

Taxon-A XML-Prototyp

Modellierungsvorschlag

Heinz Rosenkranz, Juli 2019

ÖAW
ÖSTERREICHISCHE
AKADEMIE DER
WISSENSCHAFTEN

Checklisten Biodiversität

KIÖS - KOMMISSION FÜR INTERDISZIPLINÄRE ÖKOLOGISCHE STUDIEN

Auswahl

Marchantiophyta

Lophoziaceae
Heterogemma
H. capitata
H. laxa
Lophozia
L. ascendens
L. guttulata
L. longiflora
L. ventricosa
L. wenzelii
Lophozioopsis
L. excisa
L. latifolia
L. longidens
Trilophozia
T. quinqueidentata
Tritomaria
T. exsecta
T. exsectiformis
T. scitula
Scapaniaceae
Diplophyllum
D. albicans
D. obtusifolium
D. taxifolium
Saccobasis
S. polita
Scapania
S. aequiloba
S. apiculata
S. aspera
S. calcicola
S. carinthiaca
S. crassiretis
S. curta
S. cuspiduligera
S. degenii var. *degenii*
S. degenii var. *dubia*
S. gymnostomophita
S. helvetica
S. irrigua
S. irrigua subsp. *irrigua*
S. irrigua subsp. *rufescens*

Katalog-Teil III

[Stamm] **Marchantiophyta**

[Familie] **Lophoziaceae** CAVERS

[Gattung] **Heterogemma** (JÖRG.) KONSTANT. & VILNET

[Art] ***Heterogemma capitata*** (HOOK.) KONSTANT. & VILNET

Klassifikation: Plantae | Marchantiophyta | - | Lophoziaceae | Heterogemma

H. capitata (HOOK.) KONSTANT. & VILNET – Syn.: *Lophozia capitata* (HOOK.) MACOUN, *Jungermannia capitata* HOOK.

Ökologie

Blassgrüne, mitunter rötlich überlaufene, lockere Rasen in feuchten, sandigen bis lehmigen, kalkfreien Pionierfluren in vorwiegend flachem Gelände, etwa in Sand- und Kiesgruben oder an sandigen Teichufern. An einer der beiden Fundstellen (nahe Horn) am Rand eines alten Steinbruchs auf feuchtem Silikatgras. Sporogone sind selten; Ausbreitung primär via Gemmen.

- m. azidophytisch, m.–s. hygrophytisch, m. photophytisch
- L 7, T 5, K 2, F 6, R 3

Soziologie

Die Art gilt als Element des Haplomitrium hookeri. Über die Einnischung der österreichischen Populationen ist aber wenig bekannt. An der Fundstelle bei Horn wächst sie zusammen mit *Pohlia nutans* und *Cephalozia bicuspadata*.

Verbreitung

Erst kürzlich an zwei Stellen im Waldviertel entdeckt. Weitere Vorkommen dieser Art sind im Mühl- und Innviertel wahrscheinlich. Submontan und montan, bei 440 und 900 m.

Niederösterreich: Waldviertel: bei Karlstift und W Grub NW Horn (ZECHMEISTER et al. 2013)

Weltweit

- Westeuropa, nördliches Zentraleuropa, südliches Nordeuropa, östliches Nordamerika
- nördlich subozeanisch

Statistik:
49 Art(en)
19 Abbildungen
94 Verweise auf Synonyme
428 Verweise auf Begleitflora
1071 Verweise auf Fundstellen



