An Index for Software Engineering Models

Harald Störrle[1], Regina Hebig[2], and Alexander Knapp[3]

[1] Department for Applied Mathematics and Computer Science,
Technical University of Denmark, hsto@dtu.dk
[2] Sorbonne Universites LIP6, Paris, France, regina.hebig@lip6.fr
[3] Institut für Informatik, University of Augsburg, Germany, knapp@uni-augsburg.de

1 Motivation

Data Sets are important 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*" (cf. [1, p. 1]). Surely, Model Driven Software Development (MDSD) is no exception here, there are many situations where model corpora are helpful:

- **Benchmarking:** new approaches and algorithms ought to be validated against their predecessors to be able to accurately assess their contribution.
- **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.
- **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.

Unlike other branches of science and engineering, software engineering (and in particular, MDSD) has not yet produced an accepted way of publishing models as data; there are no data journals and conferences. In fact, there are not many models freely available, and those that exist are hard to find, and not very rich in content. Of the few repositories in existence, most are relatively small and provide data without adequate meta-data or not in a machine-readable format. The purpose of this paper is to summarize the knowledge about existing model repositories, and distribute it to the community as an index to existing models.

To this end, the Free Models Initiative (FMI[1], see [3]) has been founded. On the inaugural workshop in March, 2014, researchers convened to each share and pool their knowledge regarding model repositories. The outcome of this workshop is presented in this poster, including the Software Engineering Model Index (SEMI), a catalog of model repositories that we are building up currently. SEMI is supposed to serve as a common entry-point for researchers in need of models, and those that have access to models that they want to share.

In this paper we present use cases and challenges for model repositories as identified by the FMI workshop, together with an initial list of known model

[1] http://www2.compute.dtu.dk/~hsto/events/fmi_14/index.html

repositories with preliminary assessments of their contents. We hope to attract more contributions from the community to grow the index and encourage more researchers to release their models to the public domain.

2 Challenges

A first outcome of the workshop is the insight that there are still several challenges associated with building up successful model repositories. In the following, we present the identified challenges that need to be addressed to increase the amount and quality of models published.

1. **Archiving:** The obvious technical challenge at the back-end is how to archive data with very high reliability, for very long time, yet readily accessible, and economically viable. This challenge has been addressed by others before, so we can probably rely on existing solutions and services such as ZENODO `www.zenodo.org`.
2. **Access support:** The front-end faces a less well explored challenge: how to search for models. Obviously, models need to be stored with meta-data to be able to search for them in meaningful ways. But just which meta-data are sufficient to address the future (and thus unknown) needs of researchers with the effort of extracting meta-data from models? The right balance has yet to be found. Clearly, we should strive to extract as much meta-data from models automatically as is possible, but there are many formats and many information items that might be of interest.
3. **Intellectual property:** Models are intellectual property (IP), and many interesting models are developed in industrial co-operations which means that often industrial partners own the IP, or at least have a veto to publishing. On the one hand, this issue must be addressed by convincing industrial partners to accept co-operation agreements that allow the publication of models just like the publication of scientific articles is accepted today. One way of broadening the scope of publishable work is to offer obfuscation of models. How can usability of models for different research concerns be maintained, when model are obfuscated? Do model repositories need disclaimers to make researchers aware of threads to validity resulting from e.g. model collection or obfuscation? This is a topic that has not been the focus of much research.
4. **Incentives:** On the other hand, academic partners need to be incentivized to publish their models. If models were citable just like papers, and if publishing models were to receive recognition similar to publishing papers, we believe researchers would be motivated to contribute models when possible. Of course, the same recognition should be given on tenure approval committees and so on. So, in a nutshell, we are asking for no less than a cultural change in the community.

We reckon that there is probably a mismatch between supply and demand of models: much more models are needed than are available. So, probably the biggest challenge is to find sufficiently many models to make the idea of sharing models practically useful for a large enough community of researchers.

3 Known Existing Repositories

There are many references to existing model repositories, but frequently, there is little more evidence to their existence than hearsay. Privacy, broken links, and dead references in the literature make it hard to verify the claims raised about them. We have collected the evidence and a preliminary validation of claims below.

- **Repository for Model Driven Development (ReMoDD)**
  `www.cs.colostate.edu/remodd/v1/`
  ReMoDD currently contains around 60 models in different modeling languages. The models are available for account holders, only. Models are stored in a large variety of formats, mostly PDF but also some in XMI.
- **Open Models Initiative (OMI)**
  `http://openmodels.org/`
  Like ReMoDD, OMI offers a platform allowing researchers to share models. There are currently around 70 models of different languages in the repository. The models are mostly available as pictures, some of them include other file formats like MDL. The access to the models is CC BY NA SA. In most cases no explicit hints indicating whether models stem from industry or not.
- **BPM Academic Initiative (BPM AI)**
  `http://www.BPMAI.org`
  The BPM AI is a platform for modeling and sharing models for teaching purposes. As of writing this, it claims to contain 29,285 process models in various machine-readable formats. Apparently, most of the models are created by students as part of their assignments, but some are motivated from industrial case studies, too.
- **AtlanMod Meta Model Zoos**
  `www.emn.fr/z-info/atlanmod/index.php/Zoos`
  This is a collection of around 305 meta models. Each of them is available in multiple formats (e.g. KM3, XMI, or RDF). The access to the models is free.
- **Versicherungsanwendungsarchitektur (VAA)**
  `www.gdvonline.de/vaa`
  The VAA is a standard from the association of German insurance industry. There are around 90 use case and class diagrams, most of them as diagrams in text documents, as PNG files, but also downloadable in INNOVATOR format. The access is free, the website itself is in German.
- **Dutch municipalities**
  `http://www.model-dsp.nl/`
  A large number of Dutch communes have created a common repository of communal administrative processes, which is said to contain 700-800 business process models. Access is restricted to registered members.
- **eXperience**
  `www.experience-online.ch/cases/experience20.nsf/fallstudie.xsp`
  eXperience is is a collection of 525 business modeling case studies, each

