

# Process Mining Research at the Technical University of Lisbon

Diogo R. Ferreira

<http://web.tagus.ist.utl.pt/~diogo.ferreira/>



UTL



# UNIVERSIDADE TÉCNICA DE LISBOA

► Notícias

► Eventos

► Quicklinks

Uma Universidade no Mundo



Gabinete de Atendimento e Recepção de Candidaturas

Acesso ao Ensino Superior 2009/10

Gabinete em Funcionamento na Reitoria da UTL

## A UNIVERSIDADE

[Boas-vindas](#)

[Missão e História](#)

[Escolas](#)

[Governo da Universidade](#)

[UTL em Números](#)

[Honoris Causa](#)

[Vida na UTL](#)

[Contactos](#)

[Localização](#)

## ESCOLAS

[Faculdade de Medicina Veterinária](#)

[Instituto Superior de Agronomia](#)

[Instituto Superior de Economia e Gestão](#)

[Instituto Superior Técnico](#)

[Instituto Superior de Ciências Sociais e Políticas](#)

[Faculdade de Motricidade Humana](#)

[Faculdade de Arquitectura](#)

## ESTUDAR NA UTL

[ECTS](#)

[Acesso à Universidade](#)

[Informação Académica](#)

[Acção Social](#)

[Mobilidade Nacional](#)

[Estudante Internacional](#)

[Mobilidade ERASMUS](#)

[Outros Programas de Mobilidade Internacional](#)

[Vida na UTL](#)

[Catálogo de Cursos](#)

## INVESTIGAÇÃO

[Investigação na UTL](#)

[Prémios UTL](#)

[Factos e Números](#)

[Acções Integradas](#)

[Bolsas](#)

[OTIC | UTL](#)

## COOPERAÇÃO

[UTL / Empresas](#)

[UTL / Universidades](#)

[Prémios UTL](#)

[Mobilidade ERASMUS](#)

[Outros Programas de Mobilidade](#)

[Redes](#)

[Protocolos](#)

## INFORMAÇÃO SOBRE...

[Acesso Ensino Superior](#)

[Portugal](#)

[Lisboa](#)

[Publicações da UTL](#)

[UTL é Notícia](#)

[Espaços da UTL](#)

[Merchandising da UTL](#)

[Actas do Senado ao abrigo da LAU](#)

[Recursos Humanos](#)

[Concursos Públicos](#)

[Gripe A: Prevenção e Acção](#)




**Bem-vindo ao Técnico!**  
*adquire conhecimento + estimula a tua criatividade*

## Eventos

26 de Set. 2009 **Festival Voo Livre'09 - Covilhã**

25 de Set. 2009 **Seminário de Tecnologias de Informação na Promoção das Áreas Protegidas (TIPAP-2009)**

7 a 8 de Set. 2009 **CEI2009 - Encontro Nacional do Colégio de Engenharia Informática da Ordem dos Engenheiros**

7 a 7 de Set. 2009 **Sessão de Homenagem ao Professor João Martins**

## Notícias

**IST Press edita o Livro "Ecologia Industrial: princípios e ferramentas"**  
18 de Set. 2009

**IST Press edita o Livro "Análise Complexa e Equações Diferenciais"**  
18 de Set. 2009

**Encontro "O contributo de Bolonha na prossecução da Qualidade no Ensino Superior", dia 21 de Outubro de 2009**  
17 de Set. 2009

**3ª Edição do Livro "Vitruvius-Tratado de Arquitectura"**  
16 de Set. 2009



## Localização

Campus Alameda  
Campus Taguspark

## Iniciativas + Recursos

Observatório de Empregabilidade  
Recrutamento  
Biblioteca  
Transporte "Alameda-Taguspark"  
Webmail

 Subscrever as feeds de RSS

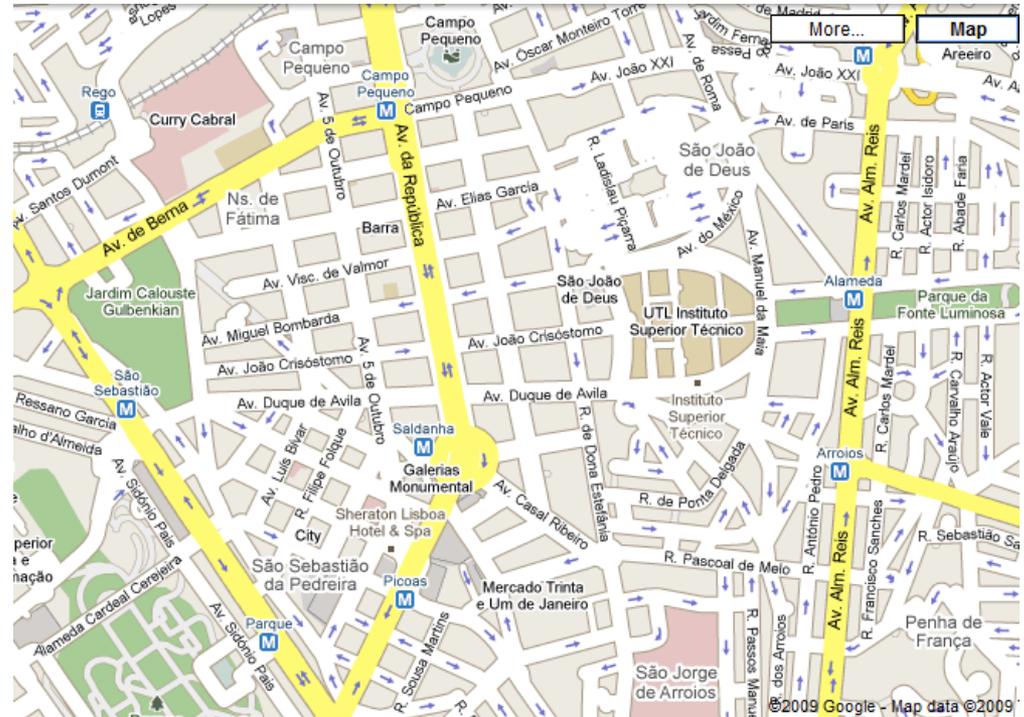
 Siga o Técnico no Twitter

# History of IST

- 1911: founded with the engineering degrees: Mining, Civil, Mechanical, Electrical, Chemical
- 1930s: construction of the campus at Alameda



"Engenho e Obra", Dom Quixote



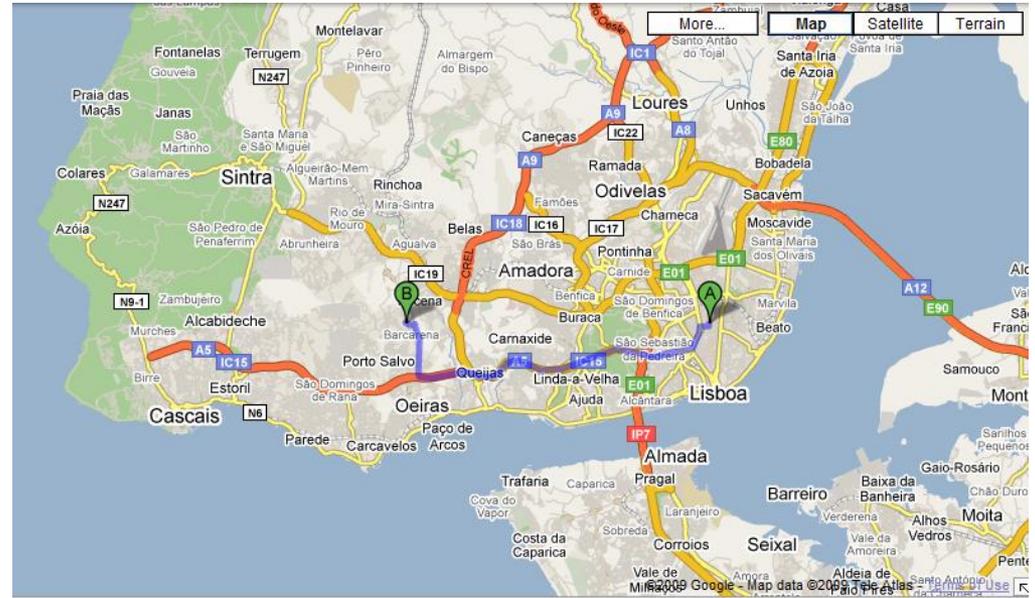
Google Maps

# History of IST

- 1930s: becomes part of Technical University of Lisbon (UTL)
- 1950-1970: development of several research centers
- 1970: engineering degrees from 6 to 5 years (currently 3+2)
- 1990s: construction of the campus at Taguspark (Oeiras)



