teamUP
ACADEMIC RESEARCH & INDUSTRY

innoviris
.brussels
empowering research

INTIMALS

Intelligent Modernisation Assistance for Legacy Software

14 September 2017 – 9:30

VUB VRIJE UNIVERSITEIT BRUSSEL

raincode LABS
compiler experts

UCL
Université catholique de Louvain

Problem & context

raincode LABS

compiler experts

migration & maintenance services

modernise legacy software

Increasing demand
for such services

Current approaches require
significant **manual work**

financial, retail, logistics, government, ...

GOAL

Towards an intelligent modernisation assistant for legacy software

KEY IDEA

Automating migration requires **pattern discovery**

consortium

raincode LABS
compiler experts

Partners know each other well

Have overlapping expertise in

- Programming languages, software and language engineering

And complementary expertise in

- **Compiler** technology and language **modernisation** (Raincode Labs)
- Pattern **mining** and software **evolution** (UCL)
- Pattern **matching** and software **analysis** (VUB)

GOAL

Towards an intelligent modernisation assistant for legacy software

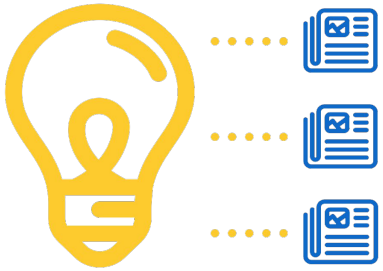
Key Idea

Automating migration requires **pattern discovery**

Mining ≠ kinds of **patterns** in **3 use cases** for **3 objectives**

	use case 1	use case 2	use case 3
Objective A			
Objective B			
Objective C			

uses cases



use case 1	use case 2	use case 3
Coding idioms and programming **conventions**	**Library usage protocols** and **violations**	**Systematic edits** or **repetitive changes**
Syntax of code	Semantics of code	Evolution of code

Objectives

Objective A **Program comprehension**	Provide insights in legacy code
Objective B **Anomaly detection**	Detect potential inconsistencies in legacy code
Objective C **Modernisation assistance**	Provide recommendations to engineers for improving legacy code

AMBITION

	use case 1 **Coding idioms** and programming **conventions**	use case 2 **Library usage protocols** and **violations**	use case 3 **Systematic edits** or **repetitive changes**
Objective A **Program comprehension**	Undiscovered syntactic patterns ***Replace*** *by macros or language built-ins*	API, library, framework usage patterns ***Refactor*** *based on* ***understanding***	Locations frequently subject to past edits ***Refactor*** *to avoid future changes there*
Objective B **Anomaly detection**	Code deviating from expected patterns ***Rewrite*** *to adhere to expected pattern*	Violations of API and library usage patterns ***Correct*** *violations to fix faulty code*	Locations similar to but not included in past systematic edits ***Reapply*** *systematic edits to these locations*
Objective C **Modernisation assistance**	Propose actions to improve respect of idioms ***Suggest*** *to complete or transform code*	API, library, framework usage ***Suggest*** *correct use of API according protocol*	Frequently executed past edits ***Suggest*** *complete edit based on partial edit*

Motivating Example


```
map STD_RTRN_INFO of SYSTEM_NAME_MNT_SRV_O
to  STD_RTRN_INFO of STD_MSG_DIS_I
```

```
map CAPTION_TEXT of STD_CLIENT_PARM_V of SYSTEM_NAME_PD_W
to  CAPTION_TEXT of STD_MSG_DIS_I
```

```
use rule STD_MSG_DIS nest
```

```
map ACTN_CD of STD_MSG_DIS_O
to  ACTN_CD of STD_CLIENT_PARM_V of SYSTEM_NAME_PD_W
```

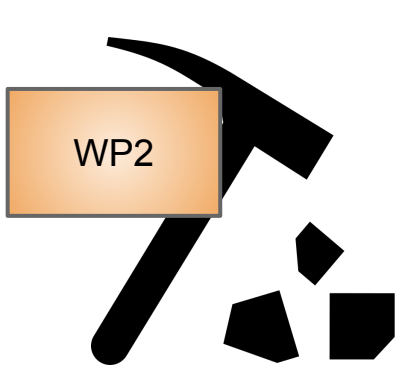
ActionCode := MessageDisplay(CaptionText, ReturnInfo);

Distinguishing features

The modernisation assistant will

- Give **proactive** advice to Raincode Labs' engineers
  - To support them in their renovation efforts
- Automatically discover *previously unknown* **patterns**
  - For our **3 use cases** and **3 objectives**
- Growing library of mined patterns
  - More patterns uncovered ⇒ **more intelligent** suggestions
- **Learn** from experience and feedback