Foto: S. Koval

Inhaltsverzeichnis

1	ALLGEMEINES	3
1.1	FALLBEISPIEL.....	3
1.2	PROTOTYP.....	3
1.3	RAHMENBEDINGUNGEN.....	3
1.4	XML.....	4
2	TAXONOMIE-MODELL	5
2.1	BUCH-KONZEPT	5
2.1.1	<i>Hierarchische Taxonomie</i>	5
2.2	OBJEKT-KONZEPT	6
2.2.1	<i>Taxon-Objekt</i>	6
2.3	INDEX.....	6
3	TAXON-MODELL	7
3.1	TAXON-ELEMENT	7
3.2	NOMENKLATUR (NAME, SYNONYM).....	8
3.2.1	<i>v-Markup</i>	8
3.3	STATISCHER TEIL (DESCRIPTION).....	9
3.3.1	<i>Abstract</i>	9
3.3.2	<i>Ecology (Ökologie)</i>	9
3.3.3	<i>Sociology (Soziologie)</i>	9
3.3.4	<i>Distribution (Verbreitung)</i>	9
3.3.5	<i>Endangerment (Gefährdung)</i>	10
3.3.6	<i>Remark (Anmerkungen)</i>	10
3.4	VARIABLER TEIL (OCCURRENCE)	10
3.4.1	<i>Location (Fundort allgemein)</i>	10
3.4.2	<i>loc (exakter Fundort)</i>	11
3.4.3	<i>Beispiel: Fundorte weltweit</i>	11
3.4.4	<i>Beispiel: Fundorte österreichweit</i>	11
3.4.5	<i>Beispiel: Fundorte Bundesland und Region</i>	12
3.4.6	<i>Beispiel: Funde eines Sammlers (oder Literaturzitat)</i>	12
3.4.7	<i>Beispiel: Fundort exakt</i>	12
4	SCHNITTSTELLEN	13
4.1	PASSIVE NUTZUNG VON DIENSTEN.....	13
4.1.1	<i>Lobid, VIAF, Wikipedia, ÖBL</i>	13
4.1.2	<i>Tropicos, GBIF, BOLD, Wikispecies</i>	13
4.1.3	<i>GeoNames</i>	13
4.2	DATENEXPORT	13
4.2.1	<i>Darwin Core</i>	13
4.2.2	<i>Jason-LD</i>	13

1 Allgemeines

1.1 Fallbeispiel

Die Prototyp-Entwicklung für die exemplarische Biodiversity-Taxonomie erfolgte auf Basis einer in Buchform publizierten Taxonomie:

Die **Horn- und Lebermoose Österreichs** von Heribert Köckinger
(ISBN 978-3-7001-8153-8, ÖAW-Verlag Wien 2017).

Diese publizierte Taxonomie enthält 264 Arten von Horn- und Lebermoosen (Taxons), deren Ökologie, Soziologie, Vorkommen und Gefährdung ausführlich beschrieben wird. Mit über 15.000 Verweise auf Synonyme, Begleitflora, Fundstellen, Sammler und Autoren stellt diese Taxonomie ein repräsentatives Fallbeispiel dar.

1.2 Prototyp

Das Taxon-A Projekt ist als „agiles Projekt“ vorgesehen, was nach allgemeinem Verständnis den Verzicht auf allzu aufwendige, vorausseilende Planung bedeutet. In diesem Zusammenhang ist „fast Prototyping“ eine sinnvolle Vorgehensweise, da ein funktionierender Prototyp

- den Nutzern die Möglichkeit für unmittelbares Feedback gibt,
- den Software-Entwicklern ein tiefergehendes „proof of concept“ ermöglicht,
- und damit einen kontinuierlichen Fortschritt des Projektes fördert

Selbst wenn der Prototyp letztendlich verworfen wird, lassen sich damit wichtige Erkenntnisse gewinnen, falsche Annahmen frühzeitig erkennen, und allzu aufwendige Nutzerwünsche behutsam in Frage stellen.

1.3 Rahmenbedingungen

Folgende Rahmenbedingungen wurden angenommen:

- Möglichst einfache XML-Konvertierung bereits publizierter oder bestehender Taxonomien
- Berücksichtigung existierender Standards, im gegebenen Fall das Darwin Core Vokabular
- Minimaler Markup-Aufwand in Hinblick auf eine möglichst einfache Workspace-Umgebung
- Freizügige Textgestaltung in Hinblick auf Journal- oder Buchpublikationen
- Ausreichende Metadaten-Anreicherung in Hinblick auf Konvertierung in andere Datenformate (z. B. Darwin Core XML, JATS, JSON-LD ...)
- Passive Einbindung von Linked Open Data (z. B. Lobid, VIAF, GBIF, Wikispecies ...)

1.4 XML

Als Dokumentsprache wird XML („**eX**tensible **M**arkup **L**anguage“) verwendet. XML ist reiner Text, zugleich aber auch Syntax, Metasprache und Anwendung:

- als **reiner Text** ist XML jederzeit mit einem einfachen Texteditor bearbeitbar – ein wichtiger Aspekt für die Langzeitarchivierung und Plattformunabhängigkeit;
- als **Syntax** gibt XML strikte Syntax-Regeln vor. Diese sogenannte „Wellformedness“ ist es, die ein XML-Dokument maschinenlesbar macht, also einem Parser ermöglicht, das Dokument in ein komplexes Dokumentobjektmodell (DOM) umzuwandeln, das mit einer Datenbank vergleichbar ist;
- als **Metasprache** dient XML dazu, neue Anwendungssprachen zu erfinden – vergleichbar mit der Metasprache SGML, mit der HTML erfunden wurde. Dazu legt XML mit einem Regelwerk (DTD oder XML-Schema) die inhaltliche Struktur einer Klasse von Dokumenten fest, z. B. juristischer Schriftverkehr, medizinische Atteste, Bilanzdaten oder Taxonomien. Jedes XML-Dokument, das sich an solch ein Regelwerk hält, ist ein gültiges („valides“) Dokument;
- Als konkrete **Anwendung** dient XML dazu, Inhalte zu repräsentieren. Seien es strikte Datenreihen, die zwischen unterschiedlichen Datenbanken ausgetauscht werden müssen, wissenschaftliche Publikationen, Lexika oder beliebige Literatur.