of which is mainly a semi-structured text with a few embedded diagrams. Access is CC:BY NC. All case studies stem from industry, e.g., construction or electronics. The majority of the items in the collection are described in German.

- **IWi Reference Model Catalog (RMC)**, [6]
  The RMC contains structured meta-information about 2290 reference models, including the VAA and the IAA mentioned in this list. The RMC does not give access to the indexed models as such, but may help finding the models required for a particular task. The meta-information is somewhat restricted, though, and seems to have not been updated since 2007.
- **Insurance Application Architecture (IAA)**, [4]
  The IAA is said to contain around 250 process models, but the link[2] reported in [4] is broken.
- **BIT Process Library**, [2]
  The BIT Process Library contains 735 process models according to [2], but the link[3] reported in [2] is broken. The collection has been cited several times.
- **Suncorp-Metway Ltd**, [5]
  The Suncorp process model repository for insurance processes contains over 6000, according to [5]. This is a purely proprietary corpus.
- **SAP R3 Process Reference Model**
  This model has been cited very many times, and although it is not in free circulation, there seem to be many copies.

We already identified more than a dozen further model collections and repository, of which many need to be checked for their status and content. Further, there are more repositories that we know of or have heard about, but could not verify. It seems that repositories sometimes get lost over time. Clearly, including such a reference in a research article is problematic, when claims are based on the availability of such model collections.

4 Contributing

In a first stage, we are calling on everybody who knows about a model repository to share their knowledge and make it available to the scientific community by adding links to SEMI. In a second stage, we will survey all known model repositories and individual models and assess them for some basic qualities, such as file format, model type, model size, origin, and so on, and publish these assessments online. In a third step, we want to develop and establish criteria for assessing models and extracting relevant meta-data. This could take a similar form than the current review process for scientific articles.

[2] `http://www.ibm.com/solutions/sg/insurance/enterprise_aa/tech_details.html`

[3] `http://www.zurich.ibm.com/csc/bit/downloads.html`

We are working on a online-system to help with providing, assessing, and using models. For all these activities, we need the support from the community. Thus, we are reaching out to everybody to join this initiative, and come forward with their knowledge, model repositories (or individual models), and expertise: share your knowledge with the community, help with the online system, and input your expertise into the model assessment/review process!

Acknowledgements

We are thankful to the participants of the FMI'14 workshop, who inspired parts of this poster with engaged and interesting discussions and helped to identify an initial list of repositories, namely Pit Pietsch, Andreas Schoknecht, Meike Ullrich, Christian Schneider, Bernhard Thalheim, Frank Wolff, Andreas Oberweis, Tom Thaler, and Mathias Weske.

References

1. Ronald Dekker. The importance of having data-sets. In *Proc. IATUL Conf.* Purdue University, e-Pubs, 2006.
2. D. Fahland, C. Favre, B. Jobstmann, J. Koehler, N. Lohmann, H. Volzer, and K. Wolf. Instantaneous soundness checking of industrial business process models. In *Proc. Intl. Conf. Business Process Management (BPM)*, volume 5701 of *LNCS*, pages 278–293. Springer, 2009.
3. Regina Hebig, Alexander Knapp, and Harald Störrle. Proc. Intl. Ws. Free Models Initiative. In Regina Hebig, Alexander Knapp, and Harald Störrle, editors, *Proc. Intl. Ws. Free Models Initiative*. DTU, 2014. Technical University of Denmark, DTU-TR-2014-15.
4. Jochen M. Küster and N. N. Detecting and Resolving Process Model Differences in the Absence of a Change Log. In *Proc. Intl. Conf. Business Process Modleing (BPM'08)*, number 0 in LNCS, pages 244–260. Springer, 2008.
5. M. La Rosa, Marlon Dumas, R. Uba, and Remco M. Dijkman. Business process model merging: An approach to business process consolidation. Technical report, QUT ePrints 38241, 2010.
6. Tom Thaler, Jrgen Walter, Peyman Ardalani, Peter Fettke, and Peter Loos. The Need for Process Model Corpora. In *Proc. Intl. Ws. Free Models Initiative*. DTU, 2014. Technical University of Denmark, DTU-TR-2014-15.