APPROACH



WP2

previously unknown software patterns

WP3

PATTERN MINING ALGORITHMS

MODERNISATION ASSISTANT

browser for patterns and instances
on-demand partial pattern matching
pro-active recommendations

WP1

DATA + METAMODEL

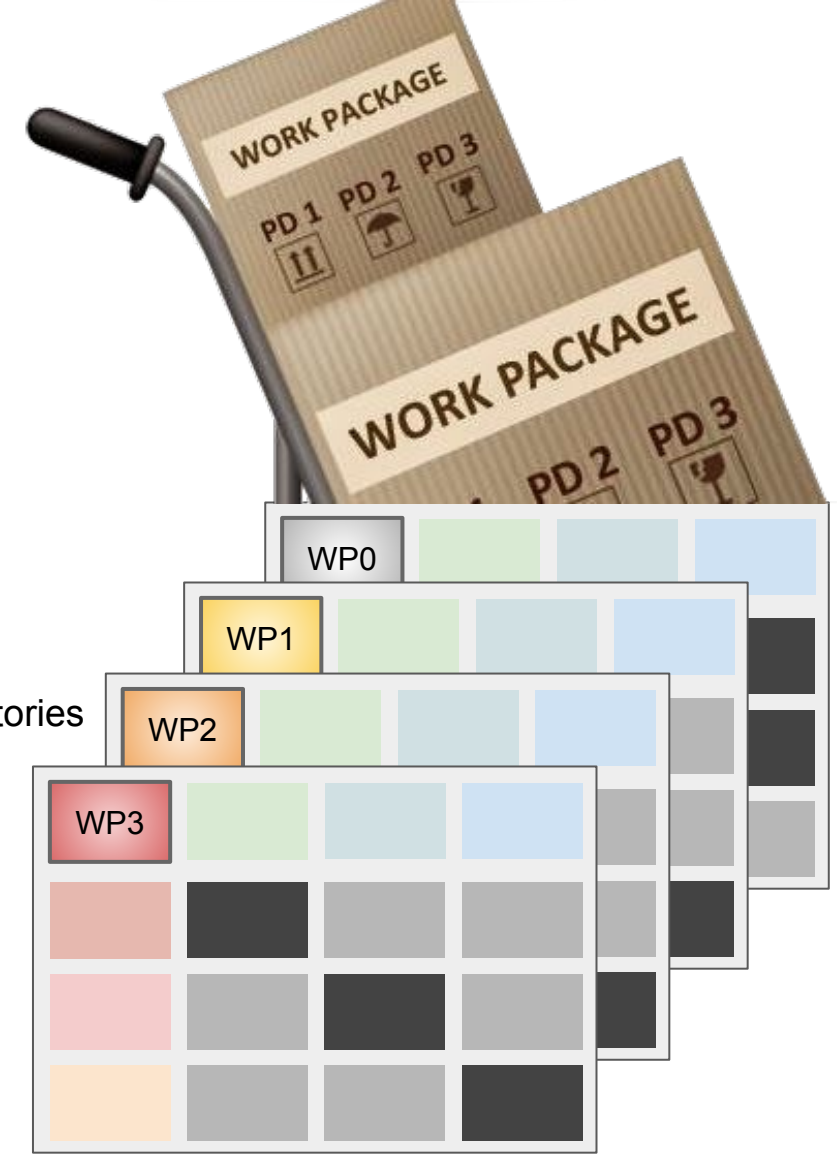
EVALUATION

WP0

IMPORTERS

- open-source software (academic)
- legacy software (industrial)

work packages



WP0

Evaluation of developed system on legacy systems **(RL)**

WP1

Data and **meta-model** for software repositories and software patterns **(**VUB, UCL, **RL)**

WP2

Pattern mining algorithms **(UCL**, VUB**)**

WP3

Pro-active **modernisation assistant** (**VUB**, UCL, RL)

