

# Modelagem e Análise de Conformidade do Processo de Atracamento Molecular

Miller Biazus, Lucinéia Heloisa Thom, Márcio Dorn

Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)  
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

mbiazus@inf.ufrgs.br, lucineia@inf.ufrgs.br, mdorn@inf.furb.br

**Abstract.** *The molecular docking process is complex, specially because it includes terms of a specialized vocabulary from the bioinformatics domain, which can lead to misunderstandings between the stakeholders. This article proposes a validated and verified molecular docking process model using the Business Process Model and Notation (BPMN) as graphical notation. The process model exposes a common understanding - based on the bioinformatics literature and on the opinion of domain experts - on how the process should be defined and executed and is used as a starting point to perform a conformance analysis of two different molecular docking softwares.*

**Resumo.** *O processo de atracamento molecular é complexo, especialmente devido ao vocabulário especialista da área de bioinformática, podendo levar a problemas de interpretação entre os stakeholders do processo. O presente artigo propõe um modelo validado e verificado para o processo de atracamento molecular utilizando a Notação e Modelo de Processos de Negócio (BPMN), reconhecida como padrão pela OMG. O modelo de processo expõe um entendimento consensual - com base na literatura de Bioinformática e na opinião de especialistas de domínio - sobre como o processo deve ser definido e executado, e é utilizado como ponto de partida para uma análise de conformidade entre dois softwares existentes para realizar atracamento molecular.*

## 1. Introdução

Bioinformática estrutural visa o entendimento de processos biológicos e a aplicação de técnicas computacionais para aprender e organizar informação estrutural associada a macromoléculas [Altman and Dugan 2005]. Atracamento molecular pode ser definido como uma metodologia que lida com o problema de prever a ligação não-covalente de um receptor (macromolécula, geralmente uma enzima ou proteína) e um ligante (molécula menor) em um nível atômico. A afinidade dessa ligação depende de uma função de avaliação e dos métodos computacionais aplicados [Guedes et al. 2013].

A modelagem de processos com vocabulário especialista e características dinâmicas, tais como os presentes em estratégias computacionais para atracamento molecular é muito complexa, principalmente porque requer o conhecimento de termos de um domínio específico que pode levar a problemas de interpretação, ambiguidades e falta de concordância entre os *stakeholders*. O processo de atracamento molecular inclui macroatividades como preparação do ligante e receptor, as quais podem ser simuladas por uma variedade de softwares (como DockThor [de Magalhães et al. 2014], Autodock Vina

[Trott and Olson 2010]). Contudo, cada software possui suas próprias técnicas para executar tais atividades, o que implica em diferentes processos de atracamento molecular, particularmente quando comparado com o processo proposto pela literatura de bioinformática estrutural .

Considerando o Gerenciamento de Processos de Negócio (BPM), a Notação e Modelo de Processo de Negócio (BPMN) é reconhecida como padrão pela *Object Management Group* (OMG) [Recker 2010]. BPMN vem sendo utilizada para uma grande variedade de propósitos de modelagem de processos, incluindo áreas como análise de negócios, análise de requerimentos, descrição, documentação, simulação e execução de processos [Mateja et al. 2015].

Este artigo propõe a modelagem do processo de atracamento molecular (validada e verificada), baseada na extração de informações da literatura do processo de atracamento molecular, representada através da BPMN. As principais contribuições do presente trabalho são a) documentação padronizada e sistemática do processo de atracamento molecular em uma notação de modelagem de processos; b) motivar o uso da BPMN para representação de processos complexos incorporados na literatura; c) apresentar um estudo de análise de conformidade entre os processos de dois softwares existentes de atracamento molecular.

O restante deste artigo está estruturado da seguinte forma: na seção 2 são introduzidos conceitos de atracamento molecular e de BPM necessários para o entendimento da abordagem proposta; na seção 3 é proposto um modelo em BPMN para o processo de atracamento molecular; nas seções 4 e 5 são descritas, respectivamente, a validação e verificação do modelo de processo; na seção 6 é apresentada uma análise de conformidade entre dois softwares de atracamento molecular; na seção 7 são apresentados trabalhos relacionados; por último, na seção 8 é apresentada a conclusão do trabalho.

