

# Managing Bio-Models: Model Storage, Retrieval, Ranking and Versioning

Ron Henkel\*, Dagmar Waltemath, Olaf Wolkenhauer

## Storage

### Background:

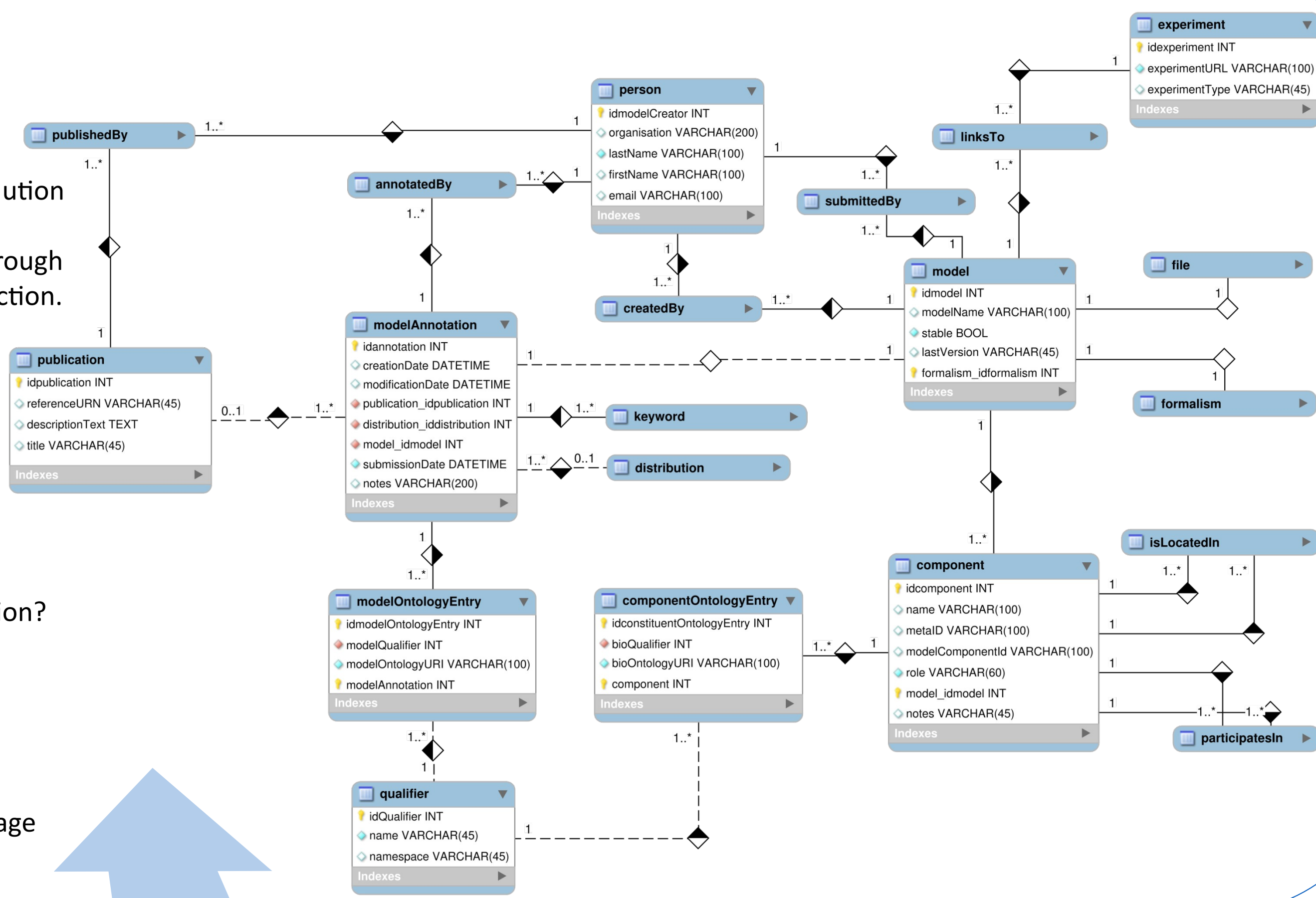
Bio-models are made available through model repositories. With the increasing importance of bio-model annotations a storage solution should not only focus on the XML format. It should also provide a thorough concept for meta-information extraction. Only pieces of information that are extracted and stored separately are later on available for model management tasks such as retrieval, visualization, and versioning [1].

### Key Questions:

What is the relevant meta-information? How can it be stored efficiently and format-independently?

