# Knowledge Management and Ontology of Forensic Anthropology Research Data From Texas State

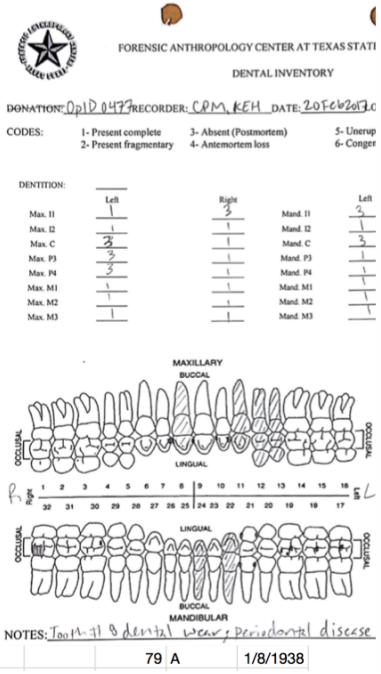
Felix Engel[1], Stephan Schlager[1],
Nicholas Herrmann[2], and Alexandria Frye[2]

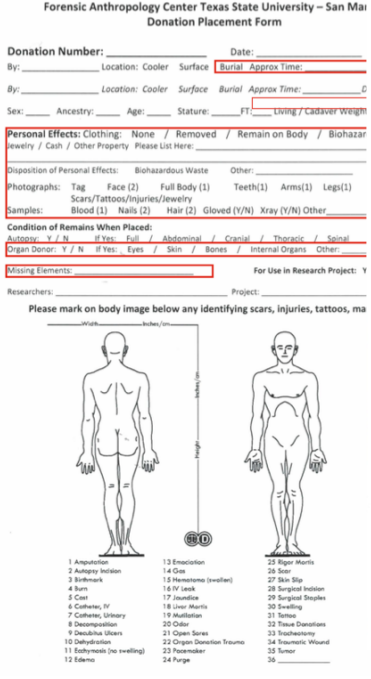
[1]Albert Ludwigs University of Freiburg; [2]Texas State University- San Marcos

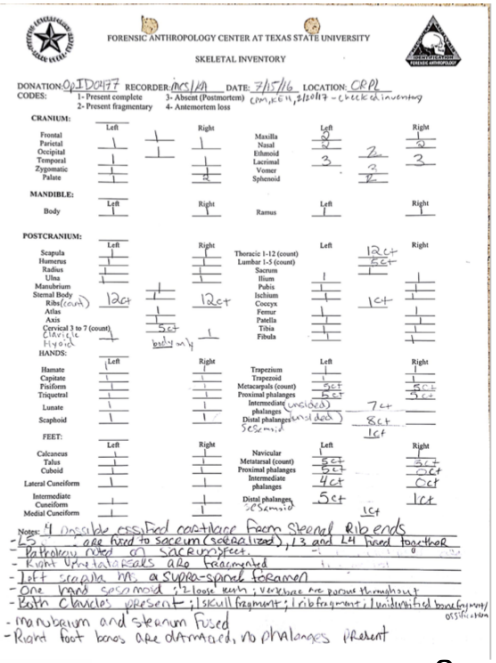
# Data Management Issue

Various data collection points

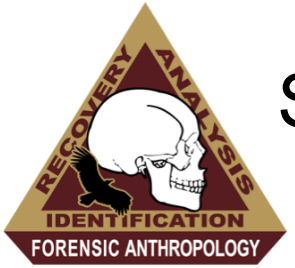
| | A | B | C | D | E | F | |
|---|---|---|---|---|---|---|---|
| 1 | Year | TxState II | Number | Received | NAME | Sex | Rac |
| 2 | 2017 | D44-2017 | 421 | 7/27/17 | | M | W |
| 3 | 2017 | D45-2017 | 422 | 7/31/17 | | F | H |
| 4 | 2017 | D46-2017 | 423 | 8/6/2017 | | F | W |
| 5 | 2017 | D47-2017 | 424 | 08/11/2017 | | M | W |
| 6 | 2017 | D48-2017 | 425 | 8/18/2017 | | M | W |
| 7 | 2017 | D49-2017 | 426 | 8/11/2017 | | M | Am. |
| 8 | 2017 | D50-2017 | 427 | 8/22/2017 | | M | W |
| 9 | 2017 | D51-2017 | 428 | 8/30/2017 | | M | W |
| 10 | 2017 | D52-2017 | 429 | 9/1/2017 | | M | W |
| 11 | 2017 | D53-2017 | 430 | 9/4/2017 | | F | W |
| 12 | 2017 | D54-2017 | 431 | 12/30/1899 | | M | W |
| 13 | 2017 | D55 2017 | 432 | 9/7/17 | | M | B |
| 14 | 2017 | D56-2017 | 433 | 9/7/2017 | | F | W |
| 15 | 2017 | D57-2017 | 434 | 10/26/1945 | | M | W |
| 16 | 2017 | D59-2017 | 436 | 10/01/2017 | | M | W |
| 17 | 2017 | D60-2017 | 437 | 10/2/2017 | | M | W |
| 18 | 2017 | D61-2017 | 438 | 10/4/2017 | | F | H |
| 19 | 2017 | D62-2017 | 439 | 10/7/2017 | | F | W |
| 20 | 2017 | D63-2017 | 440 | 10/18/2017 | | M | W |
| 21 | 2017 | 67-2017 | 441 | 11/1/2017 | | M | W |



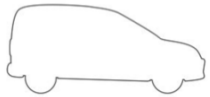



# Simplified Data Collection and Storage Process

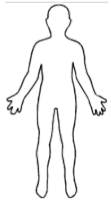



Interested Donors

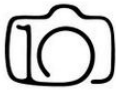
Donor Paperwork



Donation Received

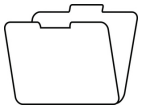


Donation Placed at FARF



Photos

Notes Taphonomy



Donation Records at FACTS

Biological Samples Collected

Donation Recovered, Processed, Inventoried

PC

Database

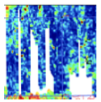
Database

Donation Curated

# OpID Data Collection and Storage Process

OpID Cases Identified

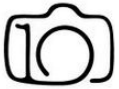
Case Paperwork and Jurisdictional Permissions



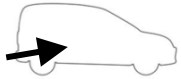
Geophys Survey

Exhumation (TxState/ UIndy)

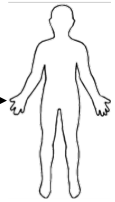
Notes; Taphonomy



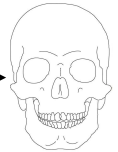
Photos



OpID Cases Transported



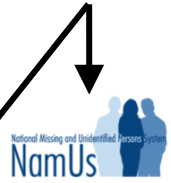
OpID Cases Placed at FARF



OpID Cases Processed and Inventoried

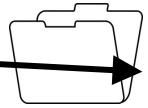
DNA samples (UNT/Bode) & other samples

Anthropology Report




NamUs

OpID Case Records at ORPL

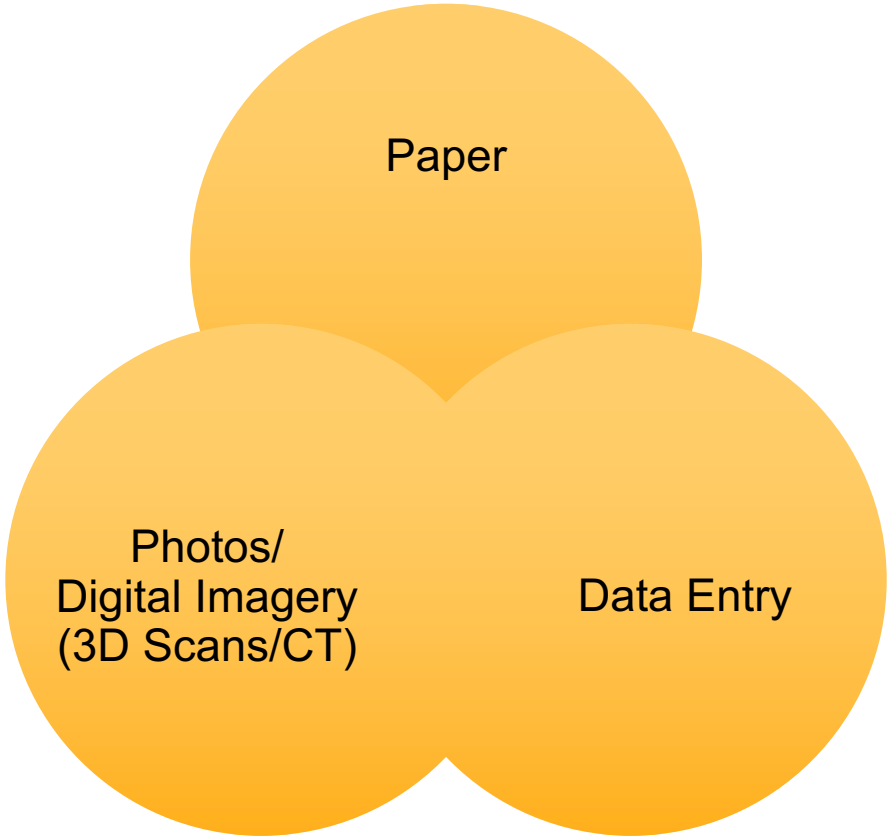


Identification and Repatriation

OpID Cases Maintained

Intake Inventory Form

Taphonomy Recording




Taphonomic Photos
Skeletal and Dental
Inventory
In-take Photos
Scan Data

Living Donor Form
Skeletal Inventory
Postcranial
and
Craniometric data

# Our Data Problem – A good thing but needs to be managed

Both a digital and paper problem

Not new problems – faced by all labs

Various collections levels (student/researcher/faculty)

How is data collected, where it's maintained, and how it's accessed.

- Several terabytes of photo data

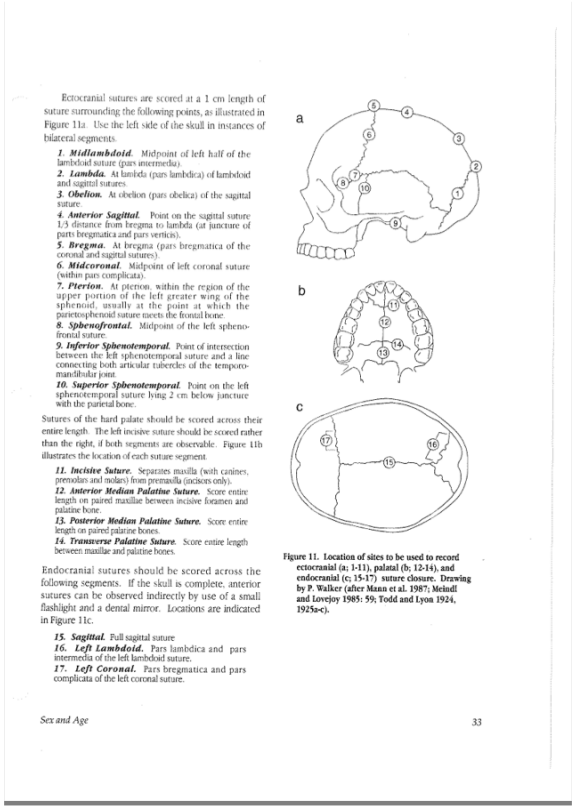
- Ex. Taphonomic data on paper forms only
- Skeletal Inventories
- What is transcribed into DBs

# Information Management in Physical Anthropology

# Data Acquisition Standards

- Implement research methods
- Explain how information is to be obtained
- Define how data are to be recorded

Buikstra & Ubelaker 1994, p.33 and appendant form



Ectocranial sutures are scored at a 1 cm length of suture surrounding the following points, as illustrated in Figure 11a. Use the left side of the skull in instances of bilateral segments.

***1. Midlambdoid.*** Midpoint of left half of the lambdoid suture (pars intermedia).
***2. Lambda.*** At lambda (pars lambdica) of lambdoid and sagittal sutures.
***3. Obelion.*** At obelion (pars obelica) of the sagittal suture.
***4. Anterior Sagittal.*** Point on the sagittal suture 1/3 distance from bregma to lambda (at juncture of parts bregmatica and pars verticis).
***5. Bregma.*** At bregma (pars bregmatica of the coronal and sagittal sutures).
***6. Midcoronal.*** Midpoint of left coronal suture (within pars complicata).
***7. Pterion.*** At pterion, within the region of the upper portion of the left greater wing of the sphenoid, usually at the point at which the parietosphenoid suture meets the frontal bone.
***8. Sphenofrontal.*** Midpoint of the left sphenofrontal suture.
***9. Inferior Sphenotemporal.*** Point of intersection between the left sphenotemporal suture and a line connecting both articular tubercles of the temporomandibular joint.
***10. Superior Sphenotemporal.*** Point on the left sphenotemporal suture lying 2 cm below juncture with the parietal bone.

Sutures of the hard palate should be scored across their entire length. The left incisive suture should be scored rather than the right, if both segments are observable. Figure 11b illustrates the location of each suture segment.

