoha databázích, archivech či jiných médiích, které nejsou přístupné interaktivnímu vyhledávání (Edwards et al., 2000). A právě z tohoto důvodu vznikají globální informační systémy, jejichž hlavním úkolem je integrace a správa dat o biologické rozmanitosti (Canhos et al., 2004). Mezi největší celosvětové datové portály patří Global Biodiversity Information Facility (GBIF), s více než jedním bilionem záznamů (Mesibov, 2013; Yesson et al., 2007) a Ocean Biogeographic Information System (OBIS), s více než 56 miliony záznamů (Costello et al., 2007). Vydavatelé těchto dat musí čelit výzvě zlepšování kvality, užitečnosti a rozsahu pokrytí. Zvláště pokud se jedná o rozvoj výzkumu, směřování vědy či sledování vývoje druhů (Faith et al., 2010). Nedostatky a rozdíly se v kvalitě dat mohou lišit z pohledu taxonomického, časového i prostorového. V uplynulých letech byly vyvinuty nástroje kontroly kvality, které jsou schopny dokumentovat úplnost a kvalitu každého distribučního záznamu. V těchto nástrojích jsou ale zatím jen základní filtry, ve kterých se podmínky vyhledávání nedají kombinovat a z důvodu množství dat, navíc není možné oslovit všechny poskytovatele. Kvalita záznamů tak zůstává napříč databázemi stále kompletně nevalidována (Vandepitte et al., 2015).

Tato práce se zabývá hodnocením kvality dat druhových databází GBIF a OBIS. Jako studovaná data byla vybrána skupina Mořských savců, kteří většinu nebo podstatnou část svého života tráví v oceánech. Jedná se o zástupce z řad kytovců, sirén a ploutvonožců. Důvodem výběru této zájmové skupiny je skutečnost, že se jedná o dlouhodobě sledované druhy (např. velryby), které jsou v mnoha případech ohroženy vyhynutím. Navíc se často jedná o charismatické druhy. V biologické rozmanitosti hrají mořští savci velmi významnou roli. Dokáží ovlivnit strukturu a funkci vodních společenstev, a to hlavně díky jejich hojnosti a velikosti (Bowen, 1997).

## 2. Cíle práce

Cílem této bakalářské práce je zhodnocení kvality záznamů mořských savců, v druhových databázích GBIF a OBIS pomocí nástrojů geografického informačního systému. Parametry pro hodnocení kvality dat se v této práci rozumí:

1. Duplicitní záznamy
2. Doba pořízení záznamu
3. Výskyty druhů vůči pobřežní linii
4. Přesnost souřadnic
5. Změny chování dat
6. Shodné záznamy GBIF a OBIS

*„Neexistuje žádný zásadní rozdíl mezi mentálními schopnostmi lidí a vyšších savců.“*

- Charles Darwin

### 3. Literární rešerše

Druhová rozmanitost se v posledních letech mění velmi rychle. A to především jako reakce vyvolaná působením člověka (Sala et al., 2002). Kombinace neudržitelné spotřeby zdrojů ve vyspělých zemích a přetrvávající chudoby v zemích rozvojových silně narušuje ekosystémové procesy (Mittermeier et al., 2011). Narušení dosahuje rozsáhlého měřítka, a tak je v současné době změna druhové diverzity považována za důležitou globální problematiku (Sala et al., 2002). Pro měření dopadu působením antropogenními vlivy je vyžadováno efektivnějšího přístupu k organizovaným datům (Robertson et al., 2014). V této oblasti jsou informační a komunikační technologie neustále rozvíjeny. Umožňují nové zkušenosti s integrací, analýzou či vizualizací informací a vedou tak k novému rozvoji výzkumu. Dostupnost těchto informací poskytují globální a regionální společnosti (Canhos et al., 2004).

#### 3.1 Primární data

Primární údaje o biologické rozmanitosti (PBD) zahrnují oblasti druhové diverzity, která vychází z biogeografie, makroekologie, krajinné ekologie a několika dalších oborů (Anderson et al., 2016). Dle Chavan et Penev (2011) jsou definovány jako základní atributy pozorování výskytu druhu. Konkrétně se jedná o zeměpisnou polohu, datum, vědecký název, fenotypová měření, hlasové záznamy a další. Správa a provoz rozsáhlých databází PBD vyžaduje překonání technických a organizačních překážek, včetně výzev pro poskytovatele a uživatele dat. Zejména pokud mají tyto záznamy heterogenní původ (Chavan et Penev, 2011; Anderson et al., 2016). Primární data se dají získávat několika způsoby i s různou mírou kvality. Prvním zdrojem jsou historická data v přírodovědných muzeích, herbářích či vládních zprávách. Kvalita těchto záznamů dosahuje nadprůměrných hodnot. Data s největší mírou kvality jsou nejčastěji publikována vědeckými pracovníky, především proto, že mají odborné znalosti. Poslední skupinou poskytovatelů dat může být kdokoliv, bez širších znalostí, vlastníci GPS. Toto z velké části zapříčiňuje sníženou kvalitu těchto záznamů (Anderson et al., 2016).

## 3.2 Global Biodiversity Information Facility

V roce 1996 provedla skupina Megascience Forum, spadající pod Organisation for Economic Co-operation and Development (OECD), výzkum v odvětví biologické rozmanitosti. Výsledkem tohoto výzkumu bylo zjištění, že existuje obrovské množství dat a informací, které se z důvodu rozptýlení nedají plnohodnotně využít. První myšlenkou vedoucí k nápravě této situace bylo založení internetového portálu, který bude data shromažďovat, spravovat a volně poskytovat všem uživatelům. V roce 2001 byla za tímto účelem založena databáze Global Biodiversity Information Facility (GBIF), která již v prvních dvou letech své existence zpřístupnila celkem 35 miliónů dat z 239 územních celků (Dooley, 2002; Saarenmaa, 2005). Podle Chavan et Penev (2011) tato databáze i nadále vykazuje velký nárůst v množství objemu dat. V současnosti obsluhuje data od více než 55 zemí, 47 mezinárodních organizací a poskytuje přístup k téměř jednomu bilionu záznamů pro více než 1,6 milionů druhů (Anderson et al., 2016; Mesibov, 2013; Yesson et al., 2007). Na první pohled se může zdát, že tato čísla vykazují značnou část záznamů o biologické rozmanitosti, přičemž ve skutečnosti pokrývají pouhých 20 %. Zbývajících 80 % dat není potencionálnímu uživateli zpřístupněno. Jedná se o data výzkumná či pouze lokální (Chavan et Penev, 2011).

V současnosti je jedním z hlavních cílů GBIF podpora vědecké spolupráce. Snaha spočívá v digitalizování záznamů ve vyspělých zemích, která byla původně shromážděna v jiných částech světa. Data pak mohou být snadno sdílena se zeměmi původu. Databáze dále umožňuje vyhledávat georeferencované záznamy výskytu druhů, které jsou přehledně popsány a vizualizovány. Kontroluje společné a vědecké názvy organismů, včetně synonym a získává seznam taxonů podle přístupu jednotlivých zemí. Veškeré tyto snahy vedou k postupnému překonávání „digitální propasti“ a hlavně k rozšíření znalostí o druhové rozmanitosti (Edwards, 2004).

## 3.3 Ocean Biogeographic Information System

V roce 1997 pořádala nadace Sloan Foundation workshop, na který pozvala vědecké skupiny z mnoha přírodovědných oborů. Řešeno bylo prostředí světových oceánů a v mimo jiné také pořizování mořských biogeografických dat (Grassle et Stock, 1999). Na základě tohoto poznatku byla v roce 1999, vědeckým programem

Census of Marine Life (CoML), založena databáze Ocean Biogeographic Information System (OBIS). Tento program funguje na principu strategického spojení lidí i organizací, které sdílejí vizi pro vytváření a publikování dat druhové diversity. Databáze OBIS tedy nepatří do skupiny vládních organizací, tak jako je tomu například u GBIF (Zhang et Grassle, 2003). V roce 2004 databáze poskytovala téměř 10 milionů dat pro 61 000 oceánských druhů z více než 100 databází (Berghe et al., 2004). Podle Zhang et. Grassle (2003) mohou uživatelé stáhnout více než 400 000 záznamů o událostech, které byly zaznamenány ve světových oceánech. Databáze OBIS je současně jedním z největších poskytovatelů dat portálu GBIF (O'Dor, 2004). Od července 2010 přispěla společnost OBIS 27,7 miliony distribučních záznamů, což představuje téměř 14% z celkového počtu dat v databázi GBIF (Vandepitte et al., 2011).

Mezi největší cíle portálu patří publikování záznamů o oceánských druzích a jejich prostředí, přičemž datové sady jsou integrovány a tudíž se dají vyhledávat podle názvu druhu, vyšší taxonomické úrovně, geografické oblasti, hloubky a času (Berghe, 2004).

### **3.4 Standardy globálních databází**

#### **3.4.1 Darwin Core a Integrated Publishing Toolkit**

Darwin Core Archive (DwC-A) je mezinárodní standard a preferovaný formát pro publikování dat o druhové rozmanitosti (Chavan et Penev, 2011). V praxi se stahuje jako komprimovaný soubor ZIP, který obsahuje propojené textové soubory a umožňuje vydavatelům sdílet svá data pomocí běžné terminologie. Textové soubory jsou rozděleny na základní a rozšiřující. Základní soubor se dále rozděluje na jádra taxonu, výskytu a události. Jádro taxonu uvádí seznam druhů typicky pocházejících ze stejné oblasti nebo sdílejících společné charakteristiky. Jádro výskytu uvádí seznam časů a míst, ve kterých byly jednotlivé druhy zaznamenány. A jádro události, které uvádí studie v terénu, včetně použitých protokolů, velikosti vzorku a umístění každého z nich. Rozšiřující soubory pak definují širší skutečnosti, které připojují k základnímu souboru vztahy 1:N. Databáze GBIF v současné době indexuje přibližně 300 milionů záznamů ve formátu Darwin Core, které vydalo více než 340 organizací ve 43 zemích (Wieczorek et al., 2012).

Integrated Publishing Toolkit (IPT) je softwarový standard, který se používá k publikování dat druhové rozmanitosti. Existující datové sady jsou kódovány do formátu Darwin Core Archive a poté sdíleny napříč databázemi (Endresen et al., 2009). Nejnovější verze IPT v2.2 zavádí Digital Object Identifiers (DOI) jako unikátní označení datových sad. Užitečnost se odráží v jednoduchém modelu pro citování a schopnosti vydavatelů sledovat, jak a kde jsou určitá data použita (GBIF, 2015). Všechny tyto přístupy pomáhají ke zlepšení interoperability dat, která je v tomto případě klíčová (Wieczorek et al., 2012).

### **3.4.2 OBIS Schema**

OBIS Schema je podporovaný datový formát neboli seznam datových polí s názvy, popisy a poznámkami (Halpin et al., 2006) Jedná se o rozšíření formátu Darwin Core verze. Oproti Darwin Core formátu obsahuje další pole pro uchovávání informací. To ale neznamená, že pokud jsou data vytvořena v OBIS Schema nebudou kompatibilní s Darwin Core. OBIS Schema má k dispozici čtyři pole, zeměpisná šířka, délka, taxonomický název, datum a čas poslední úpravy. Pokud je potřeba zahrnout jiná pole nebo přejmenovat stávající, stačí tak do schématu zadat jaká. Většina databázového softwaru umožňuje provádět automatické operace. Mezi jeho přednosti patří zadávání informací o poloze ve stupních a minutách, místo desetinných čísel, anebo zaznamenávání štítkové studie, což znamená opakovaný záznam pozorování jednoho jedince. Všechny tyto příkazy musí být v souladu s pokyny OBIS Schema (Costello et al., 2007).