### Solutions:

- annotation based model storage
- relational database for model storage



## Retrieval

### Background:

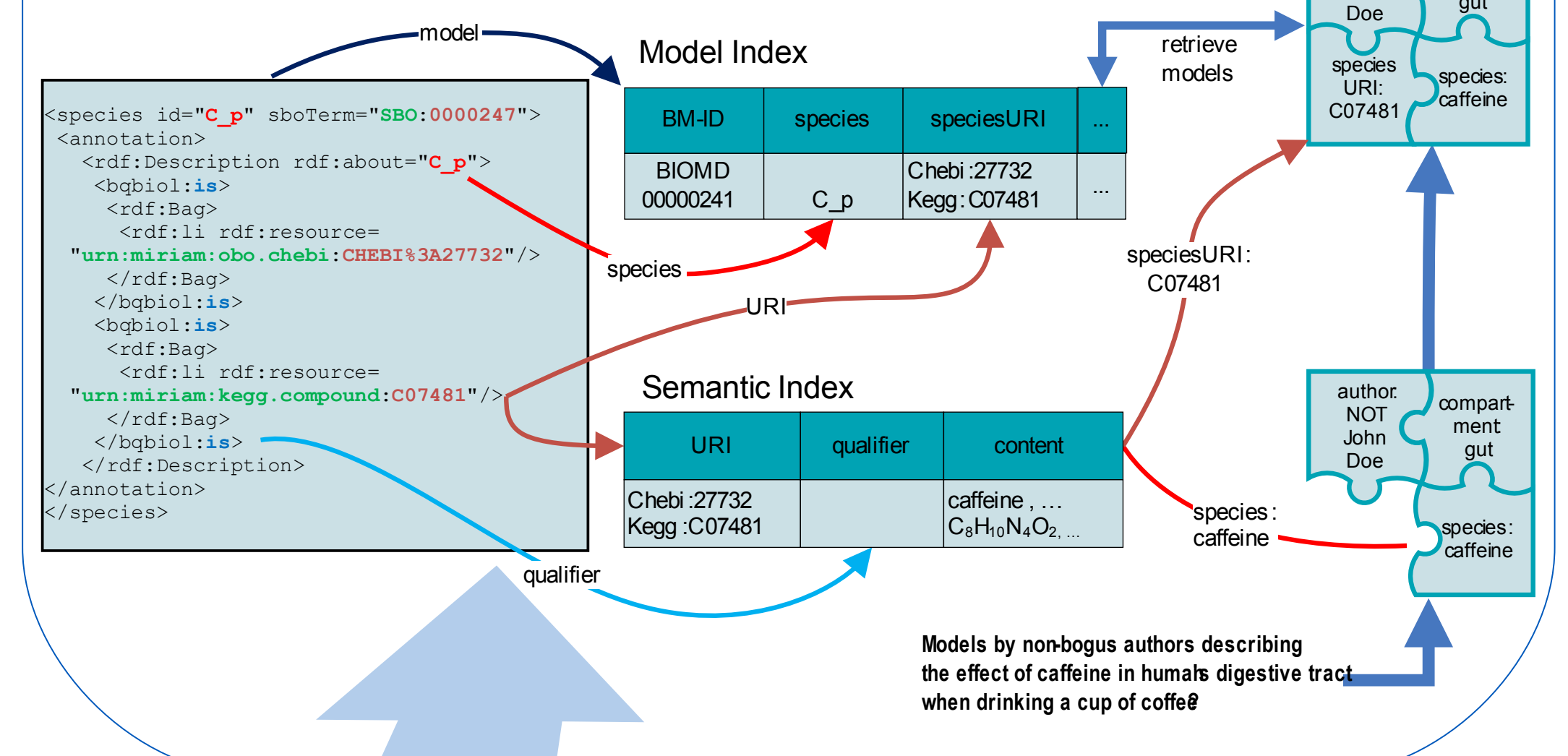
Common SQL search techniques are insufficient for bio-model search. The incorporation of Information Retrieval [2] techniques allows the advanced indexing of a set of bio-models in a repository [3].

### Key Question:

Which is the best applicable IR technique for the given set of bio-model data?