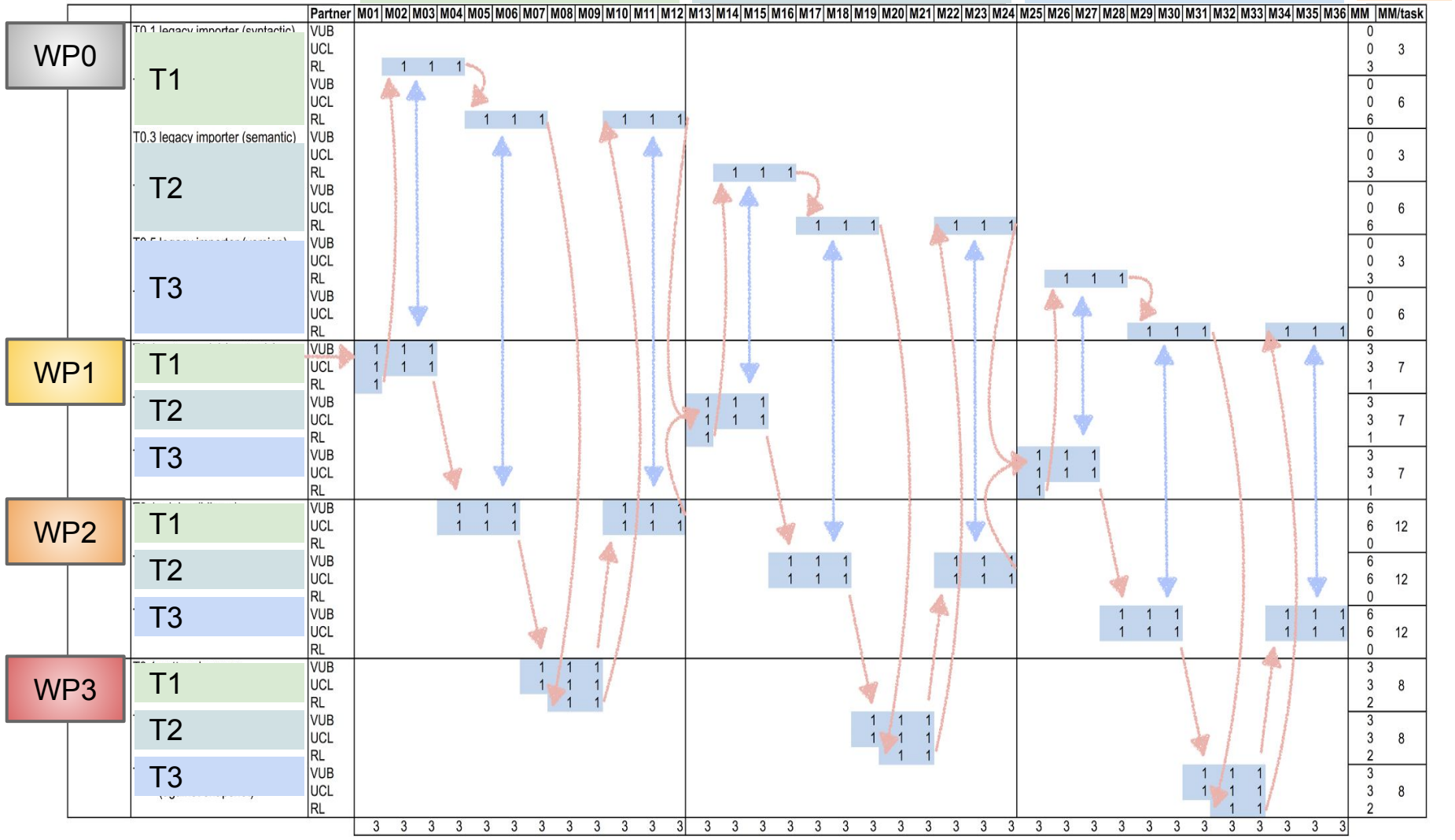
Project coordinator (**VUB**)

Iterative work plan

UC1: **Coding idioms** and programming **conventions**

UC2: **Library usage protocols** and **violations**

UC3: **Systematic edits** or **repetitive changes**



Project contributions

Scientific contributions to the **AI** field

- pattern mining over source code: *“Big Code”*
- frequent subgraph mining algorithms for property-based graphs
- supporting pattern-specific relevance constraints

Scientific contributions to the **SE** field

- language-parametric meta-model
- partial pattern matching algorithms against modernisation context
- modernisation support based on mined idioms, usage patterns, edit patterns

Technical contributions

- prototype of an intelligent modernisation assistant
- importers for legacy programming languages

Economic valorisation (Raincode Labs)

Outperform competition through unique selling features (niche market)

- **automation** of understanding, migration and renovation of legacy software
- increased **customer intimacy** because patterns can be customer-specific

Easy to address **new markets**

- other languages and technologies

Increased **quality of services** and improved **market share**

- increased **sales** and **turnover**
- increased **visibility**
- **employee retention** and **staff growth**

raincode LABS

compiler experts

Economic valorisation (Brussels-Capital Region)

Strengthen Brussels' position as strong ICT hub with high-tech companies

through close collaboration between businesses and academia (AI for SE)

and proximity of clients (banks, insurances, IT, retail, logistics, government) in need for expert services to modernise their legacy software systems