## **3.5 Chyby v datech**

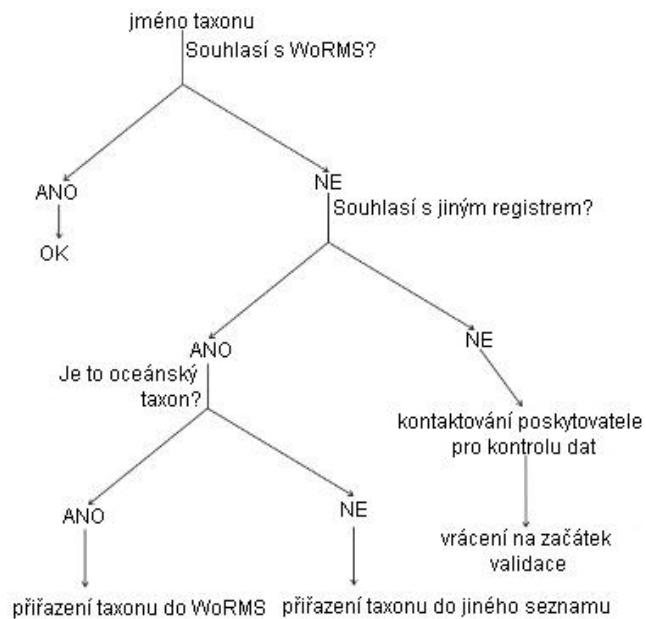
Při digitalizaci dat druhové rozmanitosti dochází k mnoha chybám (Faith et al., 2013). Dle Wheeler et al. (2004) je podstatná část exemplářů chybně taxonomicky identifikována, některá data jsou nesprávně georeferencována a objevují se také duplicitní záznamy. V takto obrovském množství dat, jako vykazuje databáze GBIF či OBIS, jsou chyby běžné a očekávané, ale zároveň nesmí být podceňované nebo dokonce ignorované. Správné pochopení chyb a nalezení jejich řešení vede k aktivní kontrole dat a následnému zlepšení požadované kvality (Chapman, 2005).

Velká část dat druhové rozmanitosti obsahuje neúplné, prohozené či špatné údaje o zeměpisných souřadnicích nebo v horším případě neobsahují souřadnice

vůbec. Dochází k tomu v důsledku nepozornosti, kdy jsou například zaměněna znaménka, anebo je vyplněná hodnota „nula“ brána jako správný údaj (Yesson et al., 2007). Pokud by tato data byla použita při biogeografických studiích, docházelo by k velkým nepřesnostem (Vandepitte et al., 2015). Z tohoto důvodu je validace dat a následný georeferencing nezbytný. Georeferencování dat je podmíněno časovými možnostmi a náklady jednotlivých společností, a to především při geokódování velkého množství dat z muzeí a herbářů. Nicméně, podle Soberón et Peterson (2004), může být 70 až 80% dat georeferencováno pomocí jednoduchých technik a automatických online přehledů. Zbývající data se buď nepodaří lokalizovat, anebo musí být podrobeny složitým analýzám a následně konzultovány s odborníky (Canhos et al., 2004).

Neúplnost či chybovost v časových údajích značí sníženou kvalitu datových sad (Amano, 2016). Jako časový údaj je brán den, měsíc a rok pořízení záznamu. Nejčastěji je zaznamenáván rok výskytu druhu, který je z hlediska zkoumání populací v jednotlivých letech zásadní. V jiných studiích, jako je sezónnost výskytu druhů, je tato informace nedostačující. Z tohoto důvodu jsou časové údaje kontrolovány a pokud možno doplňovány (Vandepitte et al., 2015).

Dalším velkým problémem v oblasti kvality dat druhové rozmanitosti je taxonomie. Při prozkoumávání záznamů bylo zjištěno, že značná část dat obsahuje neúplné taxonomické zařazení anebo jsou v něm menší či větší chyby. Tento problém vzniká nejčastěji díky neznalosti taxonomických skupin při pořizování dat. Tomuto problému čelí dobře databáze OBIS, která používá World Register of Marine Species (WoRMS) neboli Světový registr mořských druhů jako standardní seznam pro taxonomické zařazení (Obr. 1). WoRMS je autoritativní taxonomický seznam druhů vyskytujících se v celosvětovém měřítku. Všechny názvy taxonů jsou v databázi shodné, vylučují odchylky pravopisu a často řeší použitá synonyma. Naopak taxony s neznámým zařazením do druhu jsou automaticky zařazeny do první vyšší taxonomické úrovně (Vandepitte et al., 2011).



Obr. 1 Schéma schvalování dat dle WoRMS, upraveno, dostupné na adrese eurobis.org

Z těchto důvodů by měly být tyto záznamy zpřístupněné tak, aby je mohli srovnávat a upravovat kvalifikovaní vědci (Edwards, 2000). Klíčová je také zpětná vazba uživatelů, kteří mohou veškeré chyby které naleznou při používání, ohlásit (Faith et al., 2013).



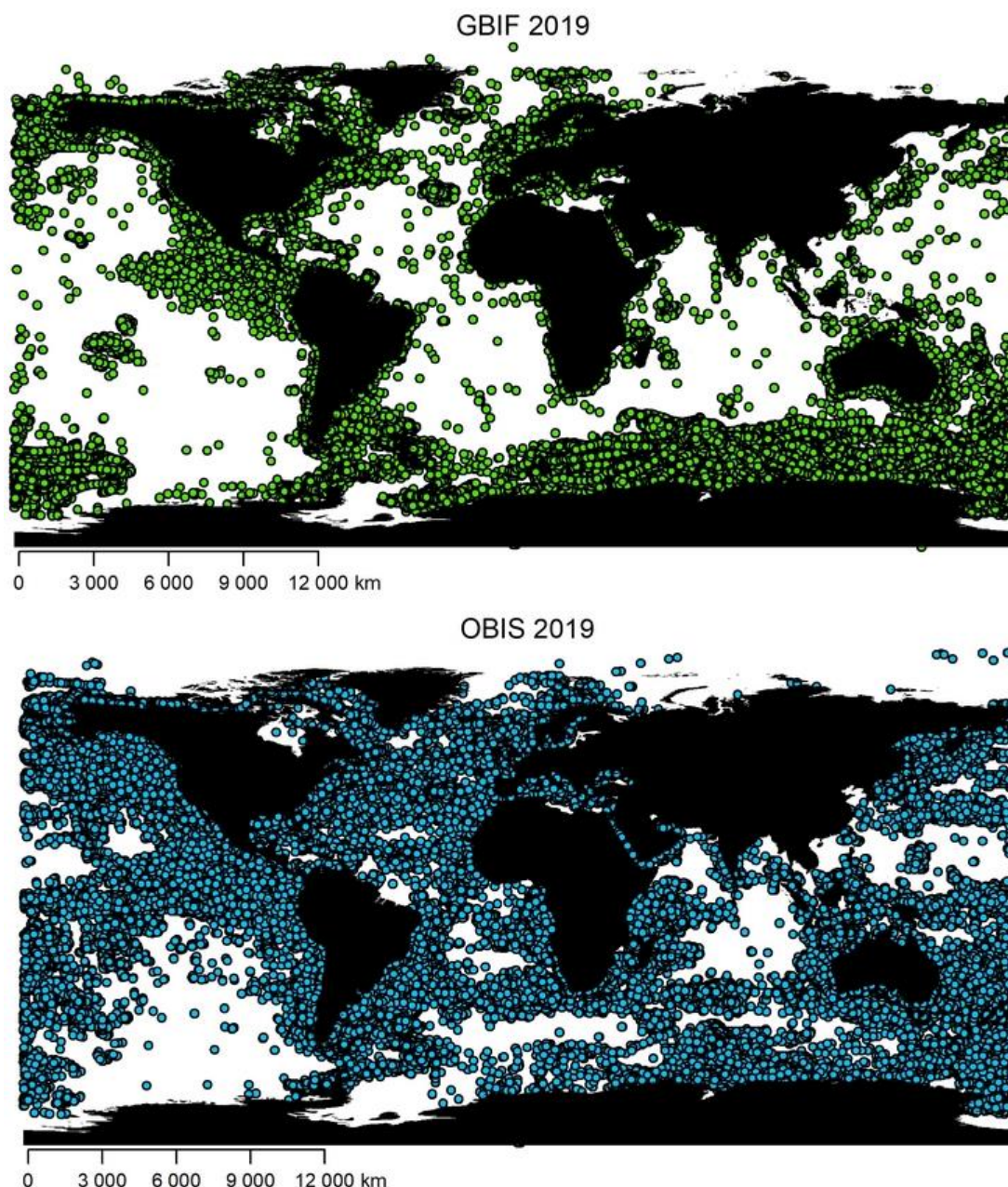
## 4. Metodika

Tato kapitola je zaměřena na charakteristiku vstupních dat a postup jednotlivých analýz vedoucích k výsledkům práce. Veškeré analýzy byly zpracovány v prostředí platformy ArcGIS for Desktop 10.6.

### 4.1 Charakteristika vstupních dat

Jako vstupní data byla vybrána skupina Mořských savců, kteří netvoří samostatnou taxonomickou skupinu ani pro ně neexistuje jednoznačná definice (Fair et Becker, 2000). Skupiny savců, které jsou takto označovány, definuje zákon o ochraně mořských savců z roku 1972. Zástupci této rozmanité skupiny patří do tří základních řádů: Kytovci (*Cetacea*), Sirény (*Sirenia*) a Šelmy (*Carnivora*). Řád kytovců je tvořen velrybami, sviňuchami a velmi atraktivní skupinou delfínů. Do řádu Sirén patří kapustňáci a dugongové. A poslední řád šelem zahrnuje vydry mořské, polární medvědy a podřád Ploutvonožců (*Pinnipedia*), do kterého patří tuleni, lachtani a mroži (O'Shea et al., 2008). Zástupci mořských savců se liší ve své evoluční historii, nutričních a energetických požadavcích, preferencích kořisti, chování a jejich anatomické a fyziologické adaptace (O'Shea et al., 2008). Fair et Becker (2000) uvádí, že řád Kytovců a Sirén žije výhradně ve vodě, zatímco tuleni a lachtani jsou přizpůsobeni životu ve vodě i na souši. Z hlediska výskytu jsou mořští savci různorodější. Velryby, delfíni a sviňuchy se vyskytují ve světových oceánech a mořích nebo dokonce i v některých sladkovodních řekách. Sirény bývají zastoupeny v pobřežních a subtropických biotopech. Ploutvonožci obývají především mírné a polární oblasti. Z globálního hlediska se tyto zástupci vyskytují téměř po celém světě.

Datové sady jednotlivých druhů byly staženy prostřednictvím geoportálů GBIF a OBIS. Stažení dat proběhlo ve dvou intervalech, v roce 2015 a 2019. Konkrétně se jedná o textové matice informací, obsahující výskyty druhů Mořských savců. Z důvodu prostorové vizualizace dat byla aplikována funkce *Create Feature Class* v souřadnicovém systému *World Geodetic System 1984 (WGS84)*, (Obr. 2).