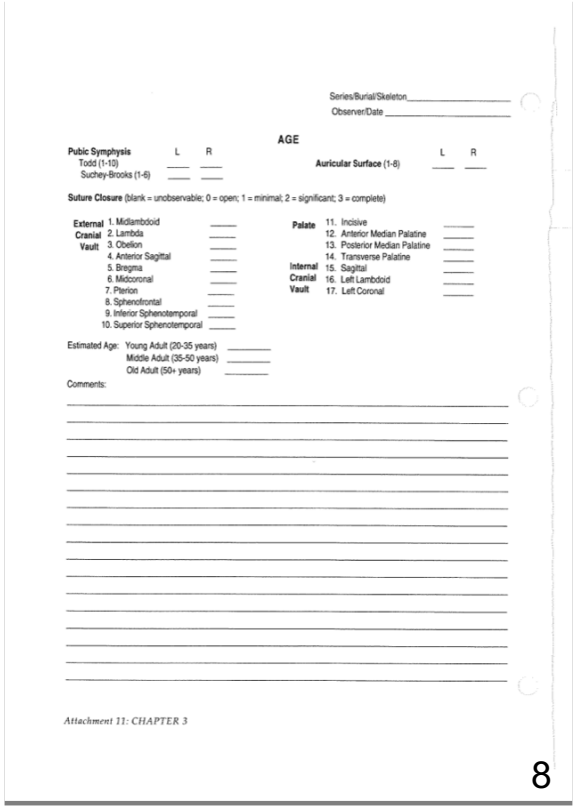
***11. Incisive Suture.*** Separates maxilla (with canines, premolars and molars) from premaxilla (incisors only).
***12. Anterior Median Palatine Suture.*** Score entire length on paired maxillae between incisive foramen and palatine bone.
***13. Posterior Median Palatine Suture.*** Score entire length on paired palatine bones.
***14. Transverse Palatine Suture.*** Score entire length between maxillae and palatine bones.

Endocranial sutures should be scored across the following segments. If the skull is complete, anterior sutures can be observed indirectly by use of a small flashlight and a dental mirror. Locations are indicated in Figure 11c.

***15. Sagittal.*** Full sagittal suture
***16. Left Lambdoid.*** Pars lambdica and pars intermedia of the left lambdoid suture.
***17. Left Coronal.*** Pars bregmatica and pars complicata of the left coronal suture.

**Figure 11. Location of sites to be used to record ectocranial (a; 1-11), palatal (b; 12-14), and endocranial (c; 15-17) suture closure. Drawing by P. Walker (after Mann et al. 1987; Meindl and Lovejoy 1985: 59; Todd and Lyon 1924, 1925a-c).**





Series/Burial/Skeleton_____
Observer/Date _____

**AGE**

**Pubic Symphysis** L R
Todd (1-10) ____ ____
Suchey-Brooks (1-6) ____ ____

**Auricular Surface** (1-8) L R ____ ____

**Suture Closure** (blank = unobservable; 0 = open; 1 = minimal; 2 = significant; 3 = complete)

**External Cranial Vault**
1. Midlambdoid ____
2. Lambda ____
3. Obelion ____
4. Anterior Sagittal ____
5. Bregma ____
6. Midcoronal ____
7. Pterion ____
8. Sphenofrontal ____
9. Inferior Sphenotemporal ____
10. Superior Sphenotemporal ____

**Palate**
11. Incisive ____
12. Anterior Median Palatine ____
13. Posterior Median Palatine ____
14. Transverse Palatine ____

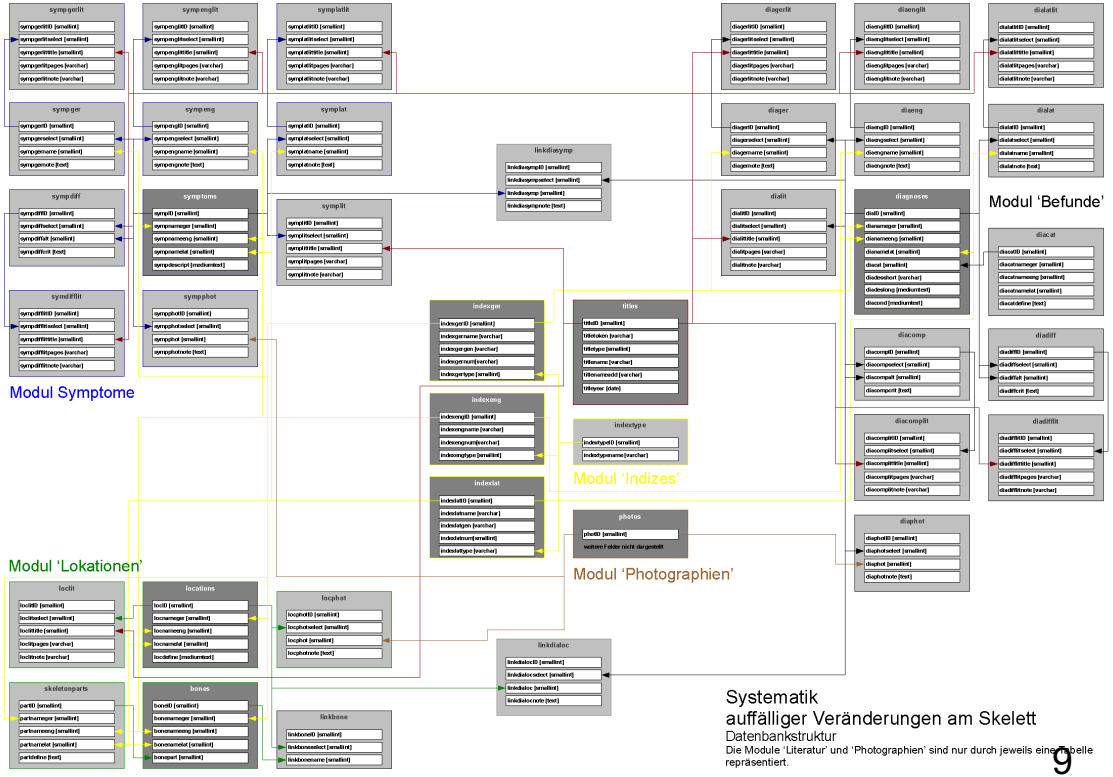
**Internal Cranial Vault**
15. Sagittal ____
16. Left Lambdoid ____
17. Left Coronal ____

Estimated Age: Young Adult (20-35 years) _____
Middle Adult (35-50 years) _____
Old Adult (50+ years) _____

Comments:

# Data Standards

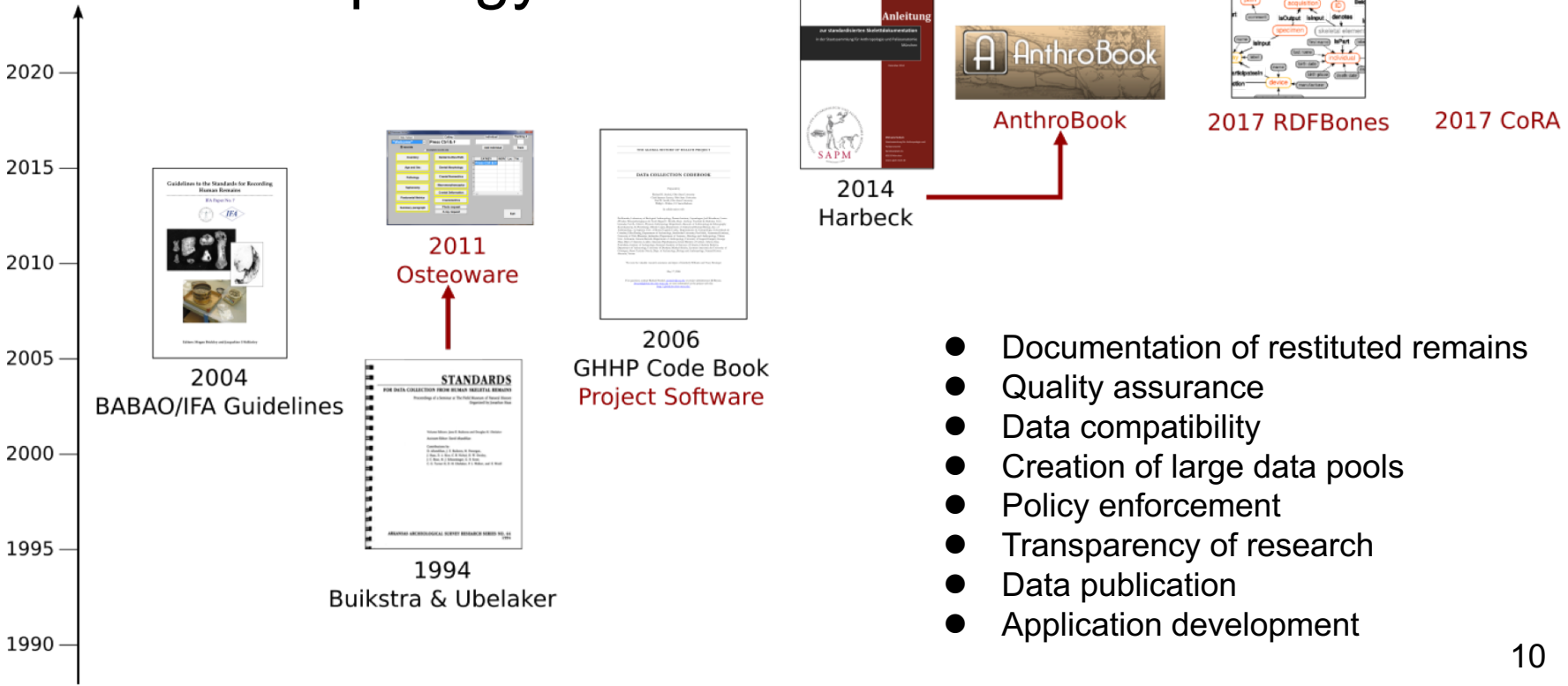
- Describe how data are structured (**data model**)
- Specify how data queries are to be formulated
- Are the basis of information systems




Simplified structure of a database for observations of skeletal features (2008)

Systematik
auffälliger Veränderungen am Skelett
Datenbankstruktur
Die Module 'Literatur' und 'Photographien' sind nur durch jeweils eine Tabelle repräsentiert.

# Standards and Software in Biological Anthropology

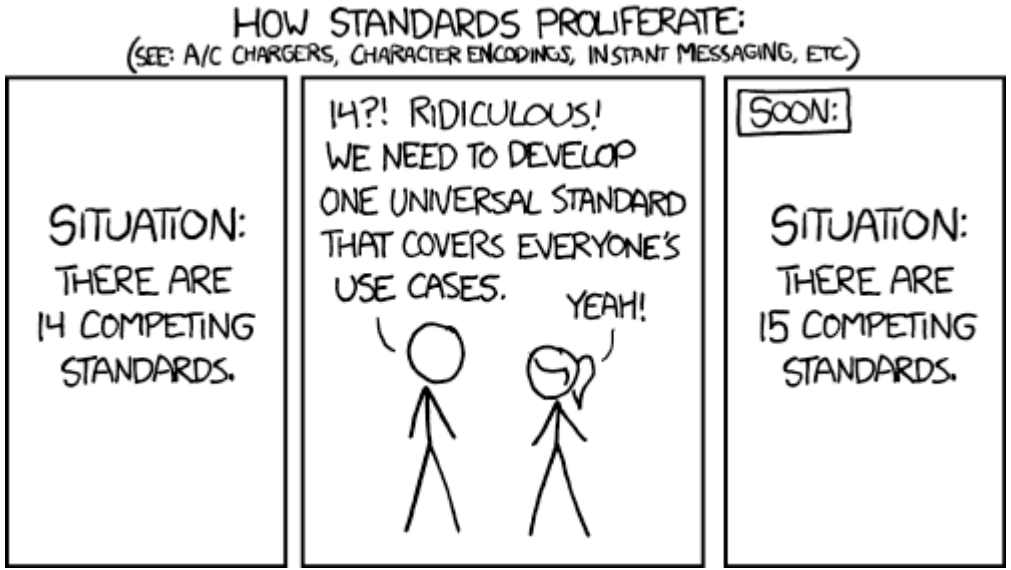



- Documentation of restituted remains
- Quality assurance
- Data compatibility
- Creation of large data pools
- Policy enforcement
- Transparency of research
- Data publication
- Application development

# Challenges

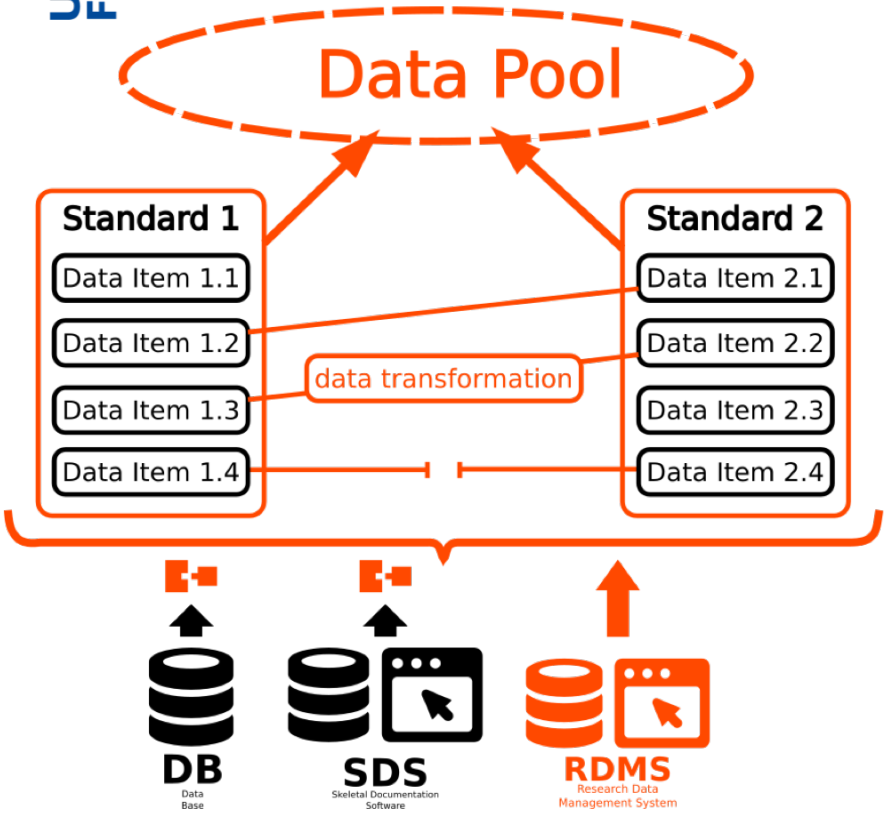
- Heterogeneous research topics
- Heterogeneous materials
- Heterogeneous methodology
- Uncoordinated data production
  - Researchers from different scientific backgrounds
  - Regional research traditions
- Little experience with data management and processing
- Incompatibility of data structures and information systems
- Curation of data standards

# Approach: One Standard to rule them all




Xkcd 927 “Standards” by Randall Munroe
https://xkcd.com/927/

# Approach: Declare compatibilities




Existing data standards
implying different data models
structuring data from different sources.

