



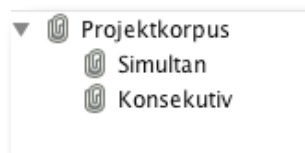
Die Organisation von Coma-Metadaten

Organisation von Coma-Metadaten

Die Struktur, in der Metadaten in Coma verwaltet werden, kennt x verschiedene vorgegebene Datencontainer: Korpora, Kommunikationen, Sprecher, Transkriptionen, Aufnahmen. Unter diesen Containern existieren verschiedene Datentypen, die in mehreren dieser Container Verwendung finden können. Da es wichtig ist, das Zusammenspiel dieser Container zu verstehen, sollen sie hier kurz beschrieben werden:

Korpora:

Korpora bilden die Container für alle anderen Metadatentypen. Korpora können entweder Korpusdaten (also Kommunikationen und Sprecher) oder aber weitere Korpora enthalten, nicht jedoch beides.



Das hier abgebildete Projektkorpus kann also keine Korpusdaten zu Sprechern und Kommunikationen enthalten, da es bereits weitere (Sub-)Korpora enthält. Wie Korpora angelegt, zugeordnet und verwaltet werden finden Sie unter (#Korpusverwaltung).

Kommunikationen

Mit Kommunikationen werden Gesprächsereignisse verwaltet. An Kommunikationen nehmen typischerweise Sprecher Teil und es kann Aufnahmen und Transkriptionen zu den Kommunikationen geben - im Datenmodell können Aufnahmen, Transkriptionen und Sprecher diesen Kommunikationen zugeordnet sein. Mit den Kommunikationen werden alle Konstellationsdaten zu dem Gesprächsereignis gespeichert, also Dinge wie Ort, Zeit, besondere Umstände, beteiligte Sprachen etc.

Sprecher

Sprecher sind Personen, die sich an der Kommunikation beteiligen. Es muss sich dabei nicht um natürliche Personen handeln (es kann beispielsweise auch ein automatisiertes Dialogsystem als Sprecher geführt werden) und die Sprecher müssen in der Kommunikation auch nicht unbedingt sprechen; wenn sie für die Kommunikation von Belang sind, so werden

sie hier erfasst. Die Daten, die zu den Sprechern erfasst werden, sind von der Kommunikation zunächst unabhängig. Es sollten keine Kommunikationsbezogenen Daten mit den Sprechern erfasst werden, damit Sprecher einfach mehreren Kommunikationen zugeordnet werden können. Erfasst werden dagegen Daten wie Geburtsort und -Datum, biographische Angaben, gesprochene Sprachen und deren Status etc.

Aufnahmen

Aufnahmen beziehen sich auf Aufnahmen, die zu Kommunikationen angefertigt wurden. Das Coma-Datenmodell unterscheidet zwar Aufnahmen und tatsächliche Mediendateien, diese Unterscheidung ist aber in der Programmoberfläche noch nicht repräsentiert. Aufnahmen lassen sich nur anlegen, wenn es bereits eine Kommunikation gibt, der sie sich zuordnen lassen.

Transkriptionen

In Transkriptionen werden Verknüpfungen zu EXMARaLDA-Transkriptionen gespeichert. Coma verwaltet dabei sowohl Basis- als auch segmentierte Transkriptionen. Zur Anzeige von Basistranskriptionen und zu ihrer Behandlung lassen sich in den #Einstellungen von Coma vorgaben machen. ComaTranskriptionen sind auch - genau wie Aufnahmen - mit Kommunikationen verknüpft - sie lassen sich ohne Kommunikation nicht verwalten. Auch mit den Aufnahmen sind sie nur über die Kommunikation verbunden.



Weitere Datentypen

Zur Erfassung der eigentlichen Metadaten existieren noch weitere Datentypen, z.B. zur einheitlichen Erfassung von Sprachen und Orten. Zwei dieser Datentypen sind dabei von besonderer Bedeutung:

Location

Eine Location bezeichnet in Coma nicht einfach nur einen bestimmten Ort, sondern einen bestimmten Ort zu einer bestimmten Zeit, da biographische und andere relevante Ereignisse häufig einen bestimmten Ort UND eine bestimmte Zeit haben.

Locations +	
Location	
City	Istanbul
Country	TR
PStart	18.12.1997 00:00:00
Description (Location)	
Typ	Geburt
Location	
City	Istanbul
Country	TR
Description (Location)	
Typ	Wohnort

In dem oben angegeben Fall wird mit einer Location beispielsweise Geburtsdatum und -Ort eines Sprechers erfasst, in einer weiteren Location nur der Wohnort. Es müssen in Locations nicht immer Ort UND Zeit vermerkt werden, es ist aber wichtig zu wissen, dass man auch eine Location verwenden muss, wenn man nur einen Zeitpunkt vermerken will.

Description

Da es extrem schwierig ist, standardisierte Metadaten für alle erdenklichen Fälle vorzusehen, wird in Coma der überwiegende Teil der Metadaten mit freien Schlüssel-Wert-Paaren erfasst. Diese Paare werden unter dem Datentyp „Description“ zusammengefasst. Descriptions existieren für die meisten der in Coma existierenden Datentypen: So lassen sich Descriptions für Korpora, Kommunikationen, Sprecher, Locations, Aufnahmen etc. anlegen.

Speaker: Kür (Kürsat) 	
Description (Speaker) 	
 Vorname	 Kübra
 SprecherNr.	 187
 Nachname	 Baydar
 Familie (Pseudo)	 Bayar
 Eigenschaft	 Geschwisterkind

In diesem Beispiel sind in einer Description Metadaten zu einem Sprecher erfasst. Dass sich die Felder innerhalb von Descriptions frei benennen lassen macht sie nicht beliebig: Jeder Anwender sollte sich für sein Korpus ein festes Inventar an Description-Schlüsseln ausdenken und dokumentieren, um später mit den Metadaten