Ergänzend dazu lassen sich einige Grundregeln formulieren, die bei der Prototyp-Modellierung zur Anwendung kommen:

- Ein XML-Dokument beschreibt den **Inhalt** und nicht das Aussehen eines Dokumentes. Dies bedeutet üblicherweise den Verzicht auf Typographie – denn alles, was typographisch hervorgehoben werden soll, sollte auch einem semantischen Zweck dienen;
- Ein XML-Dokument soll den Inhalt mit **erklärendem Markup** beschreiben, damit das Dokument für Menschen lesbar ist. Letztendlich soll aber der Inhalt gelesen werden, und nicht das Markup. Wenn Inhalte „selbsterklärend“ sind, oder bestimmtes Markup sich oft in einem Dokument wiederholt, ist eine minimalistische Namensgebung ausreichend;
- Strukturierte Inhalte werden vorzugsweise mit **XML-Elementen** notiert, und strukturlose Inhalte als **Attribute** hinzugefügt, um möglichst einfaches Markup zu erzielen. Darüber hinaus lassen sich auch Notations-Konventionen festlegen, um komplexes Markup weiter zu reduzieren. Ein typisches Beispiel dafür sind Geokoordinaten (Länge-, Breitengrade);
- Viele Informationen lassen sich auch aus dem inhaltlichen **Kontext** ableiten und müssen daher nicht extra angeführt werden. Der Verweis auf einen Taxon-Namen ist je nach Kontext entweder als Synonym oder als Begleit-Flora interpretierbar.

Ein XML-Dokument **muss** „wellformed“ sein und **sollte** auch „valide“ sein. Das lässt sich mit einem normalen Texteditor nicht erreichen. Daher ist ein entsprechender Eingabedialog (Workspace) für Autoren erforderlich, oder ein XML-Editor, der neben Syntax-Prüfung auch kontextspezifische Hilfe zu den einzelnen XML-Elementen bietet.

Für die Nutzung von XML-Dokumenten steht mit XSLT (**eX**tensible **S**tylesheet **L**anguage **T**ransform) ein mächtiges Werkzeug zur Verfügung, das neben der Erzeugung von HTML-Webseiten auch für die Transformation der Taxonomie in beliebige andere Formate genutzt wird.

2 Taxonomie-Modell

Das Taxonomie-Modell sieht grundsätzlich vor, eine Taxonomie als zusammenhängende Publikation (**Buch-Konzept**) oder als Ensemble einzelner Taxons (**Objekt-Konzept**) zu beschreiben.

Welches Konzept vorzuziehen ist, hängt im Wesentlichen von Art und Umfang der Taxonomie ab. **Kleine Taxonomien** mit bis zu 50 Taxa, die zum Beispiel nur eine Familie beschreiben, lassen sich problemlos als Gesamttaxonomie organisieren. Darüber hinausgehend wird es aber schwierig, eine nutzerfreundliche Navigation vorzusehen, und bei hohem Bildanteil werden sich auch Ladezeiten und Seitenaufbau der Webseite merklich verschlechtern.

Das Objekt-Konzept ist für **Taxonomien beliebiger Größe** vorgesehen. Hier wird jedes Taxon als eigenständiges Dokument angelegt, und über Indexdateien zu einem passenden Ensemble zusammengeführt. In Hinblick auf den voraussichtlich wachsenden Detaillierungsgrad und Versionierungsbedarf wird das Objekt-Konzept die vorzugsweise Lösung sein.

Da die eigentliche Taxon-Modellierung ident ist, kann das Taxonomie-Konzept jederzeit mit etwas organisatorischem Aufwand gewechselt werden. Dies wird zum Beispiel erforderlich sein, wenn eine Taxonomie im Zuge einer Retrodigitalisierung oder eines Datenexportes aus alten Anwendungen zuerst in eine Buchform konvertiert wird, und anschließend erst in die Objektform überführt wird.