## 2. Atracamento Molecular e Gerenciamento de Processos de Negócio

Atracamento molecular é uma abordagem *in silico* que consiste na predição de modos e afinidade de ligação entre um receptor (macromolécula), o qual possui um sítio de ligação de interesse, e um ligante (molécula), o qual atuará no sítio de ligação do receptor, inibindo ou ativando determinada função do mesmo. As moléculas receptor e ligante podem possuir um certo grau de flexibilidade, o que possibilita diferenciar o atracamento molecular em três abordagens: corpos rígidos (*Rigid-body docking*), ligante flexível (*Flexible-ligand docking*) e receptor flexível (*Flexible-receptor docking*). *Flexible-ligand docking* é a abordagem de atracamento molecular utilizada neste trabalho por ser o tipo de atracamento molecular mais comumente utilizado [Dhanik and Kavraki 2001].

O Gerenciamento de Processos de Negócio (BPM) contribui para a redução de custos, de tempo e taxas de erro em execuções de processos [Dumas et al. 2013]. Através da modelagem de processos e automação, organizações podem documentar, otimizar e controlar melhor seus processos de negócio. BPM tem sido explorado e utilizado em inúmeros domínios de aplicação, incluindo o domínio de negócios onde processos são, em geral, bem conhecidos e estáticos, e o domínio da bioinformática onde processos científicos apresentam características dinâmicas e requerem aspectos de flexibilidade e adaptabilidade quando implementados. O ciclo de vida de BPM tipicamente inclui as seguintes fases [Dumas et al. 2013]: identificação do processo, descoberta do processo,

análise do processo, redesenho do processo, implementação do processo e controle e monitoramento do processo. Neste trabalho são aplicadas as duas primeiras fases do ciclo de vida de BPM a fim de obter como resultado o processo *as-is* (ou seja, a modelagem do processo em seu estado atual) de atracamento molecular.

Considerando a fase de descoberta de processo, a notação gráfica BPMN possibilita a representação dos mais diferentes aspectos de um processo. Seu foco é ser um ponto intermediário de conexão entre a modelagem do processo de negócio e a implementação. Em vista disso, é utilizada para uma grande variedade de propósitos de modelagem de processos [Mateja et al. 2015], incluindo, principalmente, análise de negócios e documentação de processos. Diversos softwares podem ser utilizadas para modelagem de processos (por exemplo *Bizagi*, *Signavio* e *Bonita*).

### 3. Modelagem do processo de atracamento molecular

Em métodos tradicionais para realizar a modelagem o analista depende de informações obtidas e organizadas a partir de especialistas no domínio [Dumas et al. 2013]. Foram analisadas ao todo 37 referências<sup>1</sup> da literatura de bioinformática. As principais atividades de atracamento molecular foram identificadas e extraídas manualmente a partir da convergência dos dados entre as referências, ou seja, atividades que eram definidas ou pressupostas como existentes em todas as referências. Atividades específicas de técnicas particulares de atracamento molecular não foram incluídas.