Bing Maps

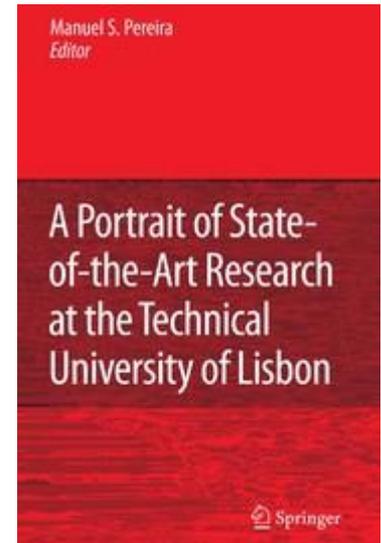


Google Maps



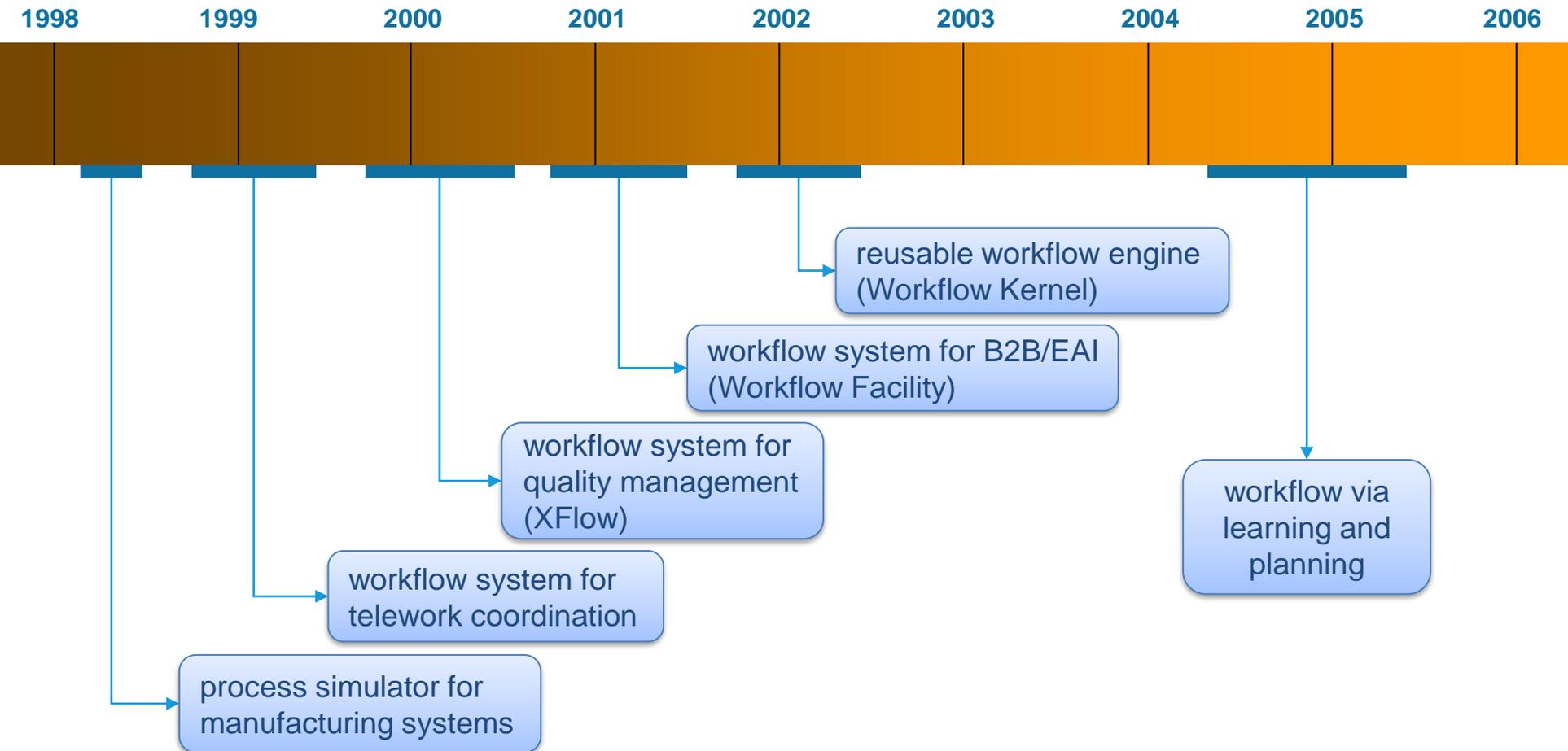
# Research at UTL

- Several schools, several disciplines
  - physics, chemistry, mathematics
  - economics, management, social sciences
  - life sciences, biotechnology, bioinformatics
  - environment and sustainability
  - health and sport sciences
  - architecture, urbanism, transportation
  - ...

