

“Can we use big data for skills anticipation and matching?”

The case of Online Job Vacancies

Fabio Mercorio, PhD in AI

Assistant Professor in AI and Data Science

University of Milano-Bicocca

Geneva, September 20th, 2019



*interuniversity
research centre
for public services*



International
Labour
Organization

The CRISP Centre @Unimib



*interuniversity
research centre
for public services*

The Interuniversity research centre on public services (CRISP) is an interdisciplinary academic network of Universities in Milan, guided by Unimib



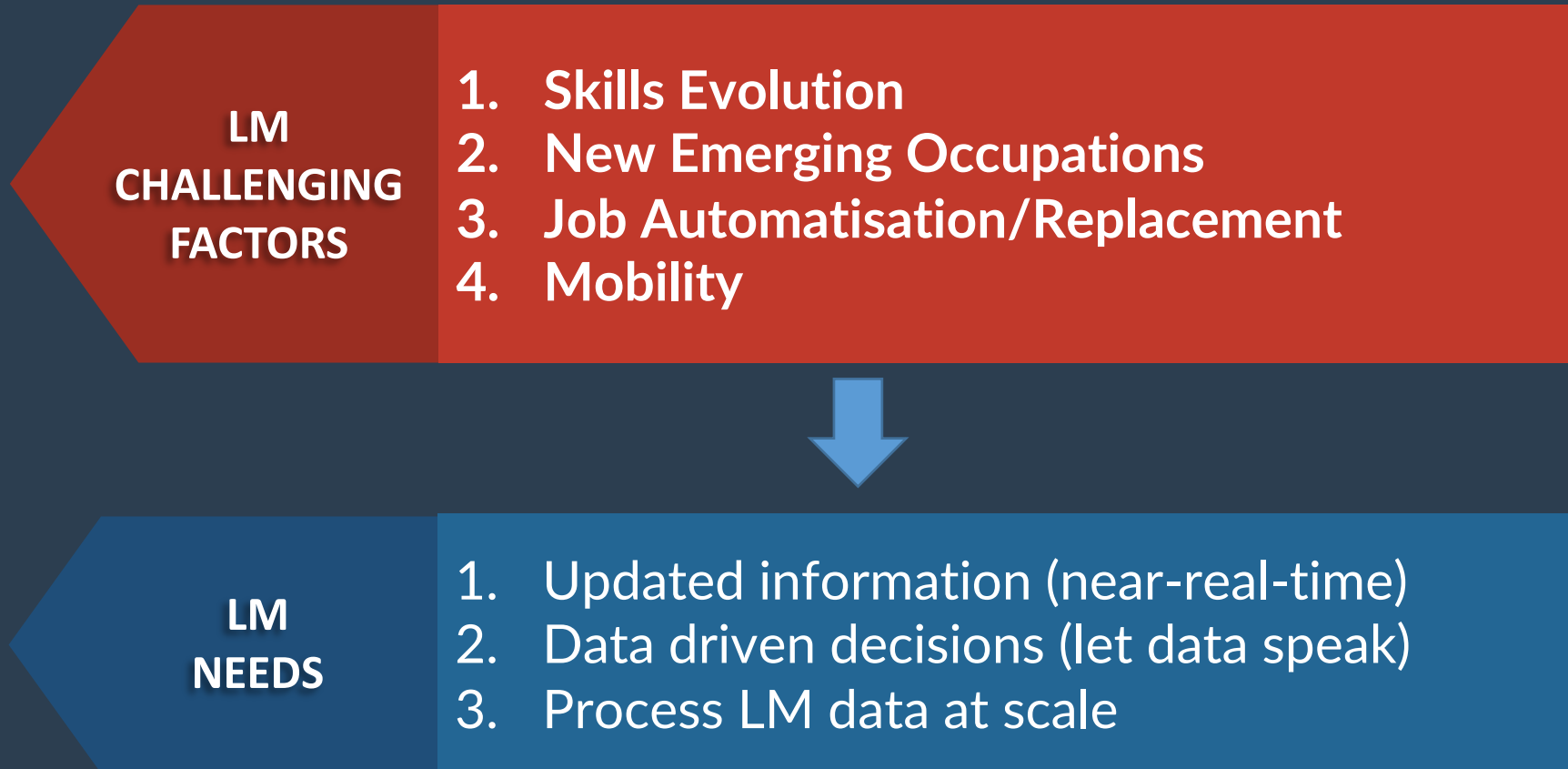
Research Goal. To study and support policy and decision makers in the analysis of socio-economic phenomena through novel AI, Big Data and Statistics algorithms and pipelines, by processing statistical, administrative and web data as well.



Some Research Projects on Big Data for LM [selection]

- (2014-2016) Cedefop I [**Prototype real-time skill and vacancies analysis**]. 5 countries (United Kingdom, Ireland, Czech Republic, Italy and Germany). Granted by Cedefop
- (2016-ongoing) Cedefop II [**Development of Cedefop I real-time system for Europe**]. 28 EU Countries, 32 languages supported. Granted by Cedefop
- (2016-ongoing) ETF [**Guide on putting Big Data into LMI + source selection for Tunisia and Morocco**], Granted by The European Training Foundation
- (2016-ongoing) Italian Digital Competences Observatory 2017, 2018 and 2019 [**Estimate Skill Impact on ICT jobs**] granted by Italian ICT Unions.
- (2018-ongoing) EXCELSIOR [**Put OJV into Official Occupation Statistics**] granted by The Italian Unions of chambers of Commerce, Industry, Crafts and Agriculture
- (2017-ongoing) REPLY-VET [**Strengthening key competencies of low-skilled people in VET to cover future replacement positions**]. Erasmus+

Labour Market Challenging Factors



Labour Market Intelligence (LMI): Design, define and implement AI-based framework and algorithms to derive knowledge from labour market information

“Can we use big data for skills anticipation and matching?”

- **Occupation and Skill Discovery:** Focus on occupations and skills requested by the online-LM
- **Soft/Digital/Hard Skill Rates:** How to estimate the impact of digitalization within occupations?
- **New Emerging Occupations** on the basis of skill (dis)similarities
- **Training Course Design** through skills identification
- **Taxonomy Extension:** Improve skills/occupations taxonomy through

**How to deal with OJVs
at scale?**

Web Job Vacancy example

Job Title: Data Scientist.