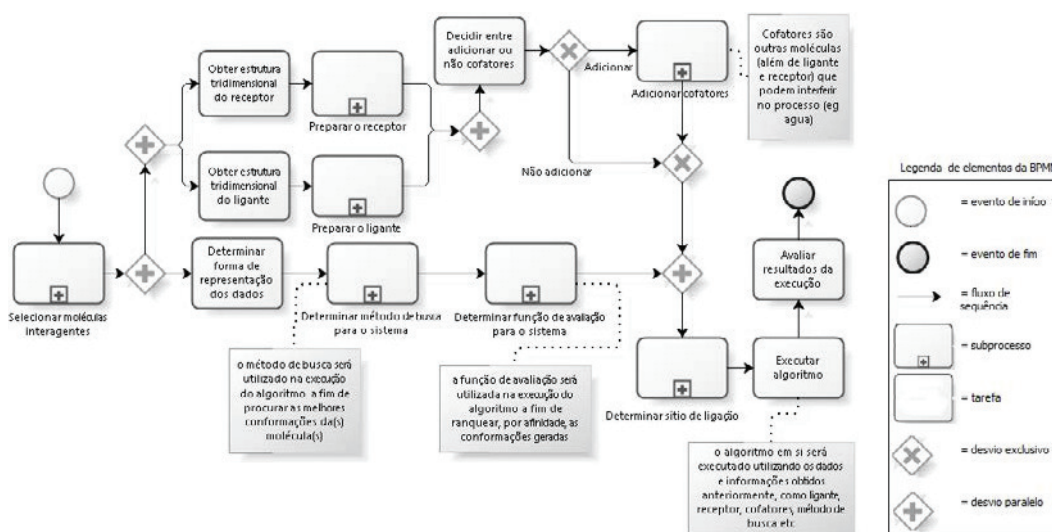


Figura 1. Modelo de processo - Atracamento molecular

Com o primeiro esboço do modelo do processo de atracamento molecular em mãos, foram conduzidos debates com três especialistas do domínio de bioinformática voltado a atracamento molecular. O objetivo foi refinar e complementar o modelo de processo de atracamento molecular obtido na literatura. O processo de atracamento molecular apresentado neste artigo foi modelado usando elementos da notação BPMN 2.0 [Allweyer 2010] e o software de modelagem Bizagi versão 2.9. A figura 1 exhibe

<sup>1</sup>Disponibilizadas em <https://goo.gl/kIMwnt>

o processo principal de atracamento molecular. A modelagem do processo foi realizada com base nas sete diretrizes de modelagem de processo, incluindo utilização do padrão verbo-objeto para descrever tarefas, eventos de início e fim únicos, entre outras [Mendling et al. 2010]. Tais diretrizes são relevantes para minimizar a probabilidade de erros estruturais no modelo, assim como *deadlocks* e *livelocks*.

#### 4. Validação da Modelagem do Processo de Atracamento Molecular

Para validar o modelo de processo de atracamento molecular apresentado na figura 1 foi utilizado um questionário<sup>2</sup>. O questionário foi respondido por 9 pesquisadores da área de biologia molecular (especificamente com envolvimento em atracamento molecular), sendo três pesquisadores com conhecimento avançado em atracamento molecular, três pesquisadores com conhecimento intermediário e três pesquisadores com conhecimento básico.

Foi disponibilizado acesso ao material de um tutorial sobre BPMN para os pesquisadores que não conheciam a notação. A principal pergunta do questionário buscava investigar se o pesquisador, de fato, determinava o modelo como correto. Os 9 pesquisadores relataram que a modelagem do processo estava correta e, portanto, sem necessidade de alterações adicionais, indicando que o processo modelado está em conformidade com o processo existente na vida real.

#### 5. Verificação da Modelagem do Processo de Atracamento Molecular

A notação BPMN, apesar de suas características favoráveis, como fácil entendimento e capacidade de representação de processos complexos, é ambígua em suas construções e deficiente em semântica formal, o que pode levar os modelos a exibirem estados indesejáveis como *deadlocks* e *livelocks*, além de inconsistência, ambiguidade e incompletude [Van Nuffel et al. 2009]. Para suprimir esses potenciais problemas, foi feita uma verificação na modelagem seguindo uma abordagem relatada por Dijkman e colaboradores (2008), onde utiliza-se um mapeamento de BPMN para Redes de Petri. Com tal mapeamento foi gerado um modelo na forma de Rede de Petri, o qual foi verificado com um software Woflan [van der Aalst 1999] a fim de analisar determinadas propriedades, como: verificação sintática (tarefas sem condição de entrada ou saída), detecção de construções suspeitas (fragmentos do modelo que podem ocasionar algum estado indesejável, como *deadlock*), detecção de transições mortas (nunca disparada em nenhuma instância do processo) e transições não-vivas (existe marca alcançável para a qual a transição é morta) e *Soundness*. Aalst (1997) define *soundness* como: “Para qualquer caso, o processo irá eventualmente terminar, e no momento que o processo terminar existirá uma marca em um lugar  $x$  e todos os outros lugares estarão vazios”.

Segundo Dijkman(2008) um processo BPMN é *bem formado* se possui as seguintes características: um evento de início possui apenas um fluxo de sequência de saída e não possui nenhum fluxo de sequência de chegada; um evento de fim possui apenas um fluxo de sequência de chegada e não possui nenhum fluxo de sequência de saída; atividades e eventos intermediários possuem apenas um fluxo de sequência de chegada e apenas um fluxo de sequência de saída; desvios divergentes possuem apenas um fluxo de sequência de chegada e mais de um fluxo de saída; desvios convergentes possuem vários fluxos de

<sup>2</sup>Disponibilizado em <https://goo.gl/aOwKQk>

sequência de chegada, porém apenas um fluxo de sequência de saída. O processo de atracamento molecular modelado, assim como cada um de seus subprocessos possui todas as condições citadas.