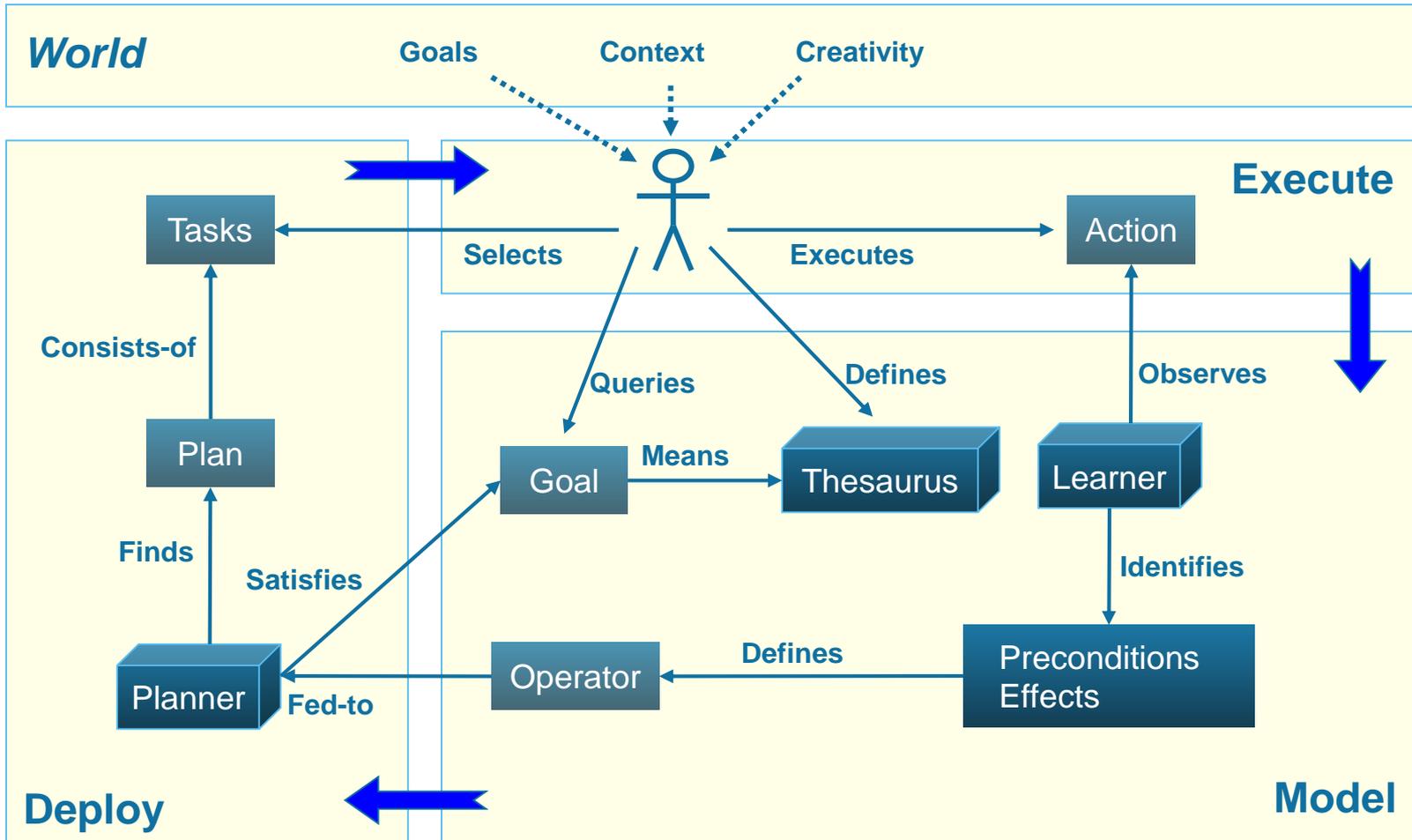


Manuel S. Pereira (Ed.)  
622 pgs, Springer 2007

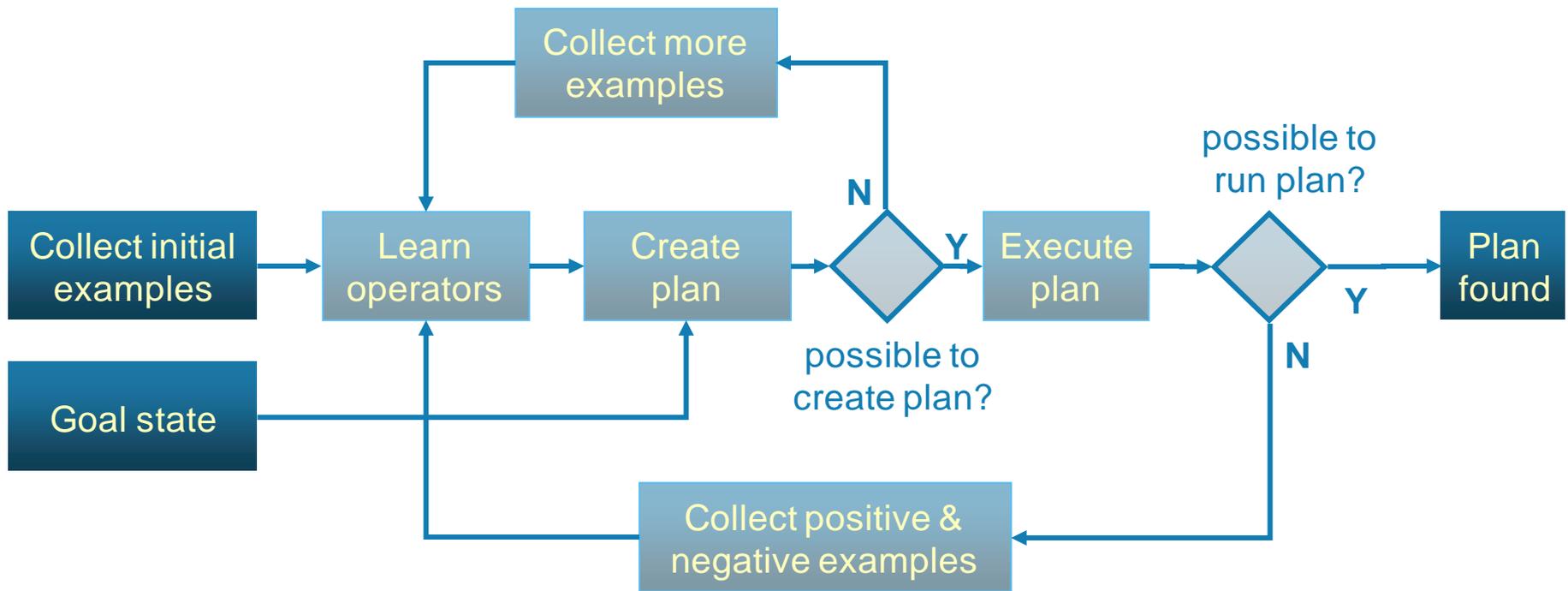
# Background



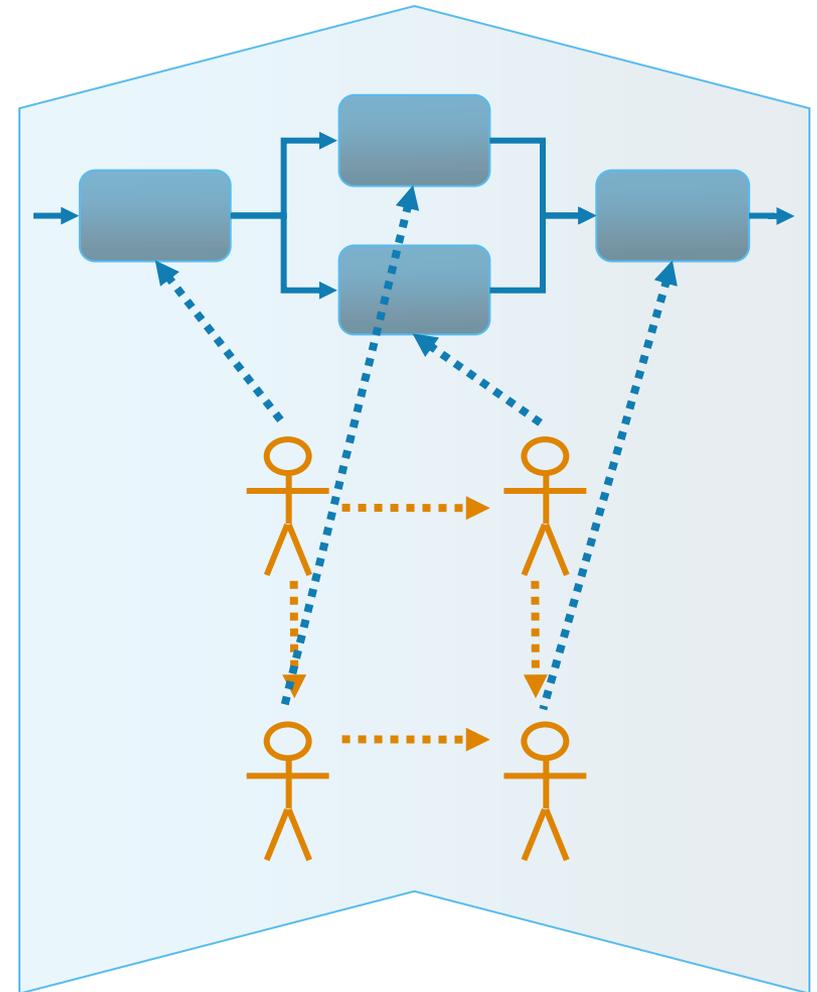
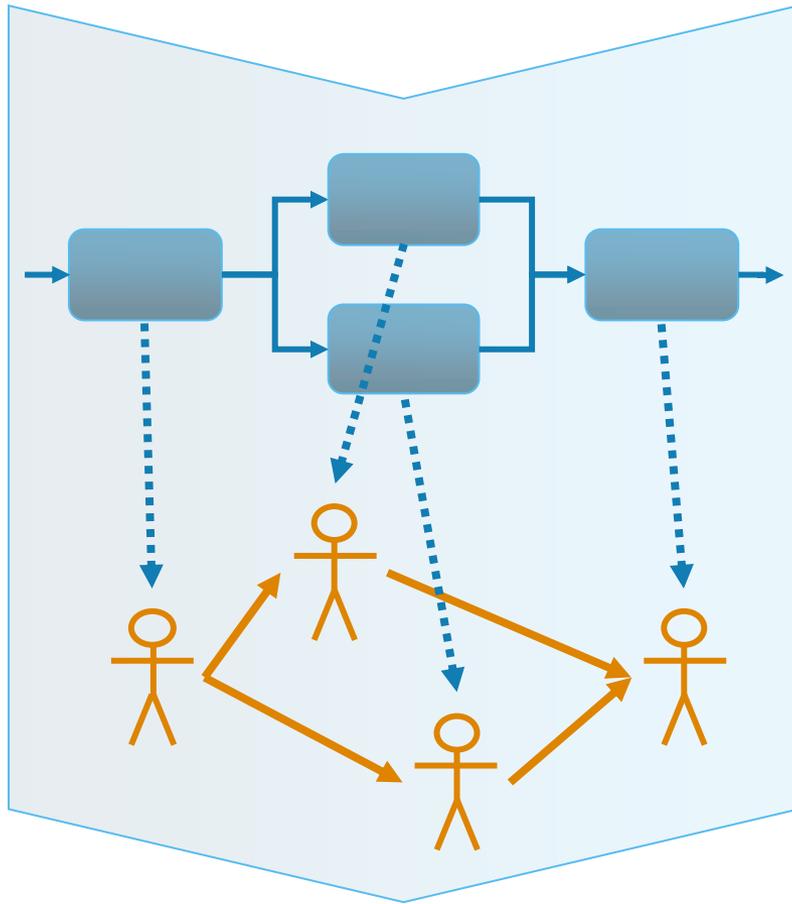
# Background



