

The Need for Process Model Corpora

Tom Thaler

Jürgen Walter

Peyman Ardalani

Peter Fettke

Peter Loos

Institute for Information Systems (IWi) at the
German Research Center for Artificial Intelligence (DFKI) and
Saarland University

- 1. Motivation**
- 2. A Vision**
- 3. The IWi Process Model Corpus**
- 4. Selected Application Scenarios**
- 5. Conclusion and Outlook**

Current Situation

- ◆ Companies use large model databases to manage their business processes.
- ◆ Methods and techniques for complexity reduction, handling and analysis of the models are needed.
- ◆ Real data material for the evaluation of these techniques is rare.

Recent Work – some trend indicators

- ◆ Other research fields:
 - ➔ Computational linguistics: The use of speech and text corpora led to high benefits in speech processing, Human Computer Interaction and automatic translation techniques
 - ➔ Biology, chemistry, medicine: Substantial progresses caused by genomic databases.
- ◆ Promising concepts for open access model repositories and efforts establishing them (e. g. afromore, openmodels, prozoom). - Propagation seems being cumbersome; only little use.
- ◆ BPM 2014: Interest of publishing source code of software and algorithms named in the proceedings

A **comprehensive open model corpus** containing models in a **standardized, digital** and **processable** format.

Focused research objectives

- ◆ Creating a consistent understanding of business application systems in different domains.
- ◆ Reusing models in other contexts.
- ◆ Creating a homogeneous data basis for different application and analysis scenarios.

Advantages of a comprehensive open model corpus are amongst others:

Practical aspects

- ◆ Replication of research findings.
- ◆ Development and evaluation of methods, techniques and algorithms.

Theoretical aspects

- ◆ Creating a consistent understanding of terms over different domains.
- ◆ Automatic identification of modeling rules and conventions.
- ◆ Improving the further development of current modeling theories.

What we have done so far

- ◆ Vertical prototyping (focus on **process models**)
- ◆ Primary modeling language is **EPC** – models available in other notations are included as well.
- ◆ Sources: **print** (books, journals, conference proceedings), **transcripts**, **audio recordings**, **face-to-face interviews**, **model files**
- ◆ Comprehensive **harmonization** in terms of the used **EPC constructs**, **syntax** and **semantic corrections**, **output format** (aml), etc.
- ◆ **Documentation** of harmonization and transformation rules.
- ◆ Three model categories: **reference models** (1919 models), **individual models** (293 models), **models from controlled modeling scenarios** (78 models)

Overview - Part 1/2

Cat.	Name	Source	Remarks	Lang.	#
Ref.	ECO-Integral	[1] – book	Information systems for environmental management	DE	85
Ref.	Retail-H 1996	[2] – book	Retail information systems, Ed. 1996	DE	110
Ref.	Retail-H 2004	[3] – book	Retail information systems, Ed. 2004	DE	234
Ref.	ITIL (SWAG)	[4-8] – ARIS-DB	IT Service Management	DE / EN	19
Ref.	SAP R/3 1998	[9] – book	SAP R/3	DE	56
Ref.	SAP R/3	unknown – epml	SAP R/3	EN	1208
Ref.	Y-CIM 2.1	ARIS-Toolset – pdf	Industrial business processes	DE	7
Ref.	Y-CIM 1994	[10] – book	Industrial business processes	EN	100
Ref.	Y-CIM 1998	[11] – book	Industrial business processes	DE	100
Ind.	Custom B2B	text	Software customizing and production of special machinery	DE	46

Legend: Cat.: model category (Ref.: reference model, Ind.: individual model, CM: model from controlled modeling scenarios),
Lang.: language (DE: German, EN: English), #: number of EPCs

