DINA collections data model

Markus Englund & Lisa Sundström

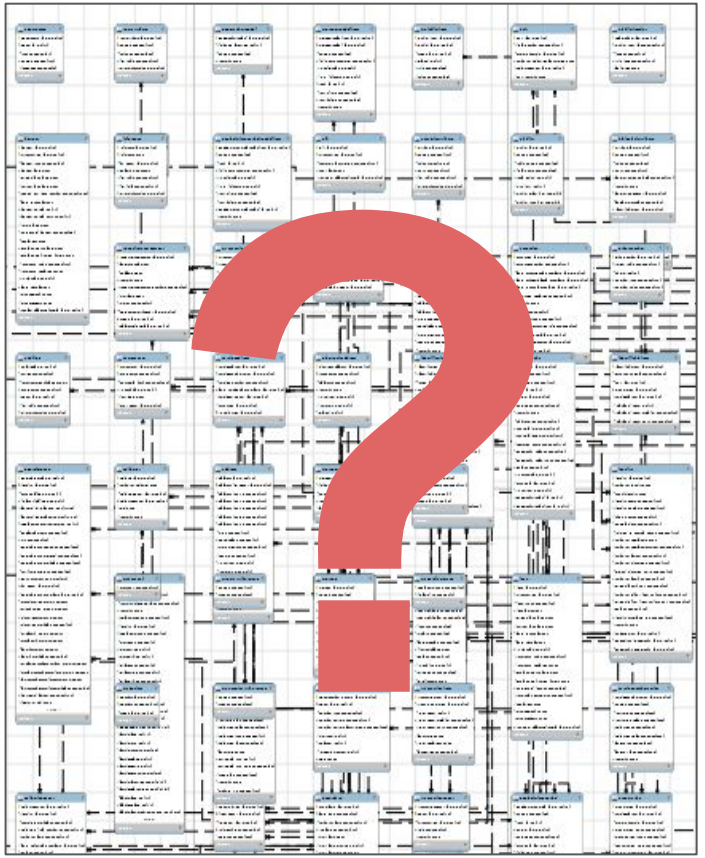
May 17, 2017

Our mission

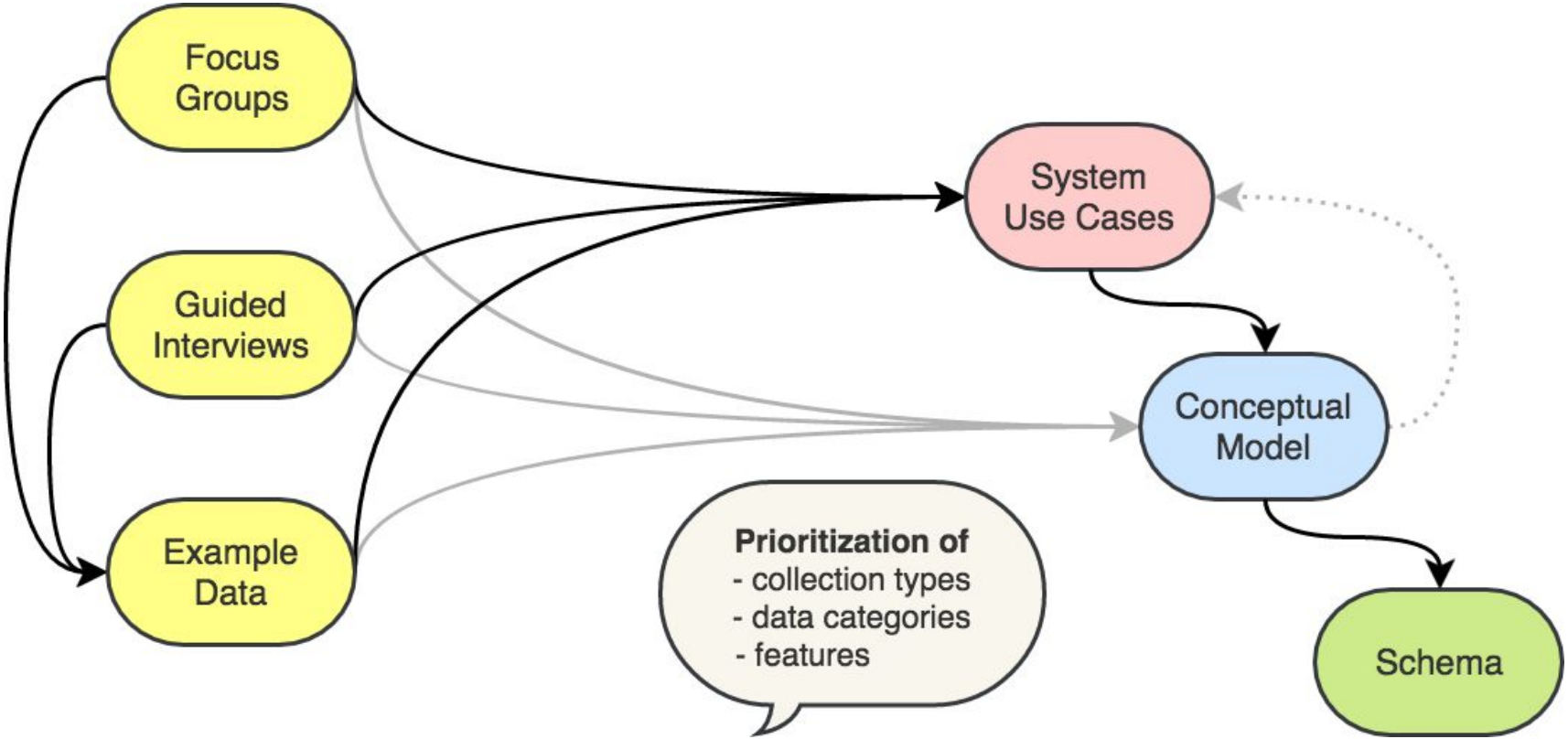
Investigate needs

Core conceptual model

Schema “starting point”



How do we want to evolve the schema?



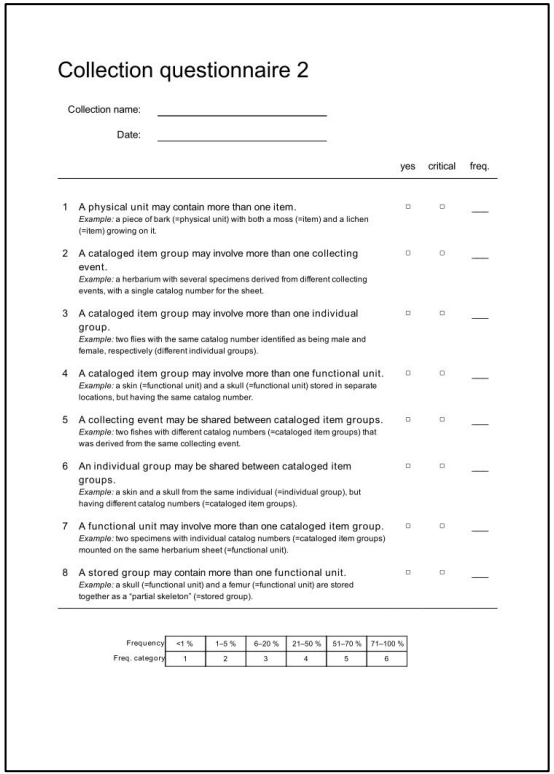
Guided interviews

A common terminology to use in our evaluation.

Guided interviews based on a questionnaire with about 10 questions focusing on the “core”.

Individual interviews with people representing different collections (NRM, Uppsala Museum of Evolution).

For each question, we also asked about the commonness and importance of each “feature”.