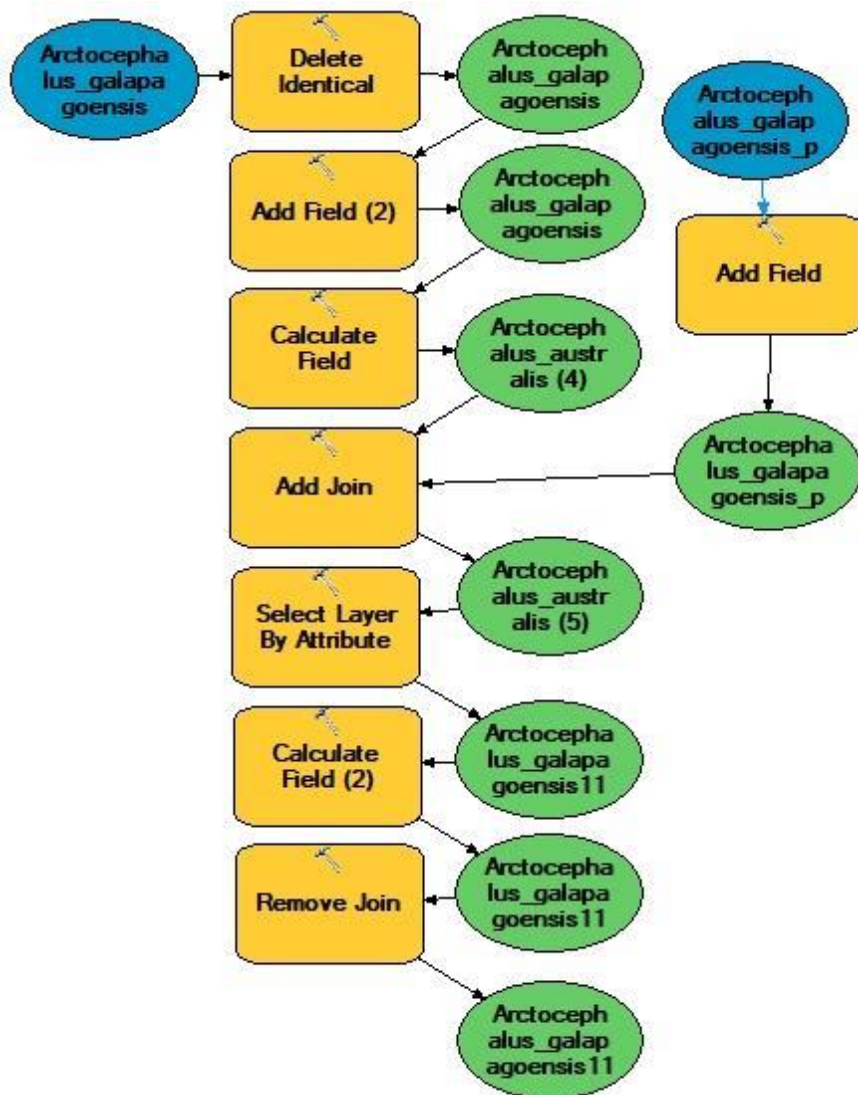


Obr.2 Vizualizace dat v databázích GBIF a OBIS z roku 2019, zdroj: autor

## 4.2 Identifikace duplicitních záznamů

Duplicitní záznamy jsou jednou z nejčastějších chyb objevující se v druhových databázích. Pro jejich odstranění byla použita funkce *Delete Identical*, která negeneruje žádné nové výstupy, ale pouze prochází a odstraňuje záznamy z původní vrstvy. K získání informace, zda je konkrétní záznam duplicitní či nikoliv musela být původní data rozšířena o nový sloupec pomocí funkce *Add Field* a následně propojena s výstupními daty funkcí *Add Join*. Propojení proběhlo na

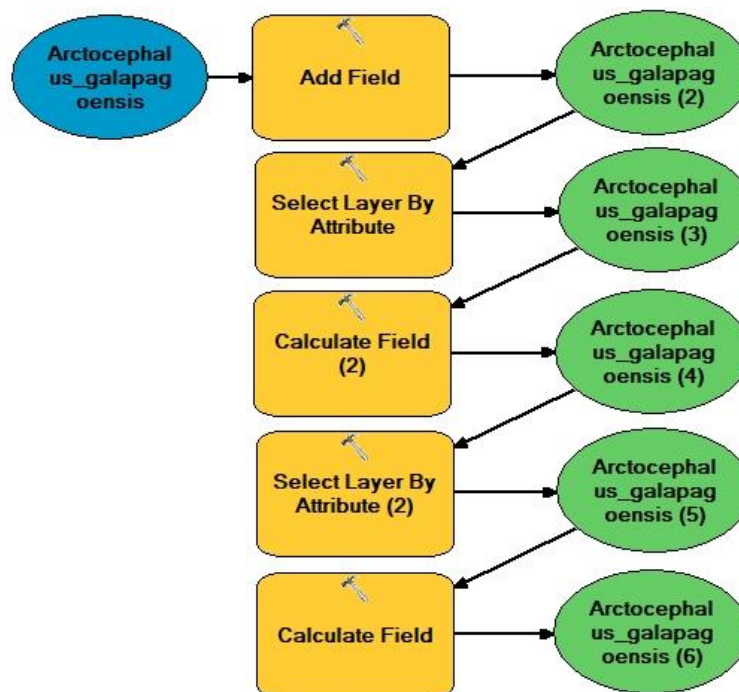
základě ID, které má každý záznam v tabulce unikátní. Tohoto faktu bylo využito při aplikaci funkce *Select Layer By Attributes*, která vybírá data podle zadaného příkazu. Pokud sloupec ID obsahoval záznam s hodnotou Null, znamená to, že je skutečně duplicitní. Pro uchování výsledků, byl použit nástroj *Calculate Field*, který vybraná data zapsal do nově vytvořeného sloupce. Selekcce dat byla následně otočena a zbylá data označena jako originál, viz Modelbuilder (Obr. 3).



Obr. 3 Modelbuilder duplicitních záznamů, zdroj: autor

### 4.3 Posouzení existence doby pořízení záznamu

Pro zjištění absence doby pořízení záznamů bylo využito jednoduchého atributového dotazování. K tomu aby mohly být informace o době pořízení zapsány a uchovány, bylo potřeba pomocí *Add Field* vytvořit nový sloupec a následně jej vyplnit na základě SQL dotazů přes *Calculate Field*. Pokud data informaci o datumu obsahovala, byla označena jako 1, pokud data informaci o datumu neobsahovala, pak byla označena jako 0. Z důvodu velkého množství dat byla analýza provedena v prostředí Modelbuilderu (Obr. 4).

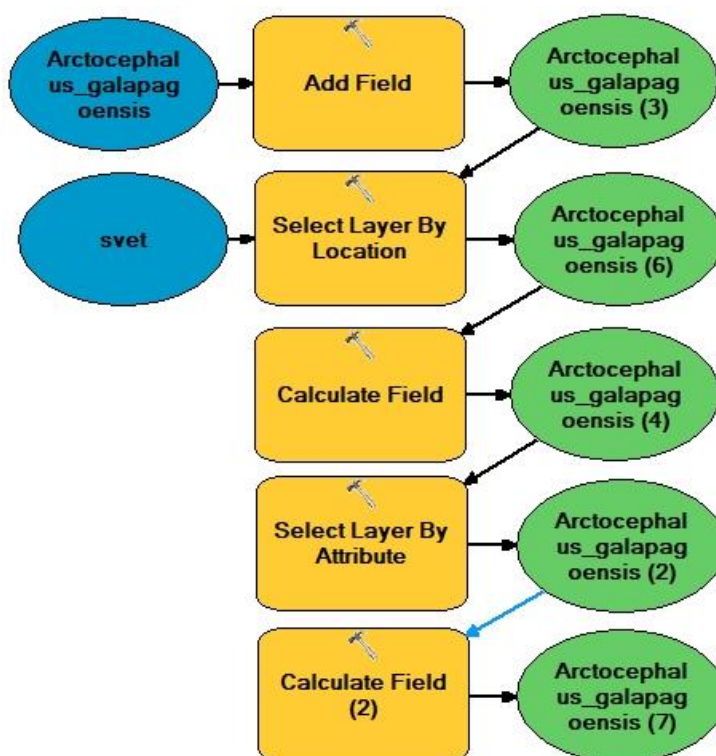


Obr. 4 Modelbuilder pro zjištění doby pořízení záznamů, zdroj: autor

### 4.4 Výskyty druhů vůči pobřežní linii

K určení výskytu druhů v moři či na pevnině bylo nutné stáhnout polygonový dataset pobřežní linie. V této práci byly použity vrstvy s největší přesností, dostupné na adrese: <https://www.ngdc.noaa.gov/mgg/shorelines/gshhs.html>. Z důvodu snadnějšího operování s vrstvami byly spojeny v jednu pomocí funkce *Merge*. V dalším kroku byla použita funkce *Add Field*, která do každé bodové vrstvy

Mořského savce přidala nový sloupec, do kterého bude následně zapsána informace o výskytu druhu. Pro získání polohy jednotlivého záznamu vůči pobřežní linii byla vybrána funkce *Select Layer By Location* s nastavením *Intersect*. Pokud se vrstvy protnuly, je zřejmé, že se jedná o záznam, který leží mimo přirozený ekosystém. Tyto záznamy byly pomocí funkce *Calculate Field* zapsány jako 1. Výběr dat byl následně otočen a zbylá data označena jako 0, tudíž ležící v oceánu (Obr. 5).