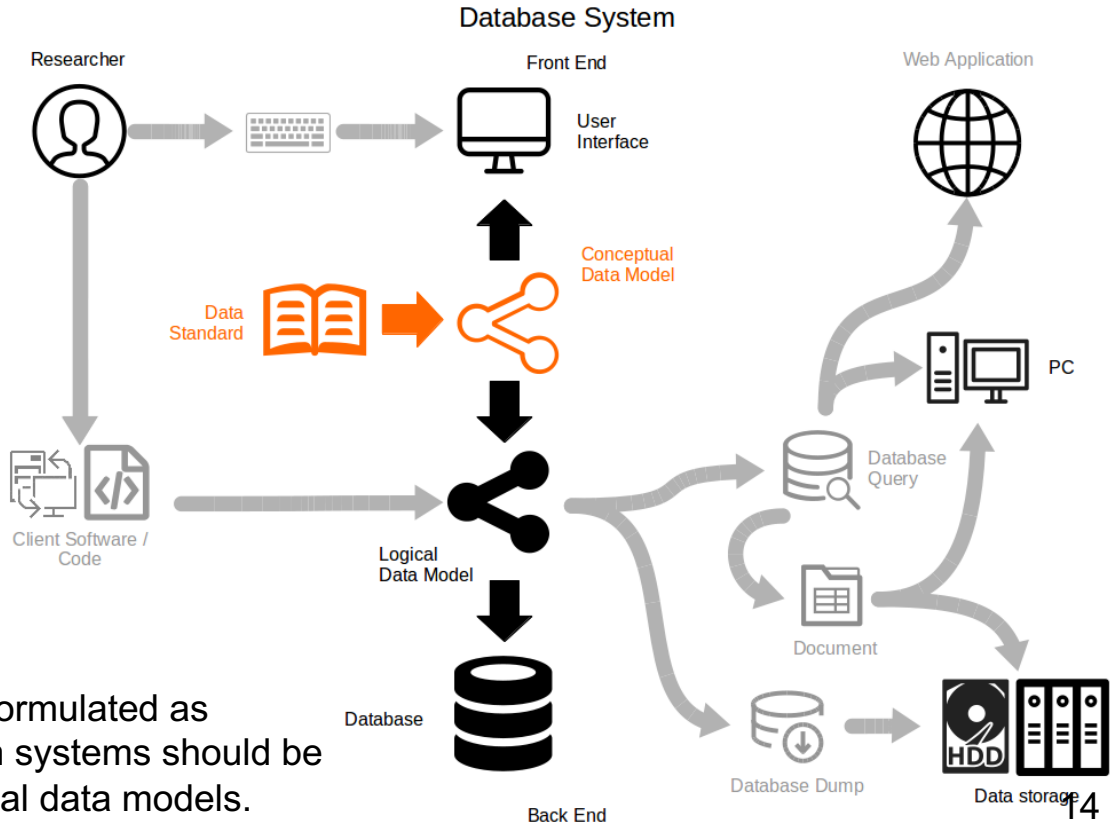
Compatibility layer
declaring how items from different standards can be translated.

# Standards and Information Systems

Scientists should be concerned with defining conceptual data models.

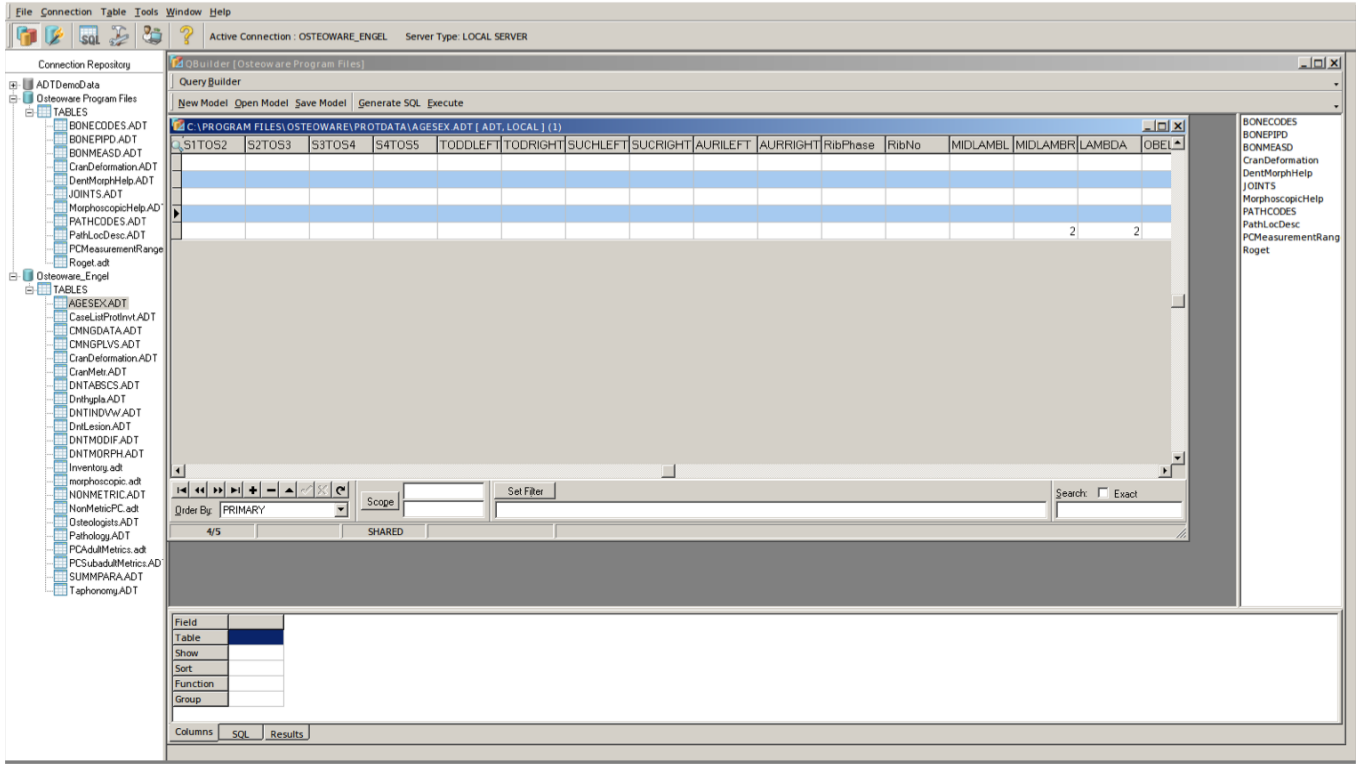
Logical data models define how data from conceptual data models are most efficiently stored in a specific database system.

Conclusion: Data standards should be formulated as conceptual data models and information systems should be able to take them as basis for their logical data models.



# Reality Check

SQL queries are directed to the logical, not the conceptual data model. They are specific to one particular database system.




Osteoware query through the Sybase Data Architect

# Requirements for Data Standards

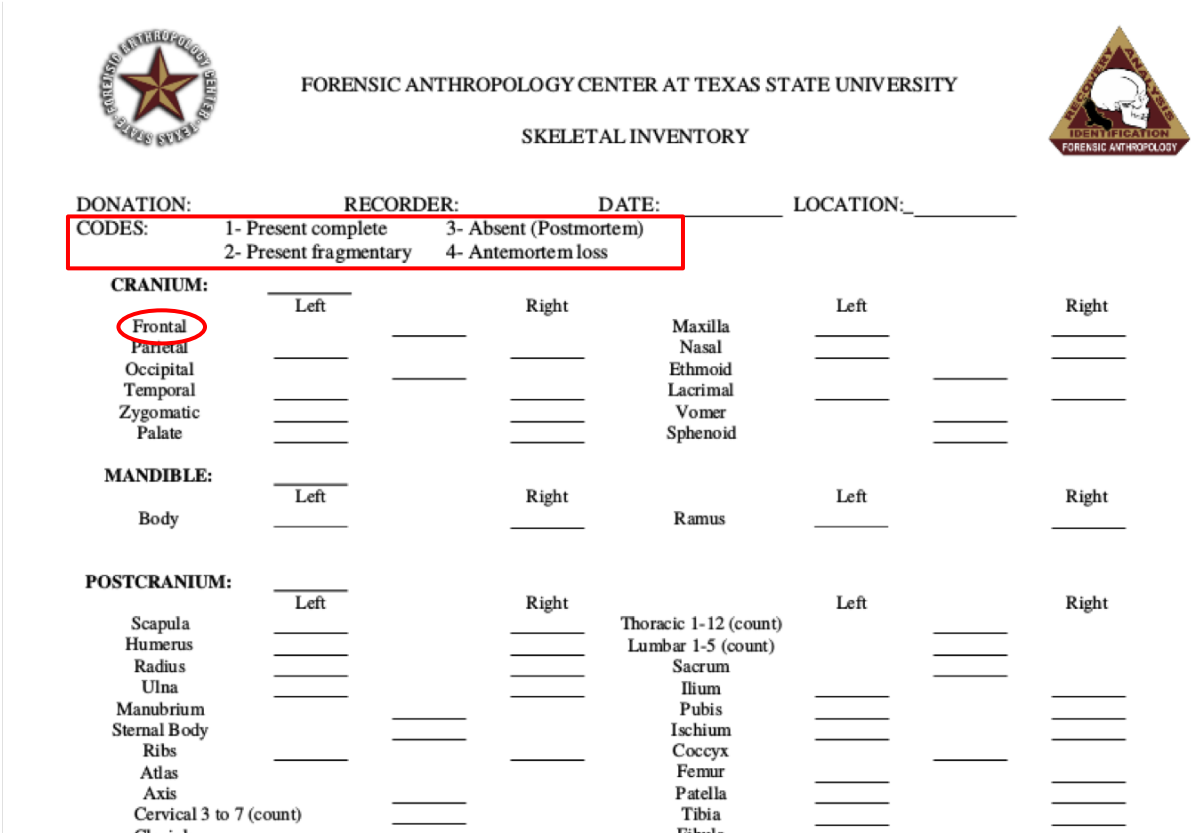
- Data standards
  - are formulated as conceptual data models
  - can be customized
  - serve as data models in database applications
  - are independent of specific database applications
- Data queries
  - are directed to the conceptual data model and can be used on many systems
- Data output
  - has a meaningful structure that can be understood without much documentation
  - can be stored in a robust, durable format

# Resource Description Framework (RDF)

## A way to describe conceptual data models

# Example: FACTS Skeletal Inventory

What is the completeness of the frontal bone?



FORENSIC ANTHROPOLOGY CENTER AT TEXAS STATE UNIVERSITY

SKELETAL INVENTORY

DONATION: RECORDER: DATE: LOCATION:

CODES: 1- Present complete 3- Absent (Postmortem)
2- Present fragmentary 4- Antemortem loss

**CRANIUM:**

| | Left | | Right | | Left | | Right |
|---|---|---|---|---|---|---|---|
| Frontal | | | | Maxilla | | | |
| Parietal | | | | Nasal | | | |
| Occipital | | | | Ethmoid | | | |
| Temporal | | | | Lacrimal | | | |
| Zygomatic | | | | Vomer | | | |
| Palate | | | | Sphenoid | | | |

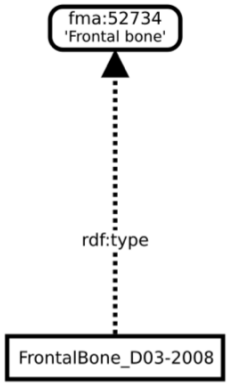
**MANDIBLE:**

| | Left | | Right | | Left | | Right |
|---|---|---|---|---|---|---|---|
| Body | | | | Ramus | | | |

**POSTCRANIUM:**

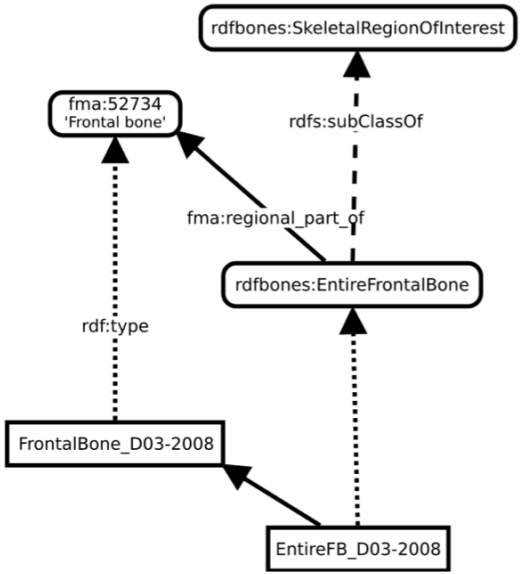
| | Left | | Right | | Left | | Right |
|---|---|---|---|---|---|---|---|
| Scapula | | | | Thoracic 1-12 (count) | | | |
| Humerus | | | | Lumbar 1-5 (count) | | | |
| Radius | | | | Sacrum | | | |
| Ulna | | | | Ilium | | | |
| Manubrium | | | | Pubis | | | |
| Sternal Body | | | | Ischium | | | |
| Ribs | | | | Coccyx | | | |
| Atlas | | | | Femur | | | |
| Axis | | | | Patella | | | |
| Cervical 3 to 7 (count) | | | | Tibia | | | |