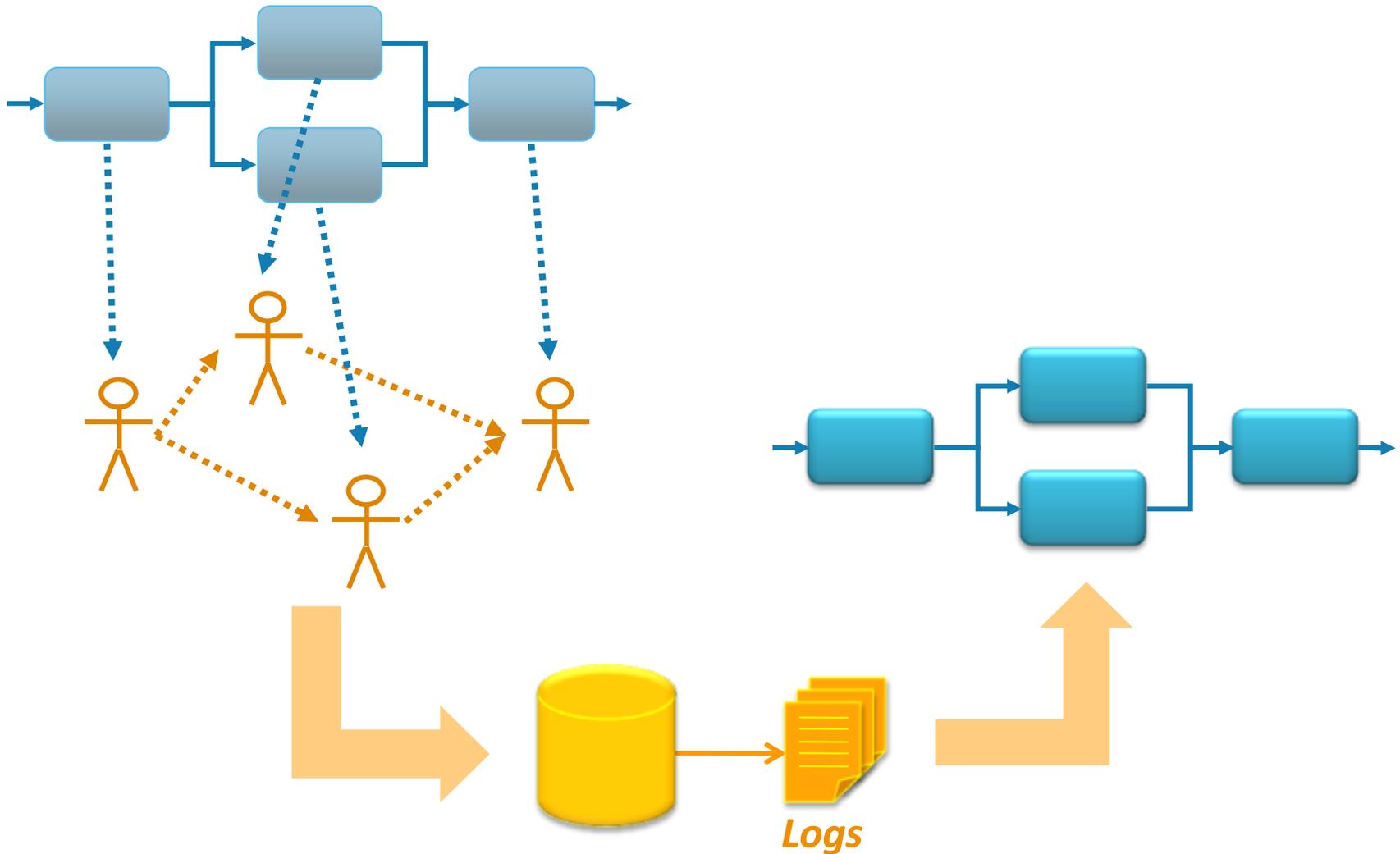
# Background

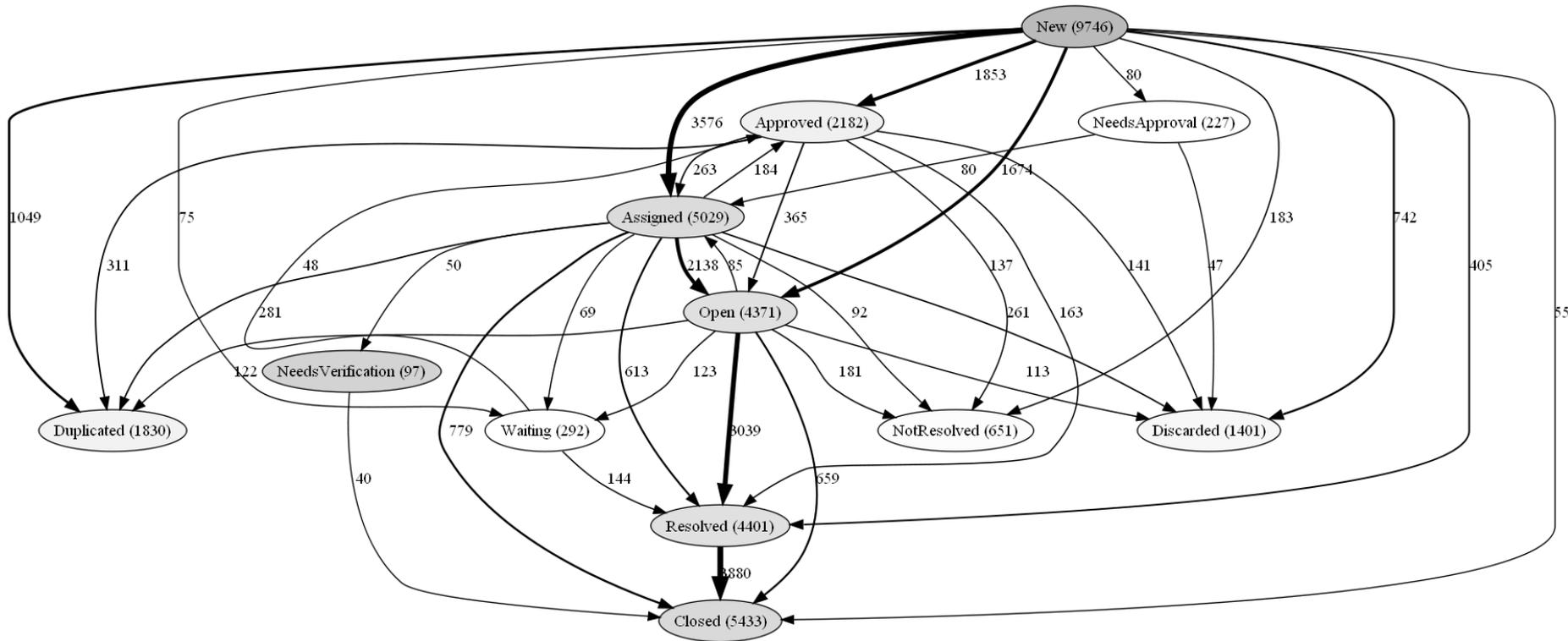
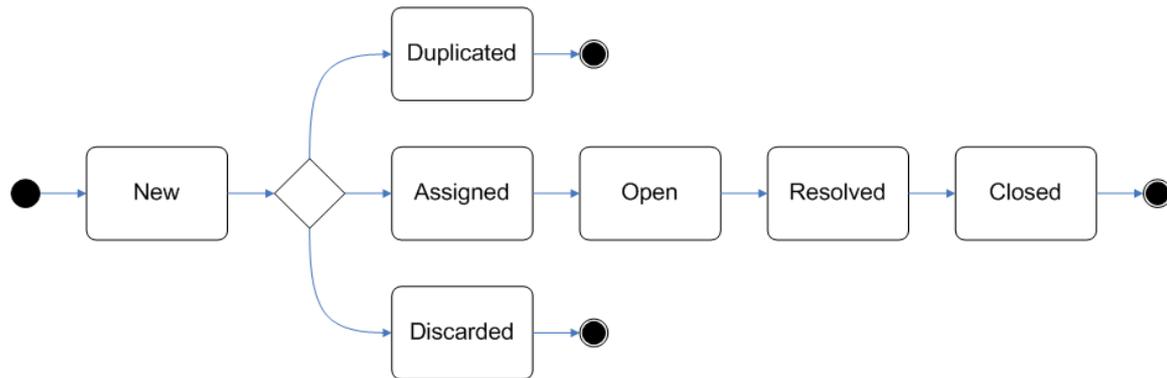


