



# VỀ KHAI PHÁ QUY TRÌNH

Trường hè Khai phá dữ liệu 2016  
<http://fit.uet.vnu.edu.vn/dmss2016/>

PGS. TS. HÀ QUANG THỤY  
PHÒNG THÍ NGHIỆM KHOA HỌC DỮ LIỆU VÀ CÔNG NGHỆ TRI THỨC  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**  
ĐẠI HỌC QUỐC GIA HÀ NỘI

17-08-2016

# Nội dung

1. Tại sao khai phá quy trình ?
2. Khai phá quy trình là gì ?
3. Một số bài toán trong khai phá quy trình
  4. Một vài trao đổi

## Tại sao khai phá quy trình ?

- Mô hình quy trình: yếu tố lợi thế cạnh tranh
- Nhu cầu tự động hóa mô hình hóa quy trình nghiệp vụ
- Khai phá quy trình với Việt Nam

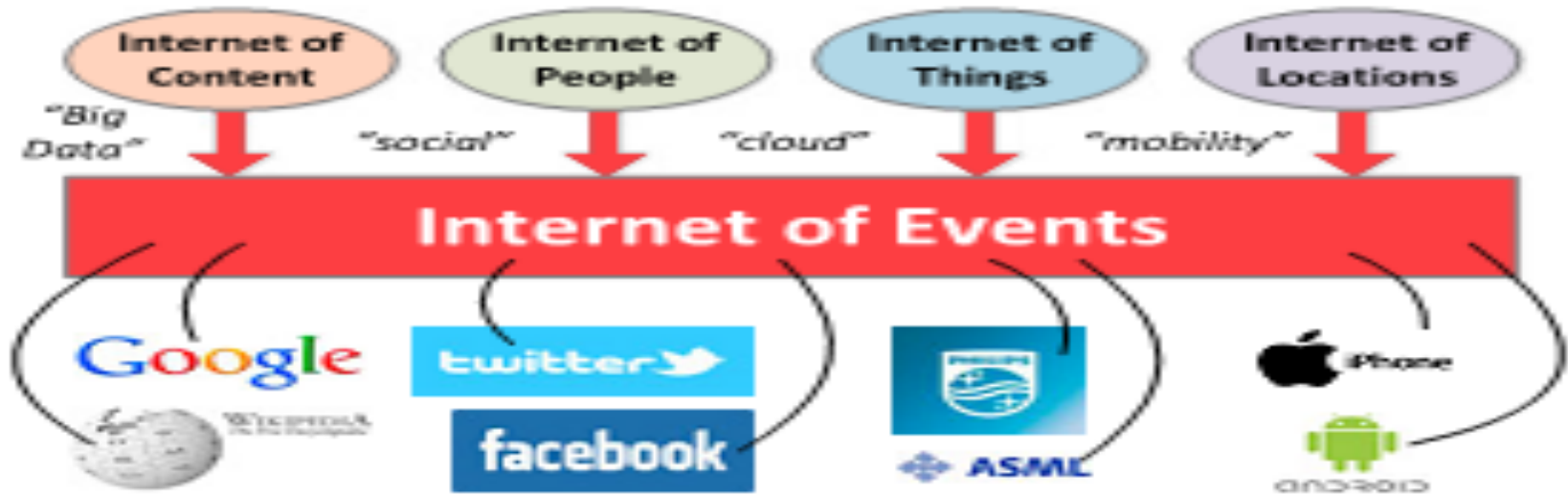


**“Process Mining: Data Science in Action”**. **Wil van der Aalst**  
<https://www.coursera.org/course/procmin>

“Data science is a hot topic, that’s proved by the large number of applications for this free online course (Massive Open Online Course, MOOC) by TU/e. The course now has more than 20,000 participants from 165 countries  
05 November 2014

<https://www.tue.nl/en/university/news-and-press/news/05-11-2014-already-more-than-20000-registrations-for-online-lecture-data-science/#top>

# KPQT: Internet of Events



**“IoE: Dữ liệu sự kiện được sinh ra từ rất nhiều nguồn được kết nối Internet”.** “IoE là thành phần của IoC, IoP, IoT, IoL

- IoC: mọi UGC nâng cao tri thức về chủ đề cụ thể (web pages, articles, encyclopedia như Wikipedia, YouTube, e-books, newsfeeds, v.v.),
- IoP: mọi dữ liệu về tương tác xã hội (e-mail, Facebook, Twitter, forums, LinkedIn, v.v.),
- IoT: mọi đối tượng vật lý kết nối mạng (mọi thứ có định danh duy nhất + hiện diện trong một cấu trúc tựa Internet),
- IoL: mọi dữ liệu có thuộc tính địa lý/không gian địa lý. Thiết bị di động (như smartphome) tăng nhanh sự kiện có định vị và chuyển động).

**[Aalst16] WMP Van der Aalst. *Process Mining: Data Science in Action (2<sup>nd</sup> edition)*. Springer, 2016**

# Trình độ cạnh tranh của nền kinh tế

## ● Giới thiệu

- Ba mức trình độ cạnh tranh kinh tế: định hướng yếu tố cơ bản, định hướng hiệu quả, định hướng đổi mới
- Hai mức phụ xen giữa ba mức chính

## ● Nền kinh tế định hướng yếu tố cơ bản

- *factor-driven economy*
- Giá trị thấp (tài nguyên thiên nhiên và lao động đơn giản) là nền tảng chi phối lợi thế cạnh tranh và xuất khẩu
- rất nhạy cảm với chu kỳ kinh tế thế giới, giá cả hàng hóa, và biến động tỷ giá

***World Economic Forum (2015). The Global Competitiveness Report 2014–2015***

[http://www3.weforum.org/docs/WEF\\_GlobalCompetitivenessReport\\_2014-15.pdf](http://www3.weforum.org/docs/WEF_GlobalCompetitivenessReport_2014-15.pdf)

# Quy trình NV: một yếu tố cạnh tranh

- **Nền kinh tế định hướng hiệu quả**
  - *Efficiency - Driven Economy*
  - Lợi thế do tạo ra sản phẩm và dịch vụ tiên tiến hơn rất hiệu quả
    - Đầu tư mạnh cơ sở hạ tầng hiệu quả,
    - quản lý chính quyền thân thiện với doanh nghiệp,
    - ưu đãi đầu tư mạnh,
    - nâng cao kỹ năng
    - và tiếp cận tốt với nguồn vốn đầu tư nhằm cải thiện lớn về năng suất
- **Nền kinh tế định hướng đổi mới**
  - *innovation-driven economy*
  - cạnh tranh bằng các sản phẩm, dịch vụ mới và/hoặc độc đáo
  - các công nghệ mới nhất và/hoặc các quy trình sản xuất/mô hình kinh doanh tinh vi nhất
  - Khu vực dịch vụ có tỷ trọng cao trong nền kinh tế
  - kiên cường trước những cú sốc từ bên ngoài

FIGURE 6.19 Cocreation-oriented management in pursuit of a series of positive service interactions. Qui14

# Cạnh tranh kinh tế quốc gia tại ASEAN



- Cột trụ: Business sophistication
- **Việt Nam**
  - Thuộc nhóm nước thấp nhất khu vực và thế giới
  - Hội nhập: kết hợp các cột trụ



# Quy trình nghiệp vụ và mô hình hóa QT

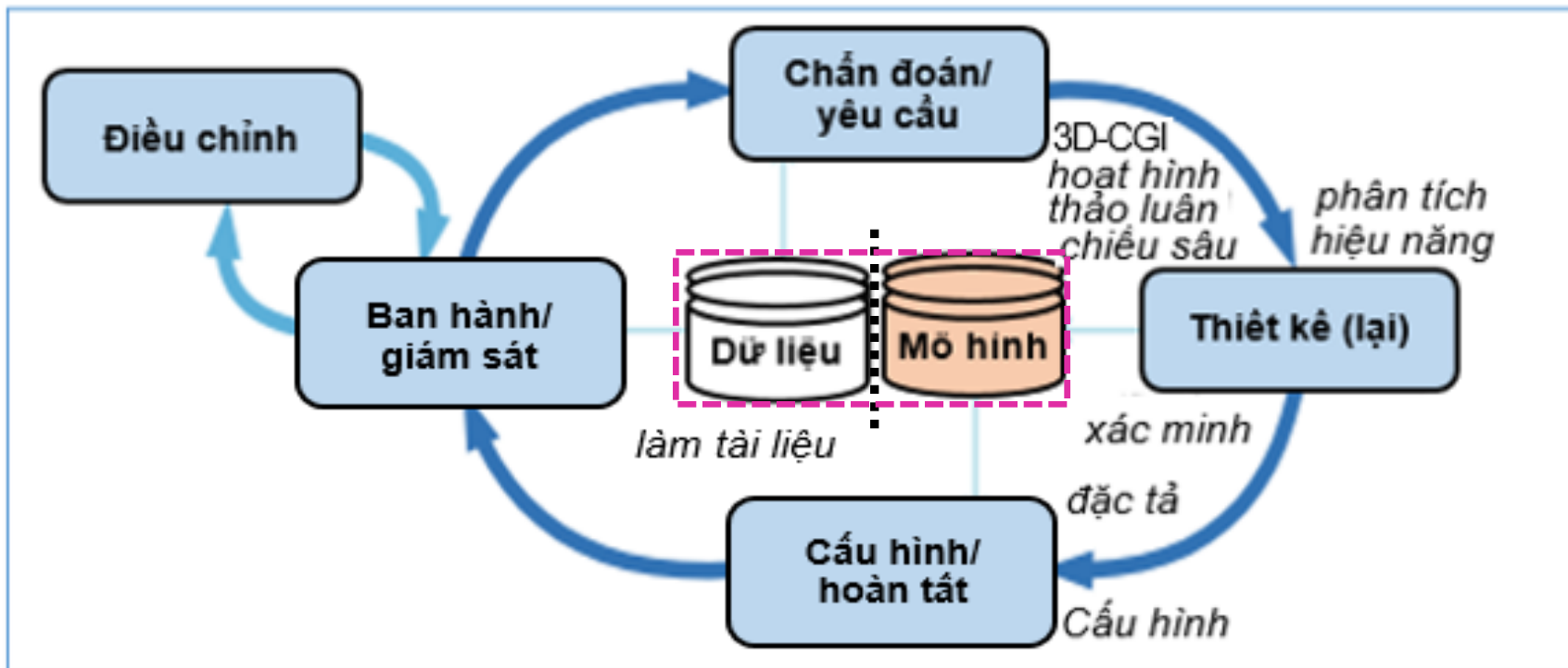
## ● Vấn đề thuật ngữ

- “Business”: kinh doanh/nghiệp vụ
- “nghiệp vụ”: toàn diện không chỉ kinh doanh
- “process”: quá trình/quy trình
- “quy trình” : dẫn dắt hành động

## ● Mô hình QTNV

- Quy trình nghiệp vụ: mọi nghiệp vụ (HTTT) trong tổ chức
- Mô hình quy trình: biểu diễn QTNV. “hiểu” và “cộng tác hành động”.
- Tổ chức nhỏ: Ngầm hiểu để cộng tác hành động. “Văn hóa tổ chức”
- Tổ chức vừa và lớn: nhiều QTNV, QTNV phức tạp hơn, nhiều người hơn → mô hình trên ngôn ngữ biểu diễn QTNV
- Mô hình QTNV: sử dụng rộng rãi, tài nguyên q/trọng mọi tổ chức
- Hoạt động liên tổ chức: hành động đúng → thỏa thuận tương tác chung (B2B)

# Vòng đời quản lý quy trình BPM



## ● Giới thiệu

- Vòng đời BPM qua năm giai đoạn quản lý quy trình
- Thiết kế** (*design/redesign*), **cấu hình/hoàn tất** (*configuration/implementation*), **ban hành/giám sát** (*enactment/monitoring*), **điều chỉnh** (*adjustment*), **chẩn đoán/yêu cầu** (*diagnosis/requirements*).
- Điều chỉnh** được coi là giai đoạn con của *ban hành/giám sát*: *ban hành/giám sát* có khoảng thời gian dài.

# Các giai đoạn trong vòng đời BPM

- **Thiết kế**
  - Một mô hình được thiết kế. Mô hình hóa quy trình
- **Cấu hình/thi hành**
  - Đưa mô hình thiết kế → một hệ thống thực thi
  - Mô hình WFM/BPM (độ sẵn sàng thực thi cao): diễn ra nhanh; mô hình chưa sẵn sàng thực thi phải lập trình (hardcode): tốn thời gian/công sức
- **Ban hành/giám sát**
  - Hệ thống được thi hành và bắt đầu được giám sát
  - Người quản lý: giám sát phát hiện bộ phận cần điều chỉnh, nếu có đi tới giai đoạn điều chỉnh
- **Điều chỉnh**
  - QT không được thiết kế lại, phần mềm mới không được tạo
  - ràng buộc sẵn có (tham số hệ thống) được dùng để điều chỉnh/cấu hình lại QT