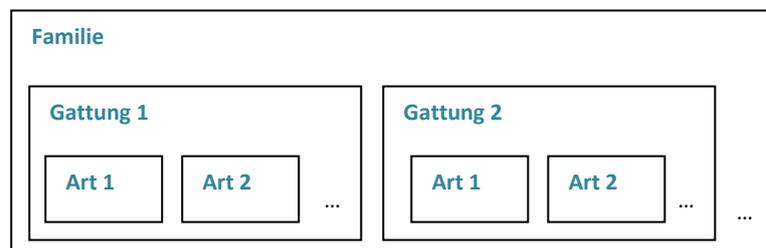
2.1 Buch-Konzept

Eine traditionelle Buch-Publikation besteht üblicherweise aus Titelei – Einleitung – Inhalt – Anhang in den verschiedensten Ausprägungen. Von Bedeutung ist hier nur die Modellierung des Inhaltes, also der eigentlichen Taxonomie, denn die restlichen Buch-Komponenten sind durch gängige Dokumentstandards gut beschrieben.

2.1.1 Hierarchische Taxonomie

XML ist objektorientiert, und eine objektorientierte Hierarchie unterscheidet sich von der Kapitel-Hierarchie eines Buches, die einfach nur eine fortlaufende Liste unterschiedlicher Hierarchie-Stufen ist (z. B. Kapitel - Subkapitel – SubSub-Kapitel – Kapitel ...). In einer objektorientierten Hierarchie kapselt ein Objekt die Inhalte, ein Objekt der obersten Ebene enthält Objekte der nachfolgenden Ebene, die wiederum Objekte der nachfolgenden Ebene enthalten (siehe Abbildung).

Damit ist der wahlfreie Zugriff auf alle Ebenen möglich, sei es eine komplette Familie, eine Gattung, oder alle Arten einer bestimmten Gattung.



Ein einzelnes Taxon ist durch seine Stellung in der Hierarchie eindeutig definiert und müsste keine Informationen über seine Zugehörigkeit enthalten. Diese könnte erst bei Bedarf, z. B. beim Datenexport in eine andere Taxonomie-Anwendung, automatisch aus dem umgebenden Kontext erzeugt werden. Aus Kompatibilitätsgründen zum nachfolgend beschriebenen Objekt-Konzept werden aber die wichtigsten Kontextinformationen explizit kodiert.

2.2 Objekt-Konzept

Das Objekt-Konzept sieht vor, jedes Taxon als eigenständiges Objekt zu behandeln, das eindeutig identifizierbar und umfassend beschrieben ist. Die Zuordnung zu einer Taxonomie erfolgt über eine Index-Datei, mit der Taxon-Ensembles für unterschiedliche Szenarien gebildet werden können, z. B.

- hierarchische Taxonomien eines Stammes, einer Klasse, Ordnung oder Familie
- Evidenzlisten verschiedener Arten einer Region
- Gruppierungen aufgrund ökologischer und/oder soziologischer Faktoren
- Neuordnungen basierend auf genetischen Merkmalen

2.2.1 Taxon-Objekt

Ein Taxon-Objekt muss eindeutig identifizierbar sein, dies geschieht üblicherweise durch einen eindeutigen **URI** (Unique Resource Identifier) in der Form Präfix:Name:Suffix.

Das **Präfix** enthält den Urheber der Beschreibung (Institution), der **Name** den eindeutigen Taxon-Namen, und das **Suffix** ein optionales Identifikationsschema (z. B. Katalog- oder Archivnummer). Alternative oder ergänzende URIs können registrierte **DOI** (Digital Object Identifier) oder akzeptierte **URN** (Unique Resource Name) anderer Institutionen sein.

Für interne Zwecke, z. B. für die Benennung der Taxon-Dateien in der Urheber-Domäne, reicht der Taxon-Name, gegebenenfalls ergänzt zu einem **URL** (Unique Resource Locator):

www.urheber.org/taxon-verzeichnis/taxon-name.xml

2.3 Index

Der Index enthält im Wesentlichen den URI, Taxon-Namen und essentielle Metadaten jedes beteiligten Taxons und verlinkt auf das zugehörige Taxon-Objekt. Ein Index ist vielseitig nutzbar und für die Erstellung des Index gibt es unterschiedliche Methoden.

Für eine Taxonomie gemäß Buch-Konzept lässt sich ein Index automatisch per XSLT-Transformation erzeugen. In diesem Fall ist der Index die Basis für das alternative Objekt-Konzept, aber auch für weiterführende Transformationen wie z. B. für die Erzeugung der Begleitflora-Grafik mittels JSON-LD.

Für Taxon-Ensembles gemäß Objekt-Konzept muss ein Index manuell erzeugt werden. In der Praxis kann aber davon ausgegangen werden, dass Taxonomien bereits in irgendeiner nutzbaren Form existieren (PDF, CSV-Datei) und daher im Zuge der XML-Konvertierung auch ein Index erzeugt werden kann.