Increased competitiveness

improving clients' legacy allows them to focus on their core business

new key functionality rather than maintaining and fixing old systems

Increased employment

modern software systems are more attractive to young developers

increases employment on the Belgian market (as opposed to outsourcing)

conclusion

AI that creates jobs

Explore novel pattern mining algorithms

by treating code and edit history of legacy software as data

Incorporate them in an intelligent modernisation assistant

to pro-actively recommend modernisation actions to software engineers

Learn from past experience and feedback from engineers

to improve modernisation assistant and discovered patterns

Through an **academic-industrial collaboration**

develop state of the art technology and validate it on industrial cases

To let industrial partner **keep and strengthen its competitive advantage**

by offering improved services and quality

directly beneficial to its clients

Questions & Answers

Backup slides

Backup slides

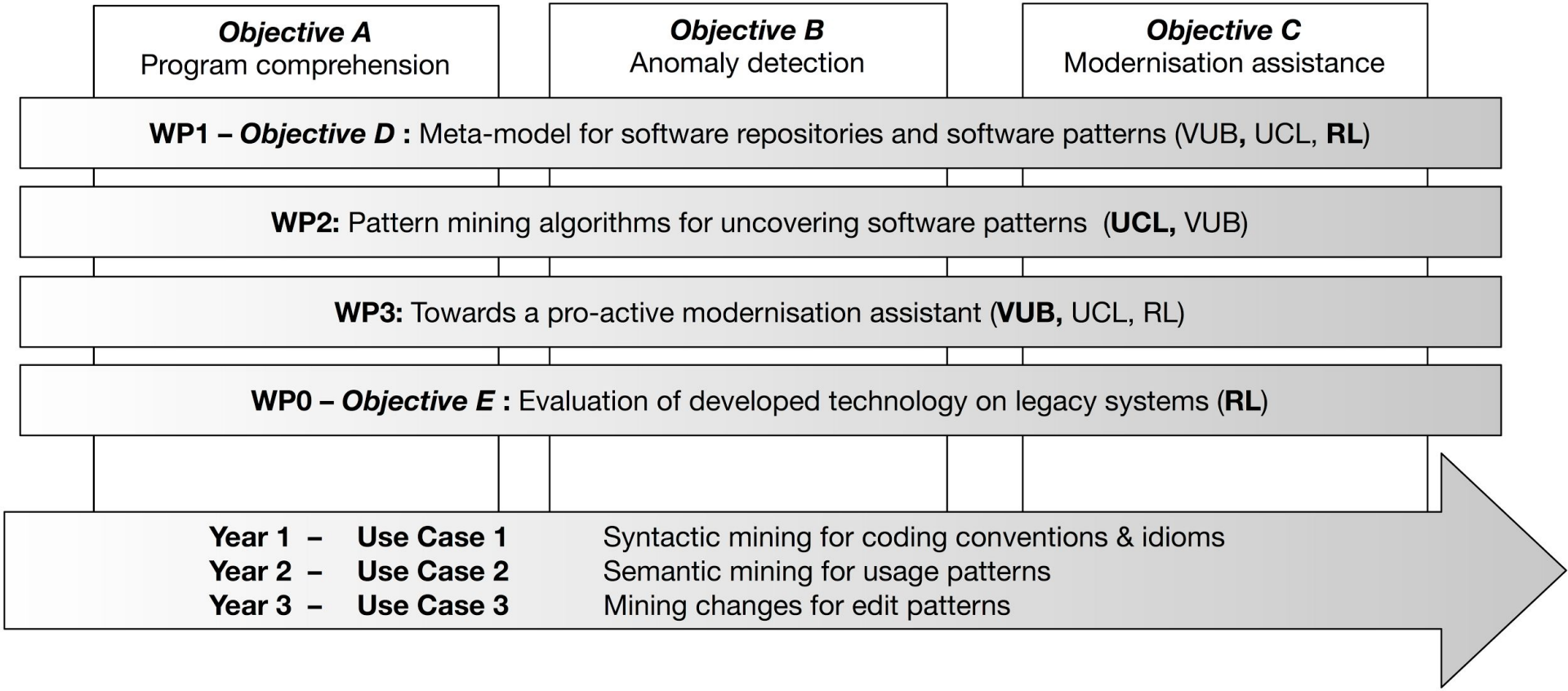
21. Use cases matrix (E.1.1)
22. Work programme (E.3)
23. WP1 Meta-model (E.3)
24. WP1 Deliverables (E.3)
25. WP2 Pattern mining (E.3)
26. WP2 Deliverables (E.3)
27. WP3 Pro-active modernisation assistant through pattern matching (E.3)
28. WP3 Deliverables (E.3)
29. WP0 Tasks (E.3)
30. WP0 Deliverables (E.3)
31. Planning (E.4)
32. Budget – Raincode Labs
33. Budget – VUB
34. Budget – UCL
35. Financial details of Raincode
36. Consortium agreement and ownership
37. AI as job creator

...