### Solution:

- annotation based query refinement and model retrieval [3]



## Bio-model management

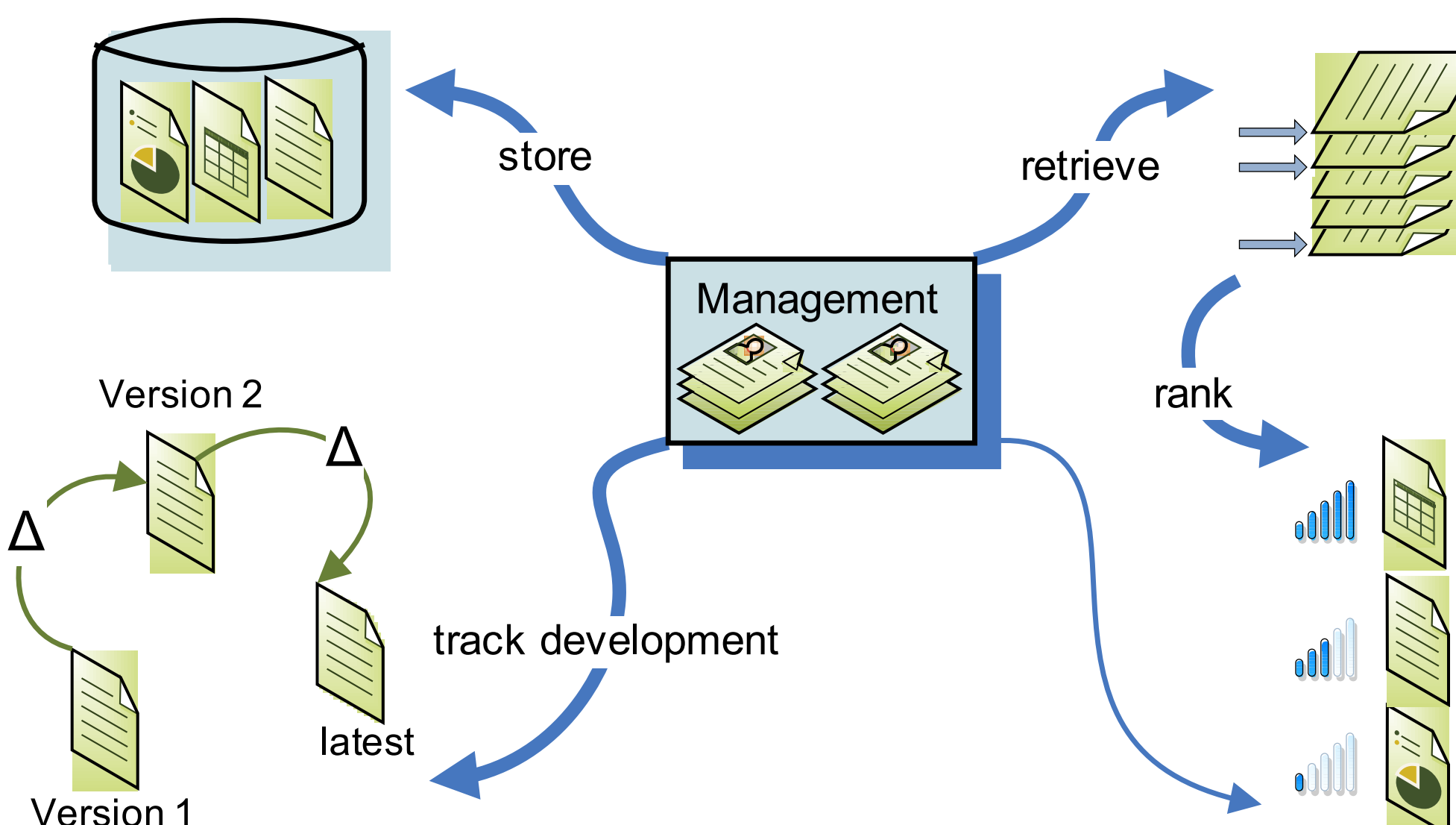
The number of publicly available bio-models is increasing rapidly. Model reuse and simulation result reproduction have therefore become important tools for the systems biology community [4]. Different model repositories provide access to model code and model meta-information. However, the application of further standard database and information systems techniques can enhance the reuse of available bio-models [5].

**Definition (Bio-Model):** A computational model of a biological system, annotated with meta-information and stored in XML format [1].

**Definition (Meta-Information):** Annotations that provide additional (third-party) knowledge about the modeled system and enhance the computer-processed understanding, e.g. model author, model entity descriptions, or modification dates.

We have in our research focused on a format-independent database design that offers a fine-grained structure for meta-information storage. The design allows the integrative storage of bio-models of different encodings (e.g. CellML and SBML models) in one place.

The availability of bio-models in all published versions is a prerequisite for successful reuse. Only if users can refer to a model in a particular version, they are able to link that model from further developments, or in literature. We propose here a thorough versioning approach for bio-models that considers changes in the mathematics, in the biology and in the meta-information.



