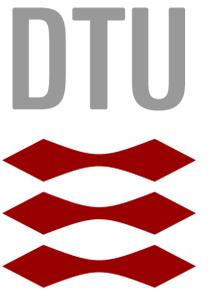


# The Free Model Initiative's SE Model Index



## Harald Störrle

Department of Applied Maths  
and Computer Science  
Danmarks Tekniske Universitet  
hsto@dtu.dk

## Regina Hebig

Sorbonne Universités,  
UPMC LIP6,  
Paris, France  
regina.hebig@lip6.fr

## Alexander Knapp

Institut für Informatik  
Universität Augsburg  
Augsburg, Germany  
knapp@informatik.uni-augsburg.de

## Motivation

In many disciplines it is common to have large collections of sample data that are accepted points of reference and freely available to all ("corpora").

The reason corpora are so widely used is that they play an important, potentially crucial role in the scientific process. In the words of R. Dekker:

*"data-sets [...] are becoming more important themselves and can sometimes be seen as the primary intellectual output of the research".*

In other areas such as the Life Sciences and the Humanities, there are large corpora of data, and it is easy to see that they are beneficial there.

Surely, in Model Based Software Development (MBSD) we could reap the same benefits. Still, we do not have such data readily available in our field. Why is this so? Can we overcome this impasse, and if so, how?

We are not the first to ask these questions, but earlier approaches have not remedied the situation. So we have launched "The Free Models Initiative".

- In April 2014, the first FMI workshop was held to kick-off our initiative. We collected contributions of participants, discussing issues and strategies.

- We are currently setting up the Model Index (MI) to disseminate the workshop results, and our follow-up activities.

- We are planning to organise a workshop at ICSE 2015 in Florence, to further raise awareness.

We hope that we have now reached the critical mass to initiate a cultural change process that will move MBSD ahead.

## Applications

In different fields, datasets are useful in various ways. For MBSD, models corpora serve these and more.

### Validation:

A body of examples that is generally accepted as being representative allows researchers to validate new models against them, as being equally valid in one aspect or another. This will allow us to expand the body of models available for research purposes. It will also allow us to assess new models and create a comprehensive overview of all pragmatic types of and phenomena in models.

### Benchmarking:

New approaches, algorithms, and tools ought to be validated against their predecessors to be able to accurately assess their contribution, so that we

avoid "redoing the same thing over and over again in insignificantly different ways". Representative data sets will be indispensable for valid evaluations. For instance, evaluating clone detection or difference computation tools requires a common benchmark.

### Best Practices:

Model benchmarks and reference models may contribute to improving the state of the practice of modeling by making good (or bad) examples widely accessible. The prerequisite for this are sample models that can be assessed by the community to develop a common understanding of model quality.

## Requirements

All existing repositories have their individual strengths and weaknesses, and all have their individual set of capabilities. But which are the right requirements?

### Durability and Integrity:

The models in a repository must be available for a long time (several decades). Special care must be taken to avoid changing archived models.

### Flexible Licensing:

There will likely be a wide range of licensing needs, as different providers have different demands. This will impact the access control.

### Access control:

A certain amount of access control (in particular in relation to licensing, and for write access) needs to be exerted.

### Search Features:

The benefit of the repository is to serve researchers in discovering the models needed for their work. So, performance, capability, and usability of the search features are of prime importance.

### Provisioning Features:

In order to attract as many contributions as possible, providing models should be made as easy and pleasant as possible.

### Non-Requirements:

In contrast, storage capacity, availability, and security are likely of minor importance.

## Challenges

There are several challenges that need to be addressed to increase the amount and quality of models published.

### Archiving:

In order to be of scientific use, models need to be stored with the same durability, reliability, and accessibility as papers. Taking advantage of more advanced disciplines, maybe ZENODO is an option.

### Searching:

What model meta-data are the right ones? Which ones can we afford/require to extract? Can we extract (some of them) automatically, and if so, how? What techniques are useful for searching model repositories?

### Terminology:

Different communities use inconsistent terminology: what is a "model" to some, is a "diagram" to others, and a "model repository" in another context.