Description: We're looking for a talented Computer Scientist to join our growing development team. Your expertise in data will help us take this to the next level. You will be responsible for identifying opportunities to further improve how we connect recruiters with jobseekers, and designing and implementing solutions. [...] Required skills and experience:

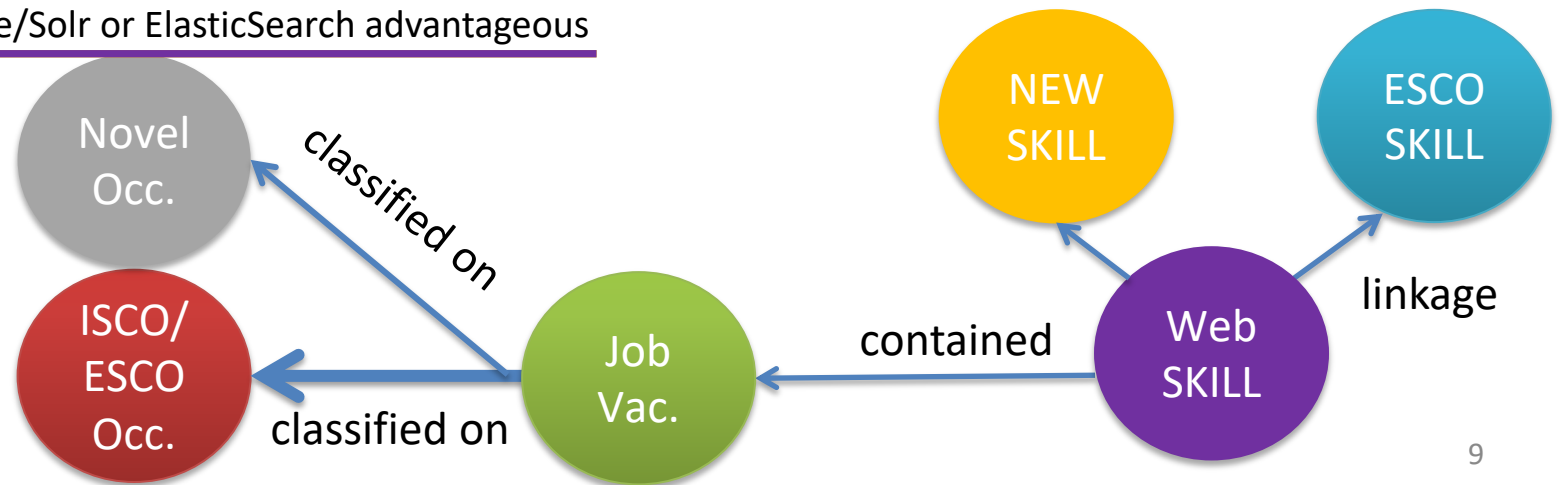
- SQL and relational databases;
- Data analysis with R (or Matlab);
- Processing large data sets with MapReduce and Hadoop);
- Real time analytics with Spark, Storm or similar;
- Machine Learning;
- Natural Language Processing (NLP) and text mining;
- Development in C++, Python, Perl;
- Experience with search engines e.g. Lucene/Solr or ElasticSearch advantageous

Web Job Vacancy example

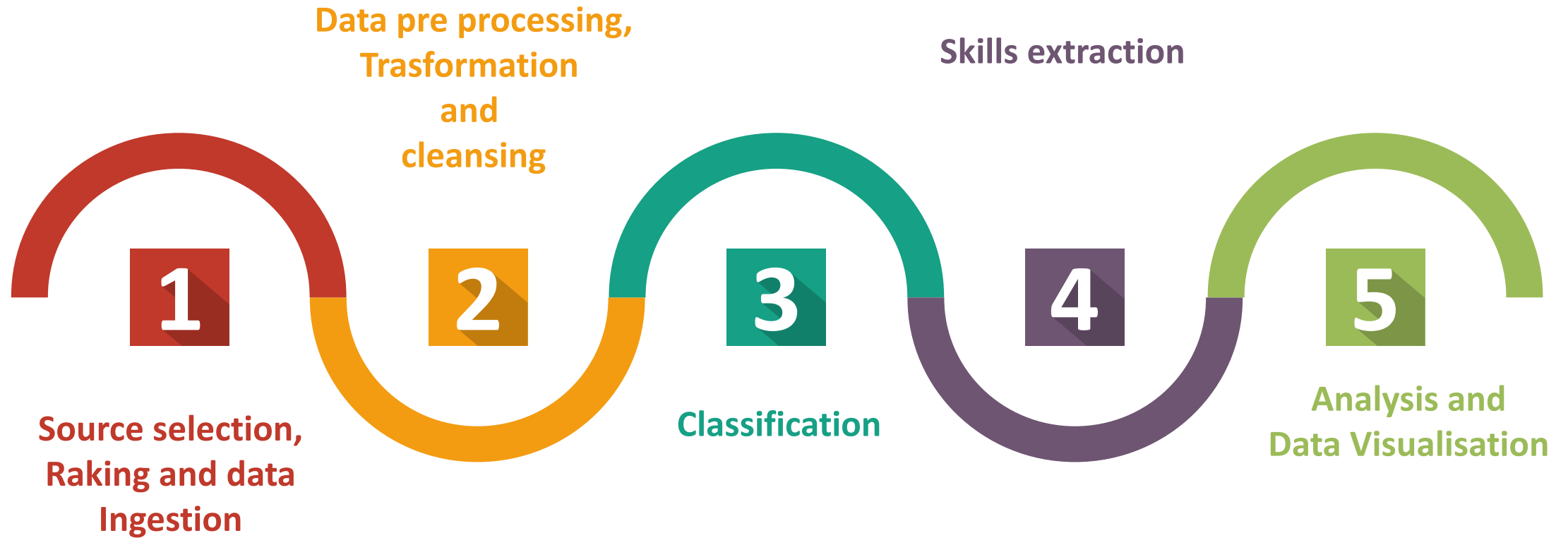
Job Title: Data Scientist.