Collection questionnaire 2

Collection name: _____

Date: _____

		yes	critical	freq.
1	A physical unit may contain more than one item. *Example:* a piece of bark (=physical unit) with both a moss (=item) and a lichen (=item) growing on it.	☐	☐	___
2	A cataloged item group may involve more than one collecting event. *Example:* a herbarium with several specimens derived from different collecting events, with a single catalog number for the sheet.	☐	☐	___
3	A cataloged item group may involve more than one individual group. *Example:* two flies with the same catalog number identified as being male and female, respectively (different individual groups).	☐	☐	___
4	A cataloged item group may involve more than one functional unit. *Example:* a skin (=functional unit) and a skull (=functional unit) stored in separate locations, but having the same catalog number.	☐	☐	___
5	A collecting event may be shared between cataloged item groups. *Example:* two fishes with different catalog numbers (=cataloged item groups) that was derived from the same collecting event.	☐	☐	___
6	An individual group may be shared between cataloged item groups. *Example:* a skin and a skull from the same individual (=individual group), but having different catalog numbers (=cataloged item groups).	☐	☐	___
7	A functional unit may involve more than one cataloged item group. *Example:* two specimens with individual catalog numbers (=cataloged item groups) mounted on the same herbarium sheet (=functional unit).	☐	☐	___
8	A stored group may contain more than one functional unit. *Example:* a skull (=functional unit) and a femur (=functional unit) are stored together as a "partial skeleton" (=stored group).	☐	☐	___

Frequency	<1 %	1–5 %	6–20 %	21–50 %	51–70 %	71–100 %
Freq. category	1	2	3	4	5	6

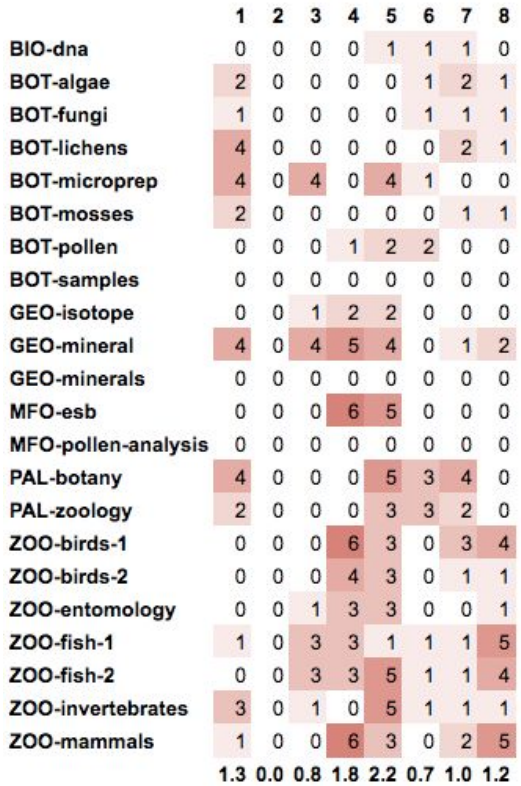
The guided interviews were rewarding

A lot of new information in a structured way.

We were also able to capture different personal views on collection management.

The result can help us prioritizing development of new features.

Our terms and the questionnaire are updated as we learn more!



