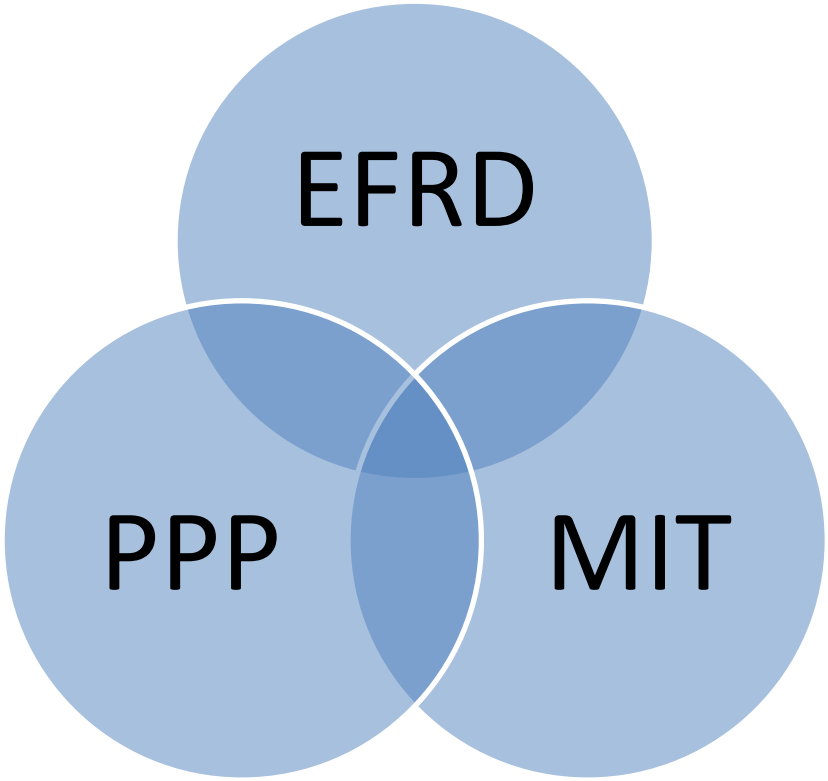
Netherlands Enterprise Agency



Comparison of R&D-topics in 3 innovation programmes

Maarten van Leeuwen

March 13 2018

Research Question

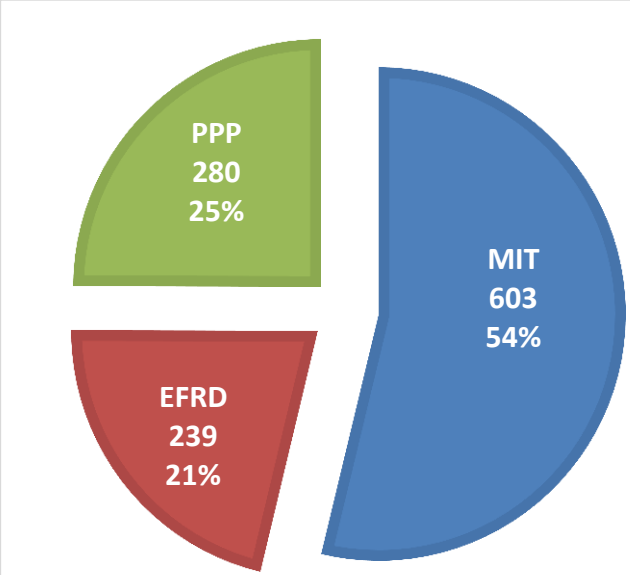
What are the differences in R&D-topics between the project portfolios of three innovation stimulation programmes in The Netherlands?