# Các giai đoạn trong vòng đời BPM

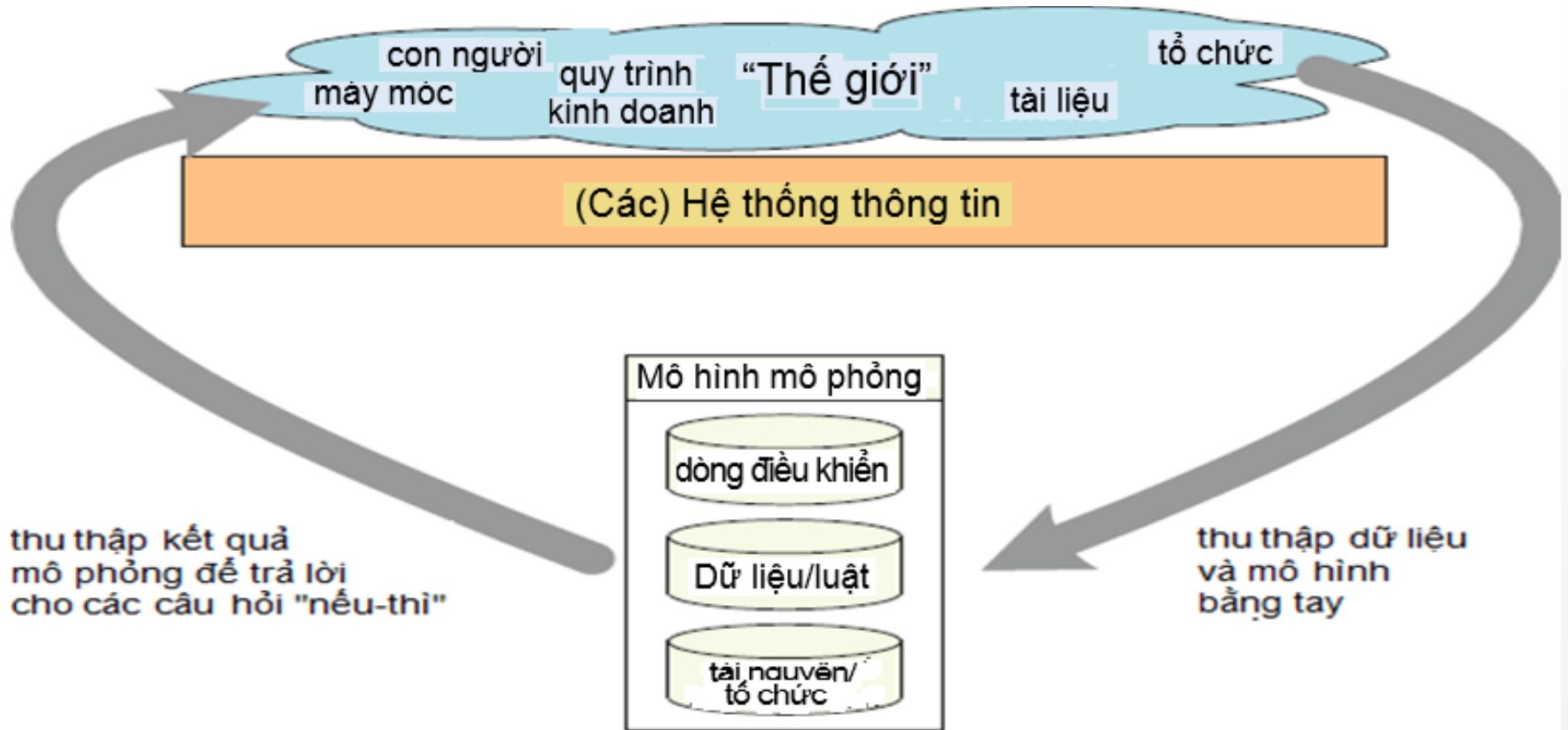
- **Chẩn đoán/yêu cầu**
  - đánh giá QT và giám sát yêu cầu mới nổi do thay đổi môi trường của QT (ví dụ, thay đổi chính sách, luật, cạnh tranh)
  - Hiệu suất kém (ví dụ, không có khả năng đáp ứng mức dịch vụ) hoặc yêu cầu mới do môi trường đặt ra
  - Lặp lại vòng đời BPM bắt đầu từ giai đoạn thiết kế lại
- **Vai trò của mô hình QT và dữ liệu**
  - Mô hình QT đóng vai trò chủ đạo trong các giai đoạn **thiết kế (lại)** và **cấu hình/thực hiện**
  - dữ liệu đóng một vai trò chủ đạo trong các giai đoạn **ban hành/ giám sát** và **chẩn đoán/yêu cầu**
- **Kết nối giữa mô hình QT và dữ liệu**
  - Hình vẽ: BPM chưa có kết nối giữa mô hình QT và dữ liệu
  - Gần đây: một vài kết nối thực sự giữa thực hiện QT và thiết kế QT
  - hầu hết trường hợp: chẩn đoán/yêu cầu không được hỗ trợ một cách có hệ thống và liên tục. ***Vấn đề nghiêm trọng/thay đổi bên ngoài lớn mới kích hoạt một lần lặp vòng đời → KPQT***

# Mô hình hóa quy trình [Aalst16]

- Nội dung mô hình hóa QTNV
  - Khởi tạo, cải tiến mô hình QTNV
  - Hoạt động tất yếu của mọi tổ chức.
- Mô hình hóa QTNV: tiếp cận truyền thống
  - Bộ phận của khoa học quản lý (management science)
  - Quản lý tác nghiệp bao gồm vận trù học (*operation research*)
  - Dùng mô hình toán học: QH tuyến tính (*linear programming*), lập KH dự án (project planning), mô hình hàng đợi (queueing models) v.v.
  - Thủ công, dựa trên kinh nghiệm hoặc kế thừa
  - Các công cụ hỗ trợ “dựa trên máy tính”

- **Để tạo mô hình tốt**

- Có tính nghệ thuật hơn là khoa học
  - ❖ Điểm chung của QLQTNV (BPM) và QLĐH (OP)
  - ❖ Khó khăn và dễ bị lỗi

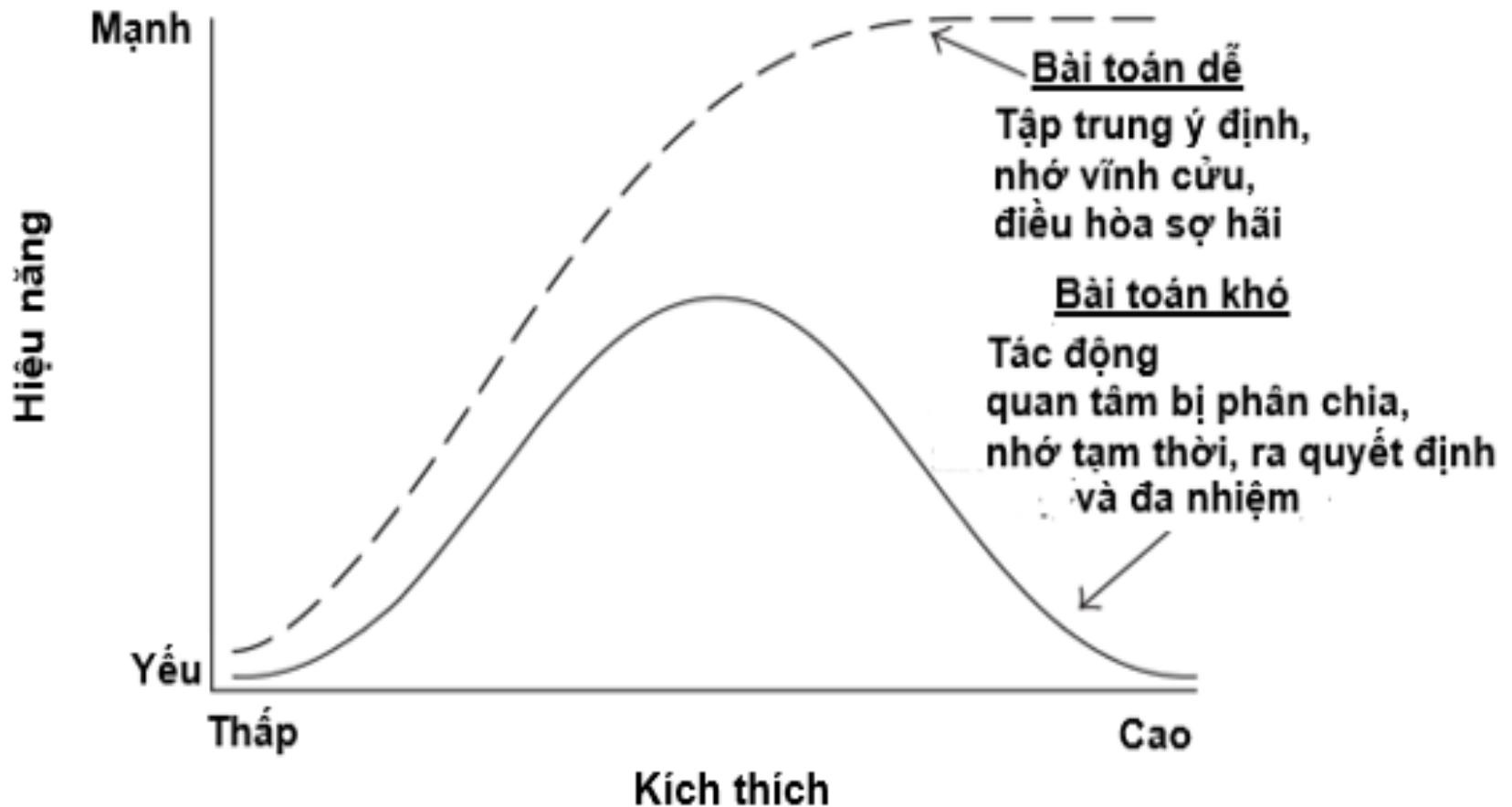


Quan niệm cổ điển về mô hình hóa: Trọng tâm là **ổn định trạng thái** và **mô hình được làm bằng tay**

# Lỗi tạo mô hình bằng tay

- Mô hình mô tả phiên bản lý tưởng hóa thực tại
  - Tập trung: “hành vi chuẩn” / “hành vi mong muốn”
  - “Mô hình chỉ cần bao gói 80% trường hợp giả định” (đơn giản hóa)
  - Không nhận thức được lệch lạc xảy ra: nhận thức có tính thiên vị
  - Mô hình thủ công có tính chủ quan: chủ yếu do mục đích “dễ hiểu”
- Không nắm bắt đầy đủ hành vi con người
  - Mô hình toán học đơn giản
    - ❖ có thể đủ mô hình hóa máy móc/con người trong dây chuyền lắp ráp;
    - ❖ Không đủ mô hình hóa sự tham gia con người với độ ưu tiên phức tạp trong quy trình phức tạp
  - Nhân viên trong quy trình phức
    - ❖ không thể phân bố quan tâm toàn bộ quy trình: mức độ đồng đều
    - ❖ không thể với năng suất không đổi: công việc khó mất nhiều thời gian và số giờ hiệu quả ít hơn (Quan sát trong hầu hết quá trình: Luật Yerkes–Dodson)

# Luật Yerkes–Dodson (Yerkes–Dodson law)



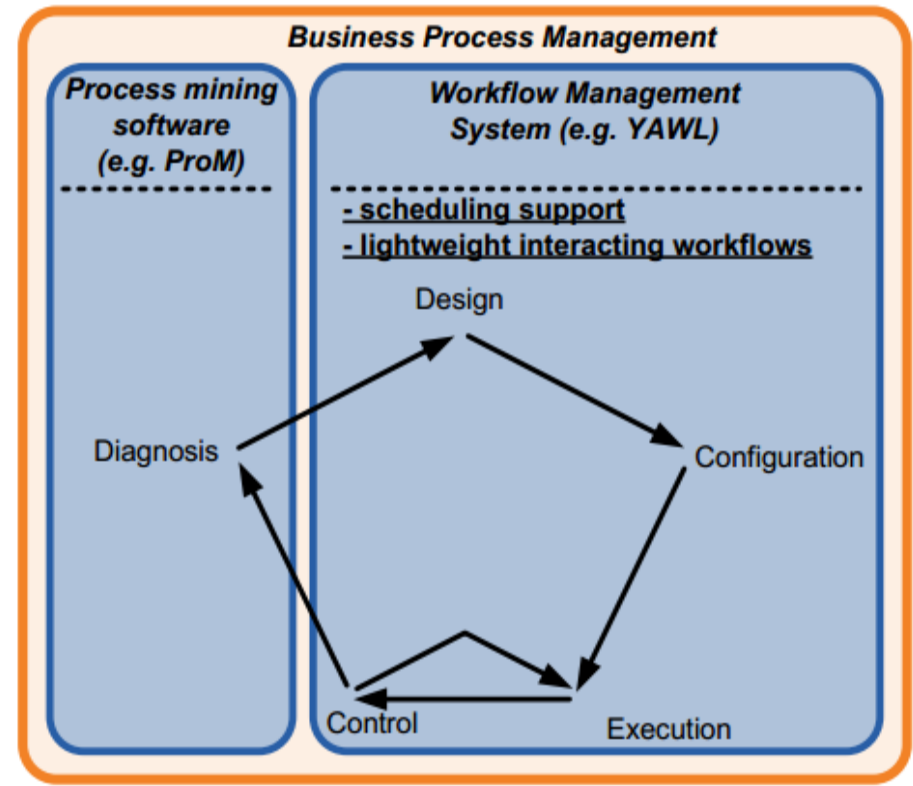
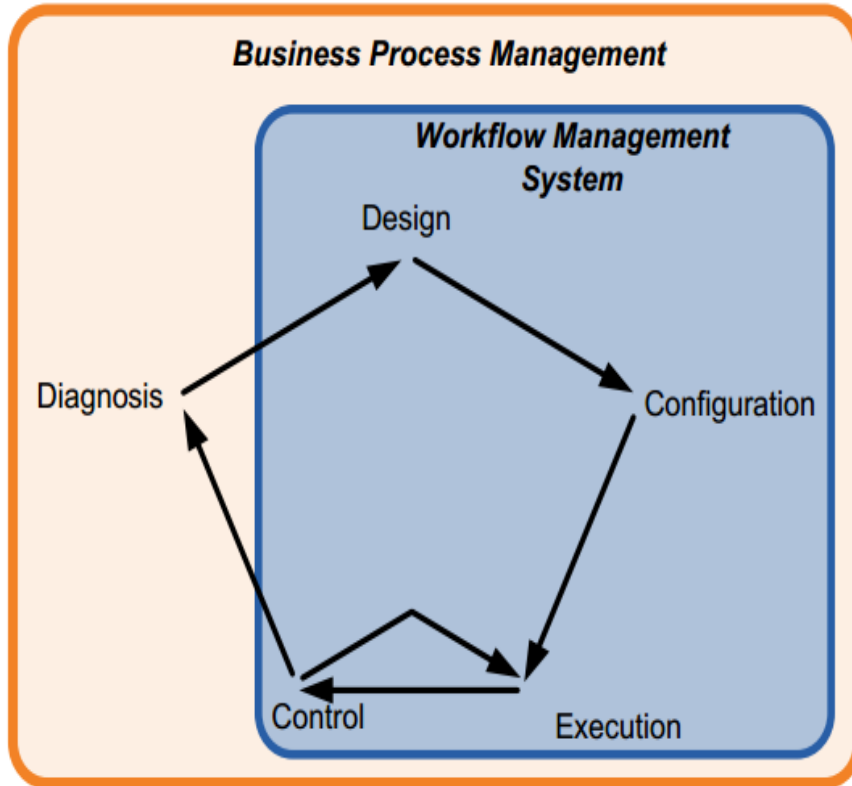
Luật Yerkes–Dodson: năng suất (*Performance*) tăng khi tăng kích thích (*Arousal*) về sinh lý hoặc tinh thần, nhưng chỉ đến một điểm nhất định. Khi mức độ của kích thích trở nên quá cao thì năng suất lại giảm.