# Triples (domain, property, range)




| subject | predicate | object |
|---|---|---|
| FrontalBone_D03-2008 | is a | frontal bone |

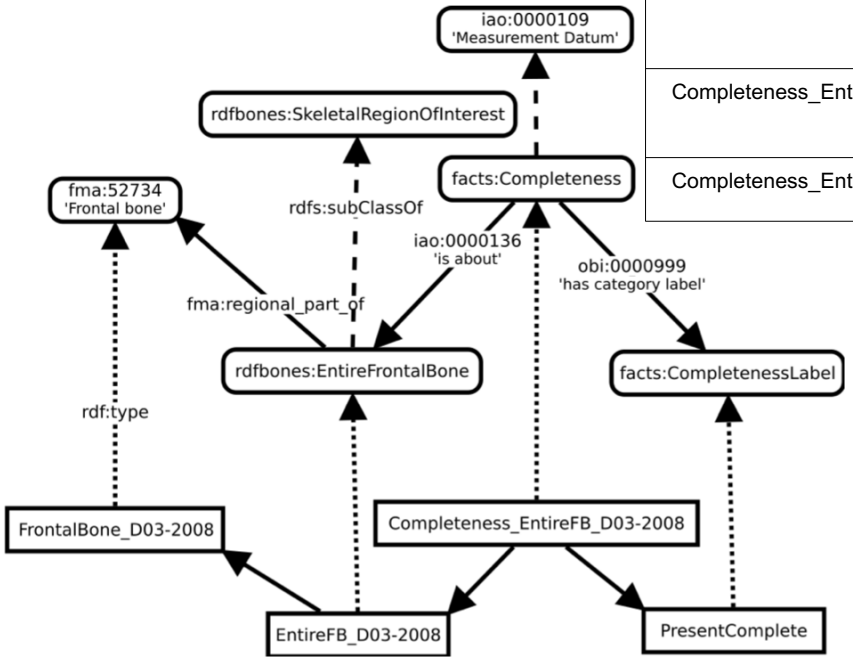
# Triples




| | | |
|---|---|---|
| EntireFB_D03-2008 | is an | entire frontal bone |
| ‘entire frontal bone’ | is a | skeletal region of interest |
| ‘entire frontal bone’ | is a regional part of | the frontal bone |
| EntireFB_D03-2008 | Is a regional part of | FrontalBone_D03-2008 |

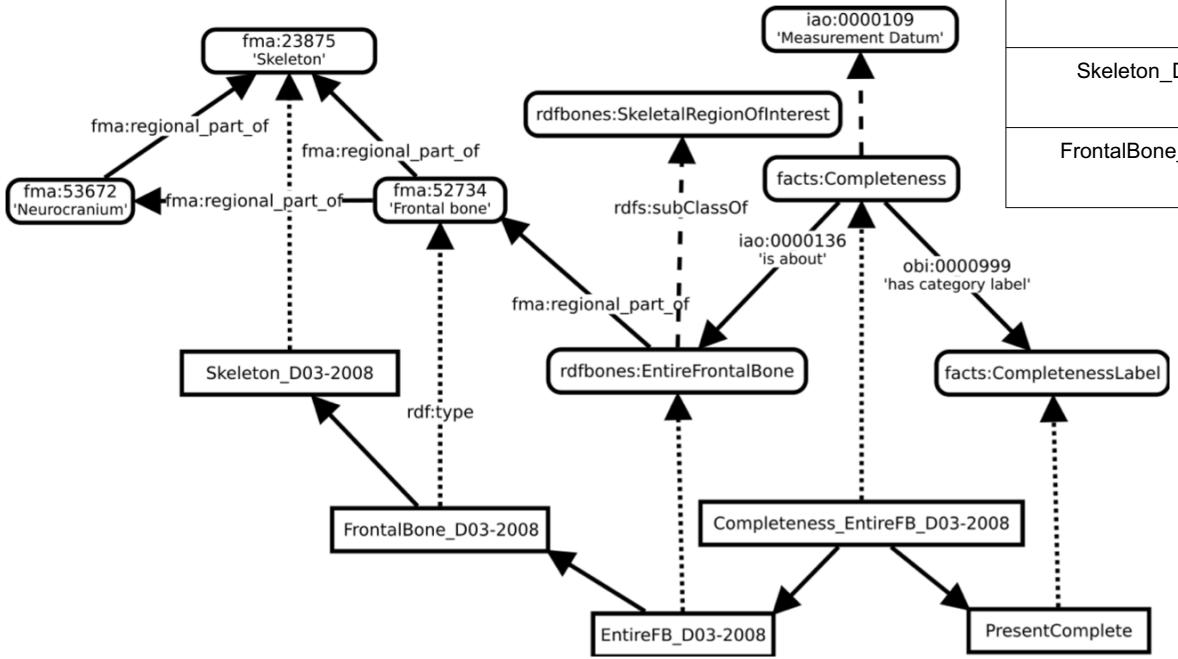
# Triples

| completeness | is a | measurement datum |
|---|---|---|
| completeness measurement data | have a category label | of type ‘completeness label’ |
| completeness measurement data | can be about | entire frontal bones |
| Completeness_EntireFB_D03-2008 | is a | completeness measurement datum |
| Completeness_EntireFB_D03-2008 | has the category label | ‘present complete’ |
| Completeness_EntireFB_D03-2008 | Is about | EntireFB_D03-2008 |

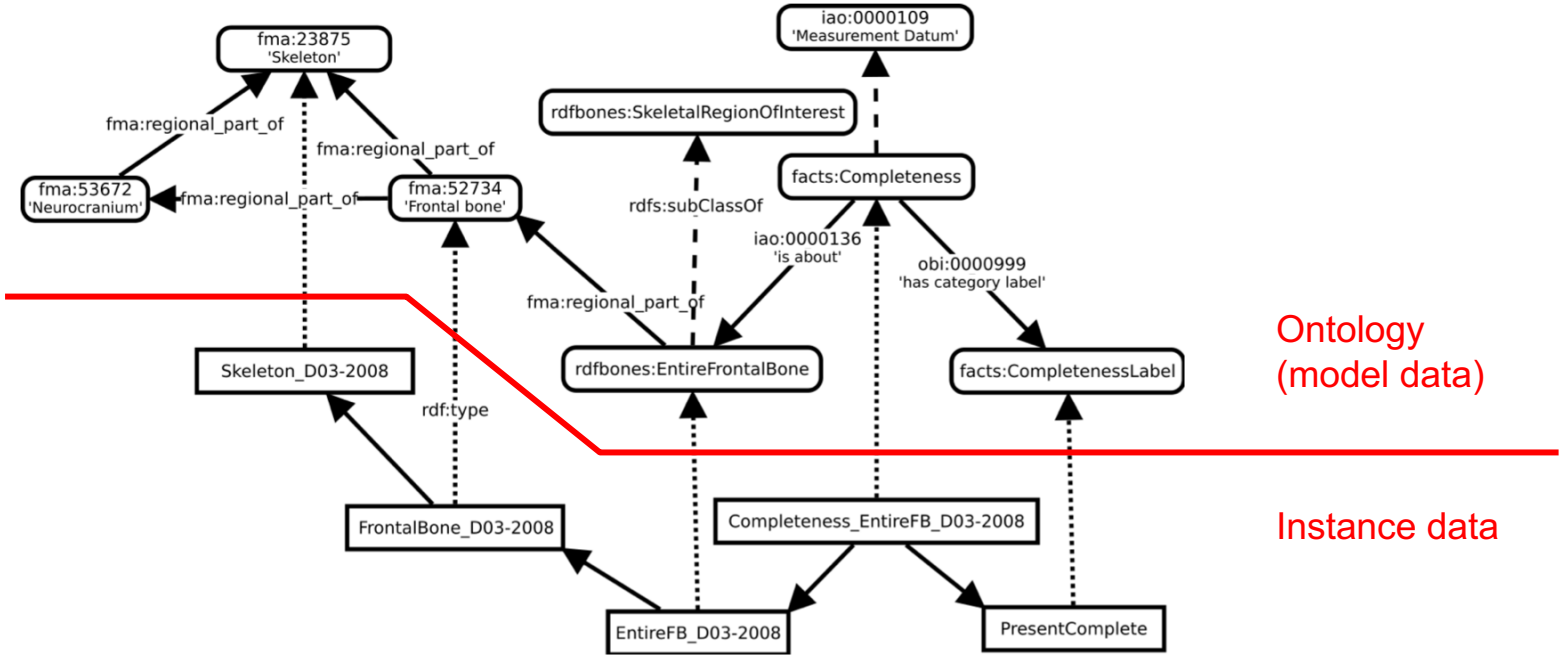


# Triples

| | | |
|---|---|---|
| the frontal bone | is a regional part of | the skeleton |
| the neurocranium | is a regional part of | the skeleton |
| the frontal bone | is a regional part of | the neurocranium |
| Skeleton_D03-2008 | is a | skeleton |
| FrontalBone_D03-2008 | is a regional part of | Skeleton_D03-2008 |



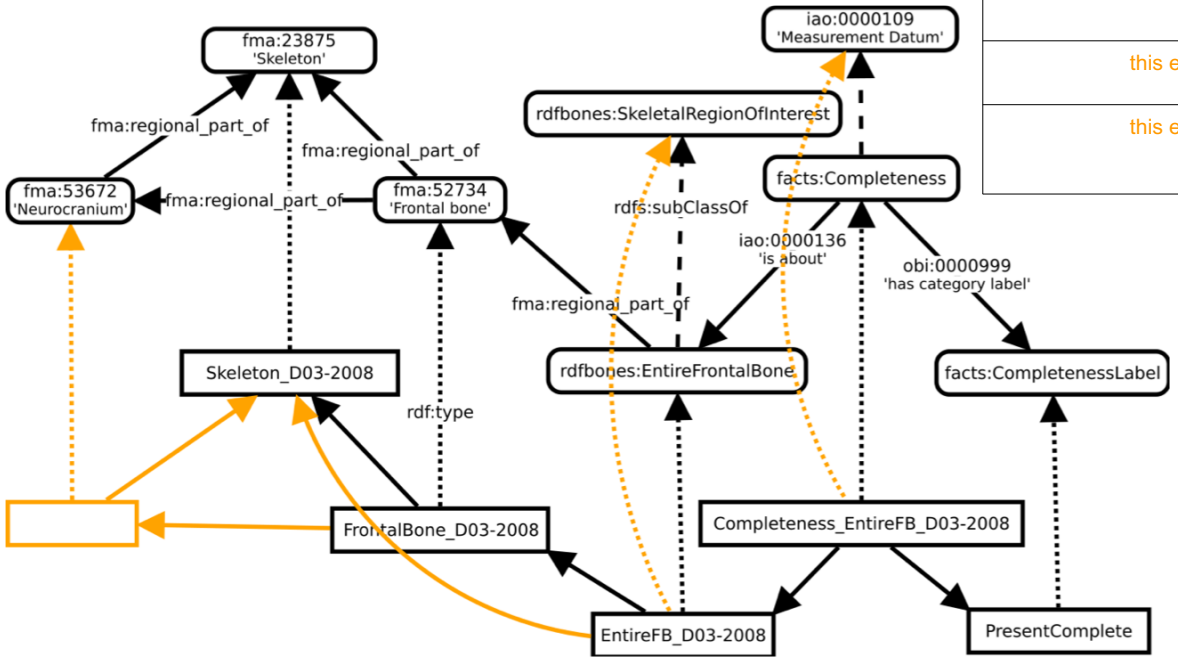
# Ontology Classes vs. Instances



# Inference

classes + instances = asserted model

classes + instances + inferences = inferred model




| | | |
|---|---|---|
| EntireFB_D03-2008 | is a | skeletal region of interest |
| Completeness_EntireFB_D03-2008 | is a | measurement datum |
| EntireFB_D03-2008 | is a regional part of | Skeleton_D03-2008 |
| FrontalBone_D03-2008 | is a regional part of | some entity |
| this entity | is a | neurocranium |
| this entity | is a regional part of | Skeleton_D03-2008 |

# Basis of Inference

- Subclass relations, statements of class equivalence
- Properties of properties (e.g. functional, transitive, symmetric, reflexive)
- Properties of classes (e.g. base, range, junction)
- Class restrictions (statement restricting the use of properties, cardinalities)

# Protocol and RDF Query Language (SPARQL)

```
PREFIX obo: <http://purl.obolibrary.org/obo/>
PREFIX rdfbones: <http://w3id.org/rdfbones/core#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX obo-fma: <http://purl.obolibrary.org/obo/fma#>
SELECT ?Subdivision ?Bone ?Element ?Segment ?Completeness
WHERE{
                ?? obo:BFO_0000051 ?MeasurementDatum.
                ?MeasurementDatum rdf:type rdfbones:CompletenessDatum .
                ?MeasurementDatum obo:IAO_0000136 ?SkeletalSegment .
                ?SkeletalSegment rdfs:label ?Segment .
                ?MeasurementDatum obo:OBI_0000999 ?CompletenessValue .
                ?CompletenessValue rdfs:label ?Completeness .
                ?SkeletalSegment obo-fma:regional_part_of ?SkeletalElement .
                ?SkeletalElement rdfs:label ?Element .
                ?SkeletalElement obo-fma:constitutional_part_of ?BoneOrgan .
                ?BoneOrgan rdfs:label ?Bone .
                OPTIONAL{
                                ?BoneOrgan obo-fma:systemic_part_of
?SkeletalSubdivision .
                                ?SkeletalSubdivision rdfs:label ?Subdivision
                }
} ORDER BY ?Subdivision
```