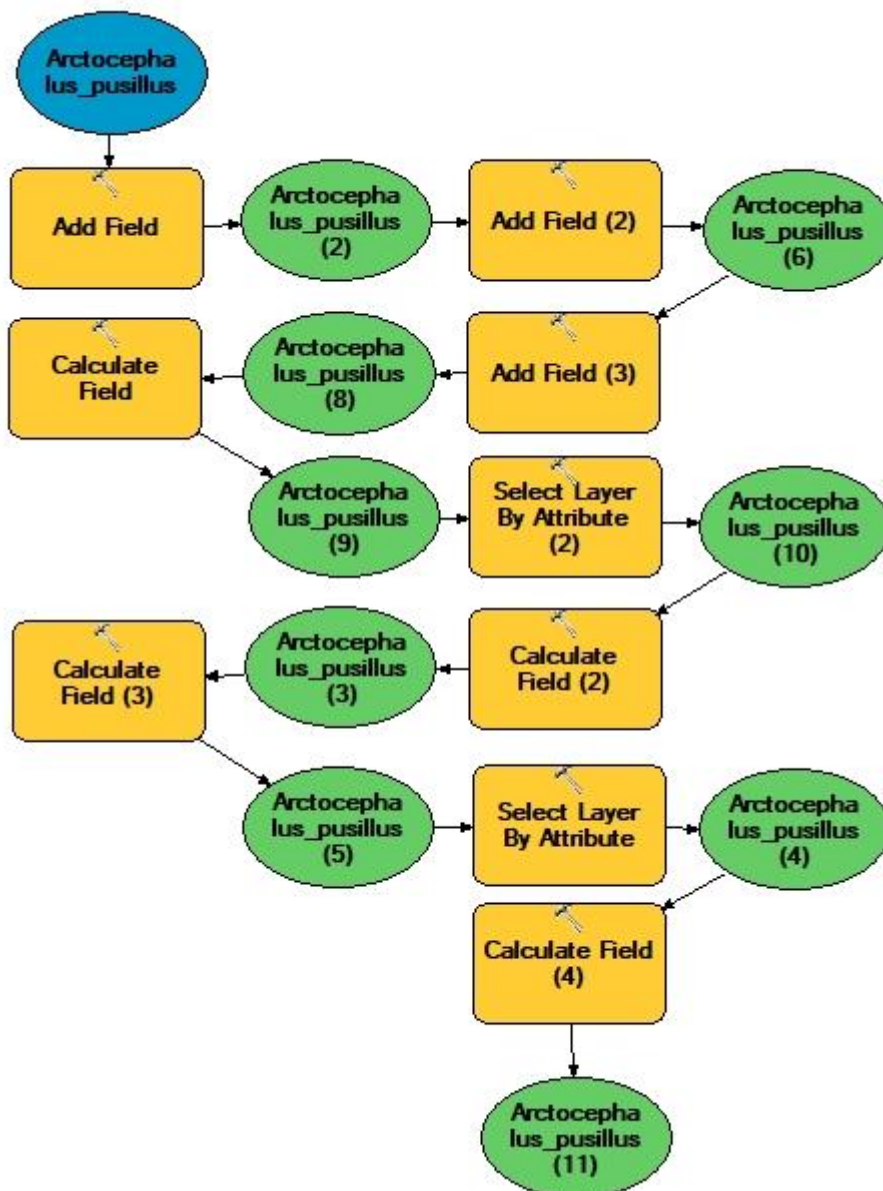


Obr. 5 Modelbuilder výskytů druhů na pevnině, zdroj: autor

## 4.5 Evaluace přesnosti souřadnic

Mezi skupinu indikátorů vypovídajících o nepřesnosti souřadnic patří počet desetinných míst zapsaných u jednotlivých záznamů. Aby mohla být nepřesnost spočtena, bylo nutné souřadnice z datového typu desetinného čísla převést na textové. K tomuto kroku bylo potřeba přidání nového sloupce pomocí *Add Field* a následný převod souřadnic nástrojem *Calculate Field*. Pro vybrání záznamů, které

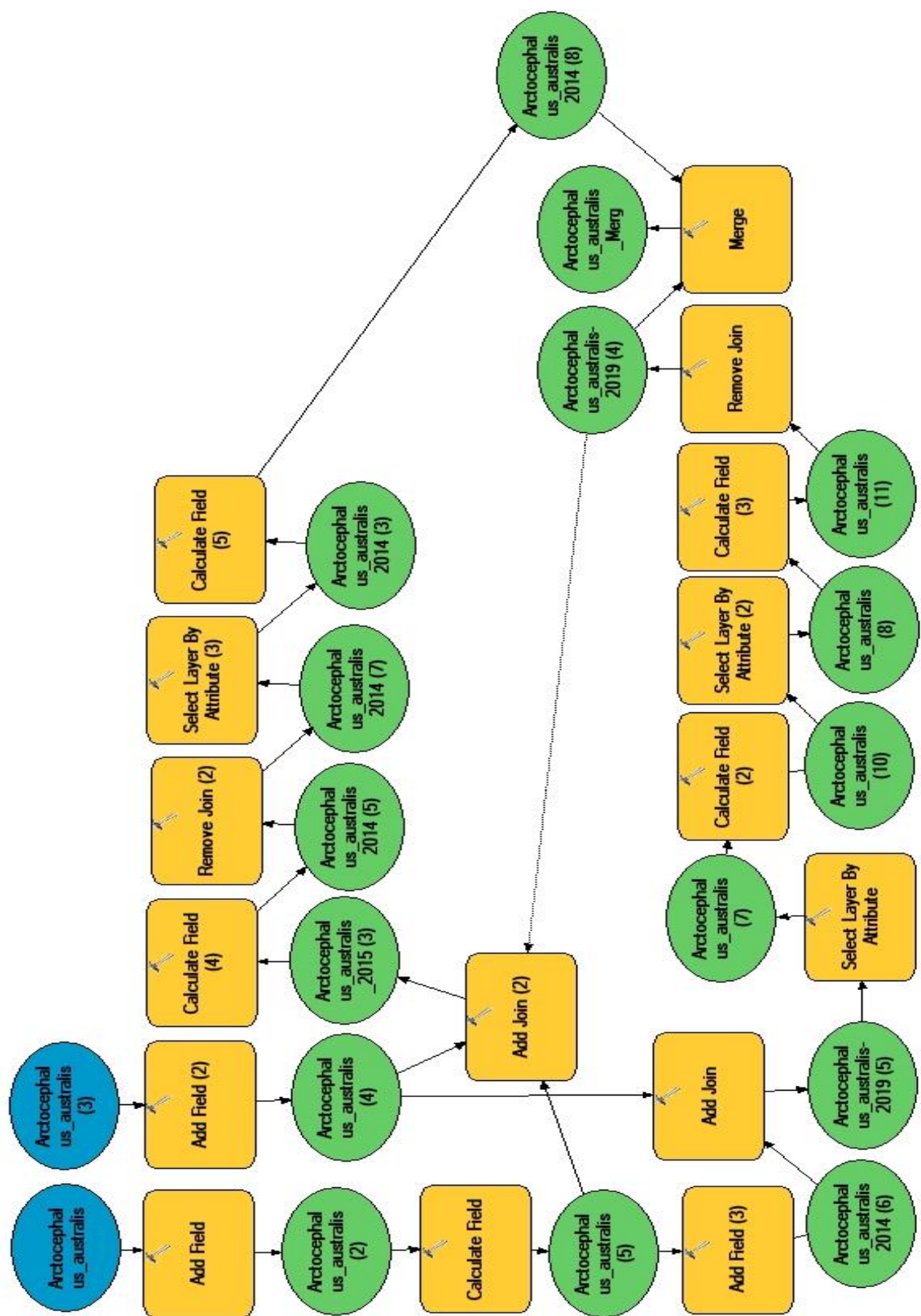
obsahovaly desetinný oddělovač (v prostředí ArcGIS desetinná tečka), bylo využito SQL dotazu v *Select Layer by Attributes*. Vybraná data byla poté překalkulována pomocí *Calculate Field* do dalšího nového sloupce, který byl vytvořen standardně nástrojem *Add Field*. Pro výpočet počtu čísel byla do nového sloupce využita funkce *Len*, která je součástí nástroje *Calculate Field*. Selekcce dat byla následně otočena a zbylá data označena jako 0, jelikož neobsahovala žádné desetinné místo. Celý postup byl následně opakován pro druhou souřadnici (Obr.6).



Obr. 6 Modelbuilder přesnosti souřadnic, zdroj: autor

## 4.6 Změny chování dat

V této práci se změnou chování dat rozumí rozdělení do tří skupin na základě odstranění, přetrvání či nového zaznamenání výskytů druhů v průběhu let 2015 až 2019. Právě z tohoto důvodu tvořily vstupní vrstvy záznamy z obou let zároveň. Aby mohly být později výsledky zapsány a uchovány, byla hned zprvu použita funkce *Add Field*, která do obou data setů přidala nový textový sloupec. V dalším kroku musela být data z roku 2019 připojena do tabulky dat z roku 2015. Připojení proběhlo na základě jejich společného ID pomocí funkce *Add Join*. Důvodem první selekce dat bylo zjistit jaké záznamy z roku 2019 obsahují ve sloupci ID hodnotu Null. Tyto záznamy byly poté označeny jako nově zaznamenané. Selekcce byla následně otočena a zbývající data zapsána jako pozůstalá neboli přetrvávající. Před aplikováním opačného propojení tabulek musela být použita funkce *Remove Join*, která zrušila propojení původní. Pak už následovala funkce *Add Join*, která fungovala na stejném principu jako v prvním případě. Další nástroj *Select Layer By Attributes* vybral data, u kterých ID v roce 2015 obsahovalo hodnotu Null. Tato data byla označena jako odstraněná, jelikož se v datové sadě 2019 již nevyskytují. Zbylé záznamy byly pomocí *Calculate Field* označeny jako pozůstalé neboli přetrvávající. Nakonec bylo funkcí *Remove Join* propojení tabulek zrušeno. Pro kontrolu správnosti byly jednotlivé skupiny následně sečteny a porovnány s celkovým počtem záznamů (Obr. 7).

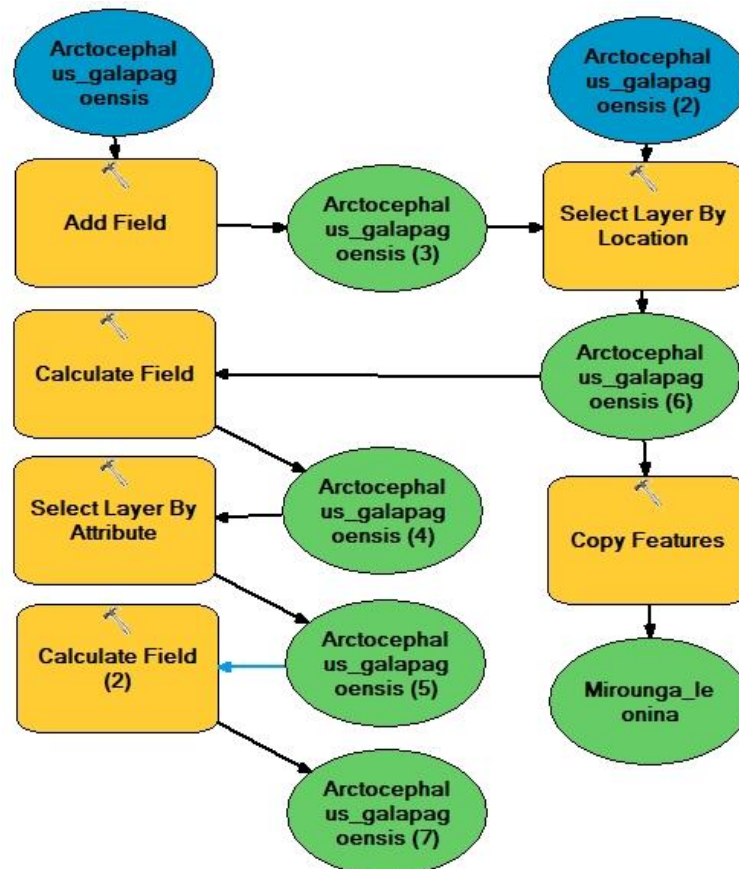


Obr. 7 Modelbuilder změny chování dat, zdroj: autor



## 4.7 Shodné záznamy GBIF a OBIS

Pro zjištění shodných záznamů, které poskytují obě databáze zároveň, bylo použito prostorové analýzy *Select Layer By Location* s nastavením *Intersect* neboli průsečíku. Data, která byla vybrána, znamenala shodu. Pro zapsání výsledků bylo využito dvou možností. V prvním případě byl do jedné z databází vytvořen nový sloupec, do kterého byla výsledná data zapsána pomocí *Calculate Field*. Selektce byla následně otočena a vybraná data označena jako neshodná. Ve druhém případě byla použita funkce *Copy Features*, která vygenerovala vybraná shodná data do nové vrstvy (Obr. 8).



Obr. 8 Modelbuilder pro zjištění shodných záznamů GBIF a OBIS, zdroj: autor

K následné vizualizaci těchto dat bylo nutné vytvořit síť polygonů zaujímavější plochu celého světa a poté spočítat počet záznamů, které připadnou na každý díl polygonové sítě. V prvním kroku byla použita funkce *Merge*, která spojila všechny bodové vrstvy záznamů v jednu. Pro vytvoření sítě, která pokryla území celého

světa, byla použita funkce *Create Fishnet*. Na obě vrstvy byl poté aplikován nástroj *Project*, který vytvořil projekci vrstev v koordinačním systému Cylindrical equal area. Tyto vrstvy byly vloženy do prostorové analýzy *Intersect*, která vygenerovala novou vrstvu obsahující průnik bodových záznamů s polygonovou sítí. Funkcí *Summary Statistics*, jejímž výstupem je atributová tabulka, byly následně spočteny záznamy, které připadly na jednotlivé díly polygonové sítě. V posledním kroku byla atributová tabulka propojena s původní sítí pomocí funkce *Add Join* (Obr. 9).