(E.1.1) Use cases matrix

	Objective A: Program comprehension	**Objective B: Anomaly detection**	**Objective C: Modernisation assistance**
Y1: syntactic mining for coding idioms	*UC1.A* Uncovered syntactic patterns	*UC1.B* Code is similar but not identical	*UC1.C* Proposing actions to adhere to idioms and conventions
Y2: semantic mining for usage patterns	*UC2.A* API, library, framework usage patterns in legacy systems	*UC2.B* Suspected violations of discovered usage patterns	*UC2.C* On-demand recommendations to correct violations
Y3: mining changes for edit patterns	*UC3.A* Frequently changed code locations	*UC3.B* Near-miss edit matches	*UC3.C* Automated proactive modernisation aspects
Industrial evaluation of results (WP0 and WP3)	T0.1 legacy importers T0.2 case study T3.1 pattern browser	T0.3 legacy importer T0.4 case study T3.2 on-demand matching	T0.5 legacy importer T0.6 case study T3.3 proactive matching

(E.3) Work Programme



WP1: Meta-model for software repositories and patterns

WP leader: RL; Overall effort: 21MMs (M01 to M27)
Effort per partner: RL 3MMs - VUB 9MMs - UCL 9MMs

Task 1.1: meta-model for representing syntactic information and coding idioms

- Task leader: RL; Overall effort: 7MMs (M01–M03)
  Effort per partner: RL 1MMs - VUB 3MMs - UCL 3MMs

Task 1.2: meta-model for representing semantic information and usage patterns

- Task leader: UCL; Overall effort: 7MMs (M13–M15)
  Effort per partner: RL 1MMs - VUB 3MMs - UCL 3MMs

Task 1.3: meta-model for representing version information and edit patterns

- Task leader: VUB; Overall effort: 7MMs (M25–M27)
  Effort per partner: RL 1MMs - VUB 3MMs - UCL 3MMs

WP1: Meta-model for software repositories and patterns

Deliverables:

- P1.1 Prototype implementation of the meta-model and an accompanying importer for syntactic information
- P1.2 Prototype implementation of its representation for coding conventions & idioms
- P1.3 Extension of meta-model and importer with support for semantic information
- P1.4 Prototype implementation of its representation for usage patterns
- P1.5 Extension of meta-model and importer with support for change information
- P1.6 Prototype implementation of its representation for edit patterns

WP2: Mining algorithms for uncovering software patterns

WP leader: UCL ; Overall effort: 36MMs (M04 to M36)
Effort per partner: VUB 18MMs - UCL 18MMs

Task 2.1: Mining to discover syntactic coding conventions and idioms

- WP leader: UCL; Overall effort: 12MMs (M04–M06 + M10–M12)
  Effort per partner: VUB 6MMs - UCL 6MMs

Task 2.2: Mining to discover usage patterns

- WP leader: UCL; Overall effort: 12MMs (M16–M18 + M22–M24)
  Effort per partner: VUB 6MMs - UCL 6MMs

Task 2.3: Mining to discover edit patterns

- WP leader: VUB; Overall effort: 12MMs (M28–M30 + M34–M36)
  Effort per partner: VUB 6MMs - UCL 6MMs

WP2: Mining algorithms for uncovering software patterns

Deliverables:

- P2.1 Frequent subgraph mining system for discovering patterns in attributed graphs, optimized for directed acyclic graphs
- R2.2 Report on its application to mining for coding conventions & idioms in syntactic information represented according to the meta-model of P1.1
- P2.3 Adaptation of the frequent subgraph mining algorithm to support usage patterns
- R2.4 Report on its application to mining for usage patterns in semantic information represented according to the meta-model of P1.3
- P2.5 Adaptation of the frequent subgraph mining algorithm to support edit patterns
- R2.6 Report on its application to mining for edit patterns in change information represented according to the meta-model of P1.5

WP3: Matching algorithms for modernization assistance

WP leader: VUB; Overall effort: 24MMs (M07 to M33)
Effort per partner: RL 6MMs - VUB 9MMs - UCL 9MMs

Task 3.1: Browser for inspecting mined coding idioms

- WP leader: RL; Overall effort: 8MMs (M07–M09)
  Effort per partner: RL 2MMs - VUB 3MMs - UCL 3MMs

Task 3.2: On-demand detection of partial matches within a software snapshot

- WP leader: UCL; Overall effort: 8MMs (M19–M21)
  Effort per partner: RL 2MMs - VUB 3MMs - UCL 3MMs