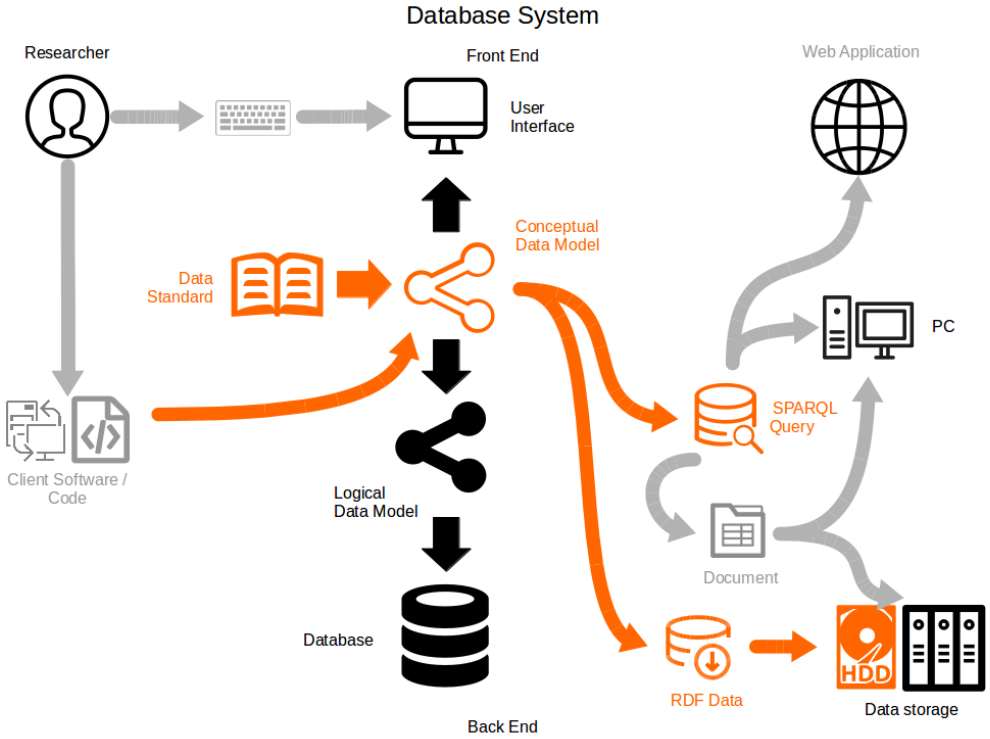
- Adopts triple structure
- Addresses the conceptual data model

# RDF Data Output

```
@prefix facts: <http://www.txstate.edu/anthropology/facts/ontology/core#>
@prefix obo: <http://purl.obolibrary.org/obo/>
@prefix obo-fma: <http://purl.obolibrary.org/obo/fma#>
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
@prefix rdfbones: <http://w3id.org/rdfbones/core#>
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
<facts:Completeness4StatesFrontalboneD2017_000032>
                a <facts:SkeletonElementCompleteness4States> ;
                rdfs:label "Completeness of frontal bone D2017_000032 in four
states" ;
                obo:IAO_0000136 <facts:EntireFrontalBoneD2017_000032> ;
                obo:OBI_0000999 <facts:presentFragmentary> ;
                obo:BFO_0000050 <facts:BSI2017_000032> .
<facts:FrontalBone2017_000032>
                a <obo:FMA_52734> ; # is a frontal bone
                rdfs:label "Frontal bone 2017_000032" .
<facts:EntireFrontalBone2017_000032>
                a <rdfbones:EntireFrontalBone> ;
                rdfs:label "Entire frontal bone 2017_000032" ;
                obo-fma:regional_part_of <facts:FrontalBone2017_000032> .
<facts:Completeness2StatesFrontalBone2017_000032>
                a <rdfbones:EntireFrontalBone> ;
                rdfs:label "Completeness of frontal bone 2017_000032 in two
states" ;
```

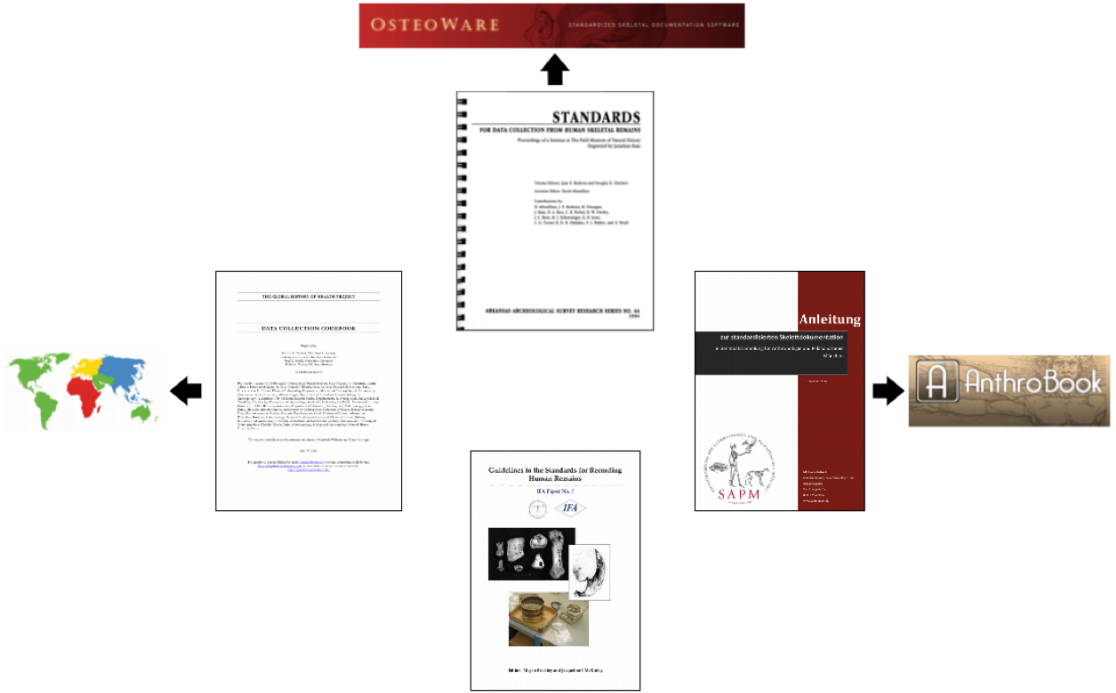
- Simple text files (durable format)
- Comprehensible without prior knowledge