# Lỗi tạo mô hình bằng tay

- **Lỗi phổ biến khi tạo mô hình bằng tay**
  - Mô hình ở một độ trừu tượng sai
    - ❖ Phụ thuộc vào: (i) dữ liệu vào; (ii) câu hỏi cần được trả lời
    - ❖ Quá trừu tượng: không trả lời được câu hỏi liên quan
    - ❖ Quá cụ thể: không đạt được câu trả lời/ quá phức tạp để hiểu
    - ❖ Ví dụ: Kho có hàng ngàn phụ tùng. Mô hình hóa mọi phụ tùng: (i) “hấp dẫn” mô phỏng so sánh các chính sách kiểm kê khác nhau; (ii) “không phù hợp” nếu không cần mô tả chi tiết phụ tùng cụ thể.
    - ❖ Khó khăn (i) khi thay đổi mức trừu tượng của một mô hình; (ii) Các câu hỏi xuất hiện lại ở mức cô đọng rất khác nhau.
  - Bàn luận
    - ❖ Còn có các lỗi khác khi tạo mô hình bằng tay
    - ❖ Mô hình có giá trị tiên đoán tốt và làm khởi đầu để thiết kế lại: Chỉ do chuyên gia giàu kinh nghiệm (nhà thiết kế/nhà phân tích)
- **Nhu cầu tự động hóa thiết kế, cải tiến mô hình nghiệp vụ khắc phục các hạn chế**
  - Thiết kế, cải tiến (thiết kế lại) mô hình quy trình
  - Khai phá quy trình: Tự động hóa

# Vai trò của khai phá quy trình



- Quản lý quy trình nghiệp vụ (**tự động hóa**): Vòng đời chờ (**pha chẩn đoán: *Diagnosis***) theo tiếp cận truyền thống (trái), vòng đời khép kín theo tiếp cận khai phá quy trình (phải)

# Điều kiện cho khai phá quy trình

## ● Dữ liệu

- Mô hình truyền thống: Thu thập bằng tay.
- Hiện tại: Quy trình NV được thi hành bằng các HTTT, ghi nhận dữ liệu vào các log file. Nhật ký sự kiện (event log).
  - ❖ Philips Healthcare giám sát thiết bị y tế do công ty sản xuất (ví dụ, máy X-quang và máy quét CT) trên toàn thế giới.

<http://www.medical.philips.com/>

- ❖ **ASML** có nhật ký lỗi HT in thạch bản khoảng 15.000 SK/máy/ngày.

<https://www.asml.com/asml/show.do?ctx=427>

- Các tổ chức chia sẻ các nhật ký sự kiện
- Big data

## ● Cộng đồng và công cụ

- Cộng đồng khai phá quy trình (Phần II)
- Công cụ khai phá quy trình: ProM

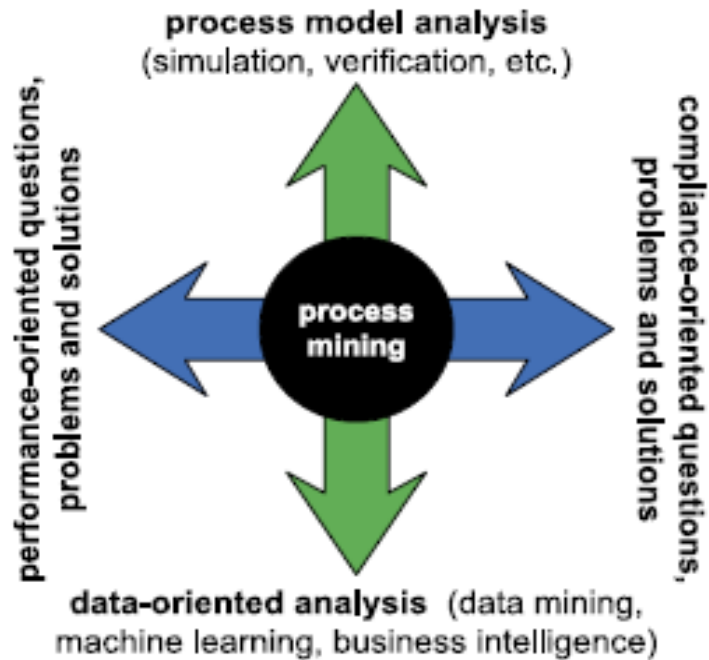
# Việt Nam với Khai phá quy trình

- **Nghị quyết số 19/NQ-CP ngày 18/03/2014**
  - Về những nhiệm vụ, giải pháp chủ yếu cải thiện môi trường kinh doanh, nâng cao năng lực cạnh tranh quốc gia
  - “rút ngắn quy trình xử lý”, “Cải cách quy trình”, “Đơn giản hóa quy trình” ...
- **Nghị quyết số 19/NQ-CP ngày 12/03/2015**
  - Về những nhiệm vụ, giải pháp chủ yếu tiếp tục cải thiện môi trường kinh doanh, nâng cao năng lực cạnh tranh quốc gia hai năm 2015 – 2016
  - *Đẩy mạnh ứng dụng công nghệ thông tin trong các ngành, lĩnh vực sản xuất kinh doanh và đời sống, coi đây là phương thức phát triển mới để **đổi mới quy trình quản lý**, sản xuất kinh doanh, nâng cao năng lực cạnh tranh của sản phẩm, doanh nghiệp, của quốc gia*
- **Nghiên cứu về khai phá quy trình tại Việt Nam**
  - Khung nhìn khoa học và công nghệ đối với các nghị quyết 19
  - Triển khai áp dụng khai phá quy trình trong tương lai

# Khai phá quy trình là gì và là một chủ đề thời sự ?

- Khái niệm khai phá quy trình
- Khai phá quy trình và khoa học dữ liệu
- Miền ứng dụng lợi thế
- Cộng đồng khai phá quy trình

# Khái niệm Khai phá quy trình

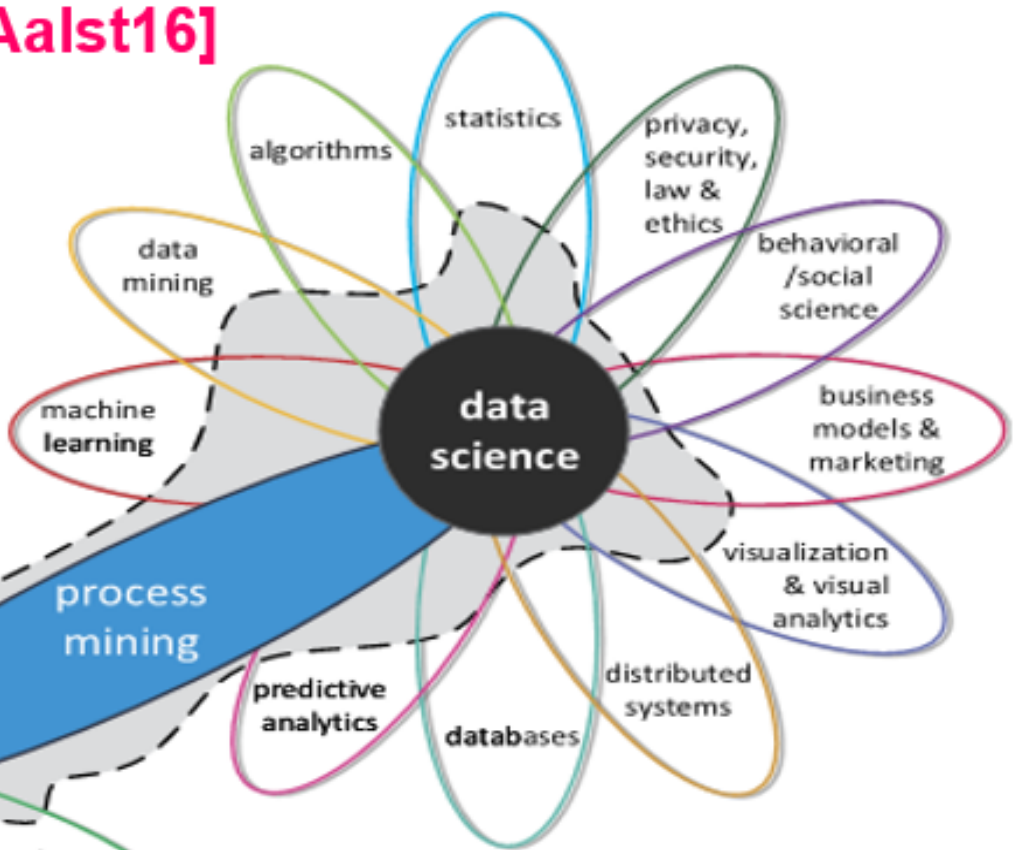
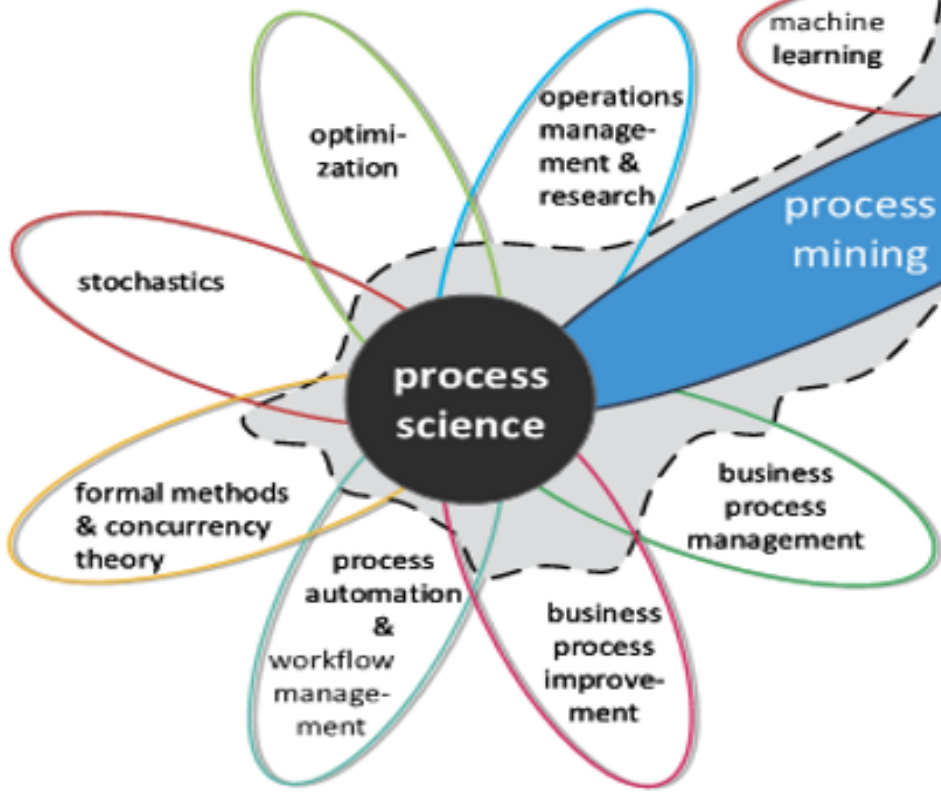


- <PT mô hình quy trình><KPQT> <PT dựa trên DL>
- PT dựa trên DL (KPD, học máy, TMKD): PT hướng đến quyết định & mẫu riêng lẻ, không toàn bộ QT (từ đầu đến cuối: E2E)
- PTMHQT (*Business Process Management: BPM* và *Workflow Management: WFM*) phân tích và tăng cường QT E2E sử dụng tri thức từ CNTT và khoa học quản lý lấy mô hình QT trung tâm
- KPQT tích hợp hai “vũ trụ” trả lời về **hiệu năng** và **tuân thủ**

# KP quy trình: cầu nối KHDL&KH quy trình

[Aalst16]

**Khoa học quy trình:** quản lý quy trình nghiệp vụ (QTNV), cải tiến QTNV, tự động quy trình và quản lý dòng công việc, phương pháp hình thức và lý thuyết tương tranh, ngẫu nhiên, tối ưu hóa, quản lý điều hành và vận trù học