3 Taxon-Modell

Für die nachfolgenden Beschreibungen wird die einfache Kurz-Notation verwendet, die auch im zugrundeliegenden Regelwerk (DTD) zur Anwendung kommt. Folgende Symbole sind dabei wichtig (sie werden in der Kommentierung auch immer wieder gezielt erklärt):

()	Umklammerung von Element-Gruppen
,	Separierung von Elementen
	Logisch ODER
	Kardinalität
?	optionales Element
+	mindesten 1 bis beliebig viele
*	0 bis beliebig viele
	exakt einmal (bei fehlender Kardinalität)
Text	Text beliebiger Länge
rot	kennzeichnet Attribut-Namen (zur Unterscheidung von Element-Namen)
fett	kennzeichnet verpflichtende Elemente oder Attribute

3.1 Taxon-Element

Das Taxon-Element besteht aus folgendem Inhaltsmodell und folgenden Attributen:

Taxon	(Taxon (Name , Synonym? , Description? , Occurrence?))*
taxonID	ID
doi	Text
urn	Text
taxonRank	(Kingdom Phylum Class Order Family Genus Species other)
higherClassification	Text
taxonRemarks	Text

Kommentierung:

- Ein **Taxon**-Element kann 0 bis beliebig oft entweder ein eingebettetes **Taxon**-Element enthalten ODER die **Taxon-Details** (Name, Synonym, Description ...)
- Eingebettete Taxon-Elemente erlauben Hierarchien beliebiger Tiefe. In so einem Fall gilt die Konvention, zuerst das Name-Element des Taxons vorzusehen, bevor mit der Einbettung tieferliegender Taxons begonnen wird
- Die weiteren Details (Synonym, Description, Occurrence) sind optional und entfallen für höhere Hierarchien, wenn z. B. nur Arten beschrieben werden sollen
- Das Attribut **taxonID** ist für jedes Taxon zwingend und muss eindeutig sein. Dafür wird der (wissenschaftliche) Taxon-Name verwendet, der innerhalb einer Taxonomie eindeutig ist
- Wichtig ist auch das Attribut **taxonRank**. Da es aber eine Enumeration von vorgegebenen Begriffen ist, aus der eine Auswahl zu treffen ist, läßt sich die Verwendung nicht erzwingen
- Das Attribut **higherClassification** beschreibt den Kontext durch Auflistung der Hierarchie
- Die Attribute **doi**, **urn** und **taxonRemarks** werden zur Zeit nicht verwendet. Hier ist zu entscheiden, ob und wie sie verwendet werden sollen

3.2 Nomenklatur (Name, Synonym)

Das **Name**-Element besteht aus folgendem Inhaltsmodell und folgenden Attributen:

Name (Text | b | v)*

scientificName	Text
authorship	Text
doi	Text
urn	Text

Kommentierung:

- Das **Name**-Element kann **Text**, das aus HTML bekannte Markup **b** (bold) und ein spezielles Markup **v** in beliebiger Anzahl und Reihenfolge enthalten
- Damit wird der sichtbare Teil des Taxon-Namens nach freiem Ermessen gestaltet, z. B. mit abgekürztem Gattungsnamen, wobei ausnahmsweise auch die typographische Angabe Fettschrift zulässig ist
- Das **Synonym**-Element weist das gleiche Inhaltsmodell auf wie Name, allerdings ohne Attribute, da es nur als Verweis auf andere Namensgebungen verwendet wird
- Das Markup **v** hat eine besondere Funktion, die nachfolgend erklärt wird
- Das Attribut **scientificName** ist zwingend erforderlich und muss den korrekten Namen des Taxons enthalten, also bei Arten mit vollständigem Gattungsnamen
- Das Attribut **authorship** enthält die vollständige Autorenschaft mit allen Zusätzen. Es ist eigentlich immer erforderlich, zumindest in den unteren Taxonomie-Hierarchien
- Die Attribute **doi** und **urn** werden zur Zeit nicht verwendet. Hier ist zu entscheiden, ob und wie sie verwendet werden sollen

3.2.1 v-Markup

Der Name eines Taxons besteht aus wissenschaftlichen Namen und Autorenschaft und bildet somit sowohl semantisch wie auch typographisch eine facettenreiche Struktur, z. B.:

Fuscocephaloziopsis albescens **var. albescens** (HOOK.) VAÑA & L. SÖDERSTR.