# RDF Database Systems

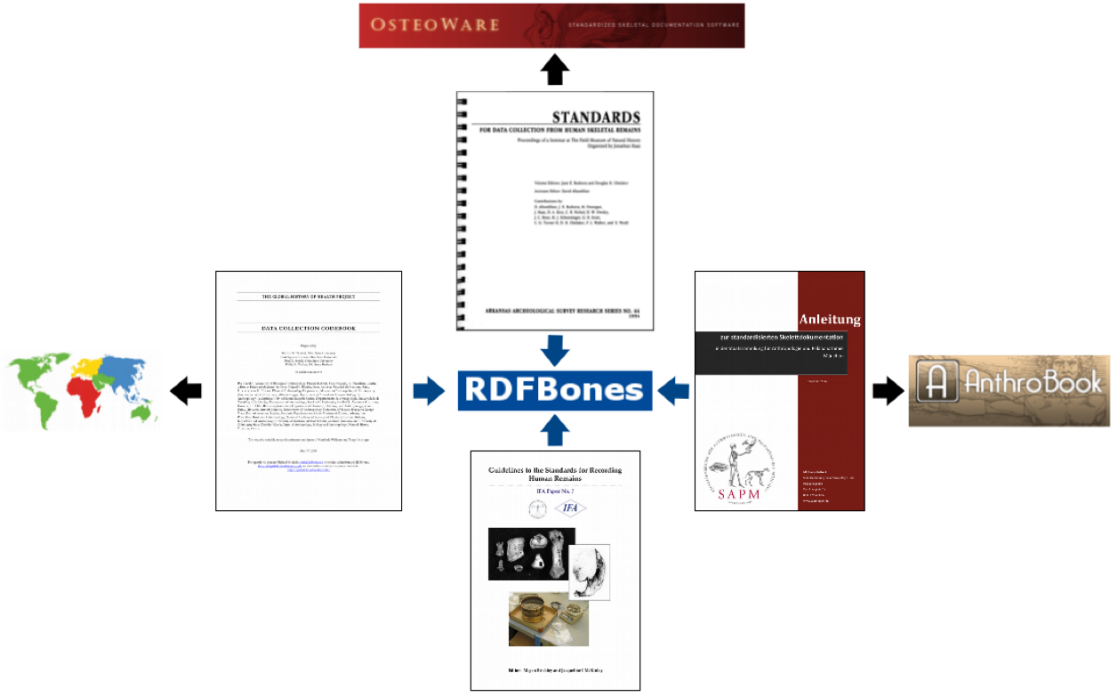


# RDFBones

# Standards and Software Implementations



# Compatibility Through a Common Framework

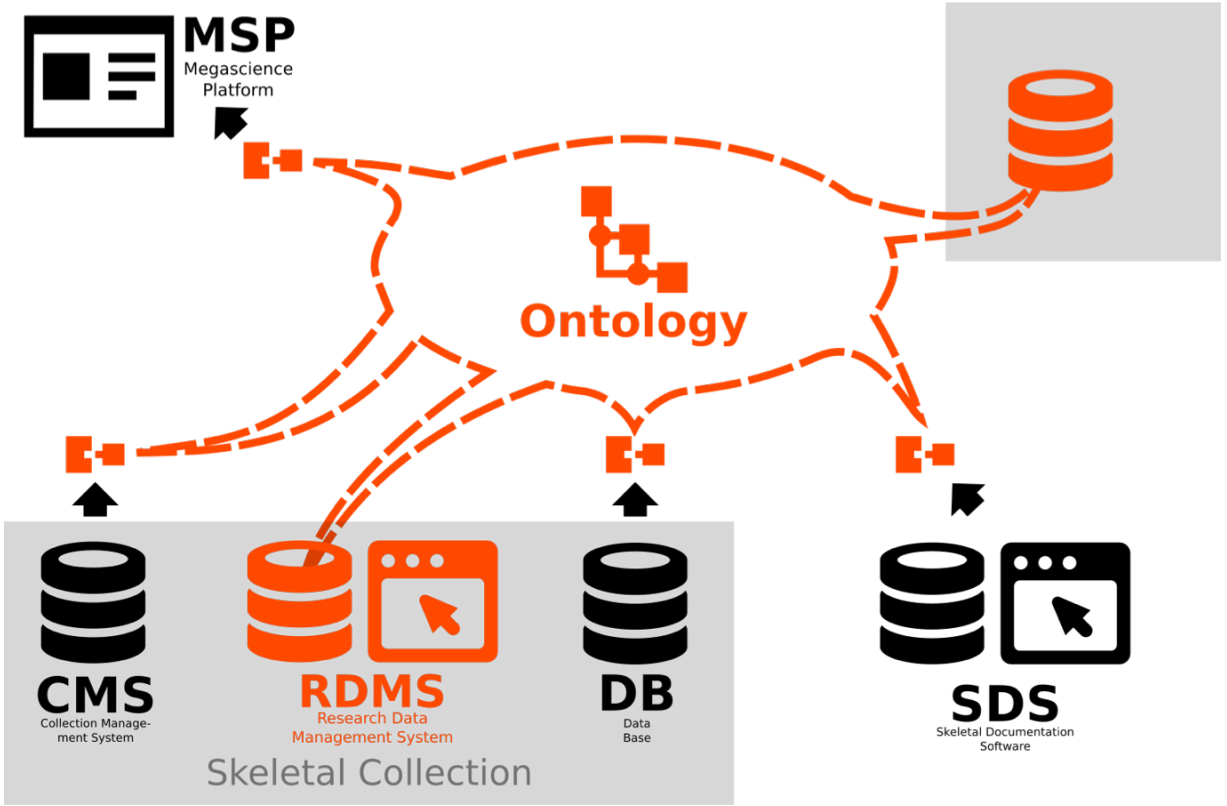


# Basic Concept

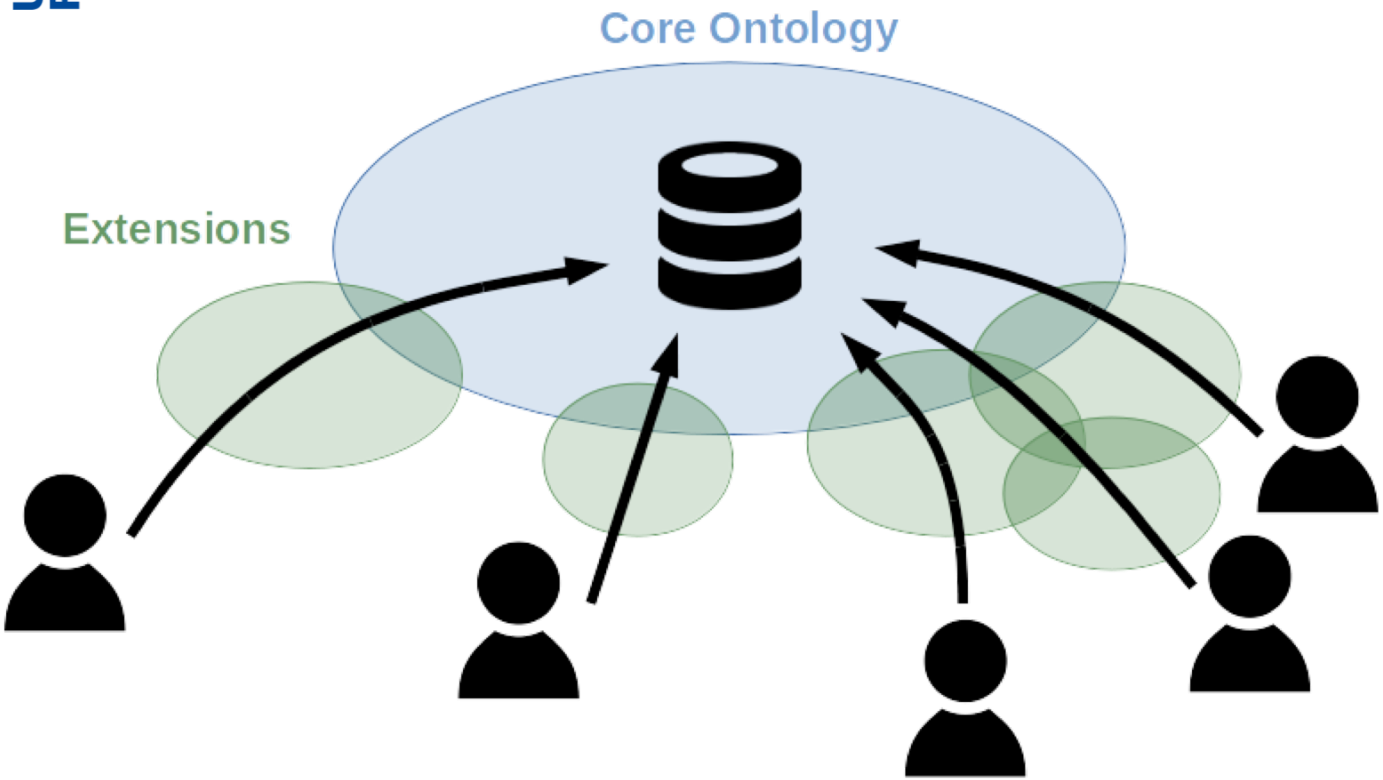
- users can customise existing standards and/or create their own new standards
- Authors of data standards can declare to what degree these are compatible with existing standards
- Data from different standards can be pooled according to compatibility declarations
- Data standards are coded in a format that can serve as a conceptual data model in information systems
- The data standard is independent from any specific software