Task 3.3: Pro-active detection of matches within current modernisation context

- WP leader: VUB; Overall effort: 8MMs (M31–M33)
  Effort per partner: RL 2MMs - VUB 3MMs - UCL 3MMs

WP3: Matching algorithms for modernization assistance

Deliverables:

- P3.1 Prototype offering assistance for program comprehension in the form of a browser for mined patterns
- P3.2 Prototype offering assistance for anomaly detection in the form of a browser for partial matches of mined patterns
- P3.3 Prototype offering assistance for program modernisation in the form of edit recommendations that complete a partial match for a mined pattern

WP0: Evaluation of Technology on Legacy Systems

WP leader: RL; Overall effort: 27MMs (M02 to M36); Effort per partner: RL 27MMs

Task 0.1: Importer of syntactic information for meta-model

- Task leader: RL; Overall effort: 3MMs (M02—M04); Effort per partner: RL 3MMs

Task 0.2: Evaluation by mining legacy software for coding idioms

- Task leader: RL; Overall effort: 6MMs (M05–M07 + M10—M12); Effort per partner: RL 6MMs

Task 0.3: Importer of semantic information for meta-model

- Task leader: RL; Overall effort: 3MMs (M14–M16); Effort per partner: RL 3MMs

Task 0.4: Evaluation by mining legacy software for usage patterns

- Task leader: RL; Overall effort: 6MMs (M17–M19 + M22–M24); Effort per partner: RL 6MMs

Task 0.5: Importer of version information for meta-model

- Task leader: RL; Overall effort: 3MMs (M26—M28); Effort per partner: RL 3MMs

Task 0.6: Evaluation by mining legacy software for edit patterns

- Task leader: RL; Overall effort: 6MMs (M29–M31 + M34—M36); Effort per partner: RL 6MMs

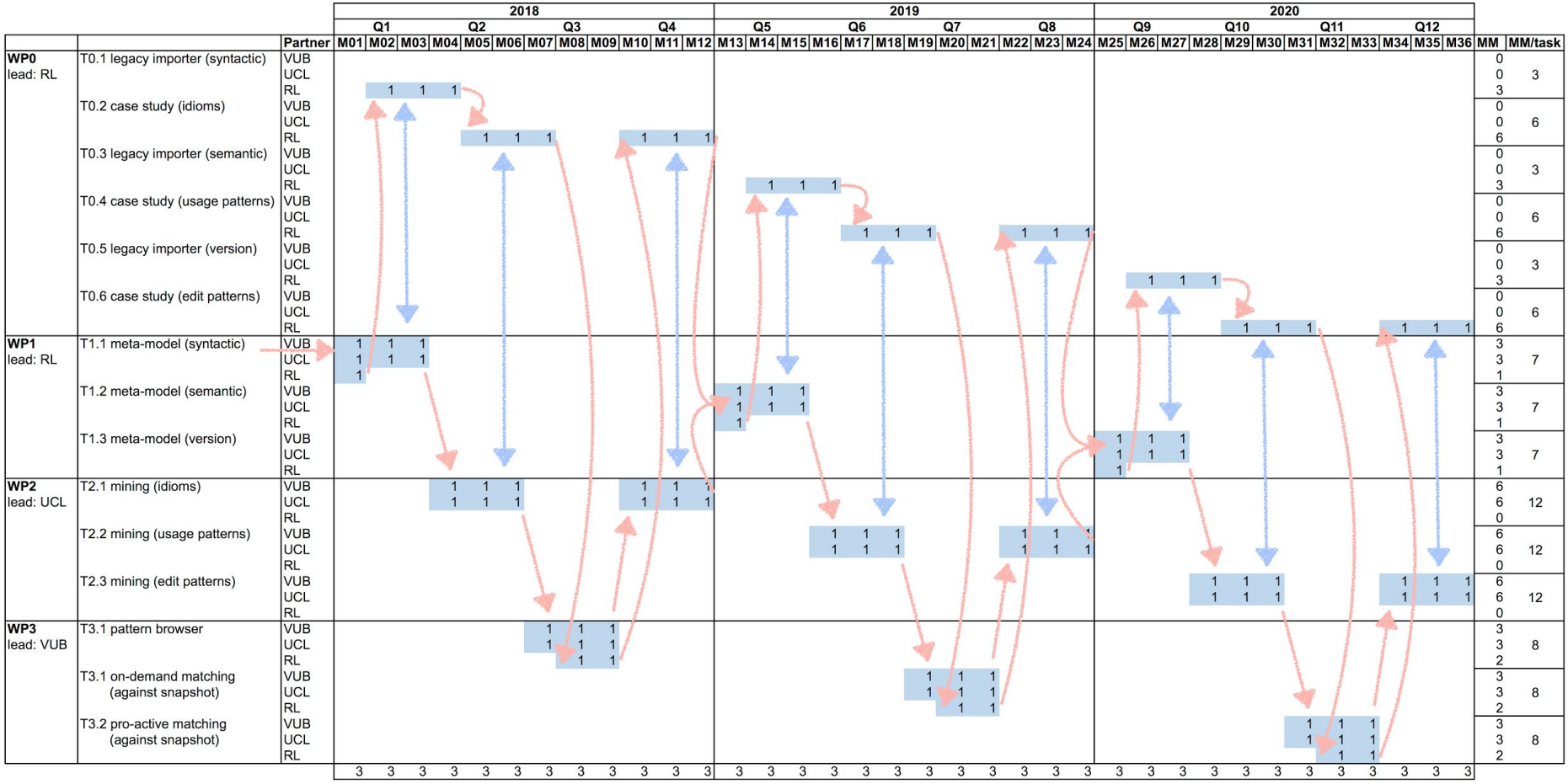
WP0: Evaluation of Technology on Legacy Systems