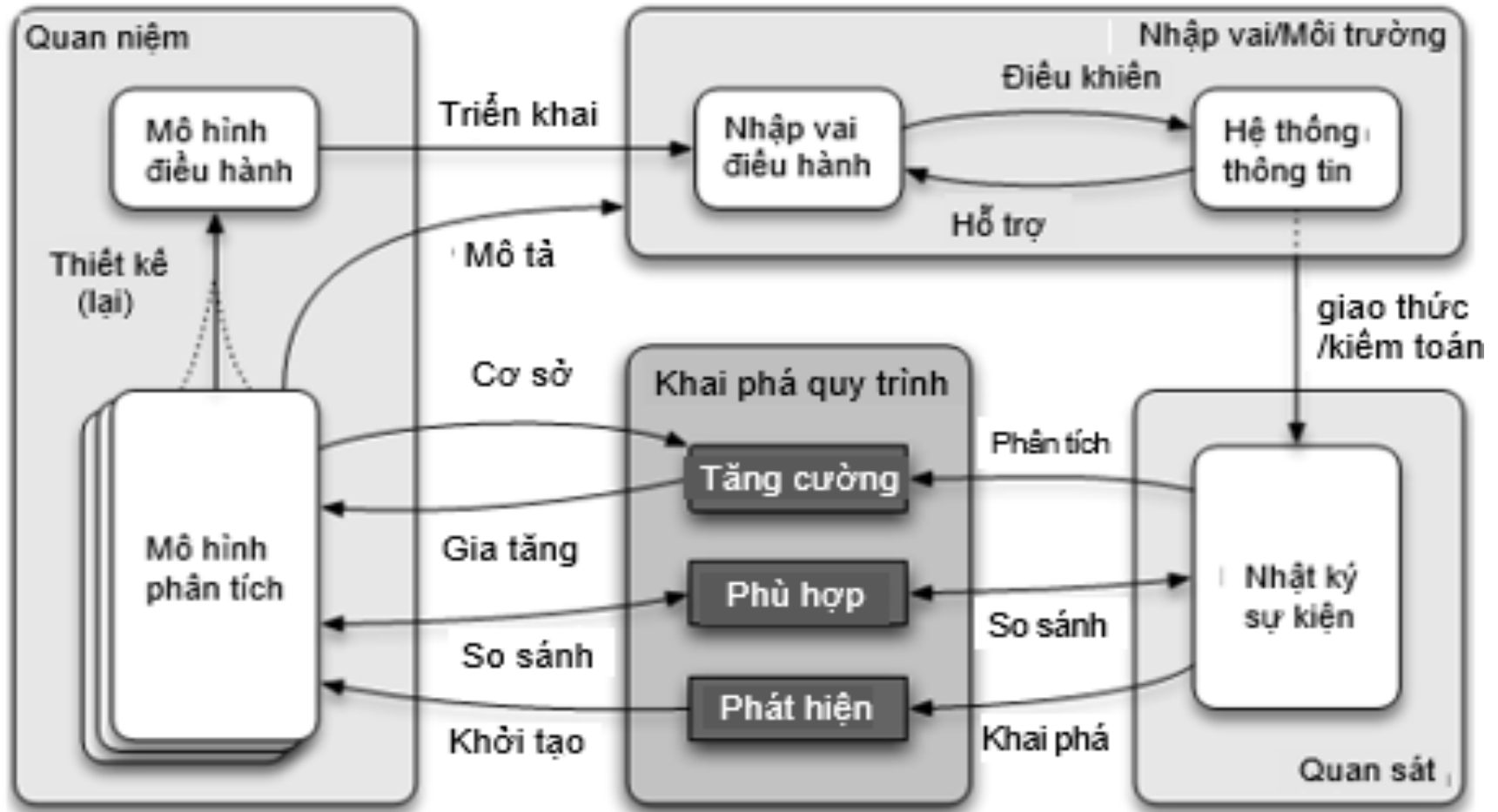
**Khoa học dữ liệu:** học máy, khai phá dữ liệu, thuật toán, thống kê, bảo mật-an ninh-luật và đạo đức, khoa học hành vi/xã hội, mô hình kinh doanh và tiếp thị, trực quan hóa và phân tích trực quan, hệ thống phân tán, phân tích dự báo

# Triết lý “đường mòn” [Aalst11]

- **Khái niệm khai phá quy trình**
  - Process Mining: chiết xuất thông tin có giá trị, liên quan đến quy trình từ nhật ký sự kiện, bổ sung tiếp cận hiện có để quản lý quy trình nghiệp vụ
  - Kết hợp giữa khai phá dữ liệu và quản lý quy trình nghiệp vụ
- **Triết lý “đường mòn” của khai phá quy trình**
  - Đường mòn (desire line) / đường xã hội (the social trail)
  - được hình thành từ xói mòn bởi bước chân người hoặc động vật
  - “Đường mòn là ngắn nhất/thuận tiện nhất giữa hai điểm
  - Xói mòn càng cao  $\Leftrightarrow$  Sử dụng càng thường xuyên
  - Hành động trong kinh doanh hình thành đường mòn kinh doanh ~ nhật ký sự kiện phản ánh quy trình tốt/thuận tiện

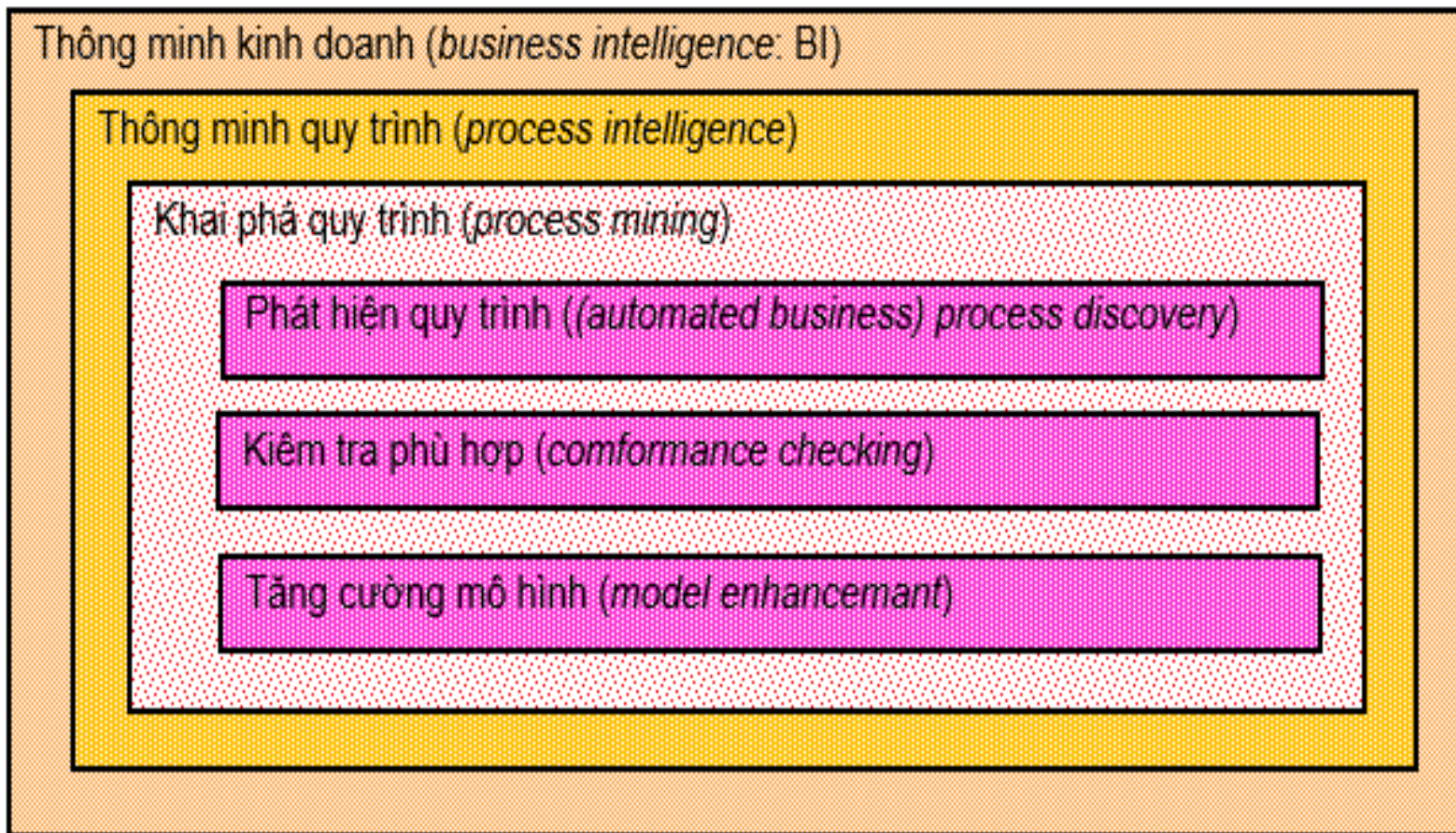


# Khai phá quy trình trong thi hành



- Khai phá quy trình khép kín tự động hóa vòng đời quản lý mô hình quy trình
  - Bộ công cụ tự động hóa

# Khai phá quy trình & thông minh kinh doanh



## ● Mỗi quan hệ

- Khai phá quy trình  $\subseteq$  Thông minh quy trình  $\subseteq$  Thông minh kinh doanh
- Khai phá quy trình: Phát hiện, Phù hợp, Tăng cường

# Miền ứng dụng điển hình

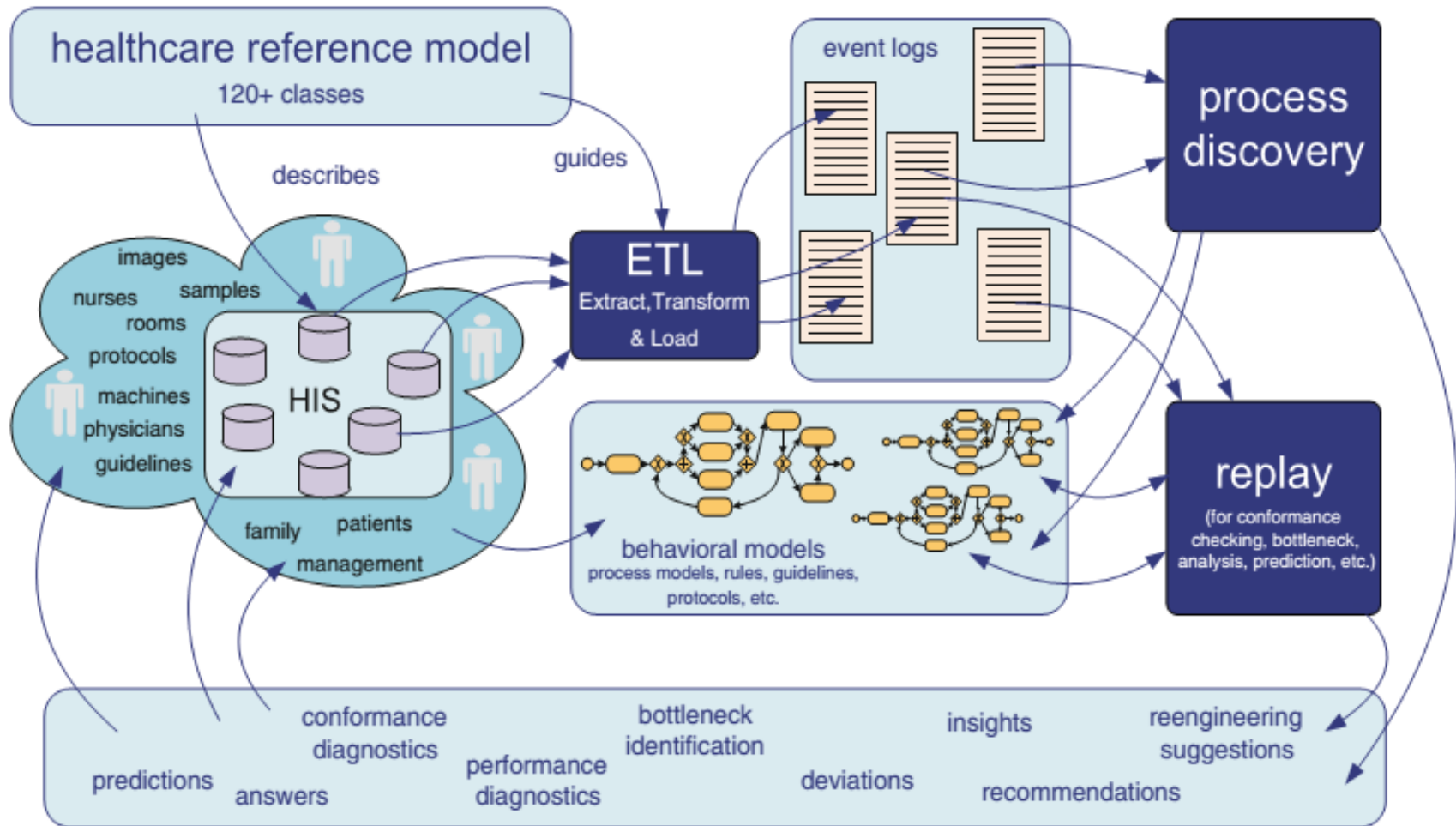
- **Đặc trưng**

- Miền ứng dụng: linh hoạt cao, cân đối hiệu năng và tuân thủ
- Chăm sóc sức khỏe, quản lý quan hệ khách hàng, viễn thông, sản xuất công nghiệp, v.v.
- Ví dụ: phân tích các quy trình điều trị tại các bệnh viện, cải tiến quy trình dịch vụ khách hàng của một công ty đa quốc gia, hiểu được hành vi duyệt web của khách hàng sử dụng trang web đặt vé, phân tích thất bại của một hệ thống xử lý hành lý, và cải thiện giao diện người dùng của một máy X-quang.

- **Miền ứng dụng lợi thế: Y tế**

- Chăm sóc sức khỏe [Mans15]: mục đích không là phân tích dữ liệu mà để cải thiện quy trình chăm sóc sức khỏe

# Khai phá quy trình trong y tế



[Mans15] Ronny S. Mans, Wil M. P. van der Aalst, Rob J. B. Vanwersch. *Process Mining in Healthcare: Evaluating & Exploiting Operational Healthcare Process*. Springer, 2015

# Cộng đồng khai phá quá trình

- Wil van der Aalst

- ĐHCN Eindhoven (Eindhoven Univ. of Technology (TU/e))
- Khởi xướng và duy trì phát triển khai phá quá trình
- <http://www.wis.win.tue.nl/~wvdaalst/>
- Master's thesis “Specificatie en Simulatie met behulp van ExSpect”  
*Đặc tả và mô phỏng bởi ExSpect*, 1988
- P h D t h e s i s “  
[Timed coloured Petri nets and their application to logistics](#)”, 1992
- Nhà khoa học Tin học ngoài Mỹ có chỉ số h-index cao thứ hai: 101  
(9/2013, 94: 5/2013) <http://www.cs.ucla.edu/~palsberg/h-number.html>

- Tổ chức nghề nghiệp về khai phá quá trình

- <http://www.processmining.org/>: phong phú:
  - các bài báo,
  - luận án Tiến sỹ  
[Http] <http://www.processmining.org/publications/phd>
  - bộ công cụ PRoM và các công cụ khác
  - các tập dữ liệu nhật ký sự kiện: một phần từ hơn 100 tổ chức
- Tuyên ngôn về khai phá quá trình
- [IEEE Task Force on Process Mining](#)



# H-index của WMP Van der Aalst

## The h Index for Computer Science

by [Jens Palsberg](#)

In [this paper](#), [J. E. Hirsch](#), Dept of Physics, UCSD, proposes "the index h, defined as the number of papers with citation number higher or equal to h, as a useful index to characterize the scientific output of a researcher."