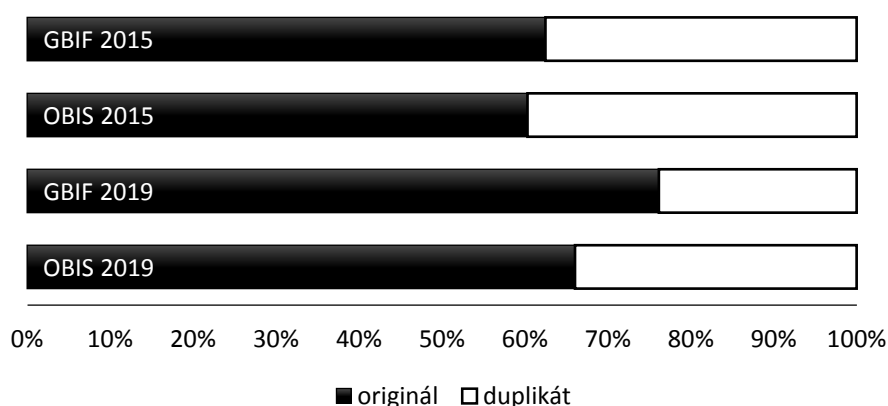


Obr. 9 Modelbuilder pro vizualizaci záznamů, zdroj: autor

## 5. Výsledky

### 5.1 Identifikace duplicitních záznamů

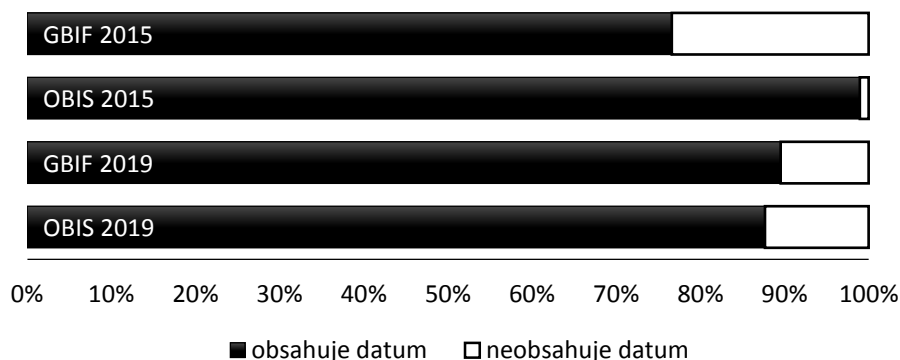
Jako duplicitní záznamy byla vybrána data, která se pro jednotlivý druh shodují v zeměpisných souřadnicích a době pořízení záznamu. Databáze GBIF v roce 2015 vykazovala cca 38 % duplicitních záznamů a OBIS přesně 40 % duplicitních záznamů. Naproti tomu obě databáze vykazují v roce 2019 významné snížení množství duplicitních záznamů: GBIF o cca 13 %, OBIS 8 % (Obr. 10).



Obr. 10 Grafické znázornění poměru originálních a duplicitních záznamů, zdroj: autor

### 5.2 Posouzení existence doby pořízení záznamu

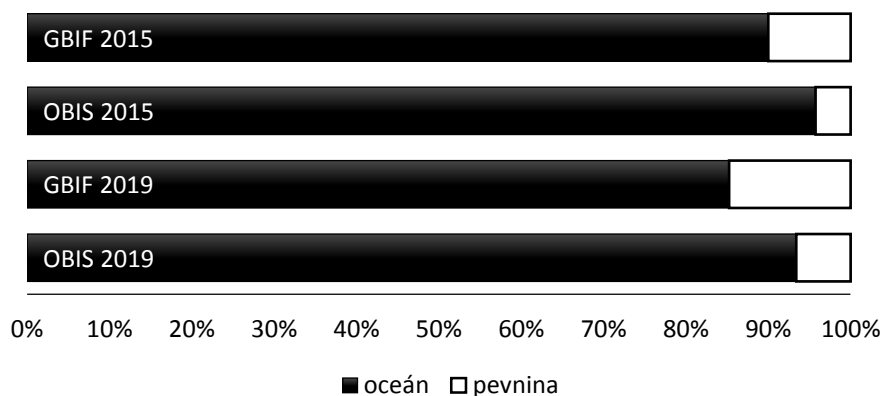
K posouzení existence doby pořízení záznamů bylo nutné vyhodnotit, zdali data obsahují informace o dni, měsíci a roku výskytu druhů. V případě databáze OBIS byla data v roce 2015 zaznamenána s 99 % úplností, což značí vysokou kvalitu dat. Databáze GBIF v tomto roce poskytovala data, která obsahovala dobu pořízení pouze ze 75 %. V dalších letech došlo u GBIF ke zlepšení zhruba o 15 %, oproti tomu databáze OBIS vykazovala snížení počtu dat, obsahující dobu pořízení záznamu o 10 % (Obr 11).



Obr. 11 Grafické znázornění poměru obsahu doby pořízení dat, zdroj: autor

### 5.3 Výskyty druhů vůči pobřežní linii

Z hlediska posouzení výskytu druhů vůči pobřežní linii si nejlépe vedl OBIS 2015, který vykazuje více než 95 % záznamů v oceánu. O něco hůř je tomu u databáze GBIF, která v roce 2015 zpřístupnila 90 % záznamů druhů v oceánu. Nejvíce výskytu druhů zaznamenaných na pevnině (15 %) zpřístupnila databáze GBIF v roce 2019, přičemž OBIS poskytovala pouhých 8 % (Obr. 12).

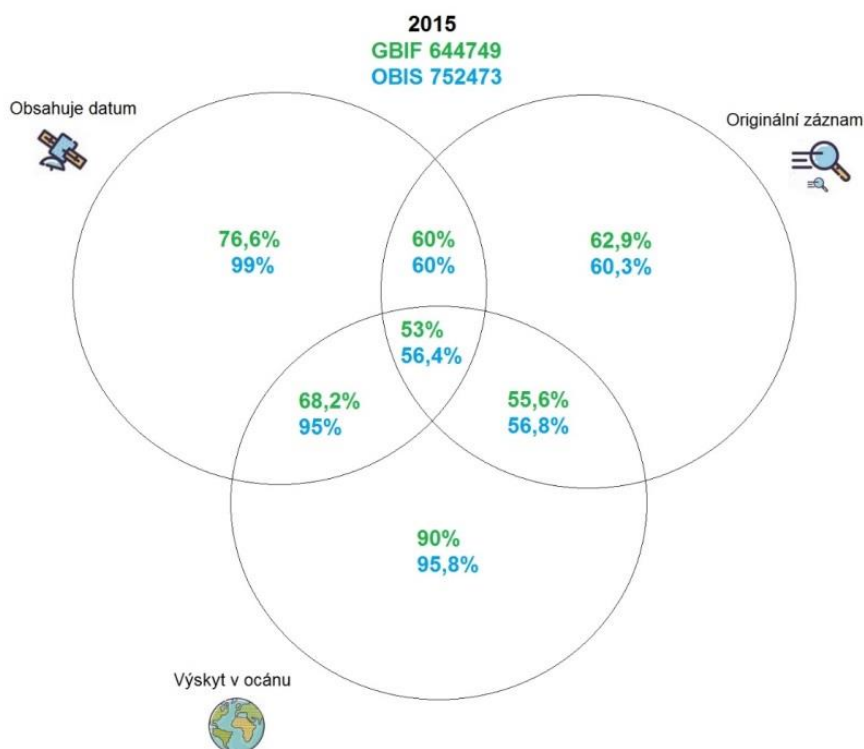


Obr. 12 Grafické znázornění poměru výskytu druhů, zdroj: autor

### 5.4 Celkové zhodnocení duplicitních záznamů, doby pořízení a výskytu druhů

Z Vennova diagramu (Obr. 13) je patrné, že v roce 2015 obsahovalo dobu pořízení záznamu a zároveň se vyskytovalo v oceánu 68,2 % záznamů databáze GBIF a 95 % z databáze OBIS. Originální záznamy, které se vyskytují v oceánu,

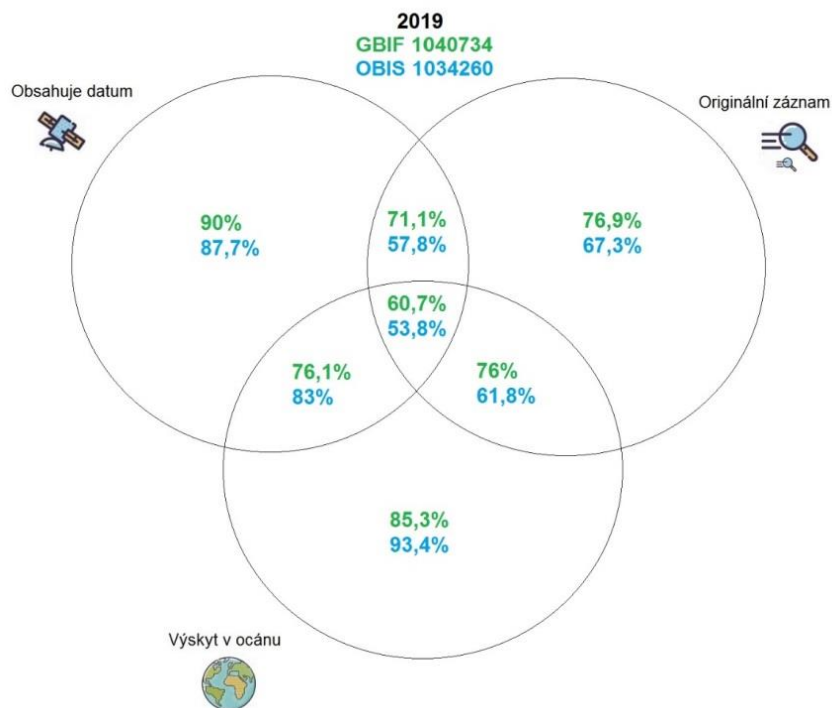
představuje 55,6 % (GBIF) a 56,8 % (OBIS). Obě databáze pak poskytují 60 % záznamů, které nejsou duplikáty a obsahují údaje o době pořízení záznamu. Při porovnání všech tří parametrů zároveň bylo zjištěno, že celkovou správnost splňuje pouhých 53 % z databáze GBIF a 56,4 % z databáze OBIS (Obr. 13). Celkově lze zhodnotit, že si záznamy z databáze OBIS vedly lépe.