Der wissenschaftliche Name wird kursiv dargestellt, mit Ausnahme von Zusätzen wie **var.** oder **subsp.**. Die Autorenschaft wird (oft) in Kapitälchenschrift dargestellt und besteht in diesem Beispiel aus drei Autoren-Kürzel, oft mit Zusätzen wie **&**, **ex** oder **auct. non** in Normalschrift. Der Name soll mit all diesen Komponenten eine zusammengehörende semantische Einheit bilden.

Hier wäre jeder Ansatz mit individuellem Markup für Autoren unzumutbar und auch mit Eingabe-Dialogfelder nur mühsam darstellbar. Es wird daher eine Notations-Konvention verwendet:

<v>Fuscocephaloziopsis albescens var. Albescens | (Hook.) Vaňa & L. Söderstr. **</v>**

Der gesamte Name wird durch ein einziges Markup **<v>** **</v>** umklammert.

- der Teil vor dem Vertikalstrich | wird semantisch als wissenschaftlicher Name interpretiert und von XSLT typographisch entsprechend kursiv dargestellt
- der Teil nach dem Vertikalstrich wird als Autorenschaft erkannt und entsprechend interpretiert und typographisch behandelt
- die oben erwähnten Zusätze werden in XSLT gefiltert und in Normalschrift dargestellt

Diese Notations-Konvention wird nicht nur für den Taxon-Namen verwendet, sondern auch bei allen Synonym- und Begleitflora-Nennungen, und bei der Vielzahl an Literaturzitaten zu den Fundstellen, wo ebenfalls Autorennamen auftreten.

3.3 Statischer Teil (Description)

Für die grundlegende Beschreibung eines Taxons wird folgendes Inhaltsmodell verwendet:

Description (Abstract?, Ecology?, Sociology?, Distribution?, Endangerment?, Remark?)
Abstract (p | list)*
Ecology (p | list)*
Sociology (p | list)*
Distribution (p | list)*
Endangerment (p | list)*
Remark (p | list)*
p (Text | b | i | sup | sub | v | loc | by | br | a | img)*
list (p*)

Kommentierung:

- Alle Elemente des Inhaltsmodells sind optional, sie werden nur bei Bedarf verwendet
- Die inneren Elemente gestattet Text-Absätze (**p**) oder Aufzählungslisten (**list**) in beliebiger Anzahl und Reihenfolge
- Das **p-Element** erlaubt folgendes Markup in beliebiger Anzahl und Reihenfolge
Typographisches Markup: **b** (fett), **i** (kursiv), **sup** (hoch), **sub** (tief), **br** (Umbruch)
Bilder und Links: **img**, **a**
Semantisches Markup: **v** (Namen), **loc** (Fundstelle), **by** (Sammler)
- Aufzählungslisten bestehen aus **p**-Elemente in beliebiger Anzahl
- Keines dieser Elemente verwendet Attribute

3.3.1 Abstract

Das **Abstract**-Element ist für eine allgemeine Beschreibung vorgesehen. Es ist damit auch der ideale Ort für Bildmaterial, wie folgender XML-Kode zeigt:

```
  
Foto: M. Lüth  
</img>
```

Kommentierung:

- Das bekannte HTML-Element **img** mit seinen Attributen wird nicht weiter erklärt
- Abweichend von HTML kann **img** auch Text für eine Bild-Legende enthalten

3.3.2 Ecology (Ökologie)

Beschreibung des Erscheinungsbildes, der besiedelten Habitate und Substrate sowie standardisierte Angaben (Ökologischer Zeigerwerte, Ökologische Amplitude). Die standardisierten Angaben erfolgen zur Zeit in Textform - ob hierfür ein spezielles Markup vorzusehen ist, muss entschieden werden.

3.3.3 Sociology (Soziologie)

Beschreibung der Moosgesellschaft und Aufzählung typischer Begleitarten. Hier ist die Verwendung des **v**-Markups erforderlich, um Begleitarten gezielt als solche zu markieren.

3.3.4 Distribution (Verbreitung)

Allgemeine Beschreibung der Verbreitung. Eine detaillierte Aufzählung findet sich im Fundorte-Teil (Occurrence).

3.3.5 Endangerment (Gefährdung)

Allgemeine Beschreibung der Gefährdungssituation.

3.3.6 Remark (Anmerkungen)

Anmerkungen aller Art.