Here is a partial list of computer science researchers who each has an h index of 40 or higher according to [Google Scholar](#). The list has about 1,000 entries and includes 1 Nobel Laureate, 34 Turing Award winners, 73 members of the National Academy of Engineering, 19 members of the National Academy of Sciences, 335 ACM Fellows, and 183 IEEE Fellows.

One can wonder how large a percentage of computer scientists are on the list. In May 2016, [DBLP](#) listed about 1.7 million authors of computer science papers. So, we can estimate that  $(1,000 / 1.7 \text{ million}) = 0.06$  percent of computer scientists are on the list.

Send comments, corrections, and new entries to [Jens Palsberg](#). I made the most recent update on Mar 19, 2016.

I do maintain the list: mostly, I add people and update numbers upon request and when I happen to notice a high h index. Some of the numbers on this page are the results of counting efforts by the listed people themselves. I have computed myself by comparing output from [Google Scholar](#) with other listings of research papers, such as from personal webpages or [DBLP](#). Whenever I notice that a person listed on this page has an entry on [Google Scholar Citations](#), I create a link.

Several websites provide easy-to-use interfaces for computing an h index, including [scHolar index](#) and [Harzing's publish and perish](#).

[161 Herbert A. Simon](#) (CMU), [Nobel Laureate](#), [Turing Award](#), [ACM Fellow](#)

[154 Anil K. Jain](#) (Michigan State U), [ACM Fellow](#), [IEEE Fellow](#), [IEEE Technical Achievement Award](#), [Member of the National Academy of Engineering](#)

[132 Terrence Sejnowski](#) (UCSD), [IEEE Fellow](#), [Member of the National Academy of Engineering](#), [Member of the National Academy of Sciences](#), [Member of the Institute of Medicine](#), [IEEE Frank Rosenblatt Award](#)

[131 Jiawei Han](#) (UTUC), [ACM Fellow](#), [IEEE Technical Achievement Award](#)

[126 Takeo Kanade](#) (CMU), [ACM Fellow](#), [IEEE Fellow](#), [Member of the National Academy of Engineering](#)

[126 Scott Shenker](#) (Berkeley), [ACM Fellow](#), [IEEE Fellow](#), [Member of the National Academy of Engineering](#)

[124 David Haussler](#) (UC Santa Cruz), [ISCB Fellow](#)

[124 Philip S. Yu](#) (UC), [ACM Fellow](#), [IEEE Fellow](#), [IEEE Technical Achievement Award](#)

[118 Hector Garcia-Molina](#) (Stanford), [ACM Fellow](#), [Member of the National Academy of Engineering](#)

[118 Geoffrey E. Hinton](#) (U Toronto), [Fellow of the Royal Society](#), [Member of the American Academy of Arts & Sciences](#), [Fellow of the Royal Society](#), [FRSC](#), [IEEE Frank Rosenblatt Award](#)

[114 Jack Dongarra](#) (U Tennessee), [ACM Fellow](#), [IEEE Fellow](#), [Member of the National Academy of Engineering](#), [SIAM Fellow](#), [AAAS Fellow](#)

[113 Don Towsley](#) (U Mass, Amherst), [ACM Fellow](#), [IEEE Fellow](#)

[112 Wil van der Aalst](#) (TU Eindhoven), [Member of Academy of Europe \(Academia Europaea\)](#), [Member of the Royal Netherlands Academy of Arts and Sciences](#), [Member of the Royal Holland Society of Sciences and Humanities](#)

[112 Steven Salzberg](#) (Johns Hopkins U), [AAAS Fellow](#), [ISCB Fellow](#)

[111 Tomaso Poggio](#) (MIT)

[110 Deborah Estrin](#) (Cornell NYC Tech), [ACM Fellow](#), [IEEE Fellow](#), [Member of the National Academy of Engineering](#)

[110 Sebastian Thrun](#) (Stanford), [Member of the National Academy of Engineering](#)

[109 Ian Foster](#) (Argonne National Laboratory & U Chicago), [ACM Fellow](#), [AAAS Fellow](#), [BCS Fellow](#)

[108 Stephen Grossberg](#) (Boston University)

[108 Michael I. Jordan](#) (Berkeley), [ACM Fellow](#), [IEEE Fellow](#), [AAAS Fellow](#), [AAAI Fellow](#), [Member of the National Academy of Engineering](#), [Member of the National Academy of Sciences](#), [Member of the American Academy of Arts & Sciences](#), [Fellow](#)

July 30, 2016

## IEEE CIS Task Force on Process Mining

Trace: • [minutes\\_bpm\\_2015](#)

[Login](#)

search



### Index


- Home
- Members
  - Organizations
  - People
  - Countries
- News
- Events
- Process Mining Manifesto
- Process Mining Movies
- Process Mining Event Logs
- Process Mining Case Studies
- Process Mining Dissertation Award
- XES Working Group
  - Participants
  - Documents
- Contact

### Minutes of the Annual Meeting IEEE Task Force on Process Mining at BPM 2015 in Eindhoven

Monday August 31th, 2015, 16.00-18.00

University of Innsbruck, Universitätsstraße 15, 6020 Innsbruck, Austria

About 40 people joined the annual meeting. Moreover, 60 people joined the process mining reception organized by the taskforce on Sunday.

 [Slides of the meeting](#)

*Please check the action points. Being part of this task force requires active participation!*

#### 1) Welcome and Overview of the IEEE Task Force on Process Mining

Short summary of the goals for the task force.

#### 2) Overview of Activities 2014-2015 (Wil van der Aalst)

See the slides for the many activities that took place. Some highlights:

- another successful BPI workshop with record numbers of participants (65) and submissions,
- the BPI challenge providing datasets widely used by PM researchers and also used in education/training,
- the process mining MOOC with 68.000 registered participants,
- the LinkedIn group (number of members more than doubled over the last year),
- the 14th translation of the PM manifesto in Thai

## IEEE CIS Task Force on Process Mining

Trace » [best\\_process\\_mining\\_dissertation\\_award\\_2015](#)

Show pagesource

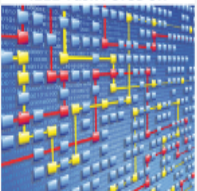
Login

search

### Index

- Home
- Members
  - Organizations
  - People
  - Countries
- News
- Events
- Process Mining Manifesto
- Process Mining Movies
- Process Mining Event Logs
- Process Mining Case Studies
- Process Mining Dissertation Award
  - Participants
  - Documents
- Contact

### IEEE Task Force on



### Best Process Mining Dissertation Award 2015

The **Best Process Mining Dissertation Award** is awarded by the IEEE Task Force on Process Mining to an outstanding PhD thesis focused on the area of business process intelligence. The award is particularly dedicated to works contributing to research in the area of process mining and/or the innovative use of process mining techniques for solving practically relevant problems.

With this award, the IEEE Task Force on Process Mining wants to draw attention to excellent works by young researchers and promote the research area as a whole.

### Winner 2015 - Jorge Munoz-Gama - Conformance Checking and Diagnosis in Process Mining

The winner of the 2014 Best Process Mining Dissertation Award is Jorge Munoz-Gama for his PhD thesis entitled "Conformance Checking and Diagnosis in Process Mining".

### Call for Nominations 2015

Below you find the call for nominations for the year 2015, which you can also [download as .pdf](#).

The Best Process Mining Dissertation Award 2015 will be conferred by the IEEE Task Force on Process Mining to the winner during the meeting of the IEEE Task Force on Process Mining at the [13th Int. Conference on Business Process Management](#) in Innsbruck, Austria on 31st August 2015. As part of this event, the recipient will be invited to give a presentation on the main results of the thesis in a form suitable for the event and audience.

The selected thesis will also be recommended for publication as a monograph in the LNBP series published by Springer. Further, the Dissertation Award will be accompanied by a monetary prize of 1,000 EUR.

### Which theses can be nominated?

- The call for nominations refers to PhD theses which have been successfully defended (and the PhD title awarded) in the period **between 1 Jan 2013 and 31 Dec 2014**. Nominations for theses that fall outside this period will not be taken into consideration.
- The thesis must be written in **English**.

### How to submit a nomination?

1. A PhD thesis is nominated for the award by their supervisors. A nomination should consist of an electronic version of the thesis itself, accompanied by a brief letter (1 or 2 pages) from the supervisor describing why the thesis should be considered for the award.
2. Nominations should highlight the quality and impact of the work. The nominator may include a pointer to research work influenced by the results of the nominated thesis, and citations to the thesis or derived works, where available.
3. Submissions should be sent to the Award Organizers at the following email addresses: Dirk Fahland, Antonella Guzzo, Marcello La Rossa.



[http://www.win.tue.nl/ieeetfpm/doku.php?id=shared:best\\_process\\_mining\\_dissertation\\_award\\_2015](http://www.win.tue.nl/ieeetfpm/doku.php?id=shared:best_process_mining_dissertation_award_2015)



## Business Process Management

> Home > Conferences and Workshops









modern Trier 1

### Business Process Management 2015: Innsbruck, Austria

-     Hamid Reza Motahari-Nezhad, Jan Recker, Matthias Weidlich:  
**Business Process Management - 13th International Conference, BPM 2015, Innsbruck, Austria, August 31 - September 3, 2015, Proceedings.** Lecture Notes in Computer Science 9253, Springer 2015, ISBN 978-3-319-23062-7 [contents]
-     Jan Mendling, Jan vom Brocke:  
**Proceedings of the Industry Track at the 13th International Conference on Business Process Management 2015 co-located with 13th International Conference on Business Process Management (BPM 2015), Innsbruck, Austria, September 2015.** CEUR Workshop Proceedings 1439, CEUR-WS.org 2015 [contents]
-     Florian Daniel, Stefan Zugal:  
**Proceedings of the BPM Demo Session 2015 Co-located with the 13th International Conference on Business Process Management (BPM 2015), Innsbruck, Austria, September 2, 2015.** CEUR Workshop Proceedings 1418, CEUR-WS.org 2015 [contents]

### Business Process Management 2014: Eindhoven, The Netherlands (Haifa, Israel)

Due to the unstable situation in southern Israel, BPM 2014 was relocated from Haifa to Eindhoven, The Netherlands

-     Shazia Wasim Sadiq, Pnina Soffer, Hagen Völzer:  
**Business Process Management - 12th International Conference, BPM 2014, Haifa, Israel, September 7-11, 2014. Proceedings.** Lecture Notes in Computer Science 8659, Springer 2014, ISBN 978-3-319-10171-2 [contents]
-     Fabiana Fournier, Jan Mendling:  
**Business Process Management Workshops - BPM 2014 International Workshops, Eindhoven, The Netherlands, September 7-8, 2014, Revised Papers.** Lecture Notes in Business Information Processing 202, Springer 2015, ISBN 978-3-319-15894-5 [contents]

## ● Giới thiệu

- Thường niên: 2000, 2003-2015
- Một số hội nghị liên quan khác

# Tài nguyên, công cụ và tầm ảnh hưởng

## ProM

IEEE Task Force on  
Process Mining

- ProM Software: [prom.sourceforge.net](http://prom.sourceforge.net)
- Process mining: [www.processmining.org](http://www.processmining.org)
- ProM 5 series nightly builds: [prom.win.tue.nl/tools/prom/nightly5/](http://prom.win.tue.nl/tools/prom/nightly5/)
- ProM 6 series nightly builds: [prom.win.tue.nl/tools/prom/nightly/](http://prom.win.tue.nl/tools/prom/nightly/)
- Converting logs (MXML-based) [promimport.sourceforge.net](http://promimport.sourceforge.net)
- XES: [www.xes-standard.org](http://www.xes-standard.org) and [www.openxes.org](http://www.openxes.org)
- Papers et al.: [vdaalst.com](http://vdaalst.com)
- IEEE Task Force on Process Mining: [www.win.tue.nl/ieeetfpm/](http://www.win.tue.nl/ieeetfpm/)



## Organizations Supporting our Task Force (1/3)



# Tầm ảnh hưởng



- **Độ phủ quốc gia**

- Châu Âu, châu Úc, châu Á (Nhật Bản ?)
- Châu Mỹ ?



# Công ty Fluxicon

## For process owners.

You are responsible for a business process. We can help you to:

- **Reduce cost** and variation, become more lean.
- **Be in control** and know what is going on — despite outsourcing.
- **Improve quality** by comparing processes beyond KPIs.

[Find out more...](#)

## For consultants.

You help your clients excel. With our technology you can:

- **Speed up** the process discovery phase.
- **Objective data** helps to support your recommendation.
- **Improve customer satisfaction** by doing more, in the same time.

[Find out more...](#)

Sure thing. We are Anne and Christian, and we met while studying software engineering at HPI Potsdam, Germany, where we first encountered process mining in 2002. We both went on to pursue our PhDs in the process mining group of Wil van der Aalst at the TU Eindhoven in the Netherlands. Our passion is to tell the world about process mining and its benefits, and to create well-designed, scientifically accurate, and high performance process mining software. This is why we founded Fluxicon in 2009.

<http://www.fluxicon.com/>. Doanh số ?

**Christian Günther và Anne Rozinat**



# Công ty Futura Process Intelligence

[PRLog Home](#) | [Business Index](#) | [Pressroom Index](#) | [Press Releases Home](#)



## Futura Process Intelligence

[About](#) | [Contact](#)

### About

[Follow on PRLog](#)

[Share This Page](#)

Futura Process Intelligence is an ICT-company focused on developing and offering products and services in the domain of Process Intelligence. Since Process intelligence is a relatively new area of expertise, there is still a lot of ambiguity about it.

[Website](#)

We see Process Intelligence positioned between the domains of BPM and BI. Because of the focus on processes and data/information about them it's definitely part of the BPM area. It's also part of the BI-arena, because we're using process- and data-mining techniques and software for further analysis to improve the processes and businesses of our customers. We also cover and fill in one quadrant of the Balanced score card (the quadrant "internal processes") with our process monitoring and compliance solutions.

Category(s):

[bpm](#)

[abd](#)

[reflect](#)

[futura](#)

[process](#)

[intelligence](#)

[saas](#)

[Software](#)

[efficiency](#)

The purpose of Process Intelligence is to unlock the hidden knowledge about processes using historical data and help customers to improve productivity, quality and profitability by making this knowledge explicit and more comprehensible so they can use it effectively.