-> infrastructure for standards development
-> infrastructure for data pools

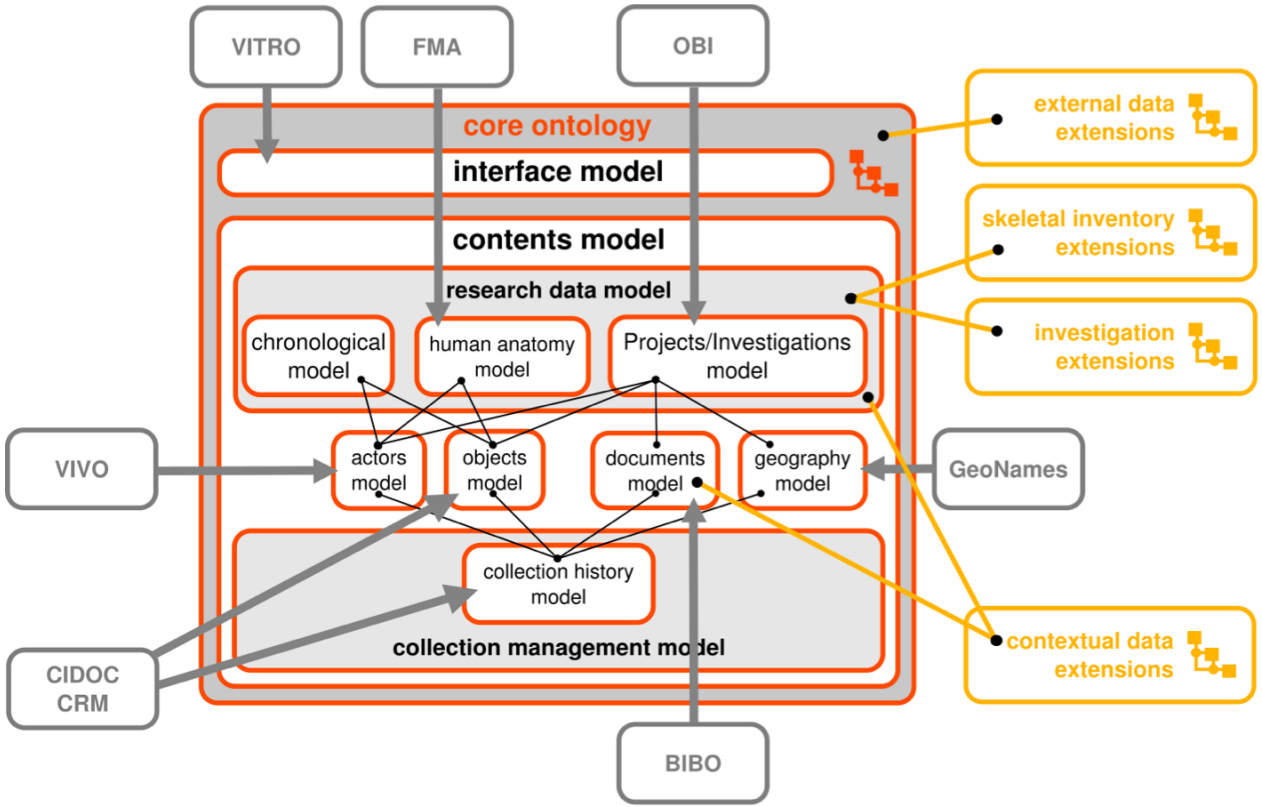
# Distributed system



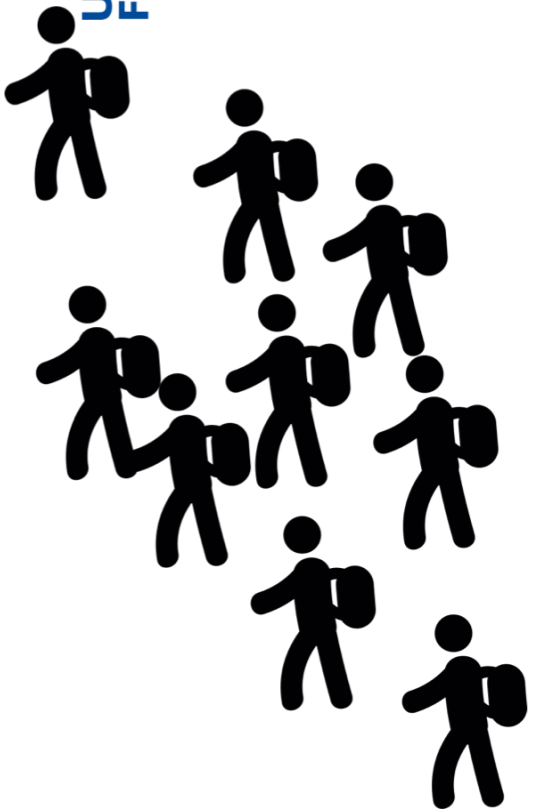
# Core Ontology and Extensions



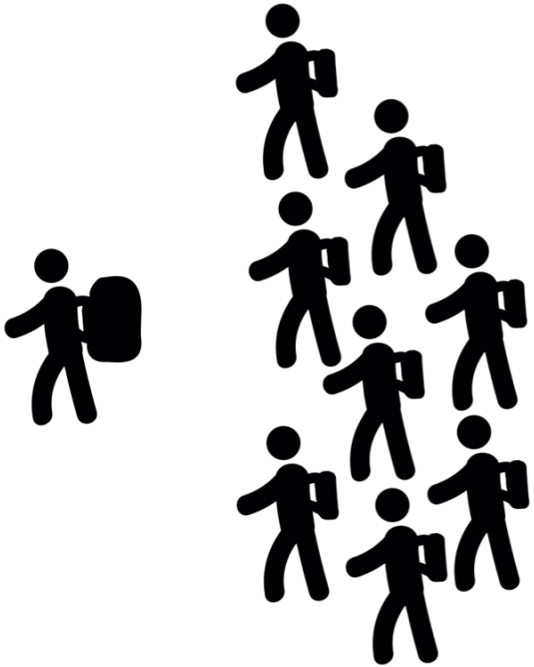
# Core Ontology and Extensions



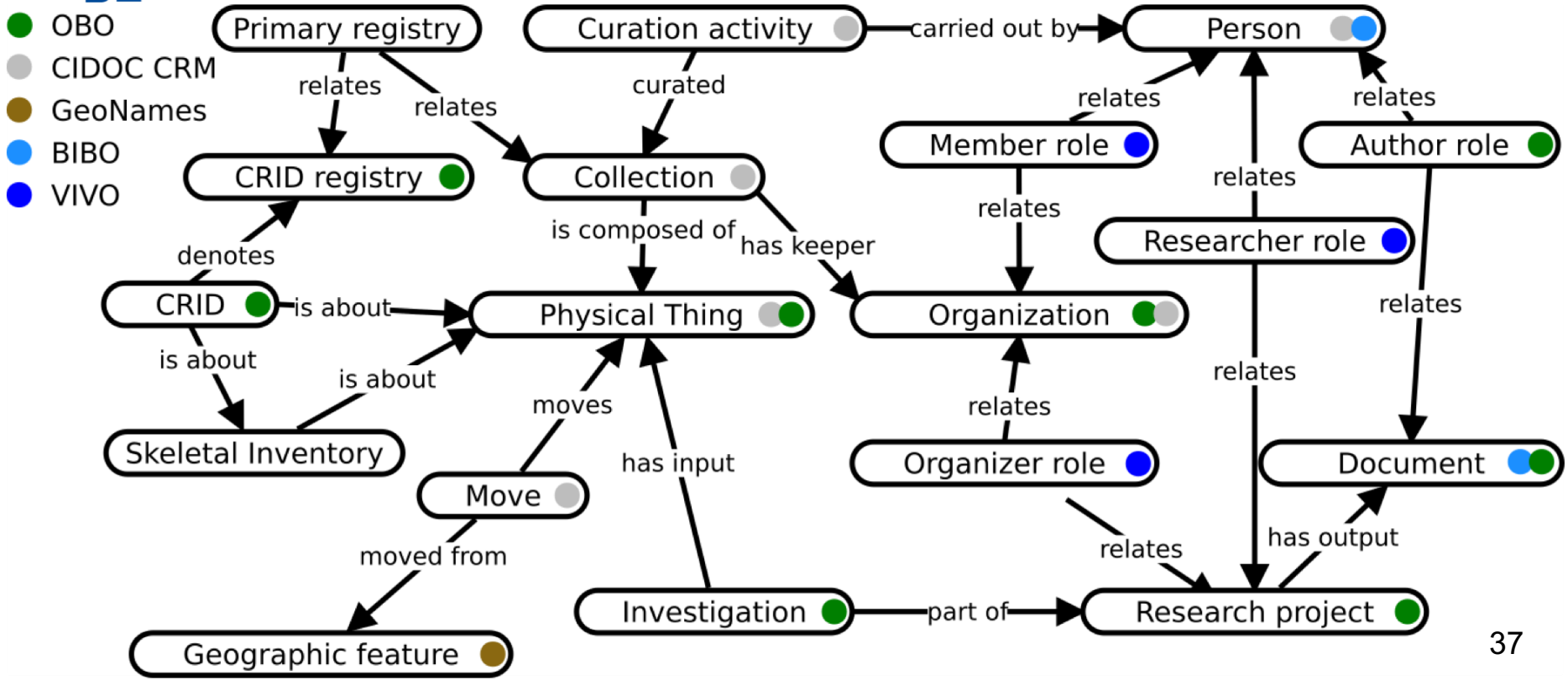
# Workload Distribution



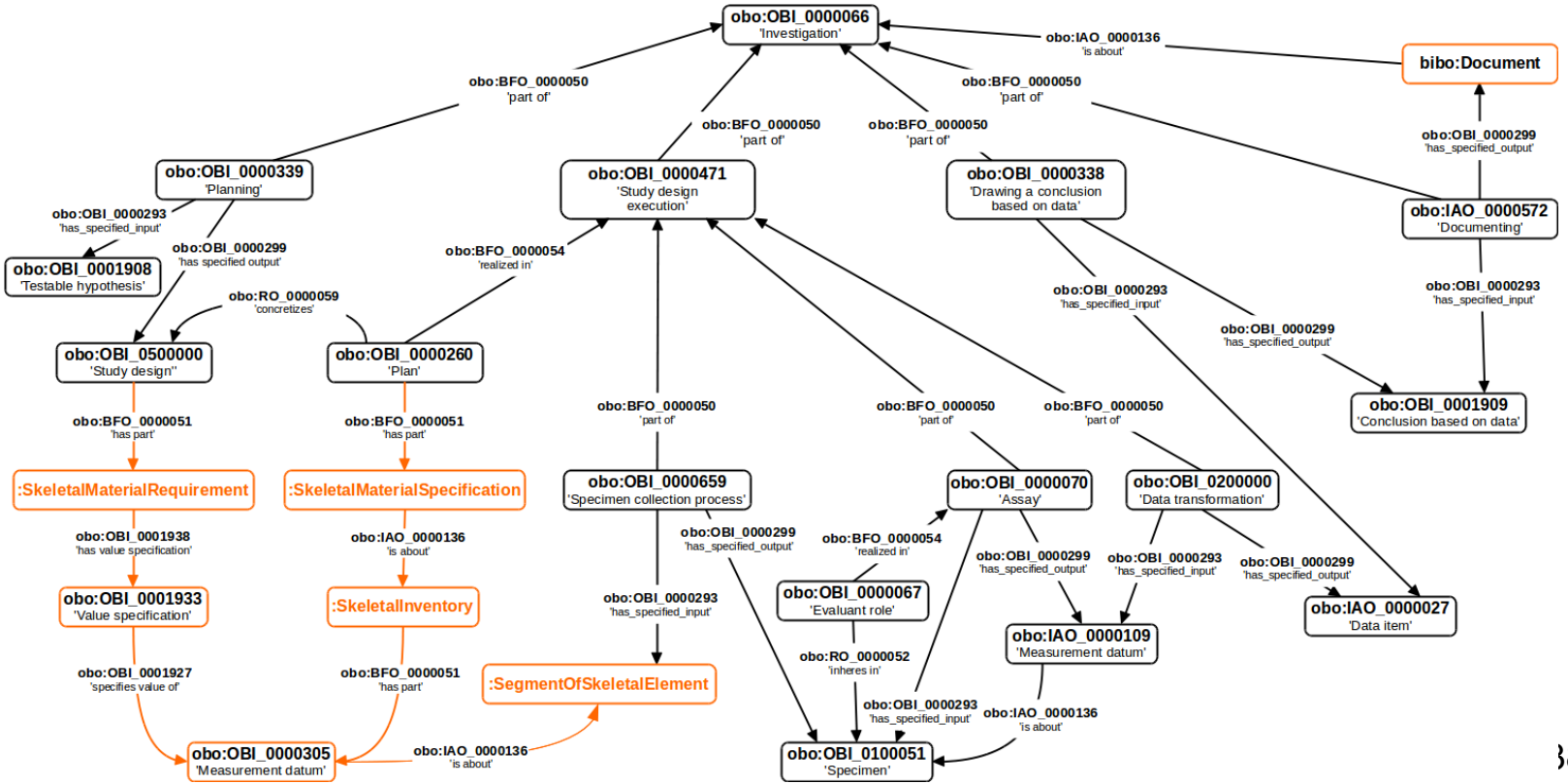
Extension development is extra work but facilitates work for potentially large groups of users.



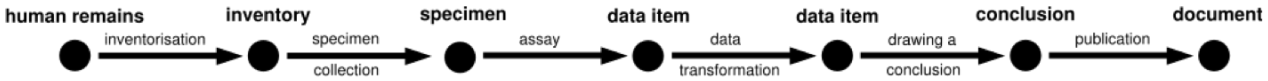
# Structure of the Core Ontology



# Structure of Investigations Model

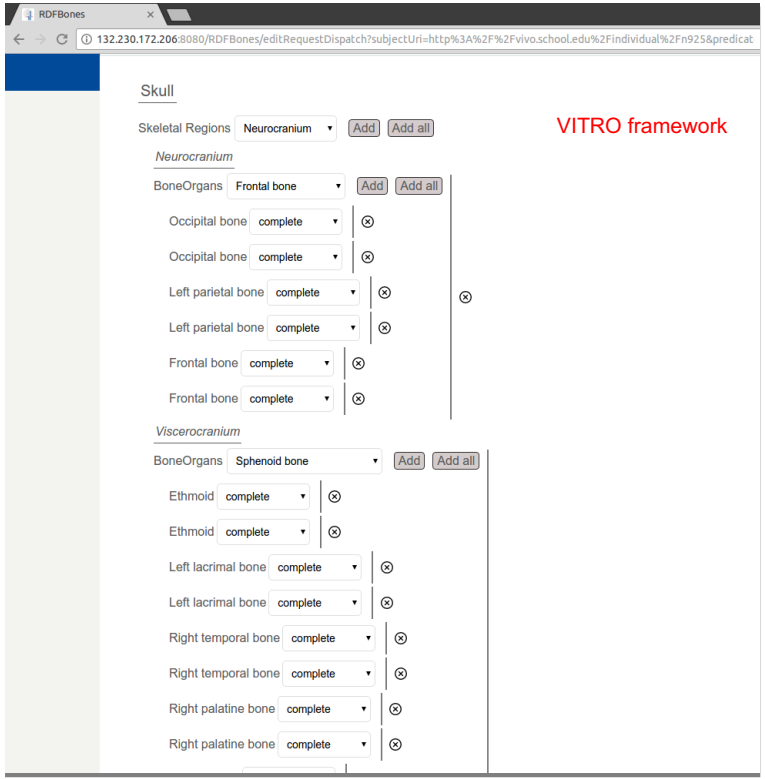


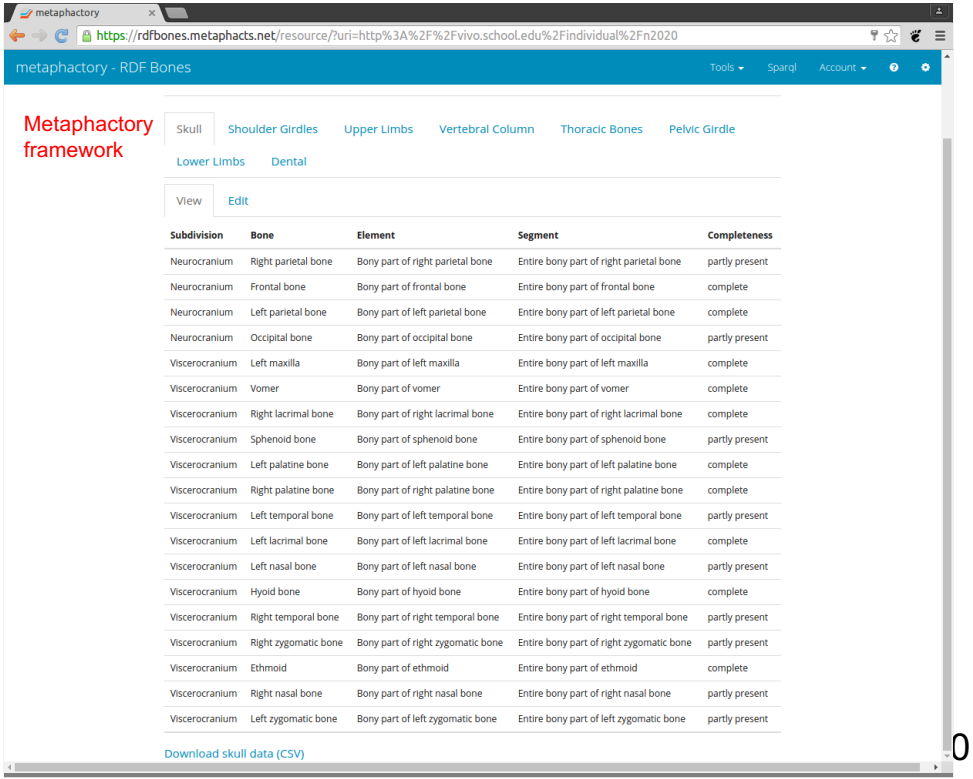
# Data Reuse


human remains
inventorisation
inventory
specimen collection
specimen
assay
data item
data transformation
data item
drawing a conclusion
conclusion
publication
document

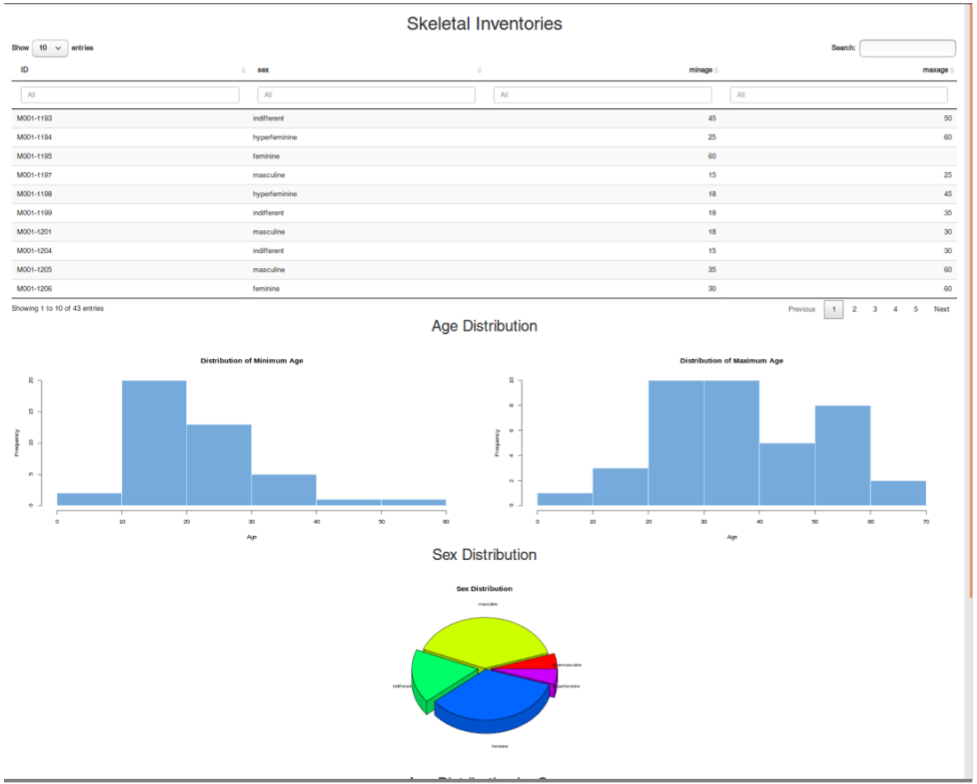

primary inventorisation
investigation 1
investigation 2
investigation 3
primary inventorisation
investigation 4
investigation 5
investigation 6
investigation 7

# Software Implementation



# Web Application with R Package ‘shiny’




http://ckan.anthropologie.uni-freiburg.de/dataset/rdfbones-demo-using-r-and-shiny

# Modelling Research Data from the FACTS

# FACTS Use Case for Donations and OpID

**Documents**

- digital photographs
- spreadsheet files (investigation data; metrics)
- paper forms (burial and taphonomic data)
- PDF forms (inventory data)
- Living Donor Forms [only Donations]

**Systems of Order**

- donors/OpID individuals
- collection
- FARF taphonomic data [only Donations]
- biological samples, laboratories
- Anthropology report [only OpID]

**People**

- grad students
- donors/OpID individuals
- (staff)
- (external researchers)

**Routines**

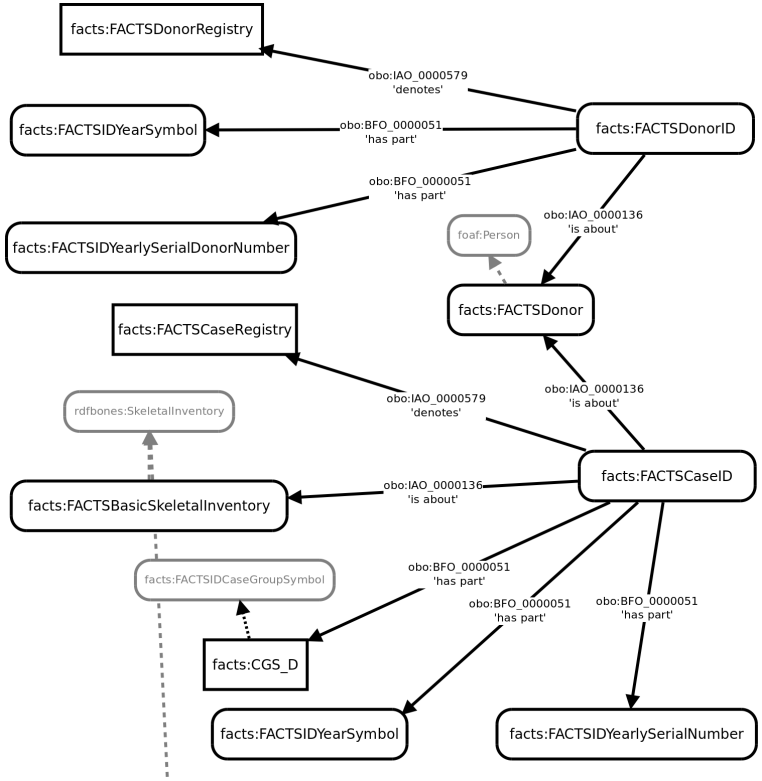
- living donor/OpID case data acquisition
- exhumation/recovery/transportation
- intake assessment
- biological sample acquisition
- FARF (time series data) [only Donations]
- disarticulation, maceration
- skeletal & dental inventorisation
- skeleton storage