Deliverables:

- P0.1 Prototype of legacy software importer (syntactic information)
- R0.2 Report on evaluation of support for mining coding conventions & idioms
- P0.3 Extension of legacy software importer (semantic information)
- R0.4 Report on evaluation of support for mining usage patterns
- P0.5 Extension of legacy software importer (version information)
- R0.6 Report on evaluation of support for edit patterns

(E.4) Planning

Total MMs		
	VUB	36
	UCL	36
	RL	36



BUDGET — RAINCODE LABS

					Year 1	Year 2	Year 3
1.	**Staff costs**				**90,000 €**	**90,000 €**	**90,000 €**
1.1	Staff				90,000 €	90,000 €	90,000 €
	Name	***Diploma***	***Function***	***Allocation (%)***	***Cost***	***Cost***	***Cost***
	NN	PhD in Computer Science	Analyst / developer	100.00%	90,000 €	90,000 €	90,000 €
1.2	Self-employed persons				0 €	0 €	0 €
	/	/	/	/	/	/	/
2.	**Operating costs**				**5,000 €**	**5,000 €**	**5,000 €**
General academic operating costs	Conference attendance	*1 conference in the field of pattern mining; 1 conference in the field of software engineering & programming languages*	2000€/conference x 2/yr x 3 yrs		4,000 €	4,000 €	4,000 €
	Additional workshops (meetings) with clients	*2 trips per year to internationally located clients with problems where the newly developed methods will be applicable, needed to obtain relevant but sensitive data on their premises and complete the feedback loop to the researchers*	500€ / year x 2/yr 3 years		1,000 €	1,000 €	1,000 €
3.	**Investment costs**				**5,000 €**	**0 €**	**0 €**
Computer equipment	Powerful high-end laptop for heavy duty data mining activities	*E.g., a fully configured MacBook Pro 15" with a 3.1 Ghz quad-core Intel core iT processor and 2TB of SSD storage*	(1 x 5000€ x 36 x 100%) / 36		5,000 €		
4.	**General costs**				**9,500 €**	**9,500 €**	**9,500 €**
	10% (Operating costs + Salaried staff costs)						
5.	**Subcontracting**				**0 €**	**0 €**	**0 €**
	/	/	/	/	/	/	/
TOTAL					**109,500 €**	**104,500 €**	**104,500 €**
		TOTAL BUDGET			**318,500 €**		
		TOTAL SUBSIDY	**60%**		**191,100 €**		
		PARTNER'S SHARE			*127,400 €*		

BUDGET

VUB

					Year 1	Year 2	Year 3
1.	**Staff costs**				**74.454 €**	**77.682 €**	**91.007 €**
1.1	Staff				74.454 €	77.682 €	91.007 €
	Name	***Diploma***	***Function***	***Allocation (%)***	***Cost***	***Cost***	***Cost***
	NN	PhD in Computer Science	Post-doctoral researcher	100,00%	74.454 €	77.682 €	91.007 €
1.2	Self-employed persons				0 €	0 €	0 €
	/	/	/	/	/	/	/
2.	**Operating costs**				**5.000 €**	**5.000 €**	**5.000 €**
General academic operating costs	Conference attendance	*1 conference in the field of pattern mining; 1 conference in the field of software engineering & programming languages*	2000€/conference x 2/yr x 3 yrs		4.000 €	4.000 €	4.000 €
	Printing and organisatorial costs for disseminatation of results	*posters, postage, workhop organisation, article publication, …*	500€ / year x 3 years		500 €	500 €	500 €
	Software	*software products, services and updates*	500€ / year x 3 years		500 €	500 €	500 €
3.	**Investment costs**				**5.000 €**	**0 €**	**0 €**
Computer equipment	Powerful high-end laptop for heavy duty data mining	*A fully configured MacBook Pro 15" with 3.1 Ghz quad-core Intel core iT processor and 2TB of SSD storage*	(1 x 5000€ x 36 x 100%) / 36		5.000 €		
4.	**General costs**				**7.945 €**	**8.268 €**	**9.601 €**
	10% (Operating costs + Salaried staff costs)						
5.	**Subcontracting**				**0 €**	**0 €**	**0 €**
	/	/	/	/	/	/	/
TOTAL					**92.399 €**	**90.950 €**	**105.608 €**