Obr. 13 Celkové zhodnocení duplicitních záznamů, doby pořízení a výskytu druhů v roce 2015, zdroj: autor

V roce 2019 bylo pozorováno hned několik změn. Data, která obsahují dobu pořízení záznamu a jsou originální, zaujmají 71,1 % v databázi GBIF a 57,8 % v databázi OBIS. Což značí od roku 2015 přírůstek zhruba o 11 % (GBIF) a úbytek cca o 2 % v přesnosti záznamů OBIS. Originální záznamy, které se vyskytují v oceánu, čítají 76 % v databázi GBIF a 61,8 % v databázi OBIS. V případě těchto dvou parametrů došlo ke zlepšení přesnosti o 20,4 % (GBIF) a 5 % (OBIS). Z hlediska doby pořízení záznamu a výskytu druhů v oceánu zaznamenává databáze GBIF 76,1 % a OBIS 83 %. U těchto dvou parametrech databáze GBIF zpřesnila své údaje o 13 %, přičemž u OBIS došlo k opačnému efektu o 12 %. Také při hodnocení všech tří parametrů si databáze GBIF vede lépe (60,7 %), přičemž OBIS vykazuje pouhých 57,8 %. Na základě těchto údajů lze konstatovat, že

databáze GBIF vykazuje lepší výsledky, než tomu bylo v roce 2015. Opačný trend je pak pozorován u druhé databáze (Obr. 14).



Obr. 14 Celkové zhodnocení duplicitních záznamů, doby pořízení a výskytu druhů v roce 2019, zdroj: autor

## 5.5 Evaluace přesnosti souřadnic

V této práci se přesnost souřadnic posuzovala na základě počtu čísel za desetinným oddělovačem. V Tab. 1 jsou zaznamenány hodnoty zastoupení každé přesnosti souřadnic z databáze GBIF. Nejvyšší dosažená přesnost souřadnic zaznamenaná v roce 2015 dosahovala počtu 14 znaků. Oproti tomu v roce 2019 dosahovala dokonce 19 znaků. V roce 2015 byla u obou souřadnic nejčastěji zastoupena hodnota 3, a to z 27,5 % u zeměpisné šířky a 28,7 % u zeměpisné délky. Nejvyšší dosažená četnost byla v roce 2019 zaznamenaná u desetinné přesnosti dosahující čísla 6 u obou souřadnic s hodnotou 24,4 %. Což značí zlepšení v kvalitě přesnosti souřadnic od roku 2015 do roku 2019. Souhrnně lze říci, že se záznamy pohybují nejčastěji od 1 do 6 desetinných míst, konkrétně v průměru 76,5 % v roce 2015 a 77,4 % v roce 2019. Souřadnice bez desetinných míst se

v obou letech pohybují vždy okolo 21,6 %, což představuje jednu pětinu ze všech záznamů.

Desetinná přesnost	GBIF								
	2015				2019				Ø [%]
	z. šířka	%	z. délka	%	z. šířka	%	z. délka	%	
0	137811	21.4	139346	21.6	224713	21.6	225083	21.6	21.6
1	24389	3.8	24395	3.8	63570	6.1	63177	6.1	4.9
2	81282	12.6	80064	12.4	93362	9.0	92656	8.9	10.7
3	177399	27.5	185258	28.7	107167	10.3	115779	11.1	19.4
4	97829	15.2	86763	13.5	108865	10.5	100188	9.6	12.2
5	110662	17.2	118030	18.3	178532	17.2	179468	17.2	17.5
6	0	0.0	0	0.0	254054	24.4	254216	24.4	12.2
7	0	0.0	0	0.0	7	0.0	0	0.0	0.0
8	0	0.0	0	0.0	236	0.0	21	0.0	0.0
9	0	0.0	2	0.0	1309	0.1	204	0.0	0.0
10	8	0.0	24	0.0	96	0.0	1425	0.1	0.0
11	39	0.0	1385	0.2	0	0.0	0	0.0	0.1
12	1117	0.2	9337	1.4	33	0.0	105	0.0	0.4
13	14212	2.2	144	0.0	201	0.0	430	0.0	0.6
14	1	0.0	1	0.0	5415	0.5	7976	0.8	0.3
15	0	0.0	0	0.0	3182	0.3	1	0.0	0.1
16	0	0.0	0	0.0	0	0.0	4	0.0	0.0
17	0	0.0	0	0.0	1	0.0	0	0.0	0.0
18	0	0.0	0	0.0	0	0.0	0	0.0	0.0
19	0	0.0	0	0.0	0	0.0	1	0.0	0.0

Tab. 1 Přesnost souřadnic pro databázi GBIF, zdroj: autor

Oproti databázi GBIF zaznamenává OBIS větší počet souřadnic s vyšší přesností. V roce 2015 zaujímala zeměpisná šířka až 37,6 % u desetinné přesnosti 13, což jí také označuje za nejčtenější a v roce 2019 pak přesnost na 10 desetinných míst s 28,6 %. Záznamy, které nevykazovaly žádná desetinná místa, a tudíž byla označena jako velmi nepřesná, zaujímala v obou letech v průměru 10,9 %. Což je zhruba jedna desetina z celkového počtu. V roce 2015 záznamy zaujímaly přesnosti od 1 do 6 z 52,9 %. Přesnost dvojnásobně vyšší zaujímalo 37,6 %. Naproti tomu se však hodnota s přesností od 1 do 6 v roce 2019 zvýšila o zhruba 10 % a došlo k mírnému snížení přesnosti souřadnic mezi lety 2015 a 2019 (Tab. 2).

Desetinná přesnost	OBIS								
	2015				2019				Ø [%]
	z. šířka	%	z. délka	%	z. šířka	%	z. délka	%	
0	54838	7.3	55768	7.4	149061	14.4	150074	14.5	10.9
1	32037	4.3	23490	3.1	43698	4.2	35276	3.4	3.8
2	48204	6.4	48142	6.4	66132	6.4	65224	6.3	6.4
3	76459	10.2	77938	10.4	118913	11.5	120514	11.7	10.9
4	71550	9.5	74777	9.9	77632	7.5	76604	7.4	8.6
5	90226	12.0	83180	11.1	126941	12.3	120941	11.7	11.8
6	82380	10.9	86622	11.5	129361	12.5	137352	13.3	12.1
7	172	0.0	142	0.0	213	0.0	652	0.1	0.0
8	2413	0.3	1699	0.2	20854	2.0	26005	2.5	1.3
9	3	0.0	1040	0.1	4486	0.4	223913	21.6	5.6
10	2	0.0	3	0.0	296260	28.6	69708	6.7	8.8
11	1744	0.2	2513	0.3	53	0.0	7333	0.7	0.3
12	9030	1.2	227510	30.2	42	0.0	54	0.0	7.9
13	282885	37.6	61440	8.2	614	0.1	63	0.0	11.5
14	373	0.0	8172	1.1	0	0.0	544	0.1	0.3
15	131	0.0	37	0.0	0	0.0	3	0.0	0.0
16	0	0.0	0	0.0	0	0.0	0	0.0	0.0
17	0	0.0	0	0.0	0	0.0	0	0.0	0.0
18	26	0.0	0	0.0	0	0.0	0	0.0	0.0
19	0	0.0	0	0.0	0	0.0	0	0.0	0.0

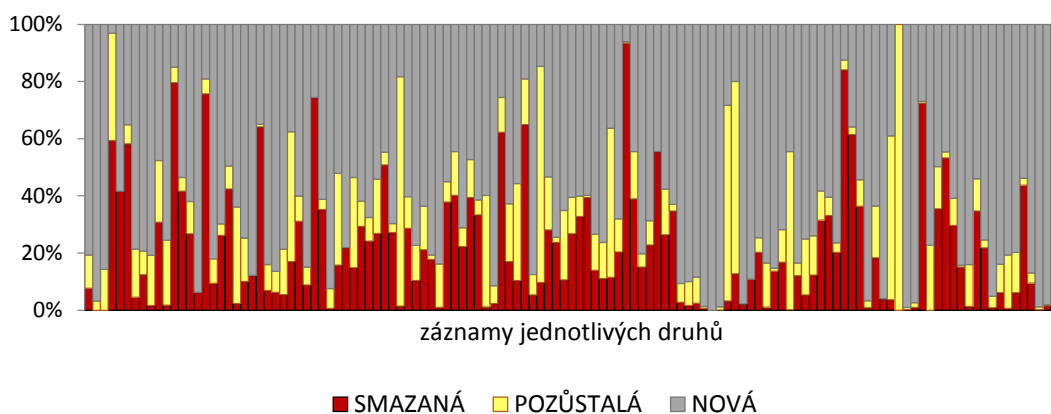
Tab. 2 Přesnost souřadnic pro databázi OBIS, zdroj: autor

## 5.6 Změny chování dat

Změnou chování dat se rozumí rozdělení do tří skupin na základě odstranění, přetrvání či nového zaznamenání výskytů druhů v průběhu let 2015 až 2019, viz kapitola 4.6. Dle grafického hodnocení (Obr. 15) je patrné, že se data do databáze GBIF častěji přidávají, než odstraňují, ačkoliv je tento rozdíl nepatrný. Oproti tomu tvoří pozůstalá data nejméně početnou skupinu.



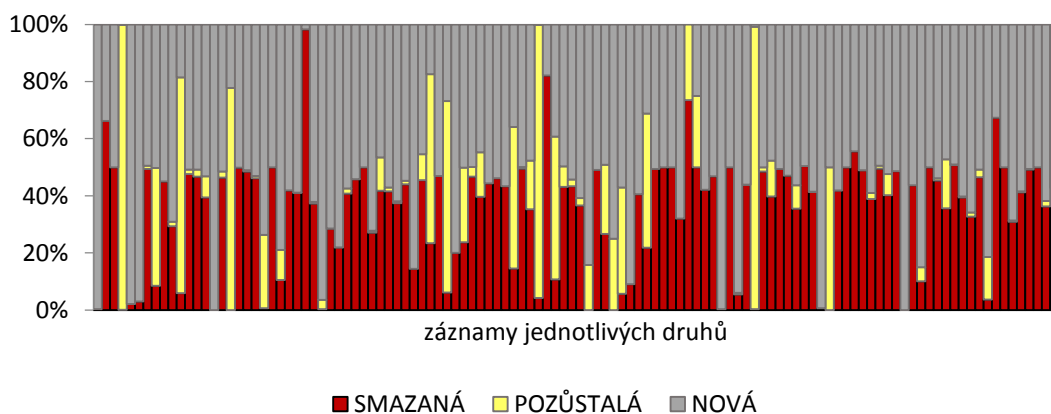
### Databáze GBIF 2015 - 2019



Obr. 15 Grafické znázornění Změn chování dat pro GBIF, zdroj: autor

Databáze OBIS zaznamenává ještě menší množství pozůstalých dat než GBIF. V podstatě lze říci, že se data od roku 2015 do 2019 rapidně změnila (Obr. 16). Nejvíce je zastoupena skupina nově zaznamenaných dat, což naznačuje trend ve sběru záznamů. Hodnota odstraněných dat dosahuje zhruba 40 %. Lze tak říci, že se databáze OBIS aktivně zapojuje do třídění záznamů.