3.4 Variabler Teil (Occurrence)

Der variable Teil, die umfangreiche Fundortliste, ist wie jede Bestandsliste periodischen Änderungen unterworfen. Daher ist sie als eigenständiges Element vom übrigen Teil der Taxon-Beschreibung getrennt. Folgendes Inhaltsmodell und folgende Attribute werden verwendet:

Occurrence (Location*)

date	Text
minElevation	Text
maxElevation	Text
status	(new stable endangered missing undefined)

Kommentierung:

- Occurrence enthält 0 bis beliebig viele Locations, also Fundorte, die in weitere Folge spezifiziert werden
- Das Attribut **date** enthält das Datum der Bestandsaufnahme
- Die Attribute **minElevation** und **maxElevation** geben allgemein den Seehöhe-Bereich des Taxon-Vorkommen als Zahl (ohne Einheit m) an
- Das Attribut **status** ist vorgesehen, um eine allgemeine Einschätzung über das Vorkommen des Taxons in Österreich anzugeben
- Hier wäre auch ein Versions-Attribut denkbar. Es ist allerdings unklar, ob eine Versionierung innerhalb des Dokumentes erfolgen soll, oder nur auf Dateiebene

3.4.1 Location (Fundort allgemein)

Ein Fundort wird durch folgendes Inhaltsmodell und folgende Attribute beschrieben:

Location (p | list)*

locationID	Text
continent	Text
countryCode	Text
stateProvince	Text (Bundesland)
locality	Text (Region)
elevation	Text (Seehöhe in m)
coord	Text (Koordinaten)

Kommentierung:

- Das **Location**-Element ist ein generalisierendes Konzept und kann wahlweise für Funde weltweit, österreichweit sowie für Bundesländer und Regionen verwendet werden
- Die Fundorte werden durch Text (**p**) oder Aufzählungslisten (**list**) allgemein beschrieben. Eine genauere Spezifikation eines Fundortes erfolgt bei Bedarf durch das **loc**-Markup, das im **p**-Element vorgesehen ist. Siehe dazu nachfolgende Beispiele
- Die Attribute sind im Wesentlichen selbsterklärend. Siehe dazu nachfolgende Beispiele

3.4.2 loc (exakter Fundort)

Ein exakter Fundort wird durch folgendes Inhaltsmodell und folgende Attribute beschrieben:

loc (Text | loc | v | by)*

id	Text	(GeoName ID)
locality	Text	(Region)
elevation	Text	(Seehöhe in m)
coord	Text	(Koordinaten)
count	Text	(Anzahl der gefundenen Exemplare)
date	Text	(Datum der Registrierung)

Kommentierung:

- Das **loc**-Element bietet die Möglichkeit, entweder einen Fundort exakt zu beschreiben, oder mehrere Fundorte einer Region, einem Sammler oder Literaturzitat zuzuordnen
- Die Fundorte werden durch Text beschrieben, in dem folgendes Markup zulässig ist: **loc** (für exakte Fundorte), **v** (für Literaturzitate), **by** (für Sammlernamen)
- Die Attribute sind im Wesentlichen selbsterklärend. Siehe dazu nachfolgende Beispiele

3.4.3 Beispiel: Fundorte weltweit

```
<Location locationID="World">
  <list>
    <p> Europa, Kaukasus, Asien, Nord- und Südamerika </p>
    <p> südlich temperat </p>
  </list>
</Location>
```

Kommentierung:

- Das Element **Location** enthält eine Aufzählungsliste mit Allgemeiner Verbreitung der Art sowie Arealtyp
- Das Attribut **locationID** enthält die Kennung „ World“

3.4.4 Beispiel: Fundorte österreichweit

```
<Location locationID="AT" countryCode="AT" stateProvince="Österreich">
  <p> B – K – N – O – ST – S – T – V – W </p>
  <p> Wien Innenstadt: keine Funde </p>
</Location>
```

Kommentierung:

- Das Element **Location** enthält ein p-Element mit einer Auflistung aller Bundesländer mit hoher Verbreitung der Art
- Das Attribut **locationID** und **countryCode** enthält die Kennung „ AT“
- Das Attribut **stateProvince** enthält die Kennung „ Österreich“

3.4.5 Beispiel: Fundorte Bundesland und Region

```
<Location locationID="N" countryCode="AT" stateProvince="Niederösterreich">
  <p>
    <loc locality="Wachau"> bei Weinzirl am Walde und Senftenberg </loc>;
    <loc locality="NA"> bei Reichenau </loc>;
    <loc locality="Wechsel "> bei Aspang und Kirchberg </loc>;
    <loc> Dunkelsteinerwald </loc>
  </p>
</Location>
```

Kommentierung:

- Das Element **Location** enthält ein p-Element mit einer Serie von loc-Elementen für jede Vorkommens-Region
- Jedes dieser loc-Elemente enthält ein Attribut **locality** mit Nennung der Region
- Ein loc-Element ist keiner Region zugeordnet
- Eine allfällige Separierung der Regionen kann durch Strichpunkt erfolgen
- Das Attribut **locationID** enthält die Kennung des Bundeslandes, z. B. „N“
- Das Attribut **stateProvince** enthält das Bundesland, z. B. „Niederösterreich“