Overview - Part 2/2

Cat.	Name	Source	Remarks	Lang.	#
Ind.	Business Registration	audio recordings and transcriptions	Business Registration processes of 8 German communes.	DE	24
Ind.	GK-Rewe	[12] – PDF	Basic course „accounting“ at Chemnitz	DE	68
Ind.	E-Payment	Text	Electronic payment processes of governance	DE	38
Ind.	PMC	[13] – PNML	Birth registration processes and University admission processes	EN	36
Ind.	Vogelaar	[14] – PDF	Dutch governance processes.	EN	81
CM.	Exams	Exams	Exams of a course at a German University between 2010 and 2012	DE	78

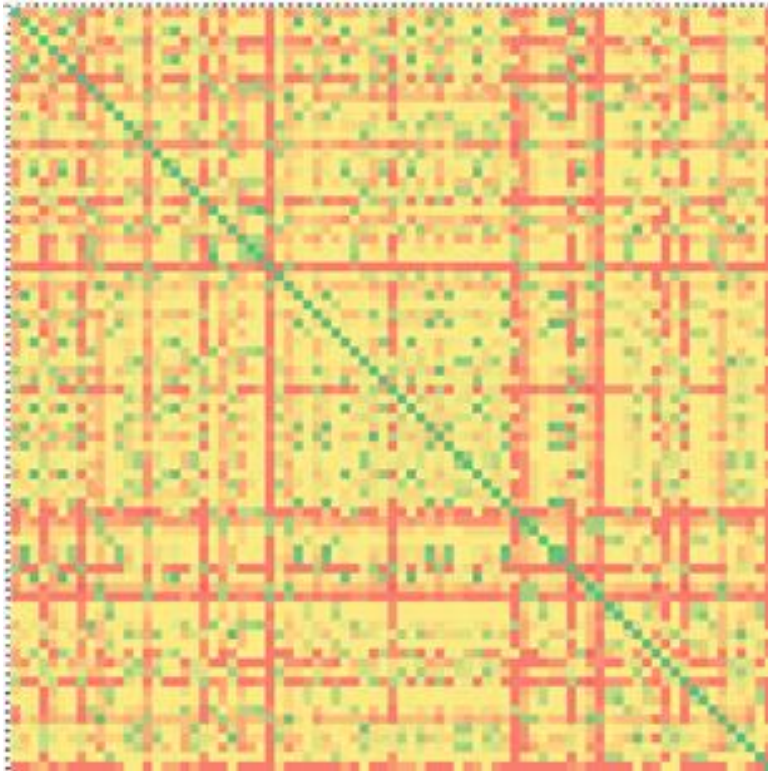
Legend: Cat.: model category (Ref.: reference model, Ind.: individual model, CM: model from controlled modeling scenarios), Lang.: language (DE: German, EN: English), #: number of EPCs

◆ **2,290 models** overall

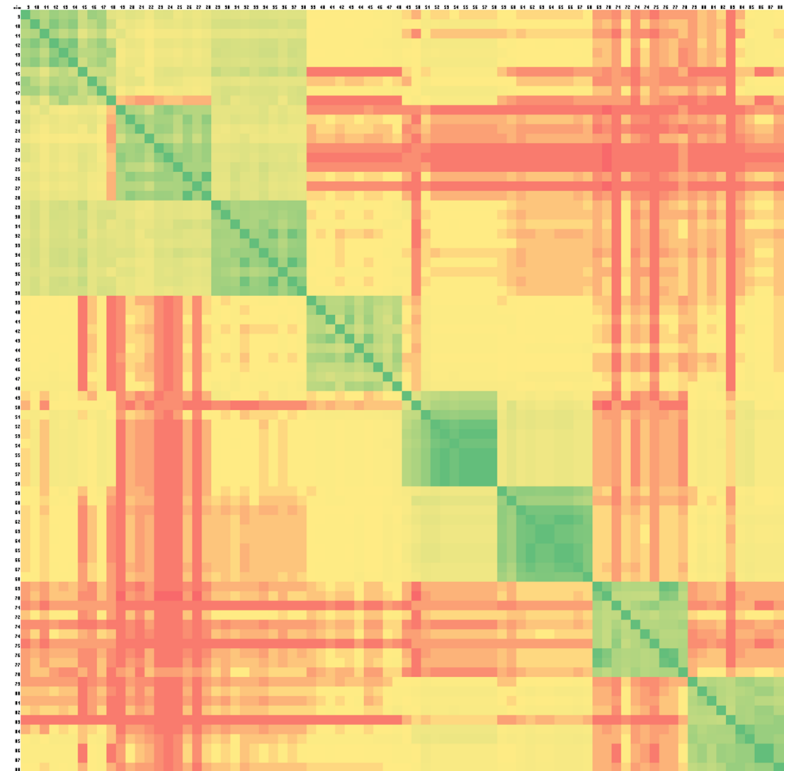
◆ not yet published, as legal aspects need to be clarified

