

Edit - feature request #3685

Open tissue sample and molecular issues

08/29/2013 01:52 AM - Andreas Müller

Status:	In Progress	Start date:	08/29/2013
Priority:	Priority11	Due date:	
Assignee:	Andreas Müller	% Done:	0%
Category:	cdm	Estimated time:	0.00 hour
Target version:	CDM UML 5.8		
Severity:	normal		

Description

this ticket holds all remaining tasks for #3360

Open fields from Excel tabel:

- Separate class for tissue samples needed? (separate recordBasis already exists) see also #4233
- (DNA) Extraction Number, maybe handled by any of the existing DerivedUnit identifiers?
- Preservation (probably handled by MaterialsAndMethods)
 - Medium, temperature, date,
 - repeatable (first condition, later second condition ,...)
- Preparation (probably handled by MaterialsAndMethods)
 - Materials&Methods, prepared by, date - #4442
- Sample Type (Tissue, DNA, RNA) !! - do we need an extra recordBasis for RNA?
- DNA Quality #4434
 - DNA Purification Method: Datatype
- Use Material&Method also for DerivationEvent (#3717)
- How to implement link to Material&Method citation
- Primer: how to implement link of reference
- DNASample do have the important field SampleProvidedBy. This could maybe be modeled as a DerivationEvent.actor of an Accessioning Event, but probably this is not clear enough. Needs to be discussed. Do we need this field for other types of Specimen too? -> do we need to store this in CDM at all (we are not a collection management system!)?
- sequence references may also include not only the publication but also additional files which hold the data itself. We need to handle this somehow
- DnaSamples may include multiple taxa, only the Consensus Sequence may belong to a single taxon (e.g. algae samples), do we need determination also for Sequence?
- see also #3713 for the CollectionUnit <-> Specimen discussion
- marker hierarchies: currently combined markers are handled as parents of each single marker, this may not work if we have 2 overlapping combined markers, therefore we need a more general handling of marker terms here (a single parent for each marker is not enough)
- MaterialSample type [#4137](#)

History

#1 - 09/12/2013 01:23 PM - Andreas Müller

- Status changed from New to Resolved

#2 - 09/12/2013 01:24 PM - Andreas Müller

- Status changed from Resolved to In Progress

#3 - 09/12/2013 01:24 PM - Andreas Müller

- Status changed from In Progress to New

#4 - 09/25/2013 10:30 AM - Andreas Müller

- Priority changed from Highest to Priority14

#5 - 03/14/2014 12:14 PM - Andreas Müller

Liebe Campanula-Kollegen,

einige Monate sind vergangen, seit wir uns durch die DNA-Elemente und die Modellierung des CDM gekämpft haben. Die Term-Liste ist inzwischen auch um Elemente aus dem Metagenomics-Bereich erweitert worden und wir betreiben im Moment ein Review der Terme. Falls Ihr mögt würde ich mich freuen auch von Euch ein Feedback zu bekommen und mir die Liste ausgefüllt zurückzuschicken. Auch im Rahmen von SYNTHESYS geht es um die Schritte zwischen Extraktion und finaler Sequenz in GenBank etc., so dass ich denke wir sollten uns hier auf jeden Fall kurzschließen wie Ihr Euch die Workflows speziell für Campanula und das CDM überlegt habt. Das wäre mir eine große Hilfe bei der Umsetzung der Bedürfnisse in unserer eigenen Bank.

Viele Grüße

Gabi

Hallo Gabi, Norbert, Tilo,

pünktlich zum 14.3. habe ich nun endlich Zeit gefunden in den Term-Entwurf zu schauen. Zu den dort gestellten fragen wie Needed, Usually Collected, Ease, ... kann ich natürlich nichts sagen, aber ich habe den Entwurf einfach mal mit dem verglichen, was wir derzeit im CDM haben oder vorhaben.

Dabei sind mir einige Fragen gekommen. Einige davon Verständnis fragen, da der Liste keine Semantik beigefügt ist und sich mir als nicht Biologen nicht unbedingt immer alle Begriffe sofort erschliessen. :