Description: We're looking for a talented Computer Scientist to join our growing development team. Your expertise in data will help us take this to the next level. You will be responsible for identifying opportunities to further improve how we connect recruiters with jobseekers, and designing and implementing solutions. [...] Required skills and experience:

- SQL and relational databases;
- Data analysis with R (or Matlab);
- Processing large data sets with MapReduce and Hadoop);
- Real time analytics with Spark, Storm or similar;
- Machine Learning;
- Natural Language Processing (NLP) and text mining;
- Development in C++, Python, Perl;
- Experience with search engines e.g. Lucene/Solr or Elasticsearch advantageous



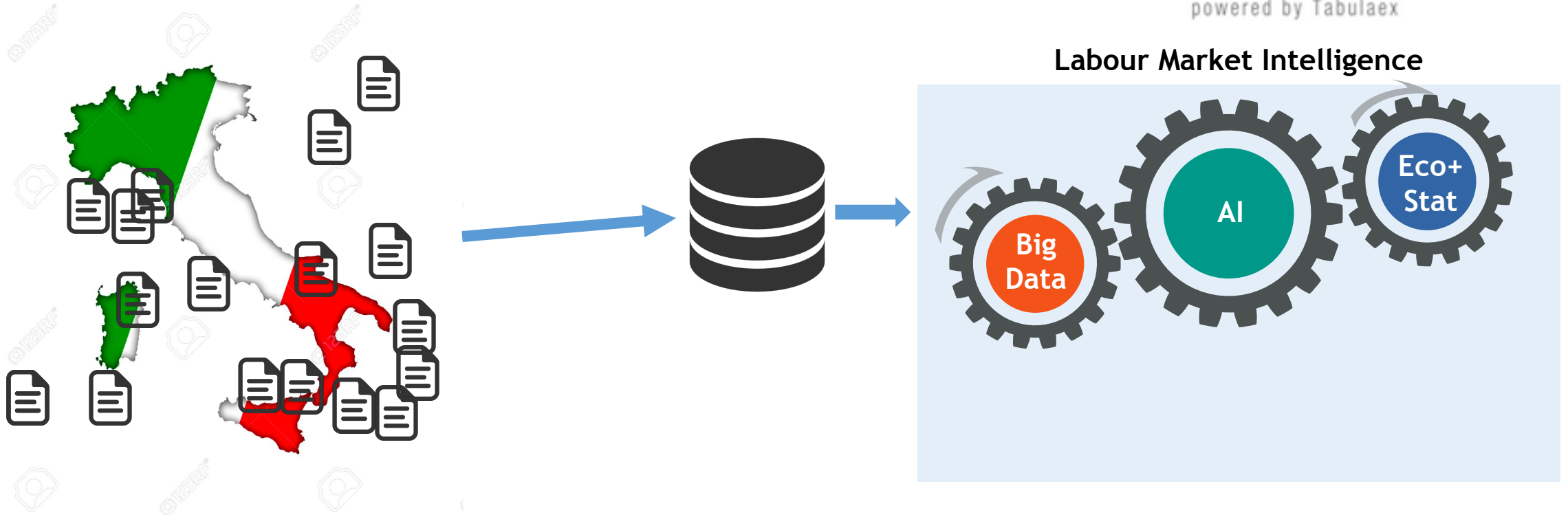
The Methodological path



ITALIAN Real-Time Labour Market Monitor



Labour Market Intelligence



OJV since 2013 – 4M+ vacancies unique

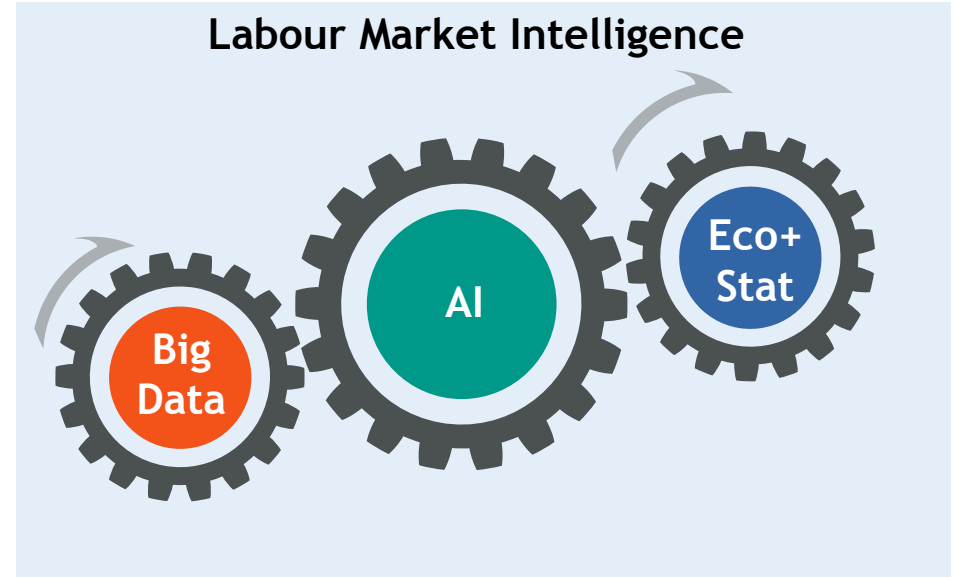
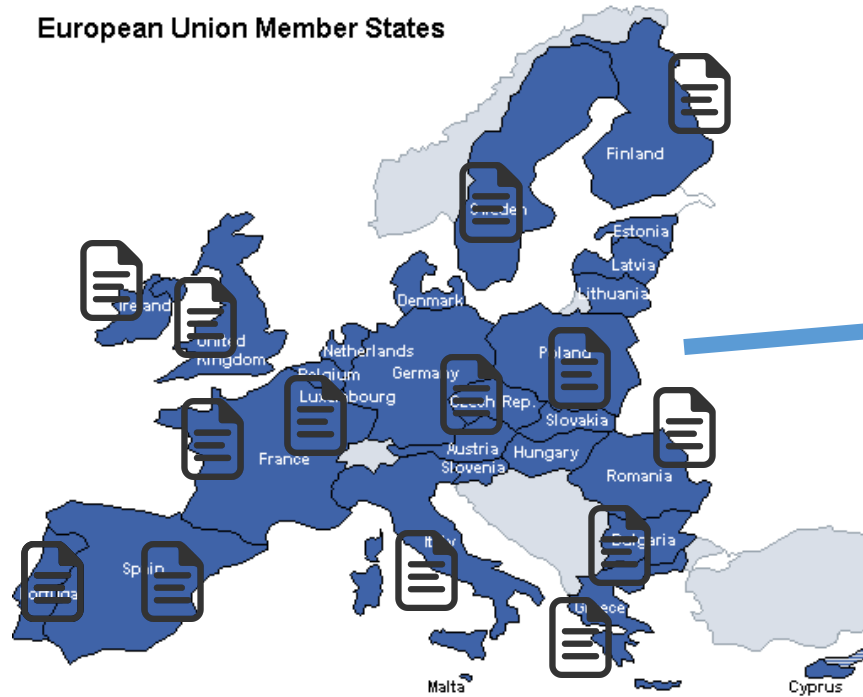
EUROPEAN Real-Time Labour Market Monitor