1. Clustering process model repositories

Unsorted Heatmap



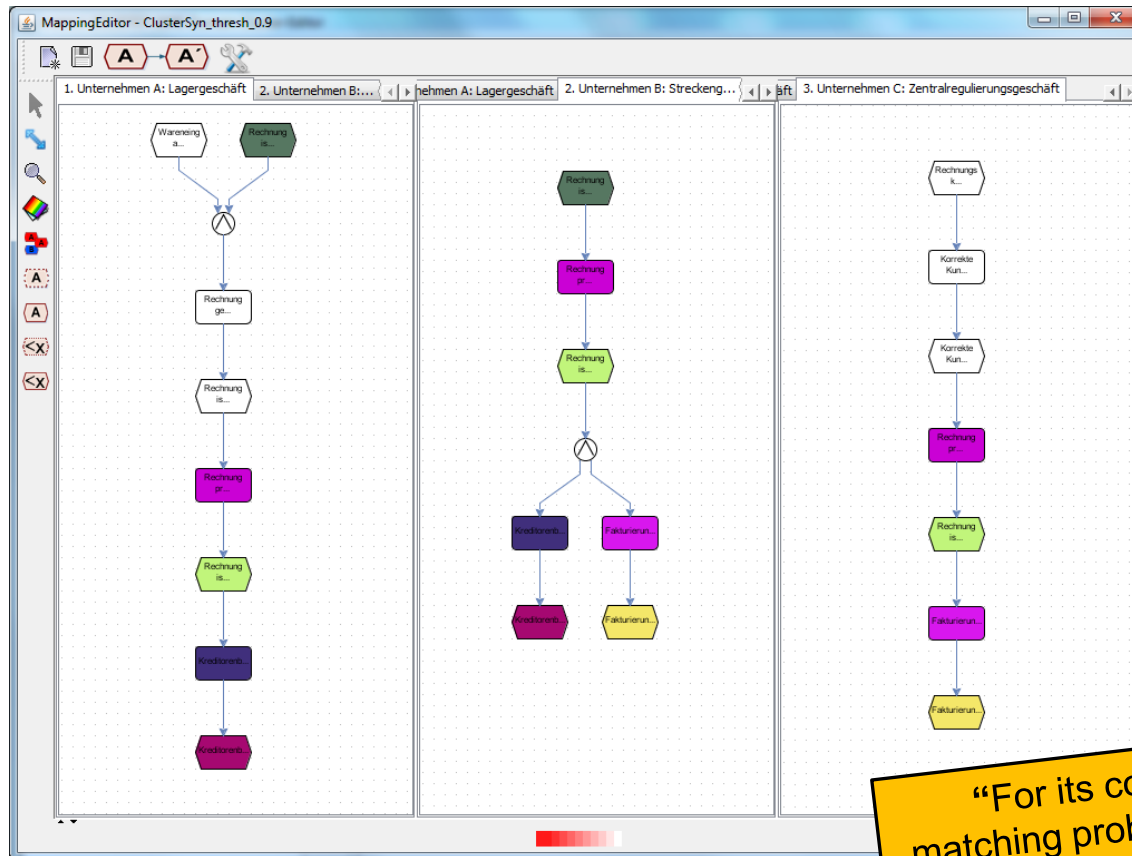
Heatmap sorted by clusters



Legend: low similarity  high similarity

Similarity matrix as a heatmap – Models from [14] (Vogelaar)