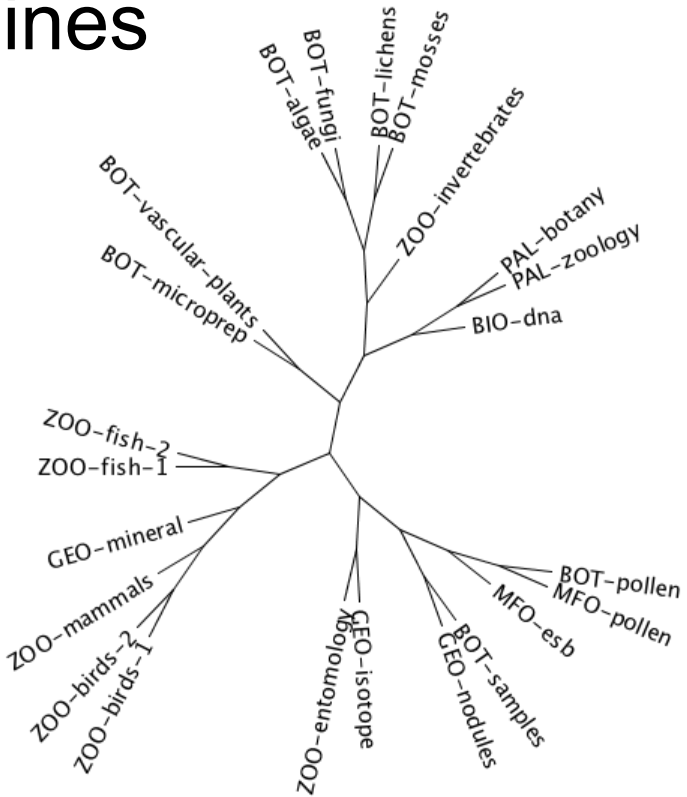
	1	2	3	4	5	6	7	8
BIO-dna	0	0	0	0	1	1	1	0
BOT-algae	2	0	0	0	0	1	2	1
BOT-fungi	1	0	0	0	0	1	1	1
BOT-lichens	4	0	0	0	0	0	2	1
BOT-microprep	4	0	4	0	4	1	0	0
BOT-mosses	2	0	0	0	0	0	1	1
BOT-pollen	0	0	0	1	2	2	0	0
BOT-samples	0	0	0	0	0	0	0	0
GEO-isotope	0	0	1	2	2	0	0	0
GEO-mineral	4	0	4	5	4	0	1	2
GEO-minerals	0	0	0	0	0	0	0	0
MFO-esb	0	0	0	6	5	0	0	0
MFO-pollen-analysis	0	0	0	0	0	0	0	0
PAL-botany	4	0	0	0	5	3	4	0
PAL-zoology	2	0	0	0	3	3	2	0
ZOO-birds-1	0	0	0	6	3	0	3	4
ZOO-birds-2	0	0	0	4	3	0	1	1
ZOO-entomology	0	0	1	3	3	0	0	1
ZOO-fish-1	1	0	3	3	1	1	1	5
ZOO-fish-2	0	0	3	3	5	1	1	4
ZOO-invertebrates	3	0	1	0	5	1	1	1
ZOO-mammals	1	0	0	6	3	0	2	5
	1.3	0.0	0.8	1.8	2.2	0.7	1.0	1.2