A Rede de Petri resultante do mapeamento é exibida na figura 2. O diagnóstico do software Woflan não revelou transições mortas ou não vivas, erros sintáticos e construções suspeitas. Logo, todas as propriedades estão corretas, e isso indica que a Rede de Petri relacionada ao processo de atracamento molecular modelado não apresenta inconsistências semânticas e está verificada formalmente e, portanto, tanto o modelo de processo de atracamento molecular mapeado para Redes de Petri quanto o modelo em BPMN não apresentem problemas semânticos e estruturais em sua modelagem.

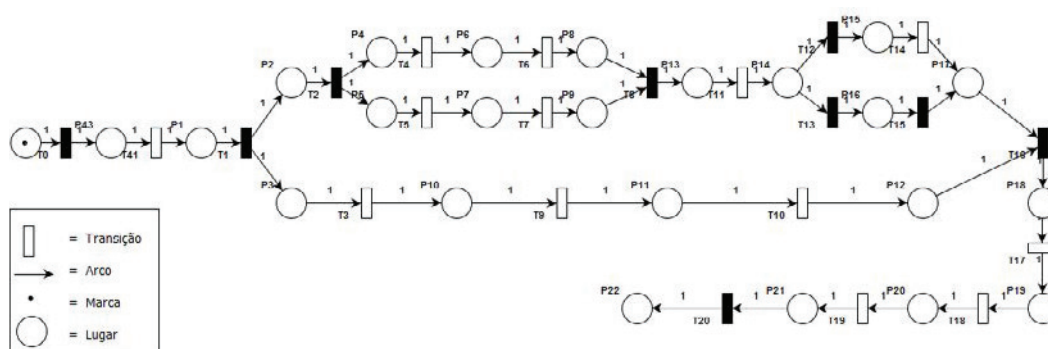


Figura 2. Rede de Petri do Processo de Atracamento Molecular

## 6. Análise de conformidade do modelo de processo de atracamento molecular com os processos de softwares de atracamento molecular

A mineração de processos representa um papel importante na área de BPM. O principal objetivo é a descoberta de modelos de processo baseada nos dados disponíveis de *logs* de eventos [Gehrke and Werner 2013]. Gehrke (2013) relata que a abordagem utilizada na mineração de processos para a obtenção de informações confiáveis sobre a execução real de um processo de negócio baseia-se na exploração de dados armazenados em sistemas, os quais são gerados durante o processamento de transações de negócios.

Foram analisados dois softwares de atracamento molecular: DockThor [de Magalhães et al. 2014] e Autodock Vina [Trott and Olson 2010]. Foram identificados 5 logs de eventos únicos em cada um dos dois softwares analisados. O *plugin* "BPMN analysis using heuristic miner" [Ayutaya et al. 2012] foi utilizado a fim de gerar um modelo em BPMN para DockThor (figura 3) e AutodockVina (figura 4) a partir dos *logs* de eventos gerados.

*Fitness* é uma medida de conformidade que indica o grau ao qual os traços de execução de um *log* podem ser associados com caminhos de execução válidos especificadas pelo modelo de processo [Rozinat and van der Aalst 2008]. No caso do DockThor, existem diversas divergências (em relação ao modelo proposto) na ordem que as atividades são executadas, porém todas as atividades existentes nos *logs* de eventos do DockThor constam no modelo de processo de atracamento molecular, e não é necessária a criação artificial de uma marca na execução do processo: os *logs* podem ser executados do começo



ao fim sem nenhum problema, se adaptando perfeitamente ao modelo de processo, o que resulta em um *fitness* de 100%.