2. Process Matching

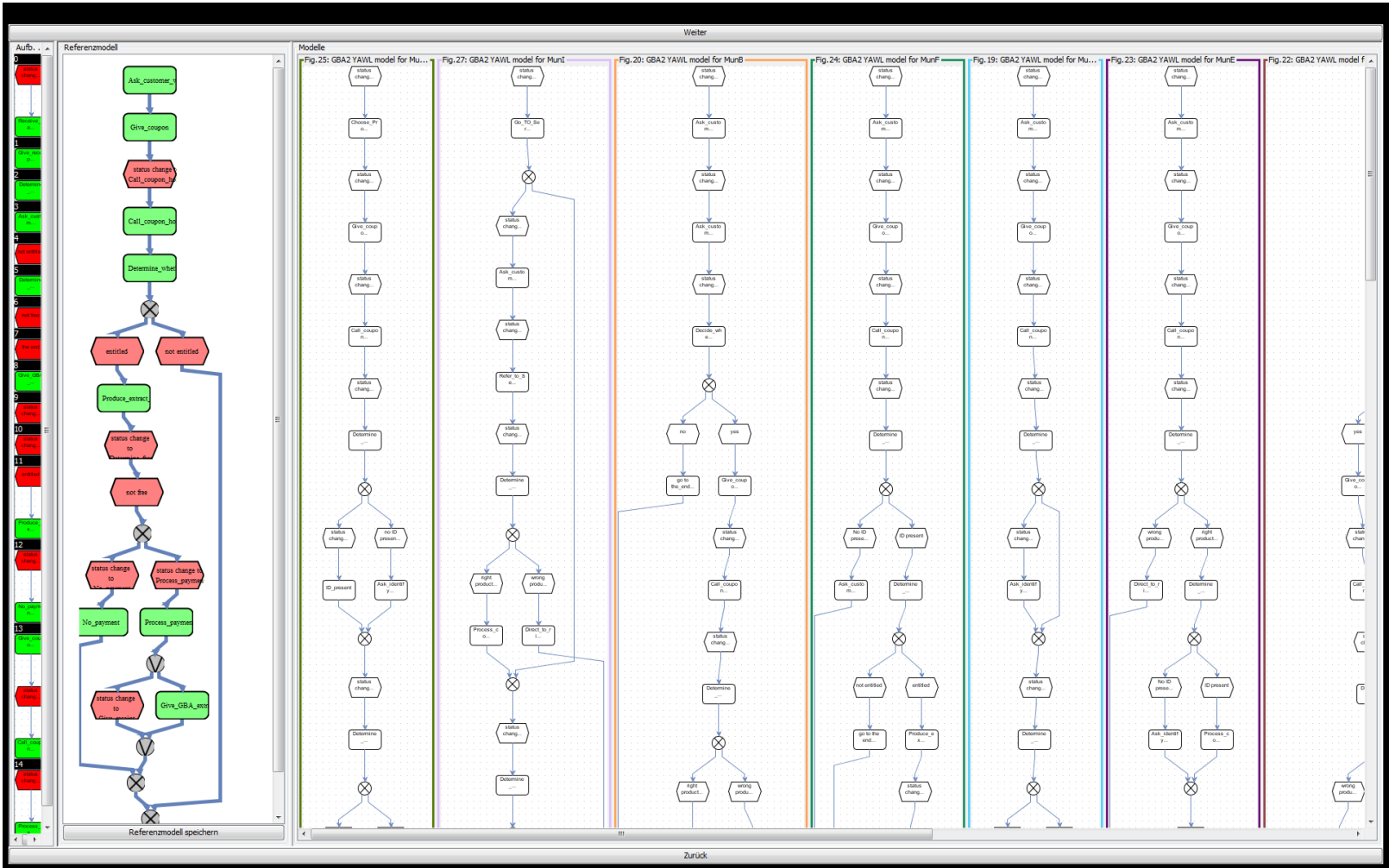


Process Matching with the RefMod-Miner



“For its consistently high performance for both matching problems, the matcher **RefMod-Mine/NSCM** ... was awarded the **Outstanding Matcher Award** of the Process Model Matching Contest 2013.”

3. Inductive reference modeling



Inductive reference modeling with the RefMod-Miner

- ◆ Current corpus contains **16 model collections** with **2,290 EPCs**.
- ◆ Models from **different domains, categories** and **national languages**.
- ◆ Containing models are **harmonized** in terms of the **modeling language, syntax, used constructs** and **format**.
- ◆ Corpus is **not representative**.

But: Already now, the corpus **improves** the **development** and **evaluation** of **methods, techniques** and **algorithms handling** and **analysing** process model repositories.

Next steps:

- ◆ Clarify legal aspects and make the corpus „open“
- ◆ Provide a platform spreading the corpus and getting feedback
- ◆ Add further models

Thank you for your attention!



- [1] Krcmar, H., et al., eds. Informationssysteme für das Umweltmanagement - Das Referenzmodell ECO-Integral. 2000, Oldenbourg: München, Wien.
- [2] Becker, J. and R. Schütte, Handelsinformationssysteme. 1996, Landsberg/Lech: verlag moderne industrie.
- [3] Becker, J. and R. Schütte, Handelsinformationssysteme. Domänenorientierte Einführung in die Wirtschaftsinformatik. 2. ed. 2004, Frankfurt am Main: Redline Wirtschaft.
- [4] Office of Government Commerce, ITIL - Service Strategy. 2010, Norwich: TSO Information & Publishing Solutions.