Furthermore, the application of Information Retrieval techniques on bio-model search has been proposed. A ranked retrieval suggests the best matching bio-model for a given search and thereby supports the users of a repository in finding the relevant information [1].

All approaches are format independent and can therefore be applied to different bio-model encodings. The code is freely available on Sourceforge from <http://bives.sourceforge.net/> and <http://sombi.sourceforge.net/>.

## Versioning

### Background:

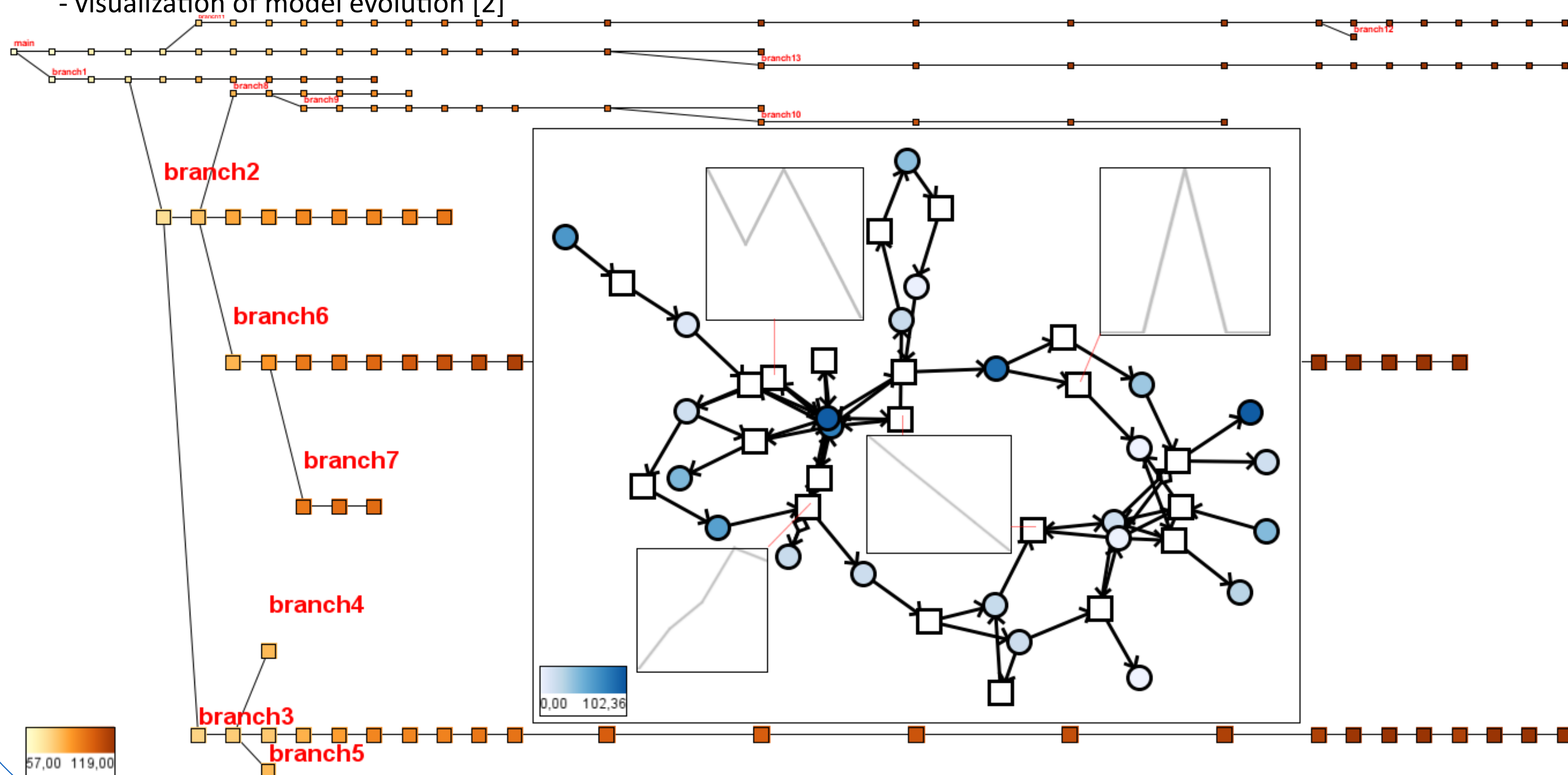
Bio-models evolve over time. Changes in bio-models need to be tracked and made available to the user. For example, sometimes a number of models co-exist in different parameterizations, or model structures are updated as new insights are revealed.