# Workflow upside down

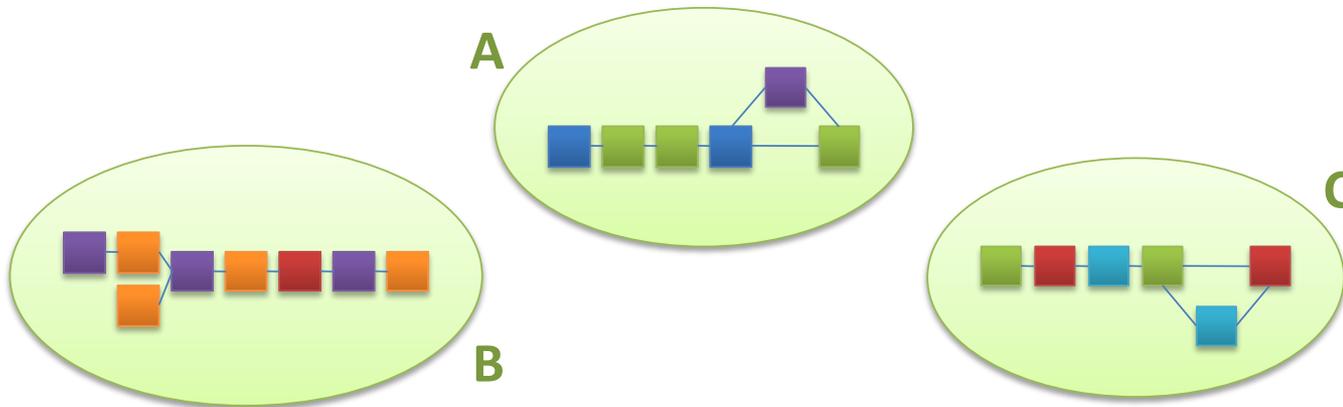
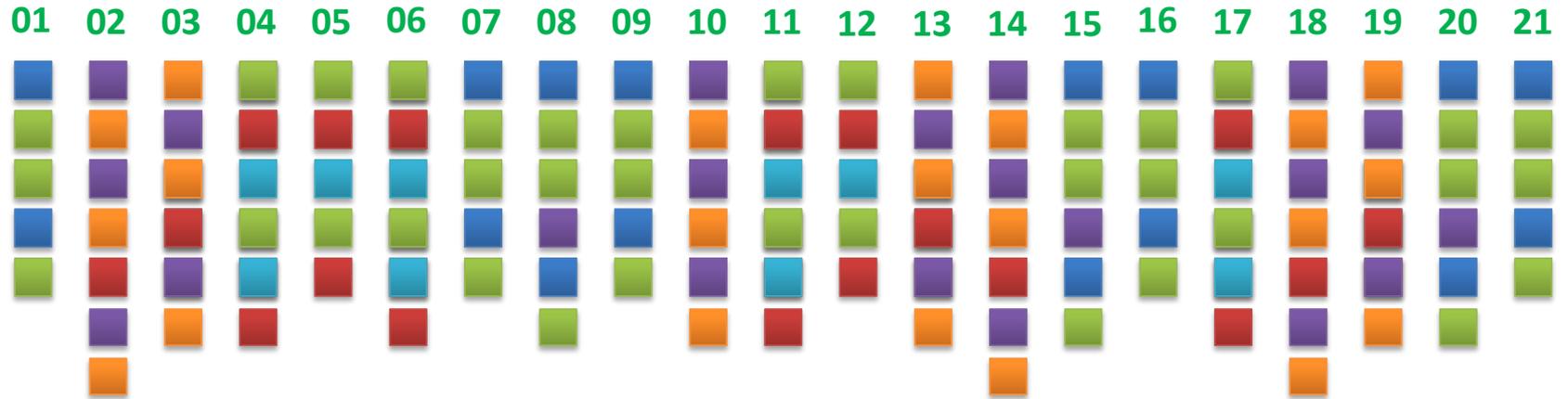


