EDIT - feature request #10194

Allow specimens as source for factual data

12/01/2022 06:29 PM - Andreas Müller

Status:	Closed	**Start date:**	
Priority:	Highest	**Due date:**	
Assignee:	Andreas Müller	**% Done:**	100%
Category:	cdm	**Estimated time:**	0:00 hour
Target version:	Release 5.35		
Severity:	normal		

Description

... and maybe others.

Think can be handled model side with existing CdmLinkSource or as explicit attribute in DescriptionElementSource

Related issues:

Related to EDIT - feature request #10202: Allow entering specimen in fact sou...	**Closed**
Related to EDIT - feature request #10203: Show specimen sources in publications	**New**

Associated revisions

Revision 861b3cdc - 12/13/2022 08:54 PM - Andreas Müller

ref #10194 Add specimen to DescriptionElementSource

Revision c6365af5 - 12/13/2022 08:54 PM - Andreas Müller

ref #10194 Add specimen to DescriptionElementSource (adapt test script and FullCoverageDataGenerator)

Revision 0905498c - 12/13/2022 11:36 PM - Andreas Müller

ref #10194 fix missing specimen_id in test data

Revision f0fda4e4 - 12/14/2022 12:07 AM - Andreas Müller

ref #10201, ref #10194 adapt testH2 to cdmlib

Revision 4f3bf065 - 12/14/2022 12:09 AM - Andreas Müller

ref #10201, ref #10194 adapt testH2 in taxeditor

Revision 1126d2a5 - 12/14/2022 12:09 AM - Andreas Müller

ref #10201, ref #10194 adapt testH2 in vaadin

Revision df718639 - 12/19/2022 03:26 PM - Andreas Müller

ref #10194 add either reference or specimen constraint

Revision 9db3e365 - 12/19/2022 03:54 PM - Andreas Müller

ref #10194 fix FullCoverageDataGenerator for new constraint

History

#1 - 12/01/2022 06:31 PM - Andreas Müller

WGB:

gibt es eigentlich eine Möglichkeit, Specimen sozusagen als Referenz (Voucher) für Faktendaten zu benutzen?
Das kommt nicht nur in (atomisierten) Beschreibungen vor, sondern z.B. auch bei Verbreitungen (Fundort des Specimens in der area) oder auch anderen Daten (Common Names), die man von einem Specimen Label bezogen hat.

Konkret soll im eFloraMex Projekt jeweils ein Specimen für einen State als Beleg angegeben werden.
Beispiel (Text Flora Mesoamericana):

1. Actinostachys germanii Fée, Mém. Foug. 11: 123 (1866). Tipo: Guadeloupe, Germain s.n. (MO!). Por R. Riba y L. Pacheco. Schizaea germanii (Fée) Prantl. Rizoma corto, tuberoso; hojas 5-15(-20) x 0.1-0.15 cm; pecíolo subterete a triangular redondeado en la base; lámina 0.1-0.15 cm

> de ancho; esporangióforos (3-)4-6(-8), 0.9-1.5 cm x 0.7-0.9 mm; esporangios en 2-3 hileras, entremezclados con tricomas filiformes. Selvas altas perennifolias, en sitios inundables. Ch (Martínez S. 14981, UAMIZ); B (Stolze, 1976: 26); G (Stolze, 1976: 26); CR (Davidse y Herrera 31455, MO) [Ch = Chiapas, B = Belice, G = Guatemala, CR = Costa Rica]