Collection management differ, even within departments and within disciplines

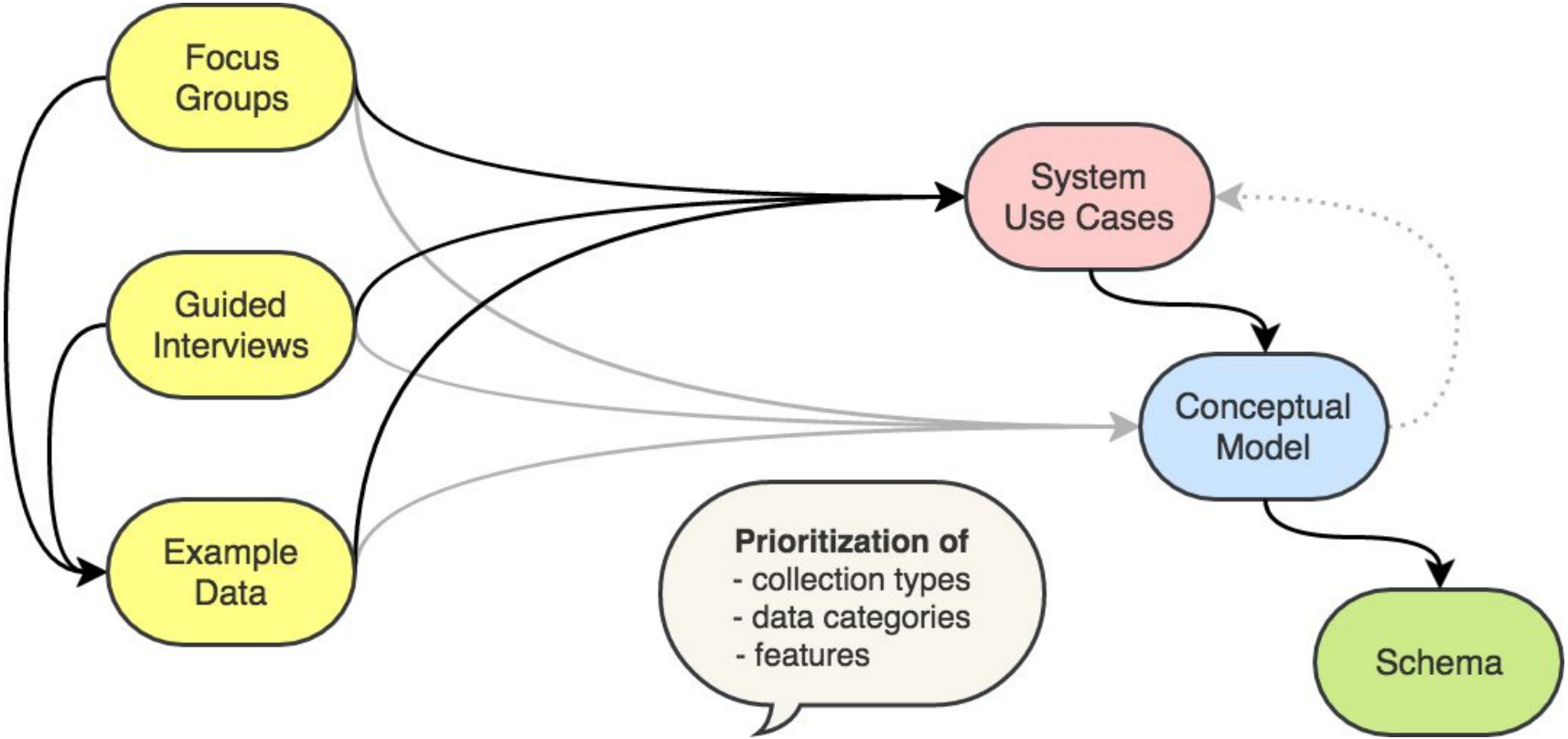
Why do management differ?

Different

- kinds of objects
- information foci
- management strategies
- software



How do we want to evolve the schema?