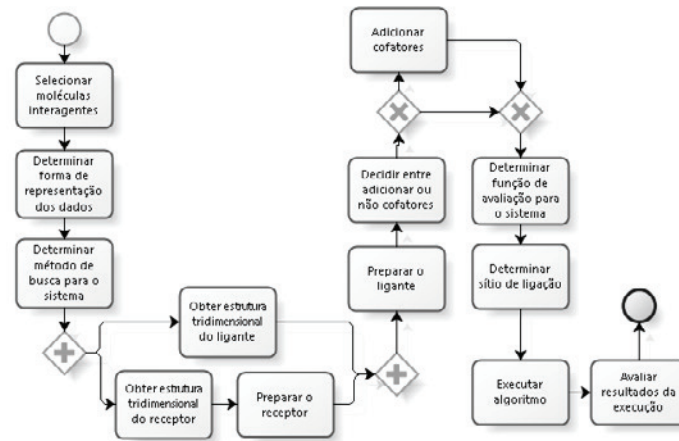


Figura 3. Processo minerado do software DokThor

Analogamente ao DockThor, todas as atividades dos *logs* de eventos do Autodock Vina constam no modelo de processo de atracamento molecular. São encontradas, no entanto, diversas divergências na ordem que essas atividades são executadas. O *replay* de todos os logs de eventos, ainda, requer a criação artificial de uma marca para a execução do processo, portanto não apresentando 100% de *fitness*.

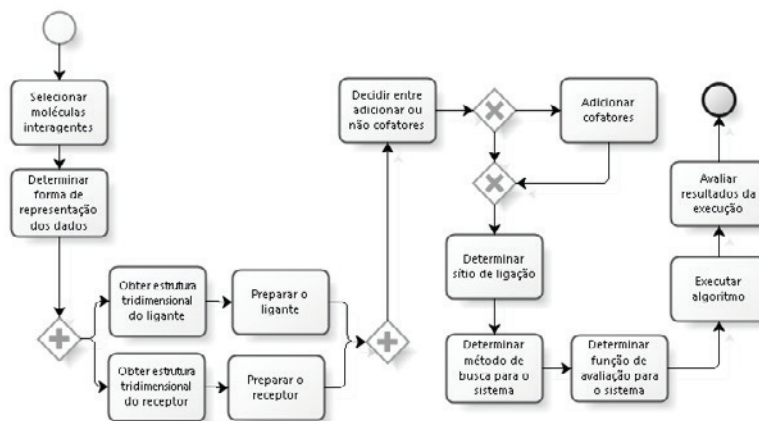


Figura 4. Processo minerado do software Autodock Vina

## 7. Trabalhos relacionados

Existem vários trabalhos que abordam técnicas para descoberta e mineração de processos. Em Gonçalves, Santoro e Baiao (2009) é explorada a técnica narrativa associada a mineração de texto e interpretação de linguagem natural para geração automática de modelos de processo. No trabalho consta que os participantes do processo descrevem a maneira de agir através de histórias de usuário.

Em Santoro e colaboradores (2010) os autores exploram um método para capturar histórias de um determinado grupo como alternativa para as entrevistas individuais

tradicionalmente utilizadas para elicitar processos. Segundo os autores, quando em coletivo, os participantes são capazes de verificar de forma cooperativa os processos descritos pelos outros integrantes por meio de suas perspectivas e adicionar novas histórias com descrições adicionais ou alternativas.

No contexto de extração de modelos a partir da literatura, existem algumas contribuições como o trabalho apresentado por Friedrich e colaboradores (2011), onde os autores apresentam um método automático para gerar modelos em BPMN a partir de textos em linguagem natural. Os métodos mais utilizados em trabalhos de elicitação de processos de negócio são a entrevista e a análise da literatura existente. Segundo Silva e colaboradores (2012), as informações obtidas a partir de diferentes técnicas de elicitação (por exemplo literatura, *brainstorms* e entrevistas) devem ser agrupadas e organizadas, e informações conflitantes devem ser resolvidas. Essas ações irão, então, refletir no modelo de processo desejado, que representa o conhecimento obtido na elicitação. A abordagem utilizada neste artigo para modelar processos a partir da literatura não é complementar às abordagens existentes, sendo que os modelos obtidos devem ser complementados (se necessário) e validados por especialistas de domínio.

Poucos trabalhos têm sido feitos em relação à documentação sistematizada de processos de bioinformática. A documentação é limitada à tutoriais e guias de usuário para softwares específicos de atracamento molecular.