CEDEFOP

European Centre
for the Development
of Vocational Training

European Union Member States



28 EU Countries – 32 Languages – more than 6M unique vacancies per month

Occupation and Skill Discovery: Focus on
occupations and skills requested by the
online-LM

Live dance: Territorial dimension



powered by Tabulaeex

AREA

Area 

Nation	Region
--------	--------

Italy

Italy

Neighbouring and other Select...Distance (km) Period (months) 1Occupational Group 

HARD SKILLS



Search skill [i](#)



Selected skills [i](#)

Next



Skill frequency 0% - 100% vacancies

Soft/Digital/Hard Skill Rates: How to estimate
the impact of digitalization within occupations?

Compute Skills Rates



Goal: Estimate the pervasiveness of ICT in both ICT and not ICT-related jobs

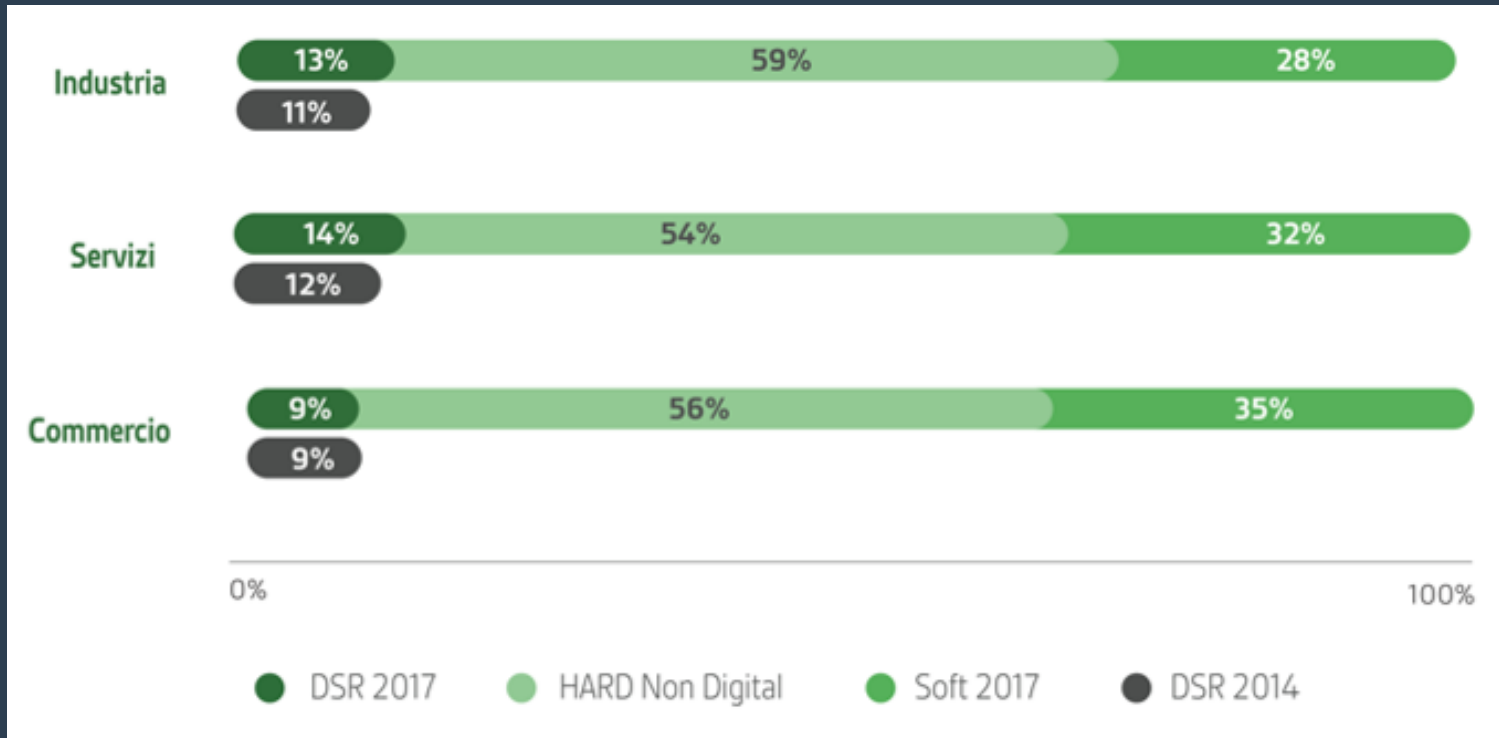
Idea: Exploit the informative power of Classified OJV for computing The Digital Skill Rate (DSR), Soft skill rate and Hard non digital Skill Rate

DSR estimates the incidence of digital skills in a single profession and comes from observing the pervasiveness of digital skills in all professions whether they are related to the ICT world or not.