Requirements are captured in system use cases

Only for developing the data model

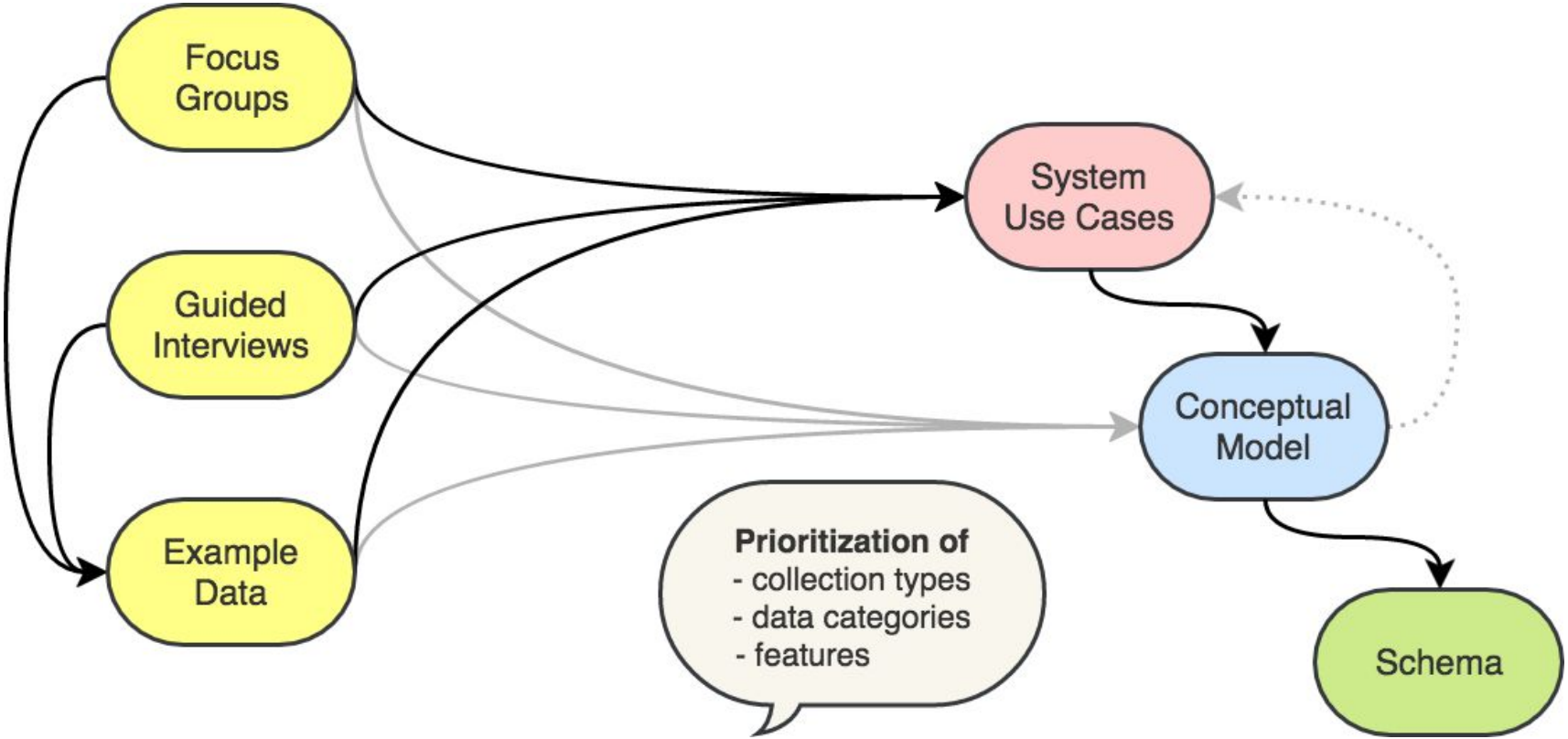
A standardized way of documenting interactions between actors and the system

Written according to a simple template:

1. Short background/motivation (why do we need this use case?)
2. Actors
3. Preconditions
4. Course of events
5. Alternative paths
6. Success post-conditions
7. Notes
8. Model treatment (how is the use case handled by different information models?)

https://github.com/DINA-Web/information-model-use-cases/blob/master/use-cases.rst

How do we want to evolve the schema?



Conceptual model development

Focusing on concepts rather than implementation details.

Use for communication.

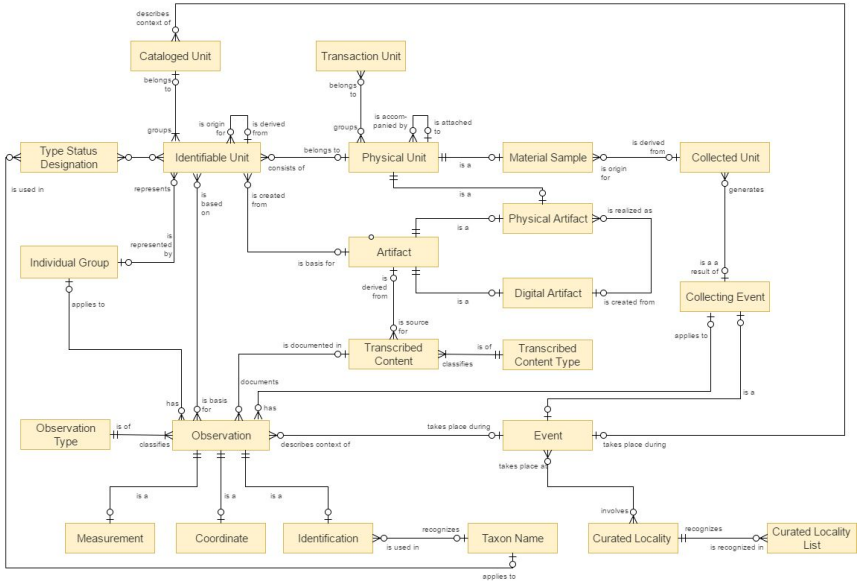
Well defined concepts.