# Process Mining





# Sequence Clustering

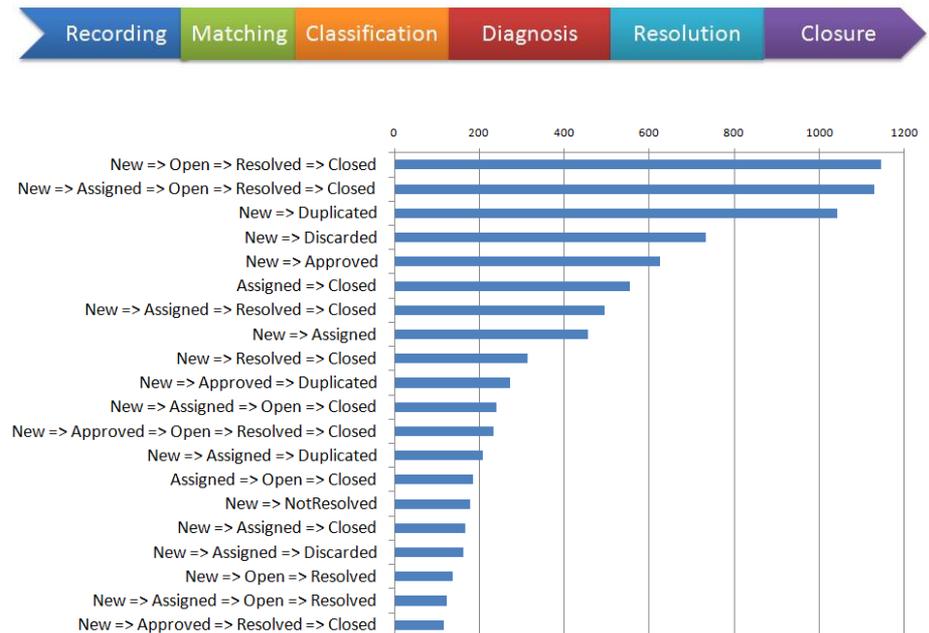
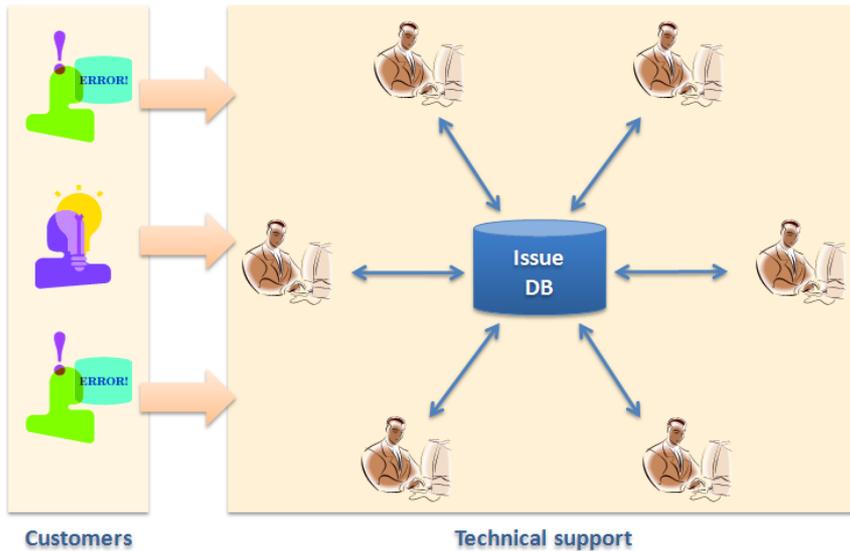
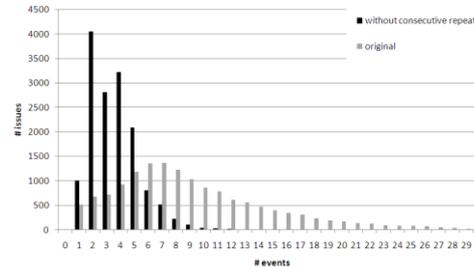




# Sequence Clustering



Miguel Mira da Silva  
IST



D. R. Ferreira, M. Mira da Silva

Using Process Mining for ITIL Assessment: a Case Study with Incident Management

Proceedings of the 13th Annual UKAIS Conference, Bournemouth University, April 10-11, 2008

# Control-flow analysis



Pedro Diniz  
IST

```

1 read i
2 read d
3 read c
4 read d
5 read c
6 write i
7 read i
8 read b
9 read c
10 read e
11 read c
12 write i
13 read i
14 read b
15 read c
16 read b
17 write i
18 read i
19 read b
20 read b
21 write i
    
```

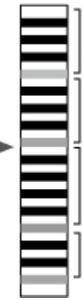
Input trace



**Phase 1**

Fourier analysis  
for identification of  
loop control variables  
and loop boundaries

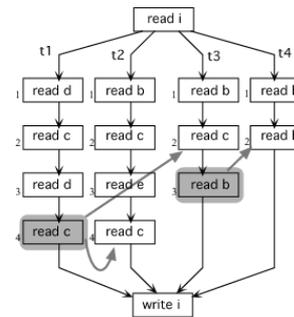
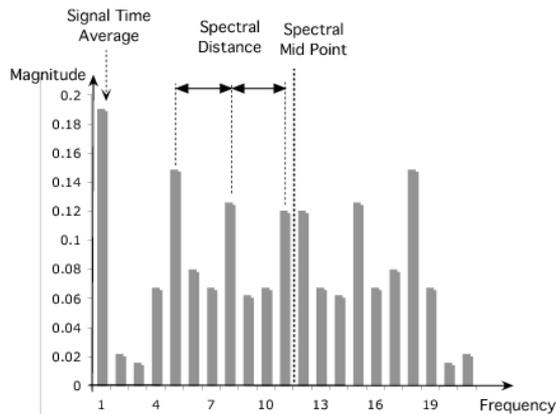
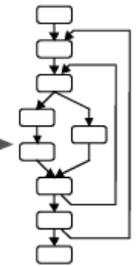
Input trace  
with loop boundaries



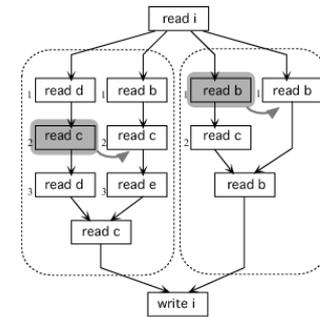
**Phase 2**

CFG extraction using  
backward and forward  
operation matching in  
subtraces

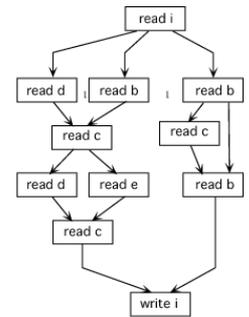
Control Flow Graph  
(CFG)



(a) Initial set-up phase CFG and matching between sub-traces.



(b) CFG after merging of nodes in traces t1 and t2 and traces t3 and t4. There is no match between t1 and t3 as they are into two distinct partitions.