Compute Skills Rates

SOURCE WOLLYBI

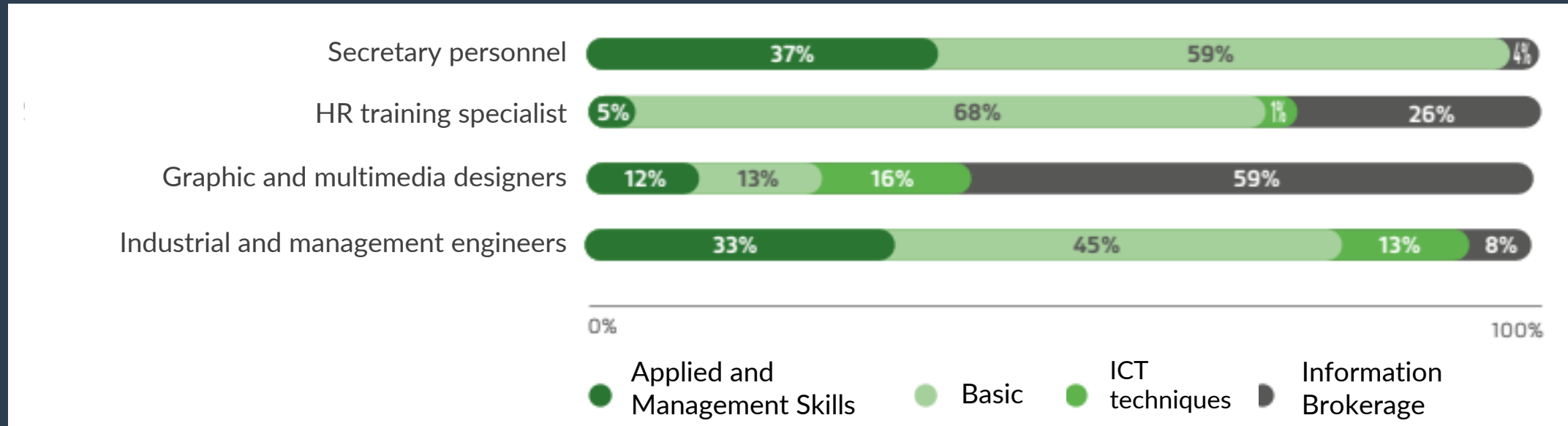
Demand of digital, specialist and soft skills - by sector



Ad hoc analyses at different level of granularity, here focusing on the «sector»....

Compute Skills Rates

SOURCE WOLLYBI



- **Applied and Management Skills** = ability to use tools and software to manage both operational and decisional processes
- **ICT Techniques Skill** = very specialized on solutions, platforms and programming languages
- **Basic Skill** = for everyday use of basic IT tools
- **Information Brokerage Skill** = for the use of IT tools aimed at corporate communication

... and more, looking at each occupation...

Credits to WollyBI: a trademark of TabulaeX

Compute Skills Rates [ESCO skills + novel]

SOURCE WOLLYBI

Applied and Management Skills				ICT techniques		Information Brokerage	
Occupation	Database usage	Programs for draughtsman	3D modelling	Front-end Website implementation	Web programming	Graphic Software Usage	SW markup usage
Graphic and multimedia designers	 25	 35	 3	 45	 2	 5	 2

Applied and Management Skills			Information Brokerage		
Occupation	Database usage	ERP	Digital data management	SEO Search Engine Optimiz.	Social Network Usage
HR training specialist	 45	 4	 45	 4	 25

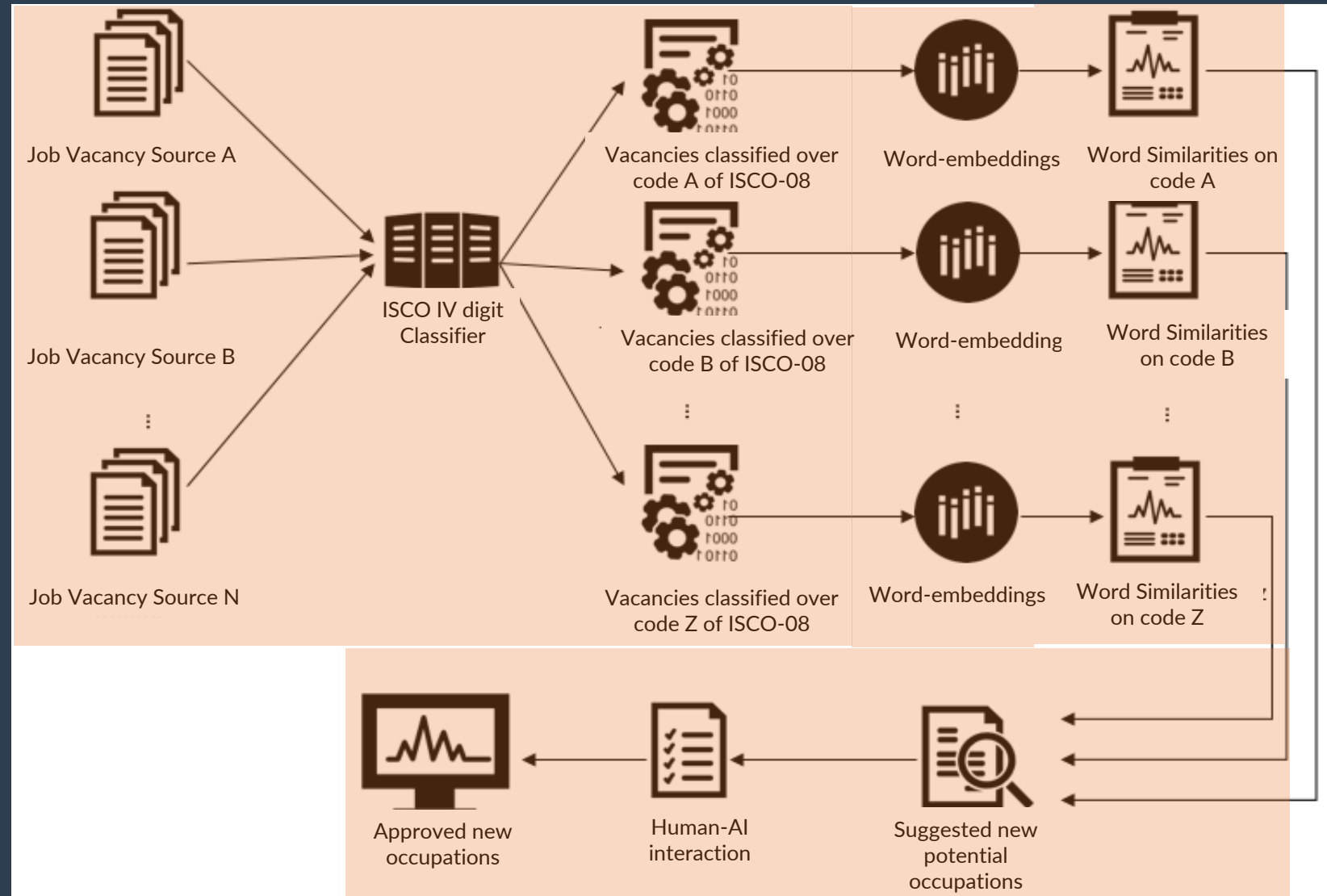
... and more, looking at elementary skills