## 8. Conclusão

No presente trabalho foi apresentado um modelo, na notação BPMN, para o processo de atracamento molecular, validado a partir de questionários com pesquisadores e validado a partir de um mapeamento para Redes de Petri, um formalismo que possibilita diagnósticos e análises formais complexas. Após, foi feita uma análise de conformidade com o modelo extraído a partir de mineração dos processos de dois softwares utilizados para atracamento molecular (DockThor e Autodock Vina). Tal análise mostra que apenas o *fitness* do processo do software Dockthor é perfeito.

O conceito de BPM, além de sua utilidade para o trabalho, foi particularmente útil para modelar um processo complexo com vocabulário especialista, como atracamento molecular, usando-se a literatura como referência e com interferência mínima de entrevistados. O modelo de processo gerado na notação BPMN apresentou alta aceitação dos usuários, os quais deram *feedback* positivo em relação ao nível de compreensão do modelo. Ainda, alguns usuários elogiaram a iniciativa e reiteraram que a modelagem utilizando BPMN é bastante clara e facilita o entendimento do processo.

Trabalhos futuros são sugeridos a fim de processar as demais fases do ciclo de vida de processo não abordadas no trabalho (por exemplo implementação e execução).

## 9. Agradecimentos

Esta pesquisa é suportada pelo Programa SticAmSud da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq e pela Fundação de Ampara à Pesquisa do Estado do Rio Grande do Sul - Fapergs.

## Referências

- Allweyer, T. (2010). *BPMN 2.0*. BoD.
- Altman, R. B. and Dugan, J. M. (2005). *Defining Bioinformatics and Structural Bioinformatics*, pages 1–14. John Wiley Sons, Inc.
- Ayutaya, N., Palungsuntikul, P., and Premchaiswadi, W. (2012). Heuristic mining: Adaptive process simplification in education. In *ICT and Knowledge Engineering 2012 10th International Conference on*, pages 221–227.
- de Magalhães, C. S., Almeida, D. M., Barbosa, H. J. C., and Dardenne, L. E. (2014). A dynamic niching genetic algorithm strategy for docking of highly flexible ligands. *Information Sciences*, pages 206–224.
- Dhanik, A. and Kavraki, L. E. (2001). Protein–ligand interactions: Computational docking.
- Dumas, M., Rosa, M. L., Mendling, J., and Reijers, H. A. (2013). *Fundamentals of Business Process Management*. Springer.
- Gehrke, N. and Werner, M. (2013). Process mining.
- Goncalves, J., Santoro, F., and Baiao, F. (2009). Business process mining from group stories. In *Computer Supported Cooperative Work in Design, 2009. CSCWD 2009. 13th International Conference on*, pages 161–166.
- Guedes, I. A., Magalhães, C. S., and Dardenne, L. E. (2013). Receptor–ligand molecular docking. *Biophysical Reviews*, 6(1):75–87.
- Mateja, K., Gregor, J., Marjan, H., and Gregor, P. (2015). Business process model and notation: The current state of affairs. *Computer Science and Information Systems*, 12:509, 539.
- Mendling, J., Reijers, H. A., and van der Aalst, W. M. P. (2010). Seven process modeling guidelines (7pmsg). *Inf. Softw. Technol.*, 52(2):127–136.
- Recker, J. C. (2010). Opportunities and constraints : the current struggle with bpmn. *Business Process Management Journal*, 16(1):181–201.
- Rozinat, A. and van der Aalst, W. (2008). Conformance checking of processes based on monitoring real behavior. *Information Systems*, 33(1):64 – 95.
- Santoro, F. M., Borges, M. R. S., and Pino, J. A. (2010). Acquiring knowledge on business processes from stakeholders’ stories. *Adv. Eng. Inform.*, 24(2):138–148.
- Trott, O. and Olson, A. J. (2010). Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Computational Chemistry* 31, pages 455–461.
- van der Aalst, W. M. P. (1999). Woflan: A petri-net-based workflow analyzer. *Syst. Anal. Model. Simul.*, 35(3):345–357.
- Van Nuffel, D., Mulder, H., and Van Kervel, S. (2009). Enhancing the formal foundations of bpmn by enterprise ontology. In *Advances in Enterprise Engineering III*, volume 34 of *Lecture Notes in Business Information Processing*, pages 115–129. Springer Berlin Heidelberg.