(c) Final CFG

# Process mining and conformance



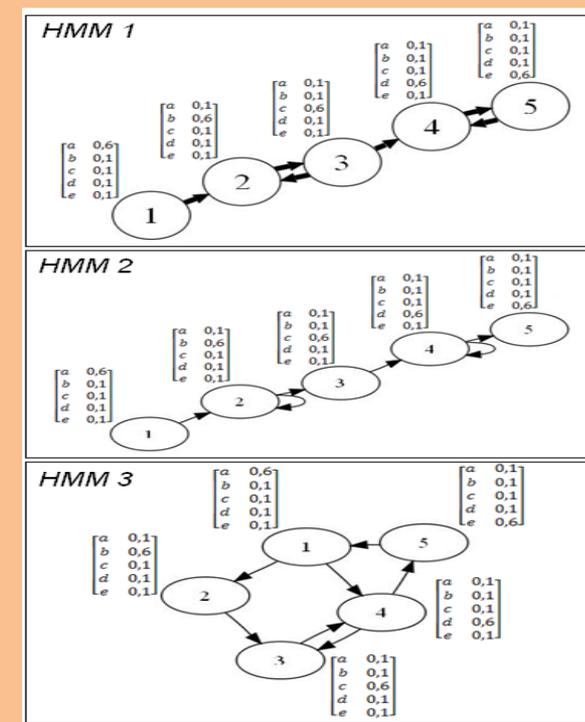
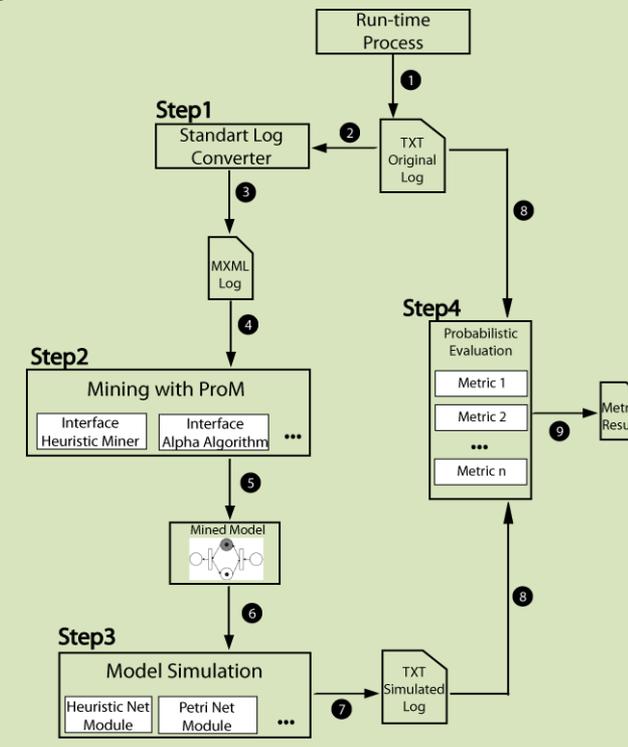
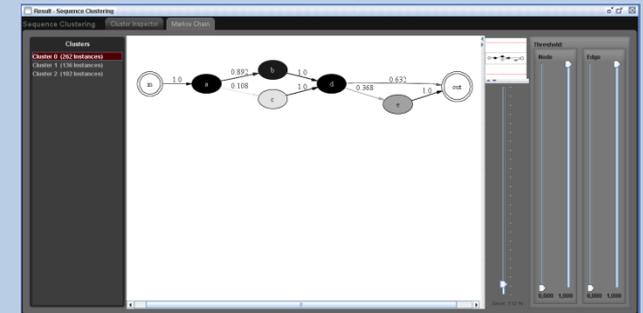
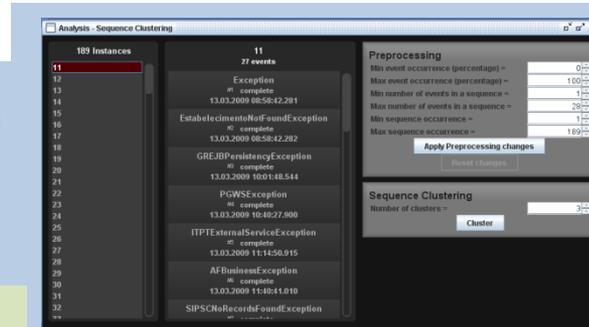
Gabriel Veiga  
MSc student



Pedro Martins  
MSc student



Gil Aires  
MSc student

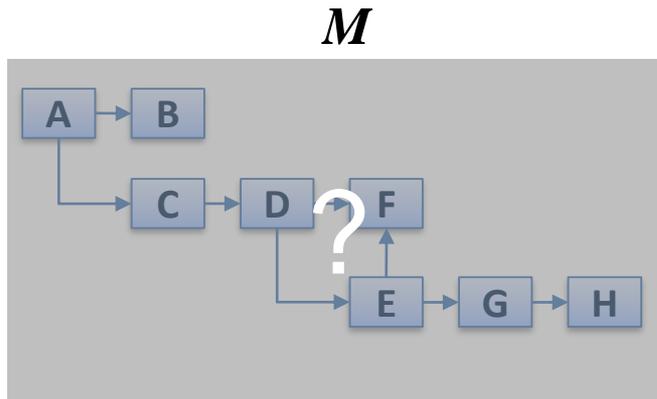


G. M. Veiga, D. R. Ferreira,  
*Understanding Spaghetti Models  
with Sequence Clustering for ProM*  
BPI 2009 Workshop, Ulm, Germany

G. Aires, D. R. Ferreira  
*Applying Hidden Markov Models to  
Process Mining*  
CISTI 2009, Portugal

# Process mining

$s$	$x$
case id	task id
1	A
1	B
?	A
?	C
2	D
2	E
1	E

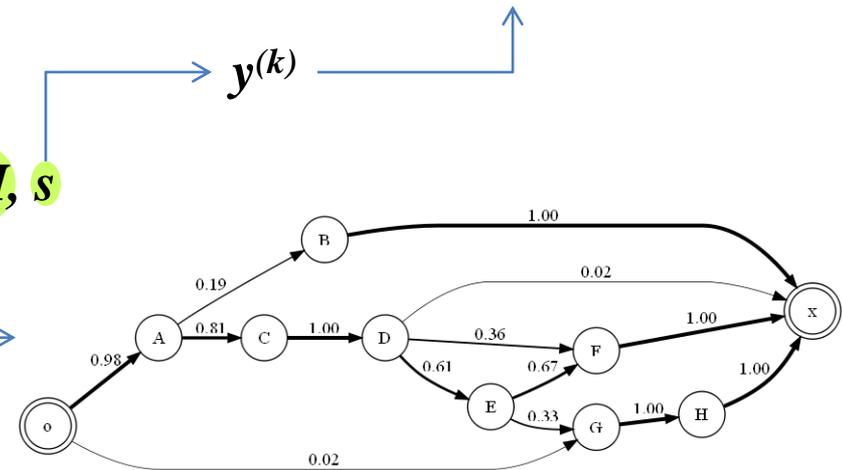
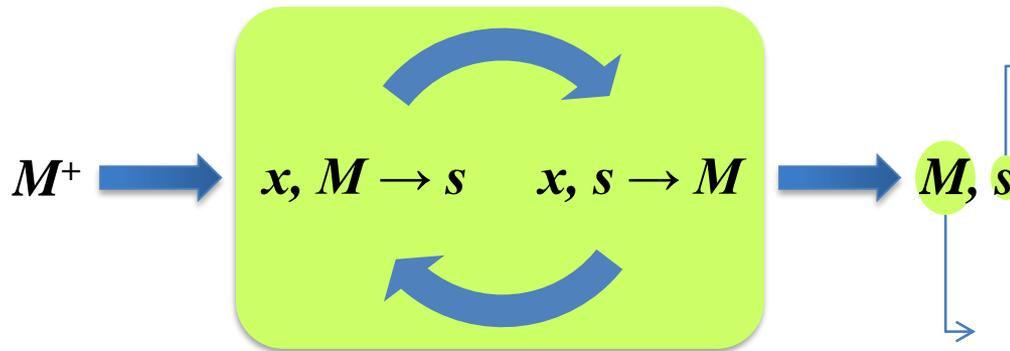


$$G(p \parallel q) \triangleq \sum_{z \in \mathbb{Z}} \sqrt{p(z) \cdot q(z)}$$

$z$	$p(z)$	$z$	$q(z)$
ACDEF	30.0 %	ACDEF	32.8 %
ACDF	30.0 %	ACDF	28.9 %
AB	20.0 %	AB	18.1 %
ACDEGH	20.0 %	ACDEGH	16.2 %
		ACD	2.0 %
		GH	2.0 %