Credits to WollyBI: a trademark of TabulaeX

New Emerging Occupations on the basis of
skill (dis)similarities

Detecting new emerging occupations through AI

1. Classify OJV over ISCO-iv digit
2. Build-up several vector-space representations of words (occupations and skills) to catch lexicon similarities between OJVs
3. Compute similarities between known terms (occupations and skills) and new ones
4. Suggest new potential occupations for Human-AI validation



(Some) New Emerging Occupations



Data Scientist



Cloud Computing



Cyber Security Expert



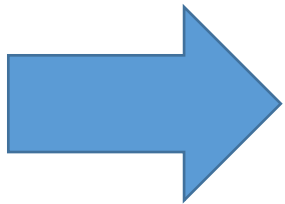
Business Intelligence Analyst



Big Data Analyst



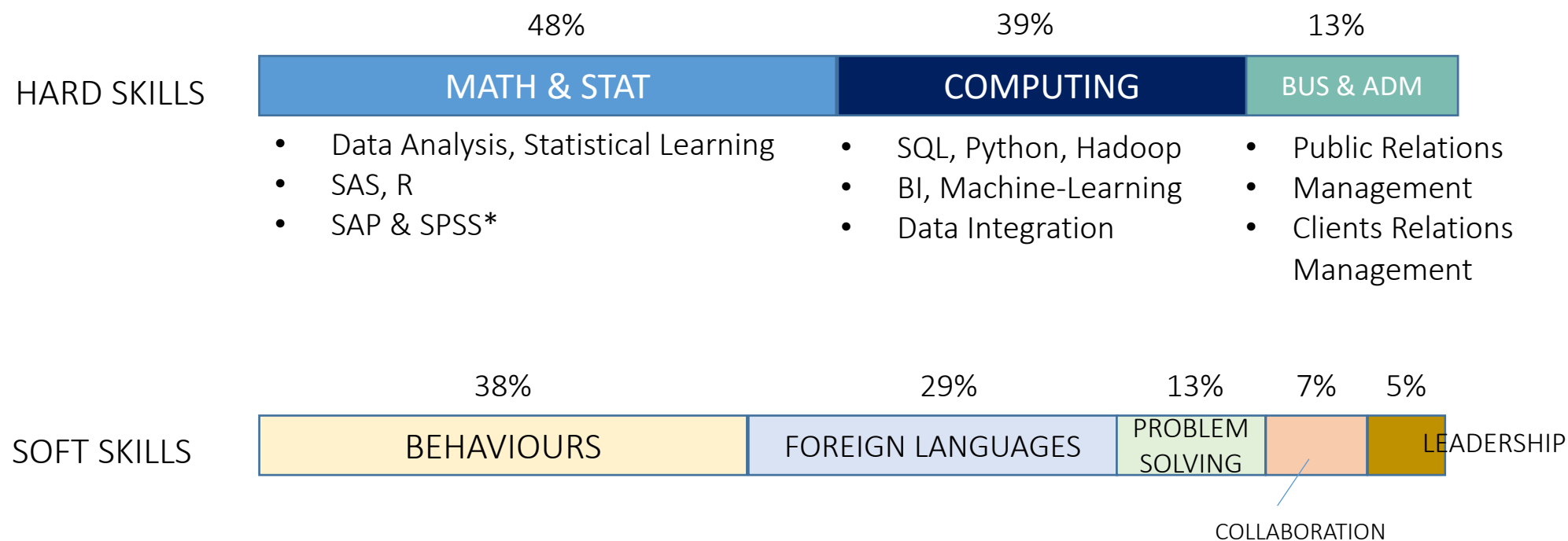
Social Media Marketing



9,000 Web Job Vacancies collected related to (some) new emerging occupations (above) between Jan-2014 and August-2019