### Key Question:

What are the specific demands of a bio-model versioning system?

### Solutions:

- XML versioning approach coupled with an ontology of change types: BiVeS
- fine-grained bio-model storage with linking between different versions
- visualization of model evolution [2]



## Ranking

### Background:

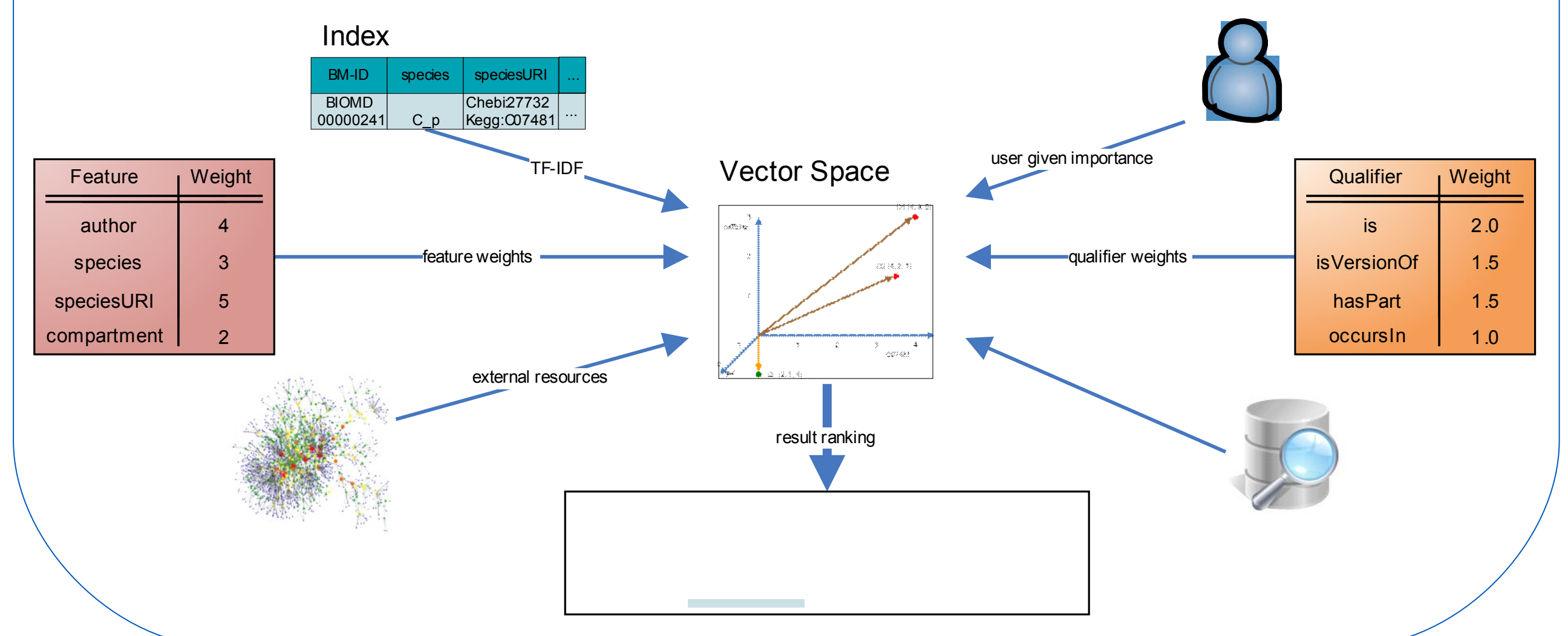
Search results are often only presented to the user in an unsorted set. With high numbers of relevant models for a query, it becomes necessary to sort the results by relevance.

### Key Questions:

How can search results be ranked according to their relevance? What are the determining bio-model features? Which is the best applicable ranking function?

### Solutions:

- ranking with a Vector Space Model
- feature set partially derived from bio-model meta-information [1]
- weight concept to satisfy different user requirements



## References

- [1] Waltemath et al. (2011) Datenbank Spek **11**(1)
- [2] Baeza-Yates et al. (1999) Addison-Wesley
- [3] Henkel et al. (2010) *BMC Bioinf*, **11**:423
- [4] Waltemath et al. (2011) *PLoS Comput Biol* **7**(4)
- [5] Köhn et al. (2009) *DILS*
- [6] Hadlak et al. (2011) *IEEE TVCG*, submitted