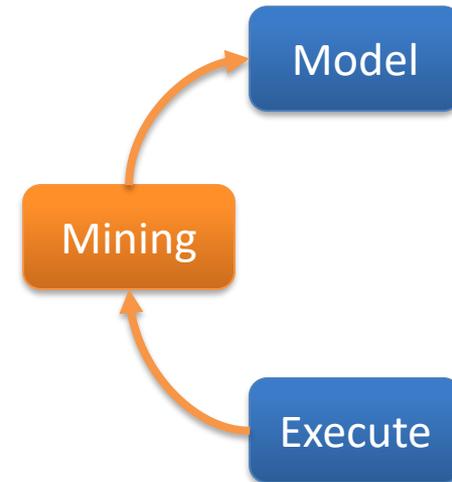


97.8 %



# Conclusion

- Process mining
  - missing link in BPM life cycle
  - valuable aid for process discovery
  - essential for conformance checking
  - very rich in practical problems



## 6. Body

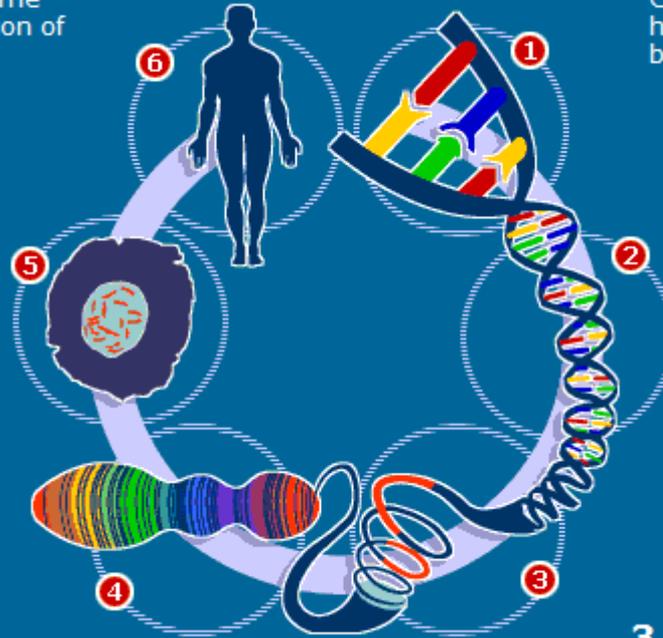
Each of the cells becomes specialised by obeying just some of the instructions in the DNA. Blood, muscle, bone, organs and so on result. The body is built from 100 trillion of these cells.

## 5. Nucleus and Cell

The 46 chromosomes are held in the nucleus found in most cells in the human body. Nearly every cell in the body contains the full DNA code for producing a human.

## 4. Chromosomes

The total number of genes is not known - estimates range from 30,000 to 120,000. However many there are, they, and all the junk DNA, are wrapped up into bundles called chromosomes. Every human has 23 pairs of chromosomes, one set from each parent.



## 1. The four letters

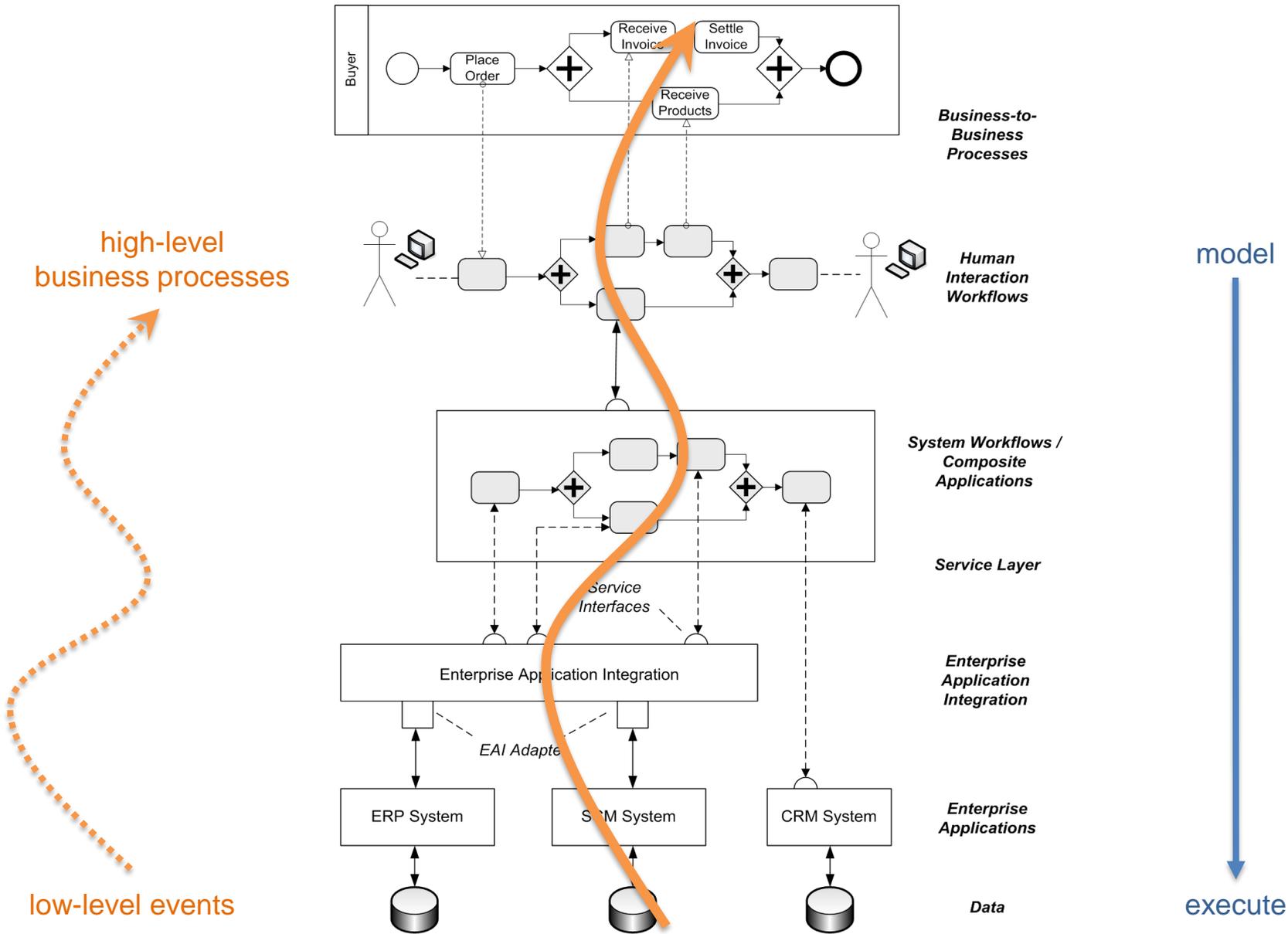
All genetic code is spelled out with just four chemical letters, or bases: adenine (A), thymine (T), cytosine (C) and guanine (G). These pair up, A with T and C with G. The human genome has between 2.8 and 3.5 billion base pairs.

## 2. DNA double helix

The base pairs form the rungs of the ladder-like DNA double helix. Running up and down the ladder are the long sequences of bases which are the code for life. Each cell in the human body contains two metres (six feet) of DNA.

## 3. Genes

As little as 3% of the total genome is made of genes - the rest is meaningless "junk". Genes are special sequences of hundreds or thousands of base pairs that provide the templates for all the proteins which the body needs to produce.





Thank you!