DATA SCIENTIST – 1,7k vacancies 2014-2019

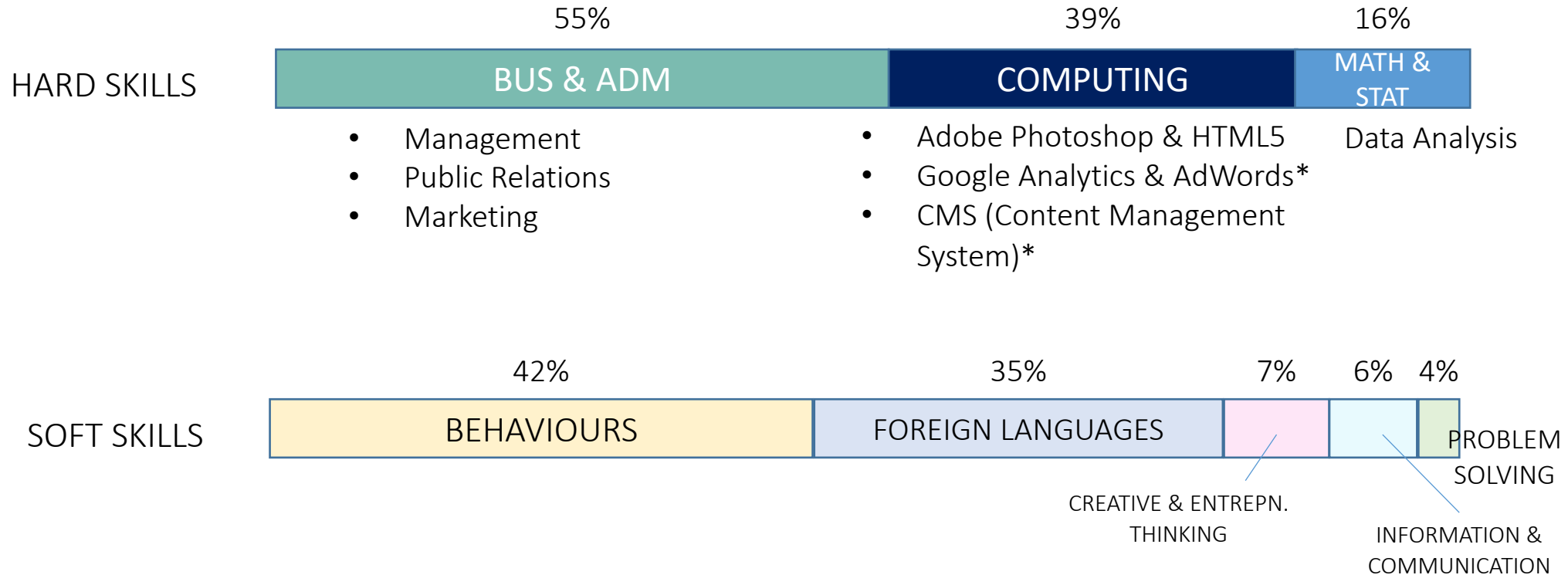


Variation 2019 vs 2018: +31%

Variation 2019 vs 2017: +149%

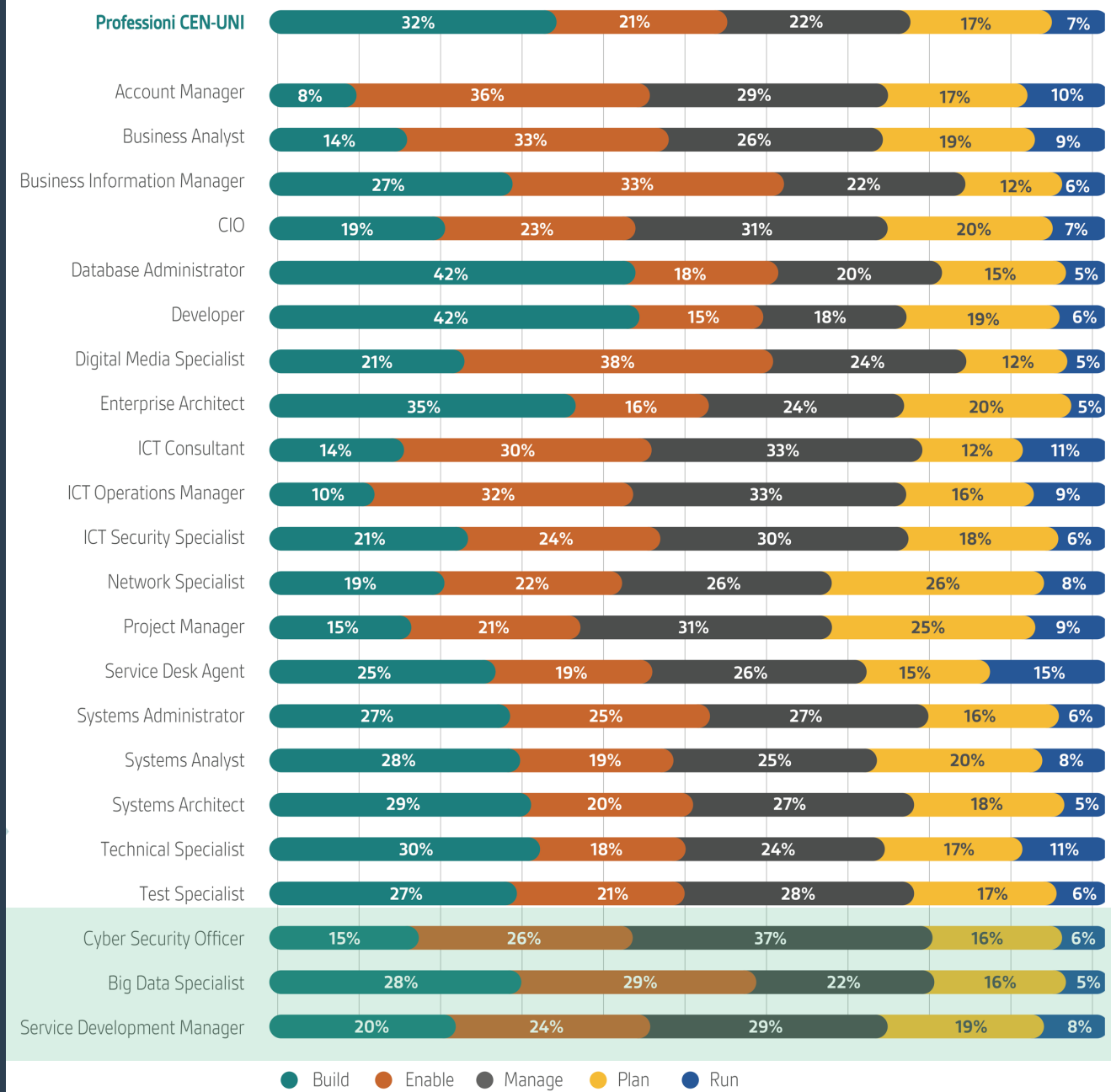


SOCIAL MEDIA SPECIALIST – 0,4k vacancies 2014-2019



Variation 2019 vs 2018: +105%

Variation 2019 vs 2017: +123%



Fonte: WollyBI



New Occupations can be compared against traditional ones using different taxonomies (the European Competence Framework in such a case)