Explicitly defined relations between entities.

A stable core.

Adding more entities/relations iteratively.

Always ahead of schema development.



Conceptual model to physical schema

https://github.com/DINA-Web/dina-collections-data-model

Design principles

Data dictionary

version_planned	version_added	version_removed	table_name	display_name	description
0.1		NA	cataloged_unit	cataloged unit	A set of identifiable units listed under the same catalog number.
0.1		NA	collecting_event	collecting event	The event (time and location) when the collecting took place.
0.1		NA	collected_unit	collected unit	The smallest unit distinguished during (or in association with) the collectin
0.1		NA	event	event	A thing that happens or takes place at a given time and location.
0.1		NA	identification	identification	A statement about a specimen's taxon name and/or type status.
0.1		NA	identifiable_unit	identifiable unit	The smallest recognized (not recognisable) "thing" in the collection, either
0.1		NA	individual_group	individual group	One or more individuals with a single set of individual characteristics.
0.1		NA	material_sample	material sample	The smallest distinguished physical "thing" in the collection, which is assur
0.1		NA	observation	observation	The data measured, collected, perceived or noticed by an agent (e.g. a per
0.1		NA	observation_type	observation type	The kind of observation.
0.1		NA	physical_unit	physical unit	A physical assemblage containing one or more "things" in the collection.
0.1		NA	transcribed_content	transcribed content	A text that is transcribed from any kind of original source (labels, field not
NA	NA	NA	measurement	measurement	
NA	NA	NA	transaction_unit	transaction unit	The participation of one or more material samples in a transaction (e.g. a l

The schema starting point

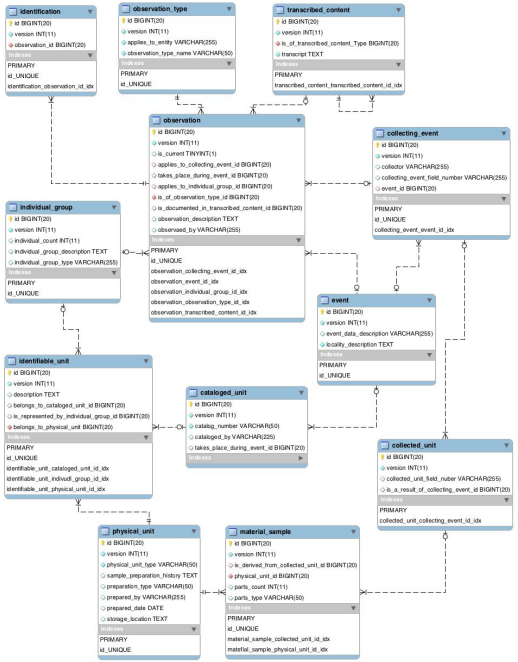
Limited to about 12 tables representing the core of the conceptual model.

Few columns – with only the most critical attributes.

Free text fields as placeholders for some tables or modules to be added later.

No verbatim fields.

An agile approach will allow us to investigate the needs for the model along with the feature development.



What next?

Decide on the schema design principles.

Decide on the schema starting point.

Decide on which parts of the data model that we want to focus on next (guided by the development of the UI and the data we choose to support).

Real data and import/upgrade tools for data together with database version (for example using Liquibase).