### Measuring:

It is not quite clear how to measure the size of models since there is no uniform terminology, and the topic of model metrics has not been exhausted yet. Similarly, it is an open question how to visualize a collection of models with a view to comparing their size, contents, and .

**Intellectual Property:** Clearly, models are IP. What licensing schemes are suitable? Can we make more models publishable by techniques like obfuscation?

**Incentives for Contribution:** Academic and industrial partners need incentives to publish. The least possible incentive is awarding the same recognition as publishing papers. This demands a cultural change in the community.

## Existing/Lost Repositories

We have collected all kinds of references, citations, links, and rumors regarding existing model repositories and have started to survey them.

### Inclusion

We include all repositories that are credibly reported somewhere (e.g., cited in a peer-reviewed article), but we could not verify the existence of them all. And many of them have disappeared.

### Assessment

We have started to assess these repositories, starting with information about their size, licensing, model type, and model origin. Our assessment is by no means complete and objective.

Repository	Size	Model Types	Access	Origin
 <b>ReMoDD</b> <a href="http://www.cs.colstate.edu/remodd/v1/thalb">http://www.cs.colstate.edu/remodd/v1/thalb</a>	60	Any models (PDF, XMI, ...)	Account	
 <b>Open Models Initiative</b> <a href="http://openmodels.org/">http://openmodels.org/</a>	70	Any models		
 <b>BPM Academic Initiative</b> <a href="http://www.BPMAI.org">http://www.BPMAI.org</a>	29285	Process Models (BPMN, EPCs, etc.) XML		
 <b>AtlanMod Metamodel Zoos</b> <a href="http://www.emn.fr/z-info/atlanmod/index.php/Zoos">http://www.emn.fr/z-info/atlanmod/index.php/Zoos</a>	305	Meta Models (KM3, XMI, RDF, ...)	Free	
 <b>Versicherungs-Anwendungs-Architektur (VAA)</b> <a href="http://www.gdv.de">http://www.gdv.de</a>	90	Class & Use Case Models (PNG, Innovator)	Free	
 <b>Dutch Municipalities</b> <a href="http://www.model-dsp.nl">http://www.model-dsp.nl</a>	>700	Process Models	Account	
 <b>eXperience</b> <a href="http://www.experience-online.ch/cases/experience20.nsf/fallstudie.xsp">http://www.experience-online.ch/cases/experience20.nsf/fallstudie.xsp</a>	525	Case Studies (PDF)		
 <b>Reference Model Catalog</b> <a href="http://rmk.iwi.uni-sb.de/catalog.php">http://rmk.iwi.uni-sb.de/catalog.php</a>	2290	Model (Structured Abstracts)	Free	
 <b>Insurance App. Architecture</b> <a href="http://www.ibm.com/solutions/sg/insurance/enterprise_aa/tech_details.htm">http://www.ibm.com/solutions/sg/insurance/enterprise_aa/tech_details.htm</a>	250	Product, Process & Information Models	?	
 <b>BIT Process Library</b> <a href="http://www.zurich.ibm.com/csc/bit/downloads.html">http://www.zurich.ibm.com/csc/bit/downloads.html</a>	735	Process Models	?	
 <b>Suncorp-Metway Ltd</b> ?	>6000	Process Models	Proprietary	
 <b>SAP R/3 Reference Model</b> ?	>1000	Process Models ER/Org. Models	Proprietary	

## Contribute!

This initiative will only succeed if many people contribute. If you think that you would benefit from having models available, make a start!

### Spread the word!

Help making this initiative known and tell your colleagues and friends via a social network, by sharing the link, or citing SEMI. The more people know about SEMI, the more can benefit and contribute.

### Share your models!

If you have models that can be shared, share them via any of the repositories indexed by SEMI.

### Use it for your research!

If you use a model from a SEMI-indexed repository, cite the Free Models Initiative to maximize publicity. Please refer to it as

Harald Störrle, Regina Hebig, Alexander Knapp (eds.): "The Free Models Initiative", DTU Compute Technical Report-2014-14, DTU, 2014

### Start today!

There is no excuse for procrastination! :-)  
Follow the link to the SEMI - do keep in mind that we've only just started. If you want to help coding or reviewing, you are also most welcome!

Visit the  
FMI  
Website!