3.4.6 Beispiel: Funde eines Sammlers (oder Literaturzitat)

```
<Location locationID="N" countryCode="AT" stateProvince="Niederösterreich">
  <p>
    <loc locality="Waldviertel">
      <loc>bei Groß-Gerungs, Hoheneich und Weitra (<v>| Heeg 1894</v>) </loc>,
      <loc>Friedental W Harmanschlag (<by>HH</by>)</loc>
    </loc>;
  </p>
</Location>
```

Kommentierung:

- Das **loc**-Element (für Waldviertel) enthält weitere loc-Elemente für Funde, die einem Literaturzitat zugeordnet werden (Heeg 1894), und Funde, die einem Sammler zugeordnet werden (HH)
- Literaturzitate werden durch die Notations-Konvention **<v>| </v>** markiert, die den Autornamen nach dem |-Symbol in Kapitälchenschrift darstellt
- Sammlernamen werden mit dem dafür vorgesehene Markup **<by> </by>** markiert
- Diese unterschiedlichen Markups können auch dazu dienen, historische Fundorte (Literatur) von aktuellen Fundorten zu unterscheiden

3.4.7 Beispiel: Fundort exakt

```
<loc locality="Lavanttal">
  <loc><loc id="2767687">Reisberg</loc> bei Wolfsberg </loc>,
  <loc elevation="550" coord="46.85, 14.8" > St. Margarethen </loc>
</loc>
```

Kommentierung:

- Im ersten Beispiel (Reisberg) wird in den Beschreibungstext ein weiteres **loc**-Element eingebettet, das mit dem Attribut **id** auf die GeoNames-Datenbank verweist
- Im zweiten Beispiel erfolgt die Angabe der Koordinaten explizit

4 Schnittstellen

4.1 Passive Nutzung von Diensten

4.1.1 Lobid, VIAF, Wikipedia, ÖBL

Die Nutzung dieser Dienste erfolgt für die Bereitstellung ausführlicher Zusatzinformationen zu den Autoren-Namen der Taxa. Die geht nicht auf direktem Weg, da die beim Taxon angeführten Autoren-Kürzel kein eindeutiges Suchkriterium sind. Die Autoren-Namen müssen gezielt recherchiert und in eine Personen-Liste nach folgendem Muster eingetragen werden:

```
<Personennamen>
  <p id="Ach." href="https://lobid.org/gnd/128995211"
    lobid="128995211" viaf="51806870" web="Erik_Acharius">
    Erik Acharius (1757–1819), schwedischer Arzt und Botaniker
  </p>
  ...
</Personennamen>
```

Kommentierung:

- Das Attribut **id** enthält das eindeutige Autoren-Kürzel
- Das Attribut **href** enthält den bevorzugten Web-Link, da das Autoren-Kürzel nur mit einem Dienst verlinkt werden kann
- Die Attribute **lobid**, **viaf** und **web** enthalten die jeweils recherchierten Links zu den einzelnen Diensten, aus denen ein bevorzugter Link gewählt wird
- Der im p-Element enthaltene Text wird als Hover-Help angezeigt
- Der Aufwand für die Erstellung solcher Personen-Listen ist erheblich, dafür könnten sie aber in jeder ähnlichen Taxonomie genutzt werden

4.1.2 Tropicos, GBIF, BOLD, Wikispecies

Die Nutzung dieser Dienste erfolgt für die Bereitstellung alternativer Informationen zum beschriebenen Taxon. Die Verlinkung erfolgt automatisch, da der wissenschaftliche Name eines Taxons ein eindeutiges Suchkriterium ist.

4.1.3 GeoNames

Die Nutzung dieses Dienstes erfolgt für die Bereitstellung exakter Geo-Informationen zu Fundorten. Neben Koordinaten und Seehöhe steht damit auch eine Google-Map zur Verfügung. Auch hier ist eine Recherche des korrekten Fundortes vorab erforderlich, was vermutlich nur für ausgewählte Fundorte Sinn macht. Z. B. für festgelegte Referenz-Fundorte, die periodisch inspiziert werden.

4.2 Datenexport

4.2.1 Darwin Core

Das Darwin-Core XML-Format wird durch XSLT-Transformation erzeugt. Zur Zeit liegt keine konkrete Spezifikation vor.

4.2.2 Jason-LD

Das JSON-LD Format wird durch XSLT-Transformation erzeugt und exemplarisch für die Erzeugung der Begleitflora-Grafik genutzt.