TOTAL BUDGET		**288.957 €**
TOTAL SUBSIDY	**100%**	**288.957 €**
PARTNER'S SHARE		*0 €*

BUDGET

UCL

					Year 1	Year 2	Year 3
1.	**Staff costs**				**90.736 €**	**92.551 €**	**99.533 €**
1.1	Staff				90.736 €	92.551 €	99.533 €
	Name	***Diploma***	***Function***	***Allocation (%)***	***Cost***	***Cost***	***Cost***
	NN	PhD in Computer Science	Senior post-doc researcher	100,00%	90.736 €	92.551 €	99.533 €
1.2	Self-employed persons				0 €	0 €	0 €
	/	/	/	/	/	/	/
2.	**Operating costs**				**5.000 €**	**5.000 €**	**5.000 €**
General academic operating costs	Conference attendance	*1 conference in the field of pattern mining; 1 conference in the field of software engineering & programming languages*	2000€/conference x 2/yr x 3 yrs		4.000 €	4.000 €	4.000 €
	Printing and organisatorial costs for disseminatation of results	*posters, postage, workhop organisation, article publication, ...*	500€ / year x 3 years		500 €	500 €	500 €
	Software	*software products, services and updates*	500€ / year x 3 years		500 €	500 €	500 €
3.	**Investment costs**				**5.000 €**	**0 €**	**0 €**
Computer equipment	Powerful high-end laptop for heavy duty data mining	*A fully configured MacBook Pro 15" with 3.1 Ghz quad-core Intel core iT processor and 2TB of SSD storage*	(1 x 5000€ x 36 x 100%) / 36		5.000 €		
4.	**General costs**				**9.574 €**	**9.755 €**	**10.453 €**
	10% (Operating costs + Salaried staff costs)						
5.	**Subcontracting**				**0 €**	**0 €**	**0 €**
	/	/	/	/	/	/	/
TOTAL					**110.310 €**	**107.306 €**	**114.986 €**

TOTAL BUDGET		**332.602 €**
TOTAL SUBSIDY	**100%**	**332.602 €**
PARTNER'S SHARE		*0 €*

Financial Details of Raincode

Année	*2012*	*2013*	*2014*	*2015*	*2016*	*2017*
Chiffre d'Affaire	€ 1.658.674,00	€ 1.744.686,00	€ 1.375.521,00	€ 1.943.948,00	€ 2.638.405,00	€ 3.200.000,00
Marge brute d'exploitation	€ 481.792,00	€ 97.138,00	€ 47.715,41	€ 326.027,00	€ 311.808,37	€ 1.200.000,00
Benef net	€ 328.510,29	€ 70.825,00	€ 42.276,03	€ 210.754,00	€ 189.055,94	€ 784.679,25

Consortium agreement & ownership

- Efforts will result in a generic framework for modernisation assistants
  - A language-parametric meta-model
  - Pattern mining algorithms
  - Pattern matching algorithms
  - Modernisation recommendations
- Framework needs to be instantiated for specific legacy languages by complementing it with dedicated importers
- Generic guidelines on IP are understood and agreed upon
  - IP for a particular task or deliverable belongs to the partners contributing to it
  - Examples according to proposed work plan:
    e.g., language-parametric meta-model is joint academic-industrial IP,
    but importer for legacy language and mined patterns for such language is industrial IP
    e.g., dedicated mining and matching algorithms are mostly joint academic IP,
    but modernisation support is joint academic-industrial IP
- Background knowledge will be listed before the start of the project
- Exploitation of joint results dependent on permission of co-owners
- Licensing of other results based on market prices

AI as job creator

Media often characterises AI as a job killer that will replace people in workplaces by robots or intelligent agents.

Industry, however, employs AI to solve challenges for which they currently lack the resources or corporate worker base to succeed.

“What makes the best AI projects stand out is that they allow for solutions that previously would have been impossible to conceive. They include what seems like human insights but at a volume humans could never achieve.”