- [5] Office of Government Commerce, ITIL - Service Design. 2010, Norwich: TSO Information & Publishing Solutions.
- [6] Office of Government Commerce, ITIL - Service Operation. 2010, Norwich: TSO Information & Publishing Solutions.
- [7] Office of Government Commerce, ITIL - Service Transition. 2010, Norwich: TSO Information & Publishing Solutions.
- [8] Office of Government Commerce, ITIL - Continual Service Improvement. 2010, Norwich: TSO Information & Publishing Solutions.
- [9] Keller, G. and T. Teufel, SAP R/3 prozeßorientiert anwenden – Iteratives Prozeß-Prototyping zur Bildung von Wertschöpfungsketten. 1998, Bonn et al.: Addison-Wesley.
- [10] Scheer, A.-W., Wirtschaftsinformatik - Referenzmodelle für industrielle Geschäftsprozesse [Studienausgabe]. 2. ed. 1998, Berlin et al.: Springer.
- [11] Scheer, A.-W., Business Process Engineering - Reference Models for Industrial Enterprises. 2. ed. 1994, Berlin et al.: Springer.
- [12] Kahlert, D. Grundkurs Rechnungswesen. 2010 [cited 2010 23.11.2010]; Available from: <http://www.tu-chemnitz.de/wirtschaft/sapr3/gkrewe/epk/>.
- [13] Cayoglu, U., et al. The Process Model Matching Contest 2013. in 4th International Workshop on Process Model Collections: Management and Reuse (PMC-MR'13). 2013. Beijing.
- [14] Vogelaar, J.J.C.L., et al., Comparing Business Processes to Determine the Feasibility of Configurable Models: A Case Study, in Business Process Management Workshops, LNBIP 100, F. Daniel, K. Barkaoui, and S. Dustdar, Editors. 2012, Springer: Berlin. p. 50-61.