At the moment we are focusing on our software product Reflect, which is a SaaS application with a business model based on pay-per-use and subscriptions. Reflect is a tool suite for process mining, analysis and charting, making process models and animations, etc. In Reflect we're using genetic algorithms, which have been researched and published at the Technical University of Eindhoven. We have a strong focus on staying ahead in the field, innovating and translating the outcomes of research into solutions that can help solve and improve real life challenges.

<http://biz.prlog.org/Futura/>. Doanh số ?

Peter van den Brand



# Các bài toán Khai phá quy trình cơ bản

- Phát hiện quy trình
- Kiểm tra phù hợp
- Tăng cường quy trình
- Các bài toán liên quan

# Phát hiện quy trình

Điểm xuất phát là một nhật ký sự kiện. Mỗi sự kiện trở lại một thể hiện (trường hợp) quy trình và một hành động. Các sự kiện có thứ tự và có các thuộc tính bổ sung (xem thời gian hoặc dữ liệu nguồn v.v.).

Nhật ký sự kiện được đọc để phát hiện vai trò trong tổ chức (chẳng hạn, một nhóm người có mẫu hành động tương tự nhau). Các vai trò này đóng để sự liên quan giữa người và hành động

case id	event id	timestamp	action	role	score	...
1	33654423	30-12-2010:11:02	register request	Pete	50	...
	33654424	31-12-2010:10:06	examine thoroughly	Sara	400	...
	33654425	05-01-2011:15:12	check ticket	Mike	100	...
	33654426	06-01-2011:11:18	decide	Sara	200	...
33654427	07-01-2011:14:24	reject request	Pete	200	...	
2	33654483	30-12-2010:11:32	register request	Mike	50	...
	33654485	30-12-2010:12:12	check ticket	Mike	100	...
	33654487	30-12-2010:14:16	examine casually	Pete	400	...
	33654488	05-01-2011:11:22	decide	Sara	200	...
33654489	08-01-2011:12:05	pay compensation	Ellen	200	...	
3	33654521	30-12-2010:14:32	register request	Pete	50	...
	33654522	30-12-2010:15:06	examine casually	Mike	400	...
	33654524	30-12-2010:16:34	check ticket	Ellen	100	...
	33654525	06-01-2011:09:18	decide	Sara	200	...
	33654526	06-01-2011:12:18	reinitiate request	Sara	200	...
	33654527	06-01-2011:13:06	examine thoroughly	Sara	400	...
33654530	08-01-2011:11:43	check ticket	Pete	100	...	
33654531	09-01-2011:09:55	decide	Sara	200	...	
33654533	15-01-2011:10:25	pay compensation	Ellen	200	...	
4	33654641	06-01-2011:15:02	register request	Pete	50	...
	33654643	07-01-2011:12:06	check ticket	Mike	100	...
	33654644	08-01-2011:14:43	examine thoroughly	Sara	400	...
	33654645	09-01-2011:12:02	decide	Sara	200	...
33654647	12-01-2011:15:44	reject request	Ellen	200	...	
6	33654871	06-01-2011:15:02	register request	Mike	50	...
	33654873	06-01-2011:16:06	examine casually	Ellen	400	...
	33654874	07-01-2011:19:22	check ticket	Mike	100	...
	33654875	07-01-2011:16:52	decide	Sara	200	...
33654877	16-01-2011:11:47	pay compensation	Mike	200	...	



Kỹ thuật phát hiện tri thức được dùng để tìm mô hình đồng điều khiển (trong trường hợp mô hình mang quyền lý quy trình định doanh) mà mô tả hành vi tuân thủ tốt nhất

Thông tin hiệu năng (như thời gian trung bình giữa hai hành động liên tiếp nhau) được trích xuất từ nhật ký sự kiện và trực quan hóa lên định mô hình.

Luật quyết định (như cây quyết định đơn trên dữ liệu đã biết tại thời điểm này) được học từ nhật ký sự kiện và được dùng để chỉ giải quyết định.

- Phát hiện quy trình là bài toán điển hình nhất
- Phát biểu bài toán
  - Input: Nhật ký sự kiện
  - Output: Mô hình quy trình với thông tin bổ sung để phân tích

# Phát hiện quy trình và biên dịch ngược

## ● Biên dịch ngược

- Decompiling/Reverse Engineering
- Phiên bản chương trình đích → phiên bản chương trình nguồn
- Bài toán khó. Mất thông tin khi biên dịch ngược.
- Một số tài liệu liên quan

## ● Phát hiện quy trình

- Phiên bản mô hình quy trình trong thi hành → mô hình quy trình (ngôn ngữ mô hình hóa)
- “Lập lại sơ đồ khối” từ một chương trình đích
- Đầu vào: không là chương trình đích, là tập các trường hợp thi hành chương trình
- “tên lệnh” và giá trị các thanh ghi ?
- thi hành “một lệnh” không là một máy mà có thể nhiều người

[Wang10] W. Wang. *Reverse engineering: technology of reinvention*. CRC, 2010

[Kadavy11] David Kadavy. *Design for Hackers: Reverse Engineering Beauty*. Wiley, 2011

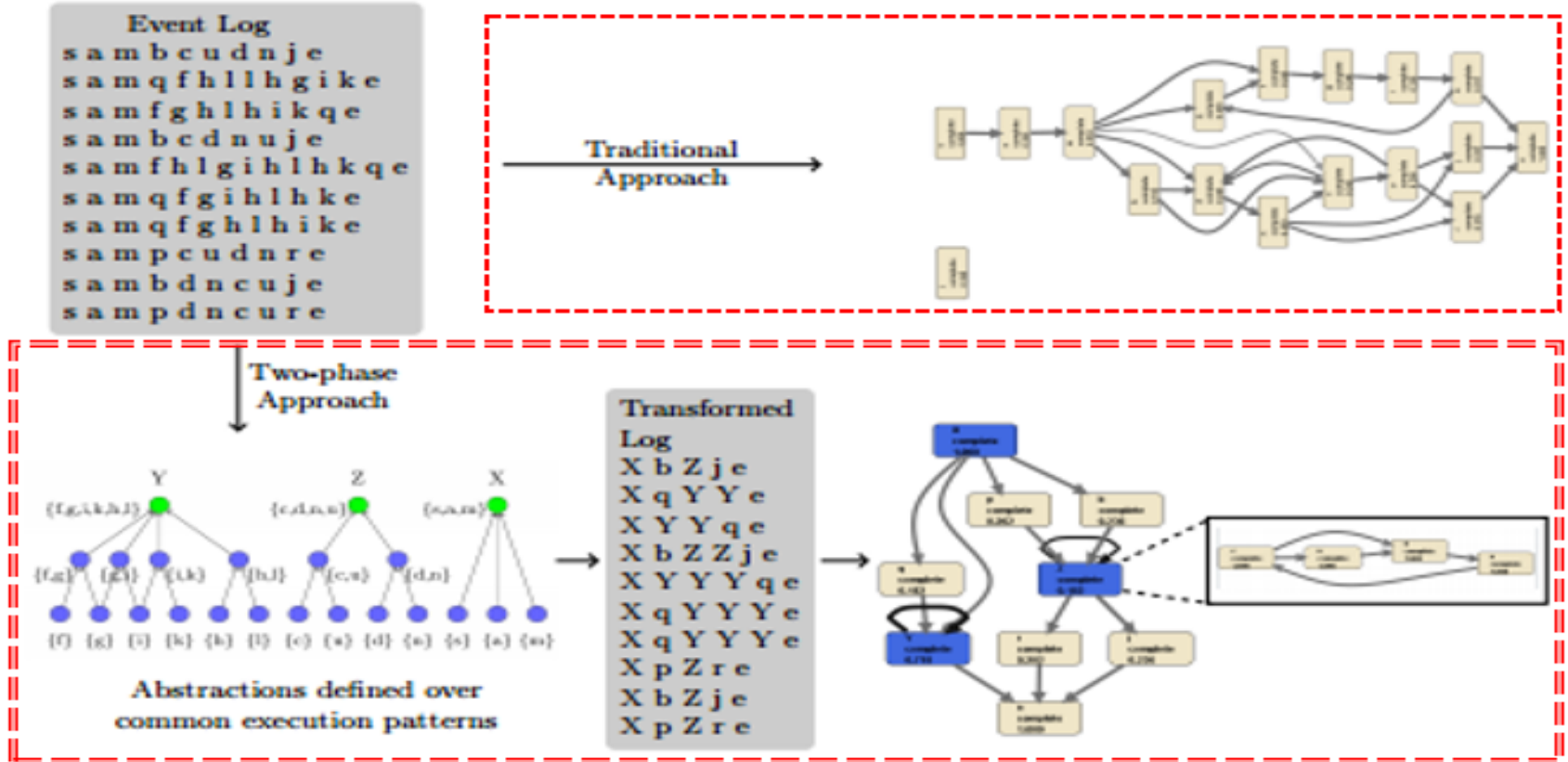
[Nolan04] Godfrey Nolan. *Decompiling Java*. Apress, 2004

[Nolan12] Godfrey Nolan. *Decompiling Android*. Apress, 2012

[Kalinovsky04] Alex Kalinovsky. *Covert Java: Techniques for Decompiling, Patching, and Reverse Engineering*. Sams, 2004



# Tiếp cận hai giai đoạn

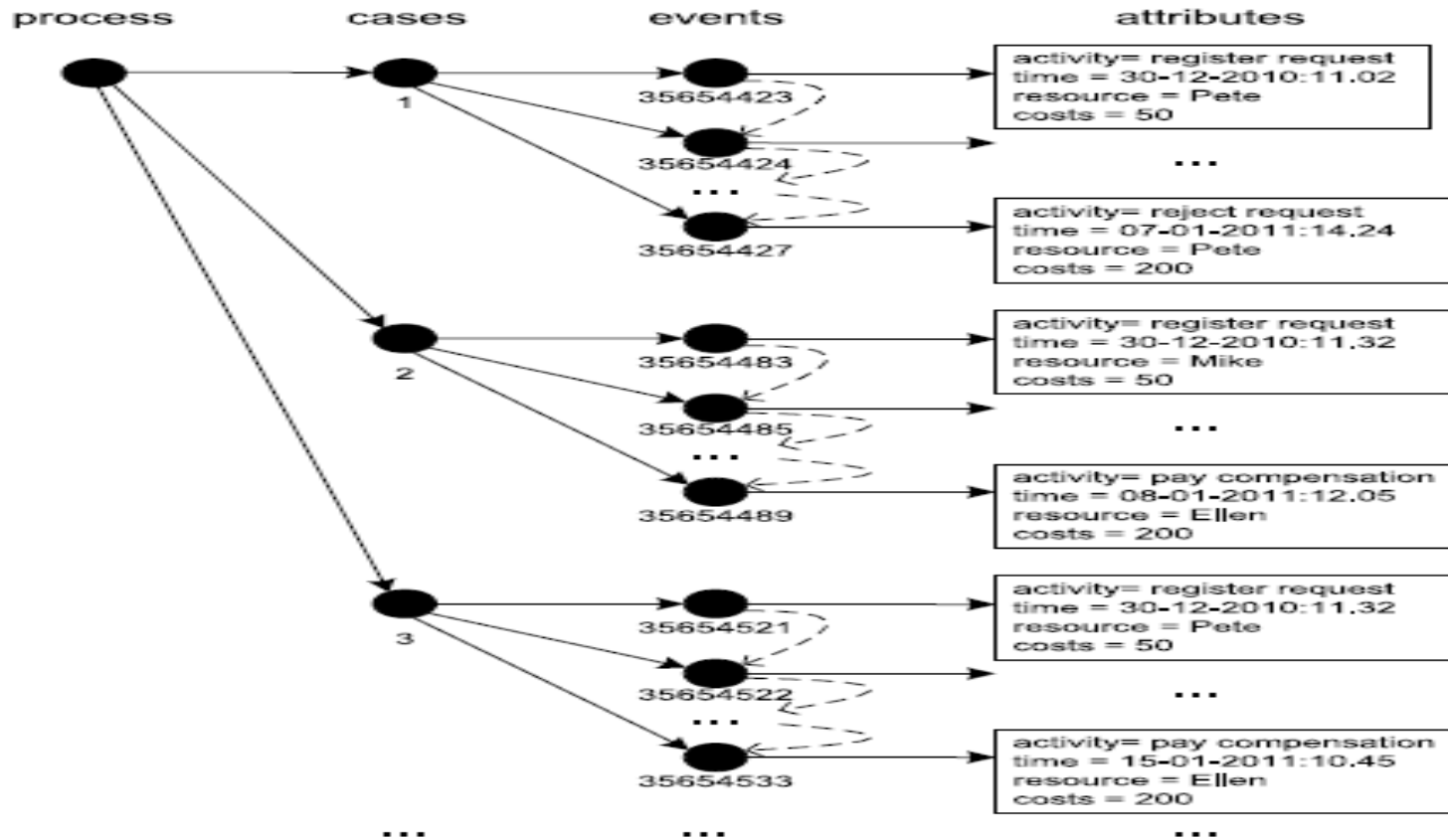


- **Mô hình hai giai đoạn+ K/phá khía cạnh bổ sung**

- Tiền xử lý → phát hiện mô hình → khía cạnh bổ sung
- Mô hình hai giai đoạn: Tiền xử lý → phát hiện mô hình
- Trình bày: Phát hiện mô hình → Tiền xử lý → Khía cạnh bổ sung

[Bose12] R. P. Jagadeesh Chandra Bose. *Process Mining in the Large: Preprocessing, Discovery, and Diagnostics*. PhD Thesis, Eindhoven University of Technology, The Netherlands, 2012

# Nhật ký sự kiện



## ● Nhật ký sự kiện

- Một nhật ký sự kiện (quy trình) gồm nhiều trường hợp (case)
- Một trường hợp: một lần thực hiện quy trình. Một trường hợp gồm nhiều sự kiện (event)
- Một sự kiện thực hiện một “hoạt động” và các thuộc tính