Data and tool

- Three innovation programmes:
  - European Fund for Regional Development (EFRD)
  - Public-private partnership programme (PPP)
  - SME Innovation programma (MIT)

- In total the portfolios include 1.122 unique projects
`

- Analysis with topic modelling technique



Approach topic modelling

- Analysis of project descriptions with an algorithm (automated) to find topics
- A topic can be defined as “a repeating pattern of co-occurring terms in a text”
- Every project description is composed of a combination of topics
- In most project descriptions 1 topic is leading

How does the algorithm work

- Defines bottom up what the topics are in a specific dataset
- For this dataset the optimum number of topics turned out to be 21
- For each topic a characterisation is given by means of word frequency (allowing to give each topic a proper name)
- Defines for each project how good it fits in each topic (this takes multiple iterations)
- Each project classified in the topic with the highest chance

A few examples of topics

Topic title	Description: projects with a high score on this topic…	Key words	Illustrative fragment from a project description	Number of projects with highest score on this topic
Health (care)	are aimed at innovation in healthcare (more care than cure)	People, care, health	"as a means to improve communication between patient and health professional"	54
Emissions fossil fuels	solve problems around emissions related to the use of fossil fuels	System, emission, fuel, problem	"while the temporary LNG infrastructure is immature, the choice mostly falls on a dual-fuel gas motor […]. Because the costs to replace the engine are too high, a "retro-fit" dual-fuel solution is necessary.	69
ICT & data	use big data en ict to find solutions for all kinds of problems	Data, platform, information, software	"the use of social data make it possible to analyse, guard and research how users behave on different social media platforms"	82

Distribution of projects on topics

	Percentage of EFRD	Percentage of MIT	Percentage of PPP
Hightech products & systems	10%	3%	5%
(Renewable) Energy	8%	7%	3%
Dutch companies	6%	4%	6%
Emissions fossil fuels	6%	7%	5%
Innovative materials	6%	6%	4%
Application virtual reality	6%	5%	4%
Life Sciences (care)	5%	5%	3%
Innovations in lifestock	5%	5%	1%
Technological applications	5%	2%	6%
Sensors en systems	5%	5%	3%
Feasability biobased	5%	10%	3%
Coating based	5%	3%	1%
Collaboration	4%	3%	2%
ICT and data	4%	9%	6%
Life Sciences (cure)	4%	3%	15%
Micro	4%	3%	7%
Cultivation optimization	3%	5%	3%
Feasability in general	3%	7%	1%
Circular in (greenhouse) horticulture	3%	3%	9%
Models	2%	1%	15%
Several innovative products	2%	3%	2%
	100%	100%	100%

	Percentage in EFRD	Percentage in MIT	Percentage in PPP
Hightech products & systems	41%	38%	23%
Innovations in lifestock	34%	76%	5%
Coating based	31%	60%	11%
Technological applications	30%	34%	39%
Dutch companies	28%	43%	30%
Collaboration	28%	58%	14%
(Renewable) Energy	28%	65%	11%
Application virtual reality	26%	54%	22%
Health (care)	24%	63%	15%
Sensors en systems	24%	64%	14%
Innovative materials	22%	63%	16%
Emissions fossil fuels	22%	65%	19%
Micro	20%	40%	42%
Feasability biobased	15%	78%	9%
Cultivation optimization	15%	67%	19%
Health (cure)	14%	29%	59%
Several innovative products	14%	72%	17%
Circular in (greenhouse) horticulture	13%	35%	52%
ICT and data	12%	68%	20%
Feasability in general	11%	82%	7%
Models	8%	15%	77%

Conclusions

- The programmes seem to be complementary
- EFRD portfolio shows relatively many Hightech projects
- Relatively many Biobased en ICT projects in SME programme
- Relatively many Health (cure) and Models projects in the PPP programma

Advantages op topic modelling

- Bottom up (topics are determined by the algorithm and are not predetermined)
- Automated so no human errors or bias in the classification

Attention point

- Algorithm is sensitive for the length of texts
- When you compare different subsets of data, make sure the descriptions are more or less equal in length

Contact

Maarten van Leeuwen

l **M** +31 (0)6 5256 2245 l **E** maarten.vanleeuwen@rvo.nl

Address: Croeselaan 15 l 3521 BJ l Utrecht l The Netherlands

P.O. box: Postbus 8242 l 3503 RE l Utrecht l The Netherlands

http://www.rvo.nl