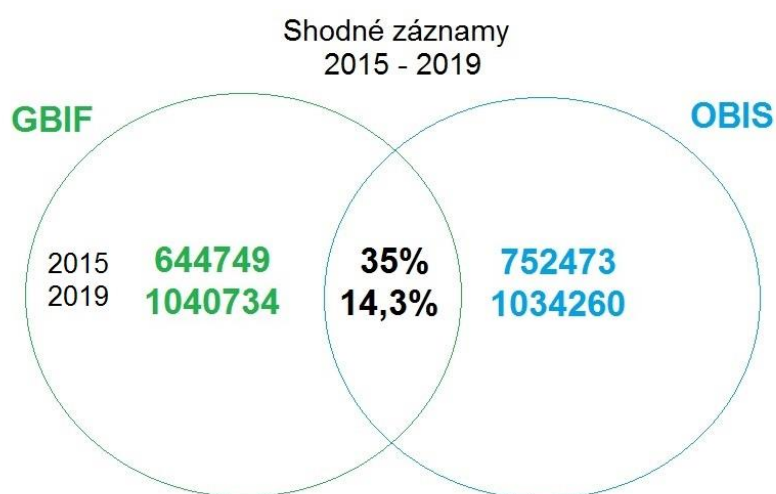
### Databáze OBIS 2015 - 2019



Obr. 16 Grafické znázornění Změn chování dat pro OBIS, zdroj: autor

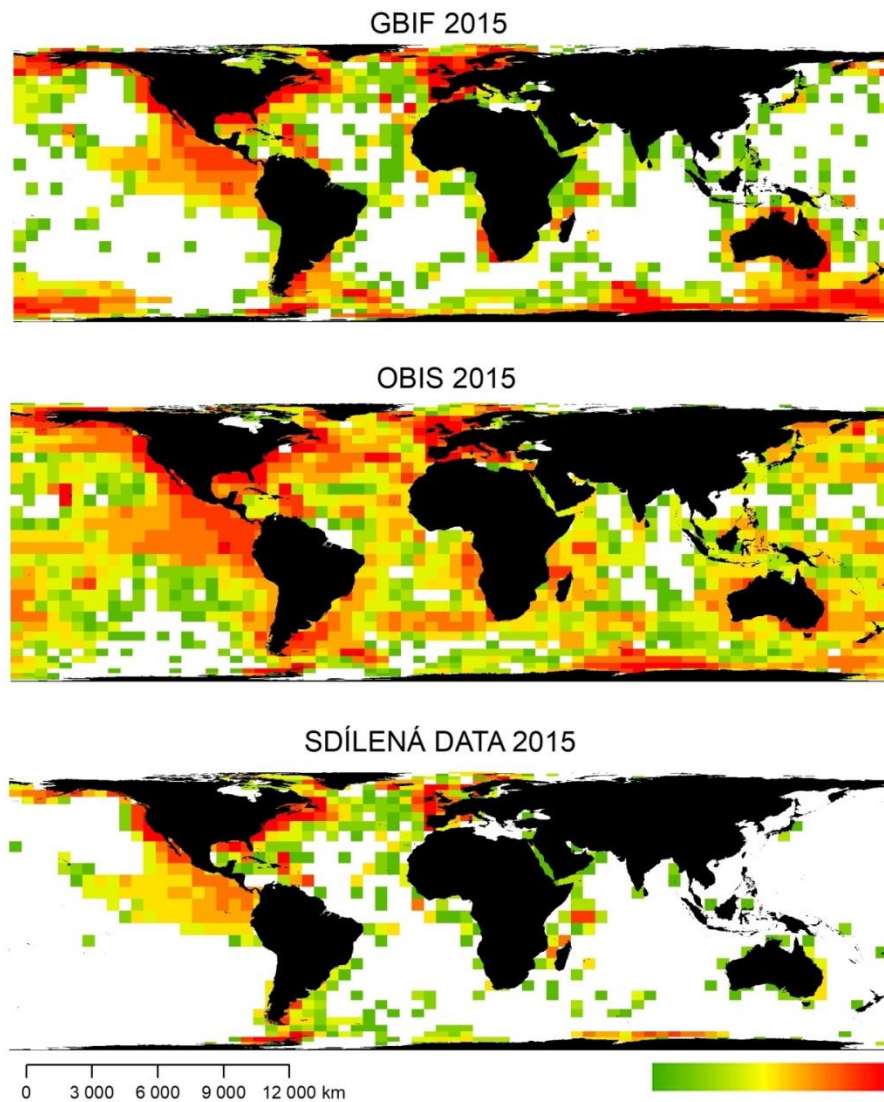
## 5.7 Shodné záznamy GBIF a OBIS

Shodné záznamy získané dle stejných zeměpisných souřadnic jsou zobrazené dvěma způsoby, (Obr. 17;18;19). V roce 2015 databáze GBIF vykazovala více než 644749 záznamů, z čehož bylo 35% poskytnuto databází OBIS. V následujících letech došlo k navýšení počtu dat zhruba o 60 %, avšak shodné záznamy se oproti tomu snížily na pouhých 14,3 % (Obr 17).



Obr. 17 Množinové zhodnocení shodných záznamů, zdroj: autor

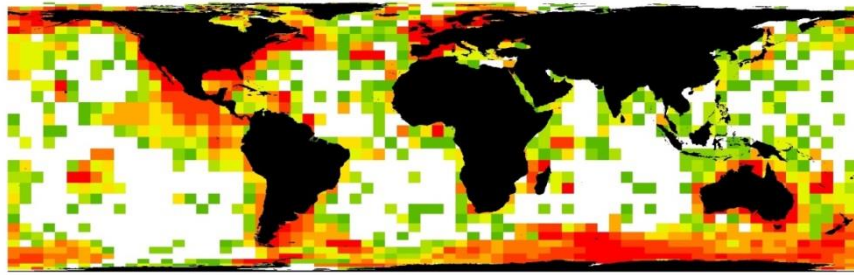
Jak je patrné z Obr. 18, databáze OBIS poskytovala lepší pokrytí světových oceánů než databáze GBIF. Shodné záznamy se v roce 2015 nejvíce vyskytovaly u pobřeží Spojených států, Kanady, Mexika, Skandinávie a Antarktidy. Nejčastěji zaznamenaný druh, shodný pro obě databáze, byl Plejtvák tmavý (*Balaenoptera physalus*) se 69688 záznamy. Dále pak Tuleň krabožravý (*Lobodon carcinophaga*) se 41826 záznamy, který se na Antarktidě vyskytuje velmi hojně.



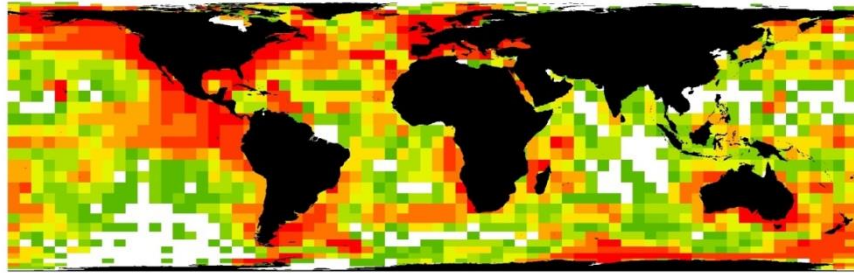
Obr. 18 Shodné záznamy pro rok 2015, zdroj: autor

V roce 2019 jsou rovněž viditelné změny v rozložení dat (Obr. 19). Obě databáze vykazují nárůst záznamů v oblasti Austrálie a Antarktidy, které pak mezi sebou sdílí. Navýšení počtu sdílených záznamů lze pozorovat i v okolí Jižní Ameriky. Oproti tomu bylo rozložení dat v blízkosti Spojených států a Mexika značně zredukováno. Nejčastěji zaznamenaný druh byl opět Plejtvák tmavý, u kterého došlo k odstranění 54130 záznamů a Tuleň krabožravý, jehož záznamy se zvýšily o 31 pozorování.

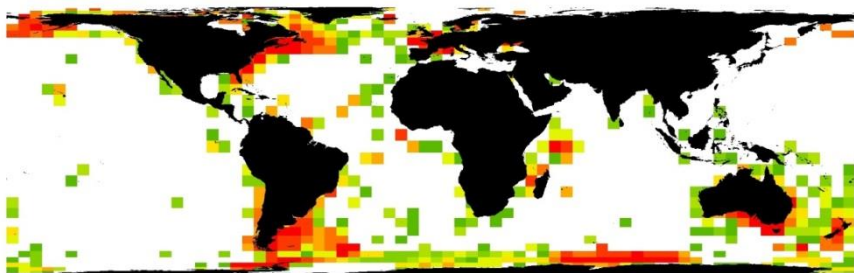
GBIF 2019



OBIS 2019



SDÍLENÁ DATA 2019



0 3 000 6 000 9 000 12 000 km



Obr. 19 Shodné záznamy pro rok 2019, zdroj: autor

## 6. Diskuze

Pro zhodnocení kvality dat v druhových databázích byly vybrány databáze GBIF a OBIS, které byly již dříve několikrát hodnoceny (Costello et al., 2007; Ficetola et al., 2014; Mesibov, 2013; Vandepitte et al., 2011; Otegui et al., 2013). Hodnocení proběhlo na základě stanovených parametrů, mezi které byly zahrnuty: duplicitní záznamy, doba pořízení záznamu, výskyty druhů vůči pobřežní linii, přesnost souřadnic, změny chování dat a shodnost záznamů z obou databází.

V této práci bylo zjištěno 37,1 % (2015) a 23,1 % (2019) duplicitních záznamů z databáze GBIF a 39,7 % (2015) a 32,7 % (2019) duplicitních záznamů z databáze OBIS. V průměru se jedná o jednu třetinu všech dat. Dle Mesibov (2013) je to způsobeno tím, že metadata dostatečně nepopisují duplicitní záznamy, které mohou být vytvořeny účelně. Konkrétně se může jednat o štítkové studie, jejichž principem je opakované zaznamenávání výskytu jednoho jedince. Nebo například sériové duplikáty, které vznikají z jednoho skupinového mapování výskytu druhů. V rozporu s tím je fakt, že jeden uživatel může přidávat záznamy pod více jmény zároveň, čímž může být množství duplikátů negativně ovlivněno.

Existence doby pořízení záznamu byla zjištěna na základě informace o dni, měsíci a roku. Tyto atributy splnilo 76,6 % (2015) a 90 % (2019), v databázi GBIF. OBIS zaznamenala 99 % (2015) a později 87,7 % (2019). Zbylá procenta znamenala neúplné časové údaje (např. pouze rok) anebo neobsahovala dobu pořízení záznamu vůbec. O něco horší výsledky uvádí studie autorů Otegui et al. (2013), kteří našli v databázi GBIF pouze 43,9 % záznamů s úplnou dobou pořízení a 41,6 % neobsahovala dobu pořízení vůbec. Zapsání neúplné doby pořízení záznamů (např. pouze rok) může být účelové, pokud se jedná o meziroční změny v druhové rozmanitosti.

Evaluací přesnosti souřadnic bylo zaznamenáno, že se v obou databázích nejčastěji vyskytují přesnosti od 1 do 6 desetinných míst. V databázi OBIS byla zjištěna i větší přesnost, 13 desetinných míst u zeměpisné šířky (37 %) a 12 desetinných míst u zeměpisné délky (30 %). Zahrnuta byla také data, která obsahovala souřadnice 0:0, přestože se může jednat o špatně vyplněné údaje místo hodnoty Null. Míra přesnosti souřadnic úzce souvisí s geografickými výskyty druhů. V databázi GBIF bylo zjištěno 90 % (2015) a 85,3 % (2019) zaznamenaných druhů v oceánu. Přičemž OBIS vykazuje až 95,8 % a 93,4 %. Tomu odpovídá závěr Yesson et al. (2007), kteří hodnotili rozložení záznamů luštěnin v databázi GBIF,

přičemž zjistili, že 13 % všech zaznamenaných dat vykazuje výskyty v oceánech, což u těchto rostlin není v podstatě možné. Tento problém přisuzují nepřesnému georeferencingu anebo právě velmi nízké přesnosti souřadnic. Na druhou stranu je třeba upozornit, že výskyt mořských savců na pevnině je v některých případech možný a ne všechny takové záznamy jsou nutně chybné.