# Nhật ký sự kiện

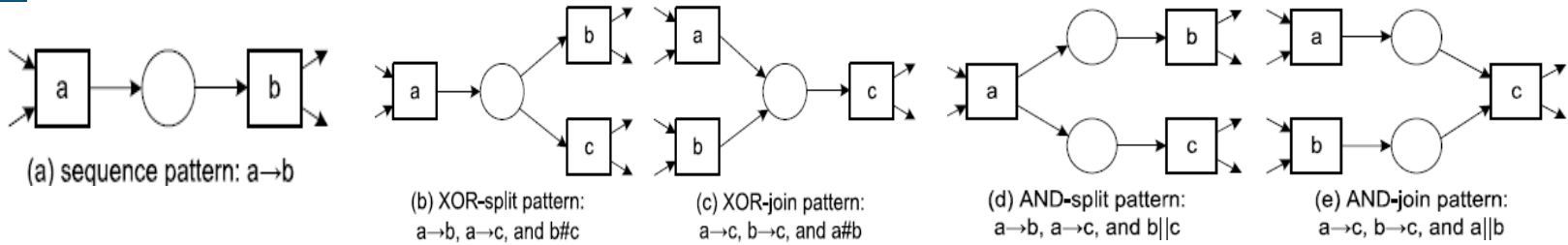
Case id	Event id	Properties				
		Timestamp	Activity	Resource	Cost	...
1	35654423	30-12-2010:11.02	Register request	Pete	50	...
	35654424	31-12-2010:10.06	Examine thoroughly	Sue	400	...
	35654425	05-01-2011:15.12	Check ticket	Mike	100	...
	35654426	06-01-2011:11.18	Decide	Sara	200	...
	35654427	07-01-2011:14.24	Reject request	Pete	200	...
2	35654483	30-12-2010:11.32	Register request	Mike	50	...
	35654485	30-12-2010:12.12	Check ticket	Mike	100	...
	35654487	30-12-2010:14.16	Examine casually	Pete	400	...
	35654488	05-01-2011:11.22	Decide	Sara	200	...
	35654489	08-01-2011:12.05	Pay compensation	Ellen	200	...
3	35654521	30-12-2010:14.32	Register request	Pete	50	...
	35654522	30-12-2010:15.06	Examine casually	Mike	400	...
	35654524	30-12-2010:16.34	Check ticket	Ellen	100	...
	35654525	06-01-2011:09.18	Decide	Sara	200	...
	35654526	06-01-2011:12.18	Reinitiate request	Sara	200	...
	35654527	06-01-2011:13.06	Examine thoroughly	Sean	400	...
	35654530	08-01-2011:11.43	Check ticket	Pete	100	...
	35654531	09-01-2011:09.55	Decide	Sara	200	...
	35654533	15-01-2011:10.45	Pay compensation	Ellen	200	...
4	35654641	06-01-2011:15.02	Register request	Pete	50	...
	35654643	07-01-2011:12.06	Check ticket	Mike	100	...
	35654644	08-01-2011:14.43	Examine thoroughly	Sean	400	...
	35654645	09-01-2011:12.02	Decide	Sara	200	...
	35654647	12-01-2011:15.44	Reject request	Ellen	200	...
...	...	...	...	...	...	

Case id	Trace
1	{a, b, d, e, h}
2	{a, d, c, e, g}
3	{a, c, d, e, f, b, d, e, g}
4	{a, d, b, e, h}
...	...

## ● Nhật ký sự kiện

- Dạng cô đọng (phải): phát hiện mô hình quy trình
- Dạng đầy đủ (trái): nguồn dạng cô đọng, khai phá khía cạnh bổ sung

# Thuật toán $\alpha$ : Các quan hệ NKSK



- Quan hệ đi trước  $>_L$ :  $\forall a, b \in T$  nói  $a >_L b$  nếu  $\exists \sigma \in L$ :  $a$  đi ngay trước  $b$  trong  $\sigma$ :  $\sigma_i = a \wedge \sigma_{i+1} = b$ . Nói  $b \succ a$ :  $\forall \sigma \in L$ :  $b$  không đi ngay trước  $a$ .
- Quan hệ nhân quả  $\rightarrow_L$ :  $a \rightarrow_L b \leftrightarrow a >_L b \wedge b \not\succeq a$ . Khi đó nói  $b \leftarrow_L a$ . Quan hệ  $\rightarrow_L$  (và  $\leftarrow_L$ ) là không đối xứng, hay quan hệ  $\leftarrow_L$  là đối ngẫu của quan hệ  $\rightarrow_L$ .
- Quan hệ đồng thời  $\parallel_L$ :  $a \parallel_L b \leftrightarrow a >_L b \wedge b >_L a$ . Quan hệ  $\parallel_L$  là một quan hệ đối xứng.
- Quan hệ không liên kề:  $\#_L$ :  $a \#_L b \leftrightarrow a \not>_L b \wedge b \not>_L a$ . Quan hệ  $\#_L$  là một quan hệ đối xứng. Liên hệ trực quan tới lưới Petri biểu diễn mô hình quy trình, khi  $a \#_L b$  thì tập mọi nút liên quan  $a$  và tập mọi nút liên  $b$  sẽ có giao là tập rỗng.

## • Nhật ký sự kiện ví dụ

- $L_1 = [\langle a, b, c, d \rangle^3, \langle a, c, b, d \rangle^2, \langle a, e, d \rangle]$

	a	b	c	d	e
a	$\#_{L_1}$	$\rightarrow_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$
b	$\leftarrow_{L_1}$	$\#_{L_1}$	$\parallel_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
c	$\leftarrow_{L_1}$	$\parallel_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
d	$\#_{L_1}$	$\leftarrow_{L_1}$	$\leftarrow_{L_1}$	$\#_{L_1}$	$\leftarrow_{L_1}$
e	$\leftarrow_{L_1}$	$\#_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$

# Thuật toán $\alpha$

- $L_1 = [\langle a, b, c, d \rangle^3, \langle a, c, b, d \rangle^2, \langle a, e, d \rangle]$

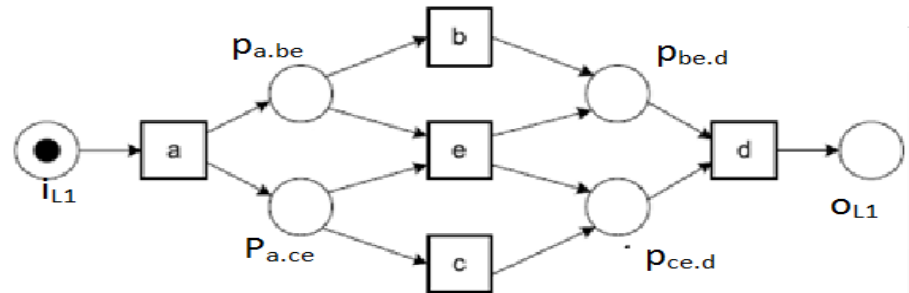
## Thuật toán $\alpha$

Input: Nhật ký sự kiện L ở dạng đơn giản

Output: Mô hình quy trình  $\alpha(L) = (P_L, T_L, F_L)$  là một lưới dòng công việc

Nội dung:

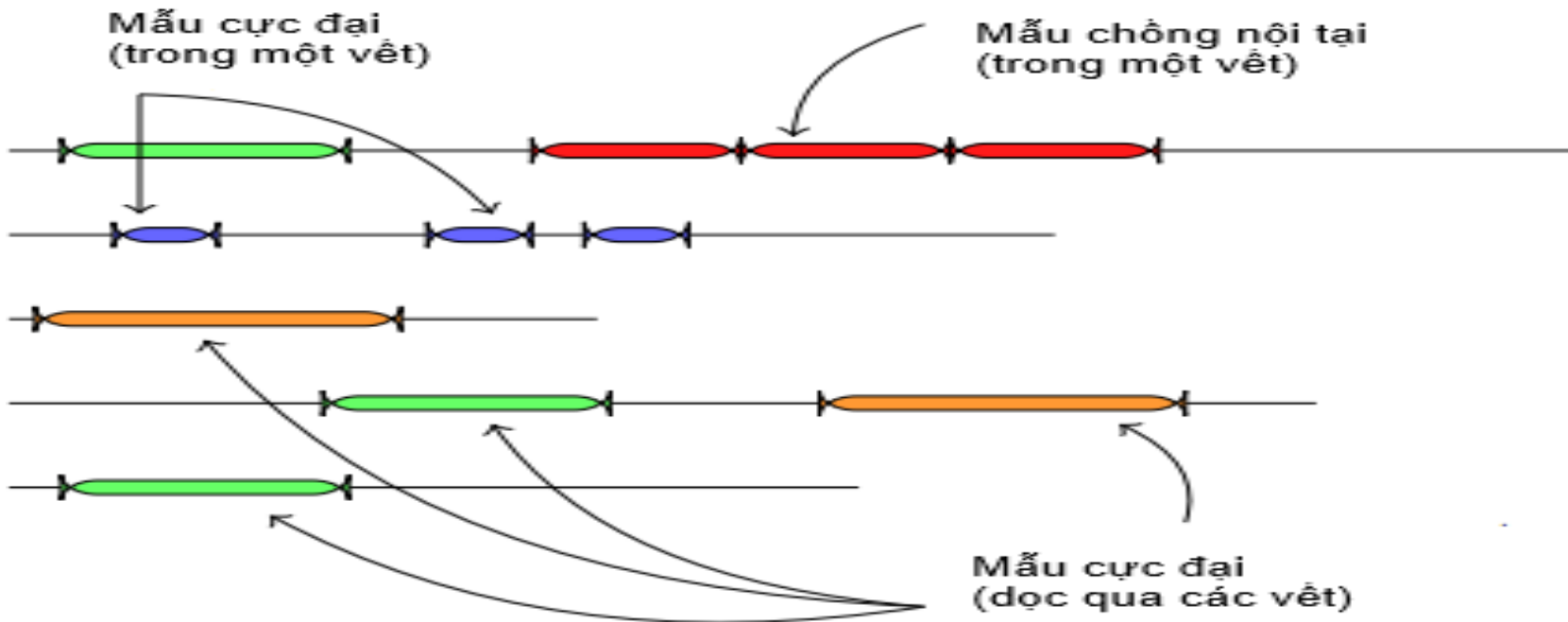
- $T_L = \{t \in T \mid \exists \sigma \in L: t \in \sigma\}$
- $T_I = \{t \in T \mid \exists \sigma \in L: t = \text{first}(\sigma)\}$
- $T_O = \{t \in T \mid \exists \sigma \in L: t = \text{last}(\sigma)\}$
- $X_L = \{(A, B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge$   
 $\forall a \in A, \forall b \in B \ a \rightarrow_L b \wedge \forall a_1, a_2 \in A: a_1 \#_L a_2 \wedge \forall b_1, b_2 \in B: b_1 \#_L b_2\}$
- $Y_L = \{(A, B) \in X_L \mid \forall (A', B') \in X_L: A \subseteq A' \wedge B \subseteq B' \Rightarrow (A, B) = (A', B')\}$
- $P_L = \{p(A, B) \mid (A, B) \in Y_L\} \cup \{i_L, o_L\}$
- $F_L = \{(a, p(A, B)) \mid (A, B) \in Y_L \wedge a \in A\} \cup$   
 $\{(p(A, B), b) \mid (A, B) \in Y_L \wedge b \in B\} \cup \{(i_L, t) \mid t \in T_I\} \cup \{(t, o_L) \mid t \in T_O\}$
- $\alpha(L) = (P_L, T_L, F_L)$



- Bước 1:  $T_L = \{a, b, c, d, e\}$ .
- Bước 2:  $T_I = \{a\}$ .
- Bước 3:  $T_O = \{d\}$ .
- Bước 4:  $X_{L1} = \{(\{a\}, \{b\}), (\{a\}, \{c\}), (\{a\}, \{e\}), (\{a\}, \{b, e\}), (\{a\}, \{c, e\}), (\{b\}, \{d\}), (\{c\}, \{d\}), (\{e\}, \{d\}), (\{b, e\}, \{d\}), (\{c, e\}, \{d\})\}$ , có 10 phần tử, mỗi phần tử là một cặp hai tập con các hoạt động.
- Bước 5:  $Y_{L1} = \{(\{a\}, \{b, e\}), (\{a\}, \{c, e\}), (\{b, e\}, \{d\}), (\{c, e\}, \{d\})\}$ , có 4 phần tử nhận được từ  $X_{L1}$  sau khi loại đi 6 phần tử  $(\{a\}, \{b\}), (\{a\}, \{c\}), (\{a\}, \{e\}), (\{b\}, \{d\}), (\{c\}, \{d\}), (\{e\}, \{d\})$ .
- Bước 6:  $P_{L1} = \{P_{a,be}, P_{a,ce}, P_{be,d}, P_{ce,d}, i_{L1}, o_{L1}\}$
- Bước 7, 8: thiết lập các cung và nhận được lưới Petri kết quả như trình bày tại Hình 6. Rõ ràng đây là một lưới dòng công việc đúng đắn.