AM:

theoretisch gibt es diese Möglichkeit modellseitig bereits, da man z.B. für Aggregationen als Quelle CDM Objekte verschiedener Klassen angeben kann als Quelle. Im TaxEditor ist das bislang aber nur so implementiert, dass dieses CDM Objekt als Quelle angezeigt wird, wenn es existiert, aber nicht editiert werden kann, da es bislang nur für berechnete Daten als Quelle hinzugefügt werden kann und somit nicht editierbar sein muss.
Das Problem ist, dass das an der Stelle sehr generisch implementiert ist, also es kann aus diversen Klassen ausgewählt werden, was beim Editieren über UI aber natürlich schwierig ist, weil man sich da natürlich entscheiden muss und nicht nur ein Label anzeigen braucht.
Die Frage ist aber, ob wir für diesen speziellen Fall nicht doch eine konkretes (nicht generisches Feld) haben wollen, also ähnlich wie die (normale) Referenz, die ja auch nicht das generische Feld verwendet. Ich tendiere glaube ich zu letzterem, da es sich hier ja um eine semantisch relativ klarer Beziehung handelt. Könnten wir Model und TaxEditor seitig sicherlich recht einfach hinzufügen. Beim Portal müsste man dann noch schauen, wie man das formattiert.
Ich nehme an das Details Feld würde dann ebenfalls verwendet werden können, z.B. für die Beschreibung des Bereichs auf dem Voucher.

Ein bisschen frage ich mich, ob man da irgendwie unterscheiden muss zw. Specimen und dem Label auf dem Specimen (als Quelle z.B. für Common Name). Das ist ja nicht unbedingt immer das exakt gleiche. Aber vielleicht müssen wir da auch nicht unbedingt so sehr ins Detail gehen.

Ein kleinbisschen denke ich auch noch darüber nach, ob zumindest bei Verbreitungsdaten die Specimens nicht noch präsenter sein könnten als „lediglich“ als Quelle der Arealverbreitung. Ich denke da z.B. an die Darstellung der Verbreitung in der Flora von Usbekistan, die dort in der Publikation primär als Specimens aufgeführt werden, mit der Arealabkürzung lediglich hinter mehreren Specimens in Klammern und teilweise auch gar keinem Subarealzugeordnet glaube ich. Um so etwas darzustellen wäre die jetzige Modellierung evtl. nicht optimal. Aber vielleicht müssen wir auch nicht alles darstellen können und uns lieber auf eine Version konzentrieren, die grundsätzlich logisch ist und die wir dann auch gut behandeln.

P.S. Vielleicht gäbe es auch noch die Möglichkeit einen Referenztyp „Specimen“ hinzuzufügen, parallel zu book, section, article, map, database, … Aber ich glaube, da gäbe es wenig Widerverwendung in der Klasse und es scheint mir auf Anhieb nicht so viel Sinn zu machen.

WGB:

ja, ich denke, als Referenztyp könnte höchstens die Field Unit gelten, da wäre das Detail dann das Specimen – aber die andere Lösungsidee scheint mir besser:

> nicht doch eine konkretes (nicht generisches Feld) haben wollen, also ähnlich wie die (normale) Referenz, die ja auch nicht das generische Feld verwendet. Ich tendiere glaube >ich zu letzterem, da es sich hier ja um eine semantisch relativ klarer Beziehung handelt.

Ja, eigentlich beruht ja alles letztendlich auf Specimen als Quelle der Information, ich habe auch in meinen Taxonomie-Vorträgen immer Literatur und Specimen als die Haupt-Quellen für den taxonomischen Arbeitsprozess drin.
Um das wirklich überall einsetzen zu können, wären aber wohl mehrfach-Nennungen nötig ...

AM:

ja, die Kardinalität war auch das nächste, was ich noch klären wollte.
Leider führt „Mehrfachnennung“ dann gleich zu einer höheren Komplexität in vielen Bereichen, daher sollte man da ein bisschen überlegen, ob und wie man das angeht. Ein ähnliches Problem haben wir bereits bei nameUsedInSource, was ja auch ein Attribut der OriginalSource Klasse ist, und wo es in Realität oft genug der Fall ist, dass in der Originalquelle eigentlich mehrere Namen stehen die die Information belegen, da die Taxonomie dort anders ist.
Für E+M haben wir das schön öfters diskutiert. Die derzeitige Lösung ist, dass in diesem Fall mehrere Quellen angelegt werden, die sich nur im NameInSource unterscheiden. Wirklich schön finde ich das dort nicht, da es ja nicht wirklich verschiedene Quellen sind. Für die Specimen/Voucher scheint mir das aber eine gute Lösung zu sein, da es ja wirklich verschiedene Quellen sind, die alle den einen Distribution-Fact bestätigen.