Změny chování dat mezi lety 2015 až 2019 byly v obou databázích posuzovány na základě ID, které mělo zaručovat shodnost. Výsledky ukázaly, že bylo zhruba 40 % dat smazáno a poté nahrazeno novými. Takto velké změny nejsou v těchto rozsáhlých databázích příliš pravděpodobné. Způsobeno to může být tím, že se klíče k označení jednotlivých dat v průběhu let mění. Do budoucna by v podobných studiích bylo vhodné použít spíše prostorovou shodu. Nicméně fakt, že identifikátory jednotlivých záznamů se v databázích změnily, výrazně snižuje možnosti validace zkvalitňování poskytovaných dat a globální databáze by se měli zapříčinit o to, aby k takovému jevu nedocházelo.

V této studii byla dokázaná shodnost 35 % v roce 2015, což znamená přesně 223156 záznamů. A poté v roce 2019 pouze 14,3 % (148278) záznamů. Studie autorů Vandepitte et al. (2011) uvádí, že se v roce 2010 databáze GBIF a OBIS shodovaly ve 14 %, což představuje 27,7 milionů distribučních záznamů. Všechna tato čísla dokazují, že jsou data v databázích GBIF a OBIS aktivně tříděna a kontrolována, ačkoliv je tento přístup vzhledem k obrovskému množství dat teprve na začátku. Vandepitte et al. (2015) uvádí, že je v databázi OBIS více jak 60 % dat podrobena nástrojům kontroly kvality a více jak 90 % záznamů splňuje potřebná kritéria.

## 7. Závěr

Globální druhové databáze jsou z pohledu kvality a rozsahu dat hodnoceny velmi často. Každý k tomuto hodnocení přistupuje s jinými kritérii a aspekty, tudíž je žádoucí provádět co nejvíce studií, které se pak mezi sebou dají porovnávat a hodnotit. Tato práce se zabývala hodnocením výskytu druhů v databázi GBIF a OBIS hned v několika aspektech, jakou jsou duplicitní záznamy, existence doby pořízení záznamu, výskyty druhů vůči pobřežní linii, přesnost souřadnic, změny chování dat a shodnost záznamů z obou databází. Poskytovaná data, která nebyla duplicitní, měla uvedenou dobu pořízení a vyskytovala se v oceánu, zaujímala 53 % (GBIF), 56,4 % (OBIS) v roce 2015 a 60,7 % (GBIF) a 53,8 % (OBIS) v roce 2019. Zeměpisné souřadnice byly nejčastěji zastoupeny do desetinné přesnosti 6. Změny chování dat vykazovaly obrovské množství smazaných záznamů, které byly v průběhu let 2015 až 2019 doplňovány novými. Shodné záznamy obou databází zaujímaly 35 % (2015) a 14,3 % (2019).

Výsledky této studie dokazují, že se v záznamech vyskytují malé i velké nepřesnosti, které ale hned nemusí znamenat chybu. Z tohoto důvodu je velmi důležité správné pochopení jednotlivých záznamů, čemuž dopomáhají úplná a správně popsaná metadata. Velmi důležitý je také zodpovědný přístup všech přispěvatelů, kteří zaznamenávají výskyty jednotlivých druhů. A v neposlední řadě také podpora vzdělanosti přispěvatelů a uživatelů těchto dat, včetně široké veřejnosti.

## Přehled literatury a použitých zdrojů

**Amano T., Lamming J.D.L. & Sutherland W.J., 2016:** Spatial Gaps in Global Biodiversity Information 350 and the Role of Citizen Science, *BioScience*, 66, 393–400.

**Anderson R. P., Araújo M., Guisan A., Lobo J. M., Martínez-Meyer E., Peterson A. T. & Soberón J., 2016:** Final report of the task group on GBIF data fitness for use in distribution modelling, Global Biodiversity Information Facility, online: [https://serval.unil.ch/resource/serval:BIB\\_768D188CEA5B.P001/REF.pdf](https://serval.unil.ch/resource/serval:BIB_768D188CEA5B.P001/REF.pdf), cit. 18.3 2019.

**Berghe E. V., Ward Appeltans, Costello M. J. & Pissierssens P., 2004:** Proceedings ocean biodiversity informatics, VLIZ, 37, ISSN: 1377-0950.

**Bowen D. W., 1997:** Role of marine mammals in aquatic ecosystems, *Marine Ecology progress series*, 158, 267-274.

**Canhos V. P., Souza R., Giovanni R. & Canhos A. L., 2004:** Global Biodiversity informatics: Setting the scene for a “NEW WORLD” of ecological modeling, *Biodiversity Informatics*, 1, 1-1.

**Chapman A. D., 2005:** Uses of Primary Species-Occurrence Data, Global Biodiversity Information Facility, online: <http://www.niobioinformatics.in/books/Uses%20of%20Primary%20Data.pdf>, cit. 18.3 2019.

**Chavan V. & Penev L., 2011:** The data paper: a mechanism to incentivize data publishing in biodiversity science, *BMC Bioinformatics*, 12, ISSN: 1471-2105.

**Costello M.J., Stocks K., Zhang Y., Grassle J.F. & Fautin D.G., 2007:** About the Ocean Biogeographic Information System, online: <http://hdl.handle.net/2292/5236>, cit. 18.3.2019.

**Dooley E. E., (2002):** ENVIRONEWS Forum: Global Biodiversity Information Facility, *Environmental Health Perspectives*, 110, A669.

**Edwards J. L., Lane M. A. & Nielsen E. S., 2000:** Interoperability of biodiversity databases: biodiversity information on every desktop, *Science*, 289, 2312-2314.

**Edwards J. L., 2004:** Research and Societal Benefits of the Global Biodiversity Information Facility, *BioScience*, 54, 485-486.



**Endersen D. T. F., Gaiji S. & Robertson T., 2009:** Darwin Core Germplasm Extension and deployment in the GBIF infrastructure, in: Weitzman A.L., 2009: Proceedings of TDWG, online:  
[https://www.researchgate.net/profile/Dag\\_Endresen/publication/258513706\\_Darwin\\_Core\\_Germplasm\\_Extension\\_and\\_deployment\\_in\\_the\\_GBIF\\_infrastructure/links/0463528c88f29345d000000.pdf](https://www.researchgate.net/profile/Dag_Endresen/publication/258513706_Darwin_Core_Germplasm_Extension_and_deployment_in_the_GBIF_infrastructure/links/0463528c88f29345d000000.pdf), cit. 18.3 2019.

**EUROBIS:** Standards and quality control, Taxonomic standard: World Register of Marine Species (WoRMS), online: <http://www.eurobis.org/standards>, cit. 18.3 2019.

**Fair A. P. & Becker R. P., 2000:** Review of stress in marine mammals, *Journal of Aquatic Ecosystem Stress and Recovery*, 7, 335-354.

**Faith D. P., Collen B., Arino A. H., Koleff P., Guinotte J., Kerr J. & Chavan V., 2013:** Birding biodiversity data gaps: recommendations to meet user's data needs, *Biodiversity Informatics*, 8, 41-58.

**Ficetola G.F., Rondinini C., Bonardi A., Katariya V., Padoa-Schioppa E. & Angulo A., 2014:** An evaluation of the robustness of global amphibian range maps, *Journal of Biogeography*, 41, 211-221.

**GBIF, 2015:** IPT software release supports consistent DOI based citation, Global Biodiversity Information Facility, online: <https://www.gbif.org/news/82383/ipt-software-release-supportsconsistent-doi-based-citation/>, cit. 18.3.2019.

**Grassle J. F. & Stocks I. K., 1999:** A Global Ocean Biogeographic Information System (OBIS) for the Census of Marine Life, *Oceanography*, 12, 12-14.

**Halpin N.P., Read A. J., Best B. D., Hyrenbach D. K., Fujioka E., Coyne S. M., Crowder B. L., Freeman A. S., Spoerri C., 2006:** OBIS-SEAMAP: developing a biogeographic research data commons for the ecological studies of marine mammals, seabirds, and sea turtles, *Marine ecology progress series*, 316, 239-246.

**Mesibov R., 2013:** A specialist's audit of aggregated occurrence records, *ZooKeys*, 18, 1–18.

**Mittermeier R. A, Turner W. R., Larsen F. W., Brooks T. M. & Gascon C., 2011:** Global Biodiversity Conservation: The Critical Role of Hotspots, *Research Gate*, online:  
[https://www.researchgate.net/profile/Frank\\_Larsen/publication/225989074\\_Global\\_Biodiversity\\_Conservation\\_The\\_Critical\\_Role\\_of\\_Hotspots/links/0fcfd503c6f3a203f2000000.pdf](https://www.researchgate.net/profile/Frank_Larsen/publication/225989074_Global_Biodiversity_Conservation_The_Critical_Role_of_Hotspots/links/0fcfd503c6f3a203f2000000.pdf), cit. 18.3 2019.

- O’Dor R., 2004:** A Census of Marine Life, *BioScience*, 54, 92-93.
- O’Shea J. T. & Odell K. D., 2008:** Large-scale marine ecosystem change and the conservation of marine mammals, *Journal of Mammalogy*, 89, 529-533.
- Otegui J., Ariño A.H., Encinas M. a & Pando F., 2013:** Assessing the primary data hosted by the Spanish node of the Global Biodiversity Information Facility (GBIF), *PloS one*, 8, e55144.
- Robertson T., Doring M., Guralnick R., Bloom D., Wieczorek J., Braak K., Otegui J., Russell L. & Desmet P., 2014:** The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet, *PLoS one*, 9, e102623.
- Saarenmaa H., 2005:** Sharing and accessing biodiversity data globally through GBIF, *Reserch Gate*, online:  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.472.6157&rep=rep1&type=pdf>, cit. 18.3 2019.
- Sala E., Aburto-Oropeza O., Paredes G.,Parra I.,C. Barrera J., Dayton K. P., 2002:** A General Model for Designing Networks of Marine Reserves, *Science*, 298, 1991.
- Soberón, J. & Peterson A. T., 2004:** Biodiversity informatics: Managing and applying primary biodiversity data, *Philosophical Transactions of the Royal Society of London B*, 359, 689-698.
- Vandepitte L., Hernandez F., Claus S., Vanhoorne B., Hauwere D. N., Deneudt K., Appeltans W. & • Mees J., 2011:** Analysing the content of the European Ocean Biogeographic Information System (EurOBIS): available data, limitations, prospects and a look at the future, *Hydrobiologia*, 667, 1-14.
- Vandepitte L., Bosch S., Tyberghein L., Waumans F., Vanhoorne B., Hernandez F., De Clerck O. & Mees J., 2015:** Fishing for data and sorting the catch: assessing the data quality, completeness and fitness for use of data in marine biogeographic databases. *Database : the journal of biological databases and curation*, 2015, 1–14.
- Wheeler Q. D., Raven P. H. & Wilson E. O., 2004:** Taxonomy: impediment or expedient?, *Science*, 303, 285–285.

**Wieczorek J., Bloom D., Guralnick R., Blum S., Doring M., Giovanni R., Robertson T. & Vieglais D., 2012:** Darwin Core: An Evolving Community-Developed Biodiversity Data Standard, PLoS one, 7, e29715.

**Yesson Ch., Brewer W. P., Sutton T., Caithness N., Pahwa S. J., Burgess M., Gray A. W., White J. R., Jones C. A., Bisby A. F. & Culham A., 2007:** How Global Is the Global Biodiversity Information Facility?, PLoS one , 2, e1124.

**Zhang Y. & Grassle J. F., 2003:** A portal for the ocean biogeographic information system, Oceanologica Acta, 25, 193–197.