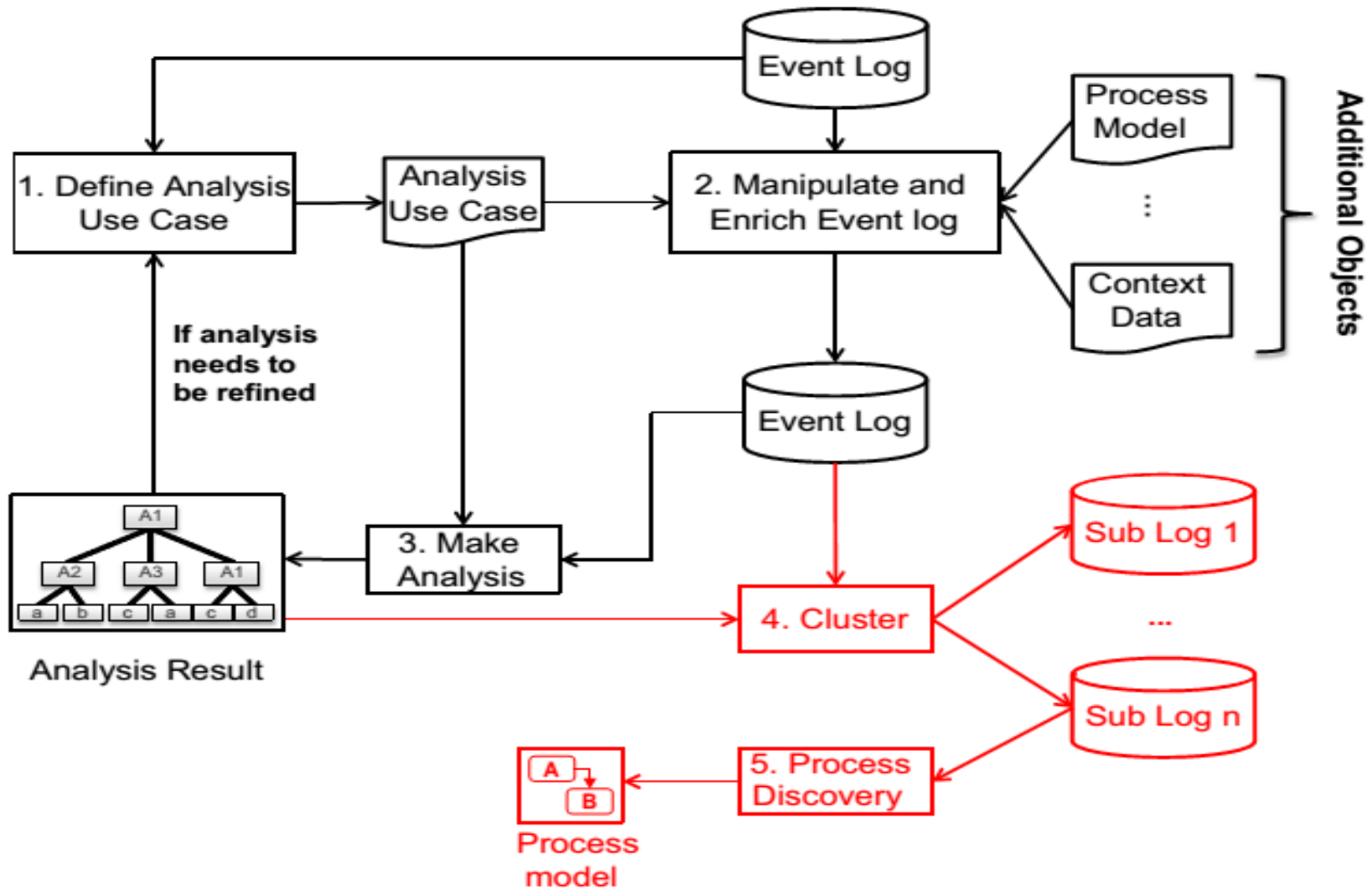
# Tiền xử lý NKSK: T/tượng hóa hoạt động

j g c i b d  
 j g c a h b d  
 u v w a h b d f k a h b d  
 p q r s u v p q  
 c j g f b d  
 m n o m n o m n o s l e



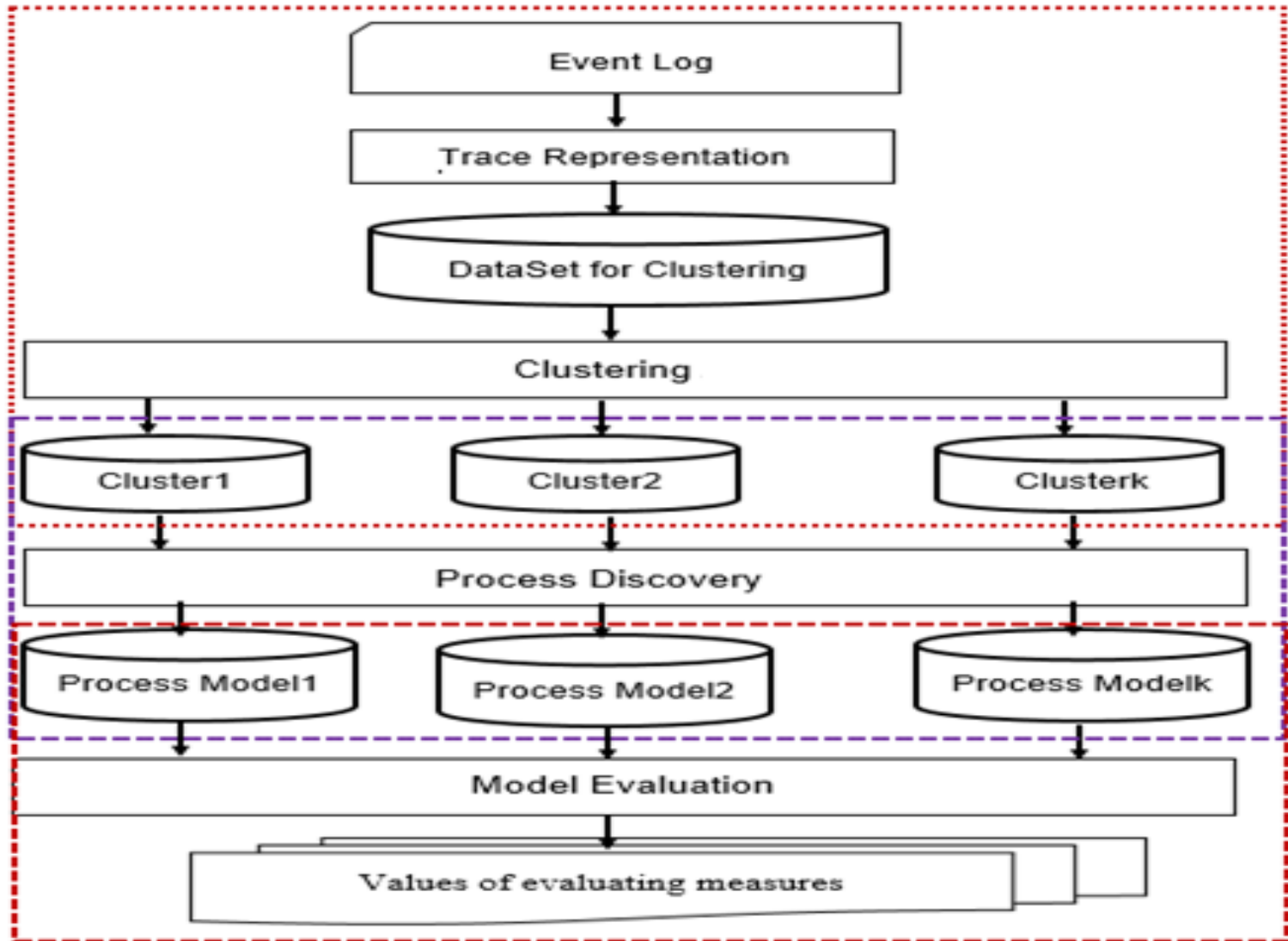
[Bose09] R. P. Jagadeesh Chandra Bose, Wil M. P. van der Aalst. *Abstractions in Process Mining: A Taxonomy of Patterns*. BPM 2009: 159-175, 2009.

# Tiền xử lý nhật ký sự kiện: Phân cụm



[Leoni16] Massimiliano de Leoni, Wil M. P. van der Aalst, Marcus Dees. *A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs*. Inf. Syst. 56 235-257, 2016

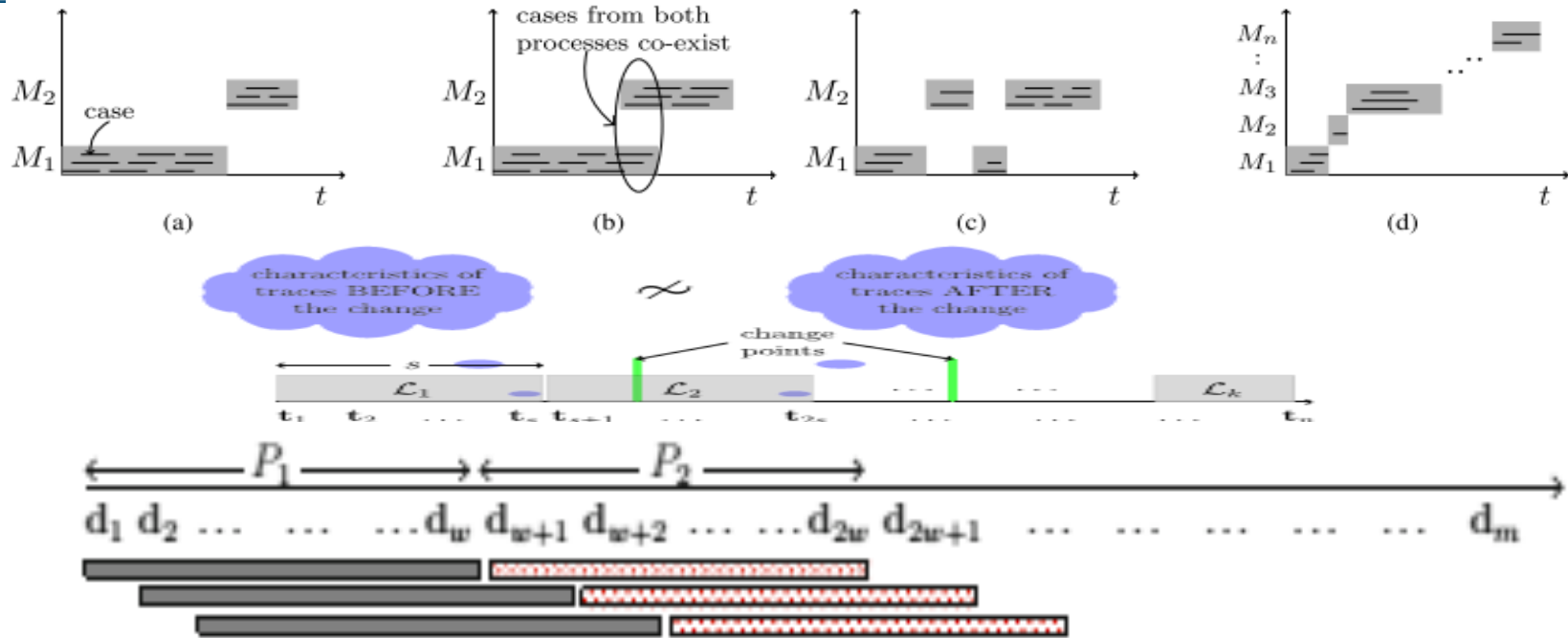
# Phân cụm theo biểu diễn đồ thị vết



[Ha16] Quang-Thuy Ha, Hong-Nhung Bui, Tri-Thanh Nguyen (2016), *A trace clustering solution based on using the distance graph model*. ICCCI 2016 (in press)



# Tiền xử lý NKSK: Trôi khái niệm



## ● Trôi khái niệm trong khai phá quy trình

- Trôi đột ngột (sudden): thay đổi lớn, trôi dần (gradual): tiếp cận nhập pha, trôi định kỳ (*Recurring*): theo mùa kinh doanh, trôi gia tăng (*Incremental*): thay đổi duy trì

[Bose14] R. P. Jagadeesh Chandra Bose, Wil M. P. van der Aalst, Indre Zliobaite, Mykola Pechenizkiy. *Dealing With Concept Drifts in Process Mining*. IEEE Trans. Neural Netw. Learning Syst. 25(1): 154-171, 2014.

[Martjushev15] J. Martjushev, R. P. Jagadeesh Chandra Bose, Wil M. P. van der Aalst. *Change Point Detection and Dealing with Gradual and Multi-order Dynamics In Process Mining*. BIR 2015: 161-178.

# Bài toán liên quan

- **Chuẩn hóa nhật ký sự kiện đầu vào**
  - Đa dạng NKSK của các tổ chức
  - Chuẩn hóa đầu vào cho các thuật toán khai phá quy trình
  - XER và XML
- **Trừu tượng hóa mô hình quy trình**
  - Các mức trừu tượng hóa mô hình quy trình
  - Thu gọn mô hình quy trình và tập mô hình quy trình trong tổ chức

# Một số trao đổi

- Khai phá quy trình tại Việt Nam
- Phối hợp

# Việt Nam: Hàn lâm & Công nghiệp

our reference

ZEIN2013ZSEL31 - 72086

Brussels

20 February 2013

SEL 31 - Bart Baesens (KU Leuven) Ha Quang Thuy (Vietnam National University)	
Improving Business Process and Complex Event Sequence Analytics using Process Mining Techniques	
+	-
<ul style="list-style-type: none"> <li>• The proposal fits into the Vietnamese country strategy, as it is focussing on ICT and IS support services.</li> <li>• Synergy: The project builds on previous and on-going research in the region and in Latin America and China.</li> <li>• Well written, but not always very concrete. The logical framework explains more in detail.</li> <li>• Both partners will benefit from this collaboration.</li> </ul>	<ul style="list-style-type: none"> <li>• The topic of process mining seems to be a global problem.</li> <li>• The needs of the Vietnamese partner are not so clear: there are some common (global) concerns, but there is no concrete connection with the region.</li> <li>• The project's results could be more academic than practical. Not clear how to "experiment" or outreach to firms.</li> <li>• The collaboration with Vietnamese local industry partners could be worked out more in detail.</li> <li>• The project proposal is rather broad; case studies are mentioned but not in detail.</li> <li>• The project's results may be too advanced for the current level of technology and staff of vast majority of Vietnamese firms.</li> </ul>

- Đề xuất dự án tại Bỉ: Đánh giá (+) và (-)
- Đề tài ĐHQGHN phát hiện mẫu: sự phối hợp của Hải Hòa

# Phối hợp các nhóm nghiên cứu

- Khai phá quy trình: đáng nghiên cứu ?
  - Cạnh tranh quốc gia và doanh nghiệp về quy trình
  - Các nghị định 19/NĐ-CP
  - Kết nối giữa chuyên gia công nghệ và nhà quản lý kinh doanh
  - Nhu cầu thực tiễn hiện tại và tương lai ?
- Nhóm nghiên cứu Trường ĐHCN
  - Khai phá quy trình và khoa học dữ liệu
  - Giảng viên, sinh viên, học viên cao học và nghiên cứu sinh
- Nhóm nghiên cứu tại các một số doanh nghiệp và các tổ chức khác
  - Nhà khoa học dữ liệu
  - Người quản lý công nghệ
  - Người quản lý nghiệp vụ

**Cám ơn !**  
**감사 !**  
**ありがとう !**  
**謝謝 !**  
**Thank you!**