Alternativ könnte es eine Liste sein, was aber die benannte Komplexität in Datenmodel und im TaxEditor-UI mit sich bringt.

Speziell bei Specimens gäbe es mit dem jetzigen Model sogar noch eine 3. Lösung, nämlich dass Specimens zusammengefasst werden über derivation events zu einer Gruppe, die dann selber auch wieder ein Beleg ist. Das gibt das jetzige Model bereits her und wird glaube ich in der Zoologie öfters gemacht. Aber auch da fehlt bislang die Implementierung im TaxEditor für die Zusammenfassung und in diesem Fall erscheint mir das auch nicht wirklich korrekt.

Ich würde also für die erste Lösung plädieren, laut der jeder Beleg als Einzelquelle hinzugefügt wird und die Verbreitung dann eben mehrere Quellen hat.

KL:

ich denke auch, dass man an dieser Stelle eher mit mehreren Quellen arbeiten sollte als mit einer Liste, da es ja in den meisten Fällen unterschiedliche Aufsammlungen sind und somit auch einzelne Quellen der Information.
Specimen in einem DerivationEvent zusammen zu fassen, würde ja nur für Specimen Sinn machen, die zu einer FieldUnit gehören, alles andere fände ich eher als unsaubere Modellierung.

WGB:

ja, ich stimme Euch zu.

> da es ja wirklich verschiedene Quellen sind, die alle den einen Distribution-Fact bestätigen

Das kann aber, wie gesagt, auch ein anderer Fact sein – z.B. ein Common Name oder ein morphologisches Feature. Perspektivisch können das auch in-Text Referenzen sein, so kommt es öfters vor, das in einer Note vom Normalfall des Taxons abweichende Specimen genannt werden. Die Analogie zu Literaturreferenzen ist also ziemlich breit.
Zur Frage Etikett / Specimen: Wir sollten aber davon ausgehen, dass (botanisches) Specimen und das dort angebrachte Label ein zitierbares Objekt sind. Falls da Interpretationen des Label Textes vorgenommen wurden, sollten die bei den Specimen Daten (auf die ja verwiesen wird) dokumentiert sein.

AM:

> Die Analogie zu Literaturreferenzen ist also ziemlich breit.

Das ist mit der Lösung modellseitig ja gegeben dadurch, dass es sich beide Male um eine „Quelle“ (OriginalSource) handelt und diese dann lediglich entweder auf eine „Referenz“ oder einen Beleg verweist. Wir unterscheiden im CDM ja Quelle und Referenz, was der normalen Semantik der Wörter vielleicht nicht gerecht wird.

#2 - 12/02/2022 10:19 AM - Andreas Müller

- Target version changed from CDM UML 5.43 to Release 5.35

#3 - 12/02/2022 09:11 PM - Andreas Müller

- Status changed from New to Resolved

- % Done changed from 0 to 70

#4 - 12/14/2022 10:36 AM - Andreas Müller

- Related to feature request #10202: Allow entering specimen in fact sources in taxeditor added

#5 - 12/14/2022 10:39 AM - Andreas Müller

- Related to feature request #10203: Show specimen sources in publications added

#6 - 12/14/2022 10:41 AM - Andreas Müller

- Assignee changed from Andreas Müller to Katja Luther

I guess we can close this ticket. Can you please do final review. Please also consider implementation in TaxEditor (#10202 - most important), dataportal and cdmlight (#10203 - for now less important).

#7 - 12/16/2022 02:19 PM - Katja Luther

- Assignee changed from Katja Luther to Andreas Müller

#8 - 12/17/2022 01:31 AM - Andreas Müller

- Status changed from Resolved to Closed

- % Done changed from 70 to 100

#9 - 12/19/2022 03:27 PM - Andreas Müller

Added constraint that either reference or specimen can have a value but not both.