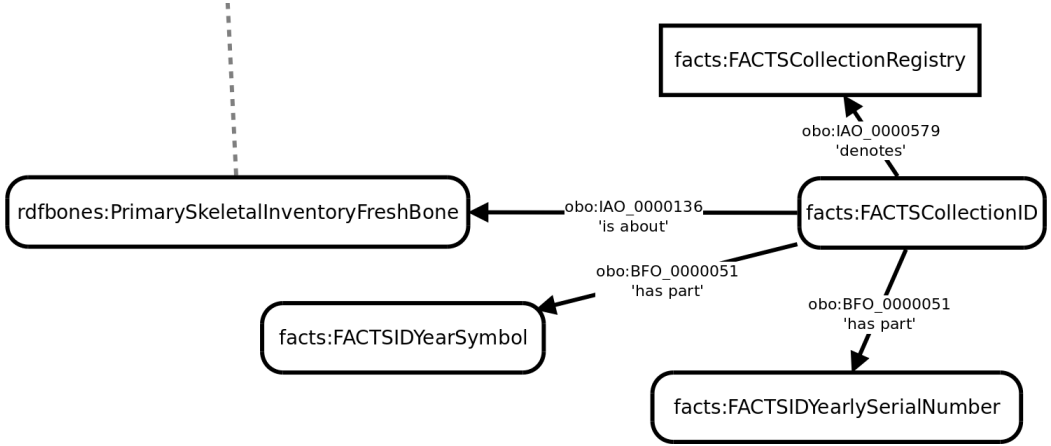


# Considerations for an Information System

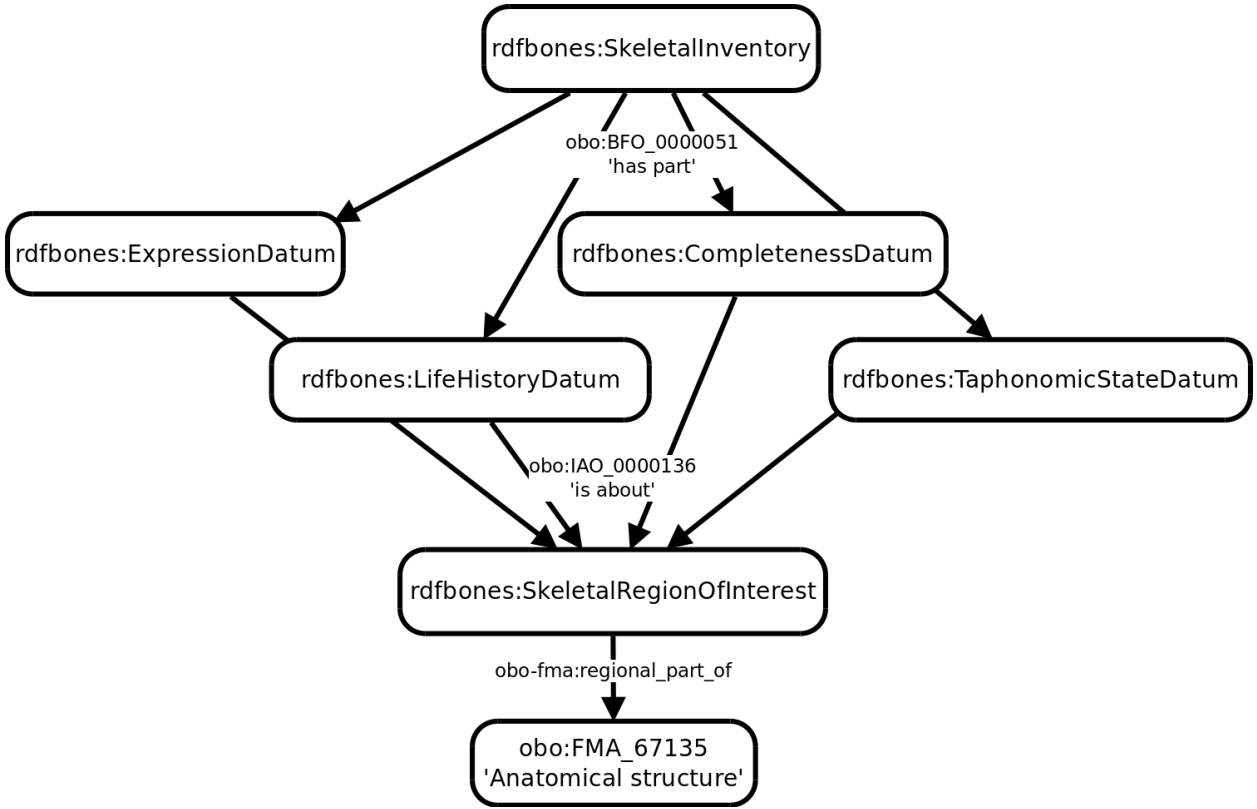
- What databases are to be kept? Should data be mirrored in the system?
- What data are to be imported?
- What documents need to be digitized and when? Is it sufficient to gather metadata first and digitized on demand?
- Is there a proven file management system? Can metadata be mirrored in the system?
- What systems of order are in operation? Do they include each other? Should the index design be simplified?
- What routines need to be modelled? How do they depend on each other?
- Will data acquisition be always digital or might paper forms be more practical?
- How will data be backed up?
- Which data need to be kept active, which can be stored to be reactivated on demand?
- Does the information need to integrate with storage systems already in operation?

# FACTS IDs

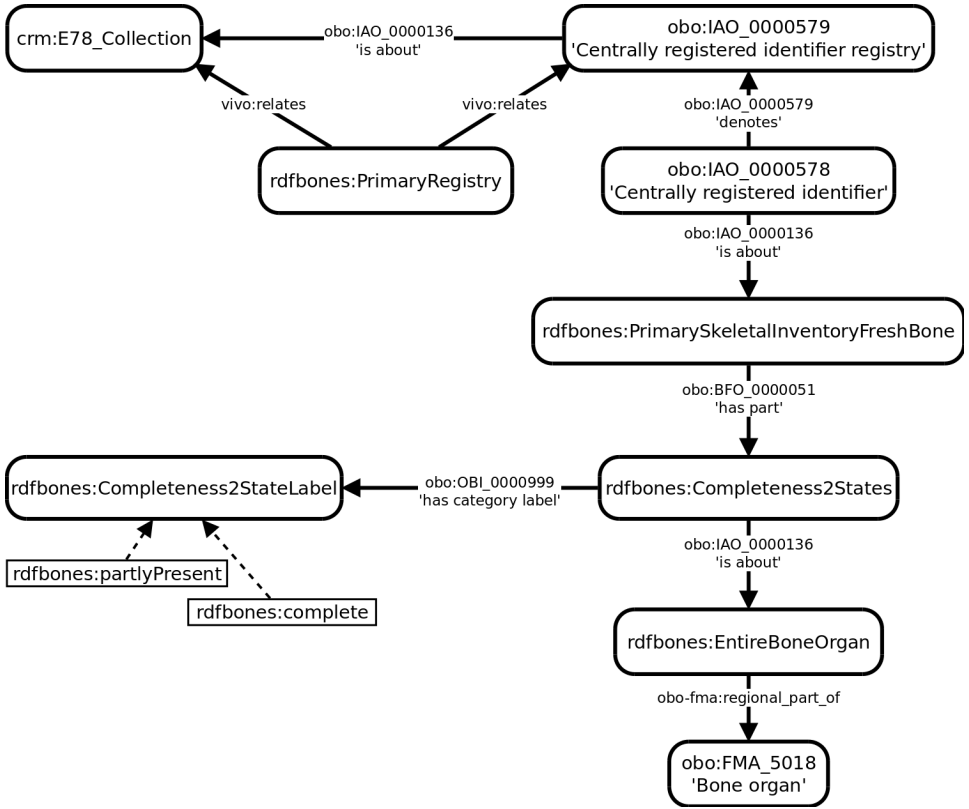

facts:FACTSDonorRegistry
obo:IAO_0000579
'denotes'
facts:FACTSIDYearSymbol
obo:BFO_0000051
'has part'
facts:FACTSDonorID
obo:BFO_0000051
'has part'
facts:FACTSIDYearlySerialDonorNumber
foaf:Person
obo:IAO_0000136
'is about'
facts:FACTSDonor
facts:FACTSCaseRegistry
obo:IAO_0000579
'denotes'
obo:IAO_0000136
'is about'
rdfbones:SkeletalInventory
facts:FACTSCaseID
facts:FACTSBasicSkeletalInventory
obo:IAO_0000136
'is about'
facts:FACTSIDCaseGroupSymbol
obo:BFO_0000051
'has part'
obo:BFO_0000051
'has part'
obo:BFO_0000051
'has part'
facts:CGS_D
facts:FACTSIDYearSymbol
facts:FACTSIDYearlySerialNumber


facts:FACTSCollectionRegistry
obo:IAO_0000579
'denotes'
rdfbones:PrimarySkeletalInventoryFreshBone
obo:IAO_0000136
'is about'
facts:FACTSCollectionID
obo:BFO_0000051
'has part'
obo:BFO_0000051
'has part'
facts:FACTSIDYearSymbol
facts:FACTSIDYearlySerialNumber

# RDFBones Skeletal Inventories

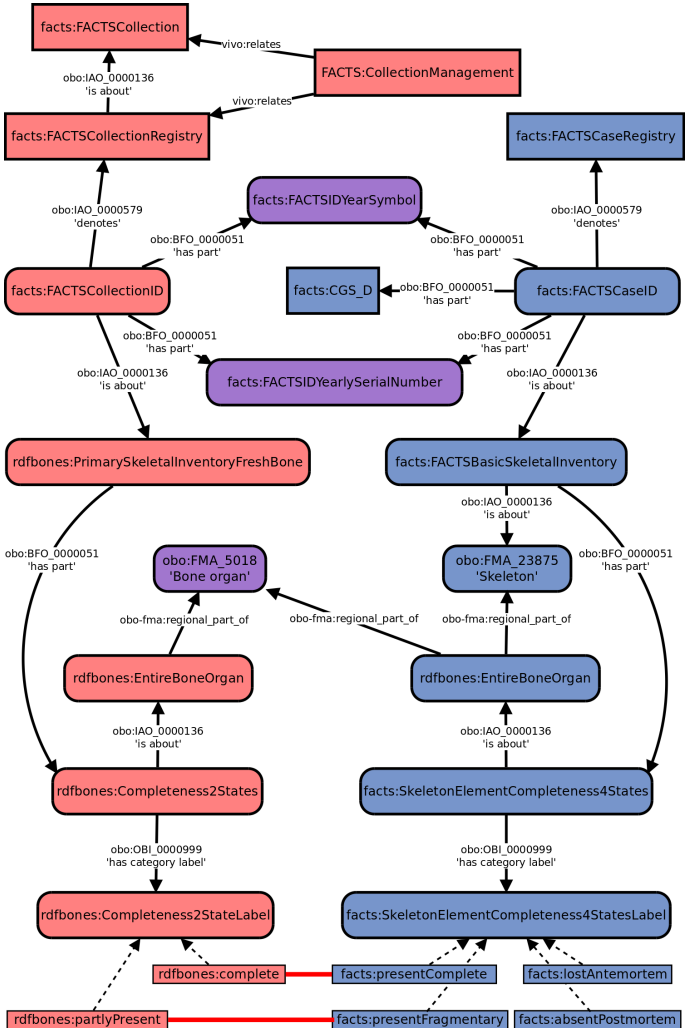


# RDFBones Primary Registries



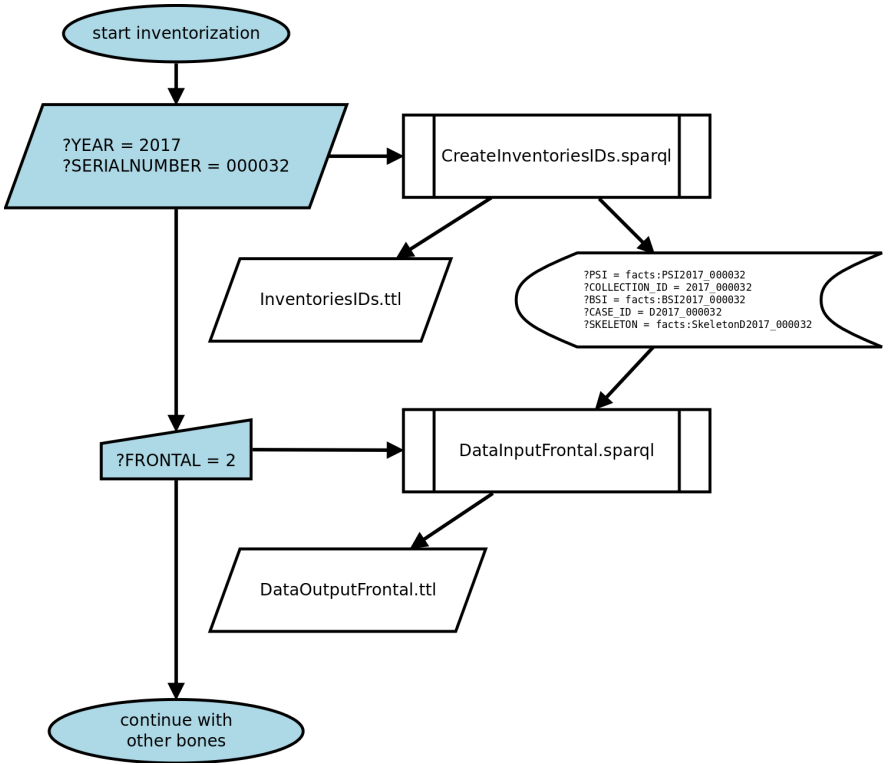
# FACTS Skeletal Inventory

- Combines primary and custom skeletal inventories



# Data Generation During Data Input

- GUI interaction triggers update queries
- Update queries create triple data and set variables



# Conclusions

# Advantages of RDF over Relational Databases

- Data in- and output are based on the conceptual data model
- Data queries are supported by semantic inferences
- Data representation is readable by both man and machine
- Data representation and queries are software-independent
- Data models are extendable
- Different data models can be used for data acquisition and analysis
- Works also on the basis of text files, no database required

# When is the RDF Approach feasible?

- If the data to be modeled are valuable.
- If the data are curated by an institution.
- If rigorous application of standardised methods is important.
- If data reuse and pooling are intended.
- If publication of primary data is intended.
- If long-term data storage is intended.

# Challenges

- Errors might go unnoticed due to the open-world assumption of the RDF.
- Researchers might experience the technology as not ‘user-friendly’ and institutions have to invest in the development of easy-to-use information systems.

# Acknowledgements

Dr. Kate Spradley for collaboration on the Operation Identification project

The development of RDFBones was funded by the German Research Foundation

DFG Deutsche Forschungsgemeinschaft