Training Course Design through skills identification

<http://tiny.cc/masterBI>

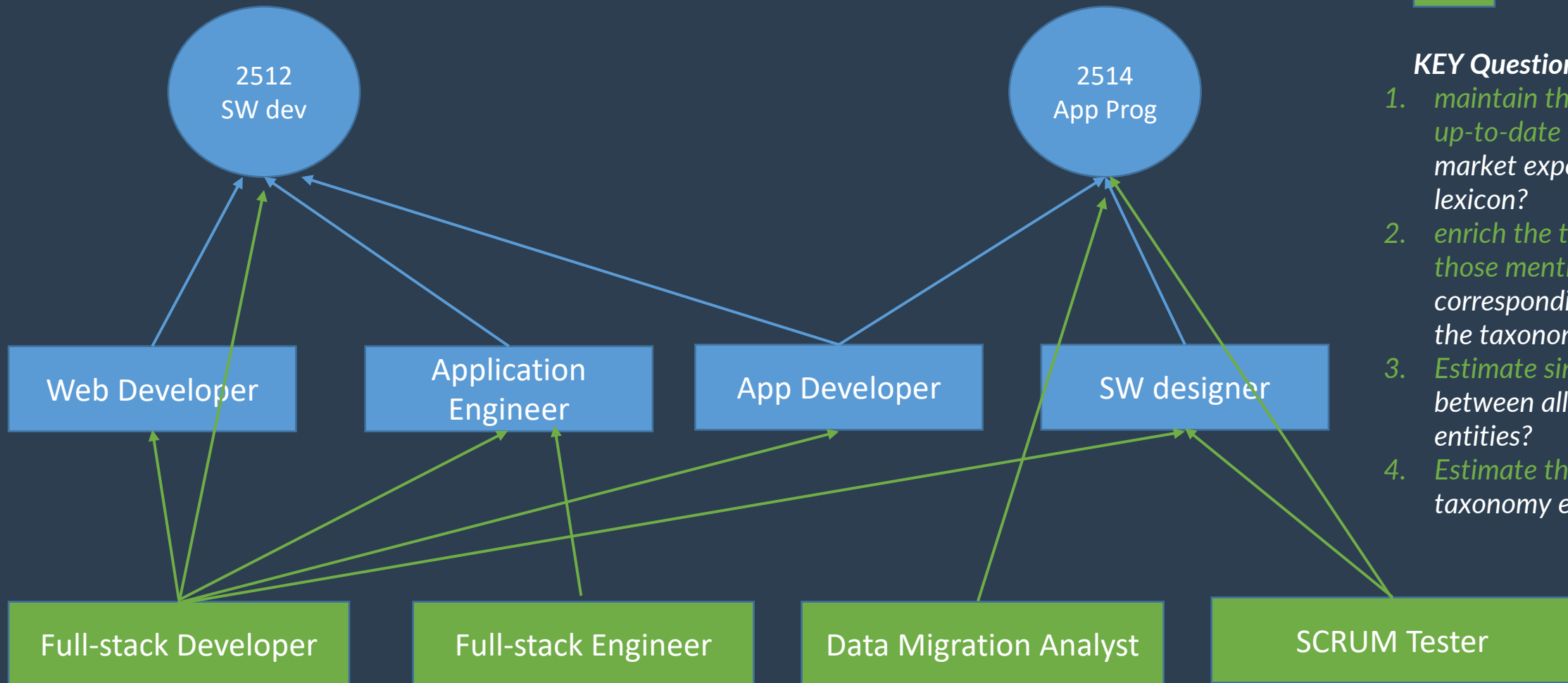
LMI for Course Design allows one to:



4. Iteratively refine and improve the whole course to stay focused on real LM expectations
5. **Design paths** according to student backgrounds guiding them through the course

Taxonomy Extension: Improve
skills/occupations taxonomy through
semantic similarities within OJVs

Main Idea



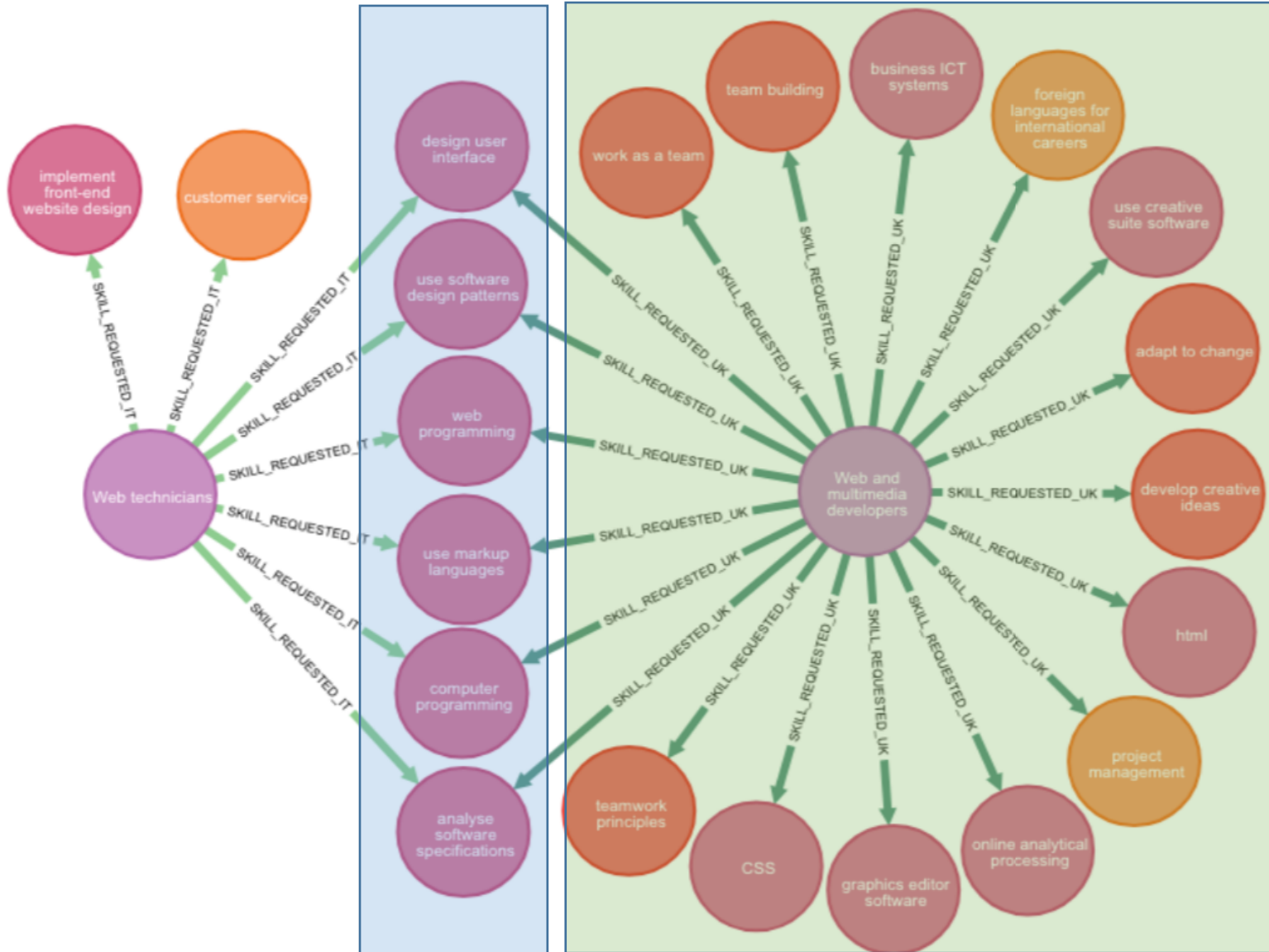
KEY Questions: How to...

1. *maintain the taxonomy up-to-date* with labour market expectation and lexicon?
2. *enrich the taxonomy with those mentions* to corresponding entities in the taxonomy?
3. *Estimate similarities* between all taxonomy entities?
4. *Estimate the relevance* of taxonomy entities?

Extending Skills Taxonomy

Skills in common

Skills GAP!!!