- 1) Was bedeutet das Feld sequence bei Tissue facts ?
- 2) Was bedeutet sequence bei DNA/RNA. Die Consensus Sequence kann hier ja wohl nicht gemeint sein, da die noch später kommt.
- 3) MethodDeterminationConcentrationAndRatios: Ist das alles eine Methode oder mehrere?
- 4) Was bedeutet CompositeDesign bei Tissue facts?
- 5) Brauchen wir CompositeDesign und SampleWeight für Tissue facts im CDM auch. Bisher wurden diese Parameter nicht genannt.
- 6) Ist PreservationType gleichzustellen mit PreservationMedium in älteren Versionen? Wenn ja, warum wurde es umbenannt?
- 7) Ist PreparationType gleichzustellen mit Preparation Materials And Methods? Wenn ja, warum wurde es umbenannt?
- 8) Im CDM brauchen wir die Angaben Volume - QualityCheckDate nicht, oder?
- 9) GelRemarks haben wir im CDM bislang nicht. Brauchen wir das oder fliegt das zusammen mit den Amplificationsdaten sowieso eher raus?
- 10) GSC related DNA facts: Ist das für CDM relevant? Sagt mir inhaltlich nichts.
- 11) ConsensusSequence: Es wundert mich, dass diese unterhalb Amplification aufgezählt wird. Sollten wir dafür nicht eine eigene Rubrik haben. Insbesondere auch, wenn man bedenkt, dass das Verhältnis nicht unbedingt immer 1:1 ist?
- 12) GeneticAccessionNumberURI: Was ist hiermit genau gemeint? Das hängt doch davon ab, auf welches System man verlinkt, daher würde ich es entweder GenBankAccessionURI/EMBLAccessionURI/... nennen oder ganz weglassen, da es leicht on the fly berechnet werden kann. Letzteres machen wir im CDM.
- 13) GSC related Amplification facts: Diese behandeln wir im CDM derzeit nicht. Benötigt?

- 13) Single Sequencing: FragmentLength. Das gibt es im CDM bislang nicht und wurde bislang auch nicht genannt. Wird es benötigt?
- 14) Primer: Adapters und MultiplexIdentifiers gibt es bislang nicht im CDM. Benötigt?
- 15) Bei der ConsensusSequence wurde auf die Angabe von Referenzen verzichtet. Bewusst, weil diese auch über GenBank zu beziehen sind oder gibt es anderen Grund?
- 16) Auch das contigFile steht nicht zur Verfügung. Gewollt?
- 17) Bei ChromatogramFileURI hatten wir uns im CDM für Pherogram als Begriff entschieden, da dieser spezifischer ist.
- 18) MaterialSampleID: Verstehe ich es richtig, dass dies die ID des Samples ist? Ist der Begriff aus einem anderem Schema abgeleitet oder für GGBN neu kreiert? (im CDM benutzen wir nur ID für alle Klassen, lediglich Fremdschlüssel werden mit semantischen Namen versehen).
- 19) Sind die Beziehungen zwischen den einzelnen Teilen bewusst weggelassen worden oder handelt es sich wirklich um eine flache Liste? Ich fände es hilfreich, wenn diese Felder auch mit auf die Liste kämen, so dass diese klar sind und auch klar ist, inwieweit Mehrfachverwendung von gewissen Elementen abgedeckt und benötigt wird.

Soviel erstmal.

@Norbert, Tilo: Hattet ihr eigentlich auch schon auf die Liste geschaut?

Viele Grüße,

Andreas M.

Hallo Andreas,

erstmal vielen Dank fürs Durchkucken. Ich bin ab Mittwoch wieder da, aber kurz eine Rückmeldung.

Alles, was unter GSC gelistet ist stammt aus der Metagenomics-Szene (Bakterien, 10.000e Sequenzen pro Sample) und ist bis auf Weiteres nicht relevant fürs CDM, z.B: Deine Fragen 4,5, 10, 13a, 14

MaterialSampleID kommt durch die Diskussion mit DarwinCore-Archive, dort wird diese ID benötigt (kann alles mögliche sein, z.B. Datenbank-ID), in ABCD wird es z.B. nicht benötigt.

Die Relationen zueinander sind bewusst unterdrückt worden. Es geht hier rein um die Terme. Wie man sie implementiert und in Relation setzt soll hier nicht Thema sein.

Dokumentation gibt's hier: <http://goo.gl/WgsAqs>

Soweit erstmal.

Viele Grüße

#6 - 10/18/2014 01:01 AM - Andreas Müller

- Status changed from New to In Progress

#7 - 03/20/2015 01:24 PM - Andreas Müller

- Priority changed from Priority14 to Priority12

#8 - 06/04/2015 02:20 PM - Andreas Müller

- Target version deleted (CDM UML 3.6)

#9 - 05/02/2016 03:45 PM - Andreas Müller

- Target version changed from CDM UML 4.0 to CDM UML 4.1

#10 - 12/01/2016 01:38 PM - Andreas Müller

- Target version changed from CDM UML 4.1 to CDM UML 4.7

#11 - 06/05/2017 05:03 PM - Andreas Müller

- Description updated

- Private changed from Yes to No

#12 - 06/07/2017 01:17 AM - Andreas Müller

- Target version changed from CDM UML 4.7 to CDM UML 5.0

#13 - 05/08/2018 06:46 PM - Andreas Müller

- Target version changed from CDM UML 5.0 to CDM UML 5.5

#14 - 01/19/2019 12:31 AM - Andreas Müller

- Priority changed from Priority12 to Priority11

#15 - 02/19/2019 10:06 AM - Andreas Müller

- Target version changed from CDM UML 5.5 to CDM UML 5.8

Files

DNA_9.9.2013.xls	37 KB	09/11/2013	Andreas Müller
GGBN Data Standard_review.xlsx	19.4 KB	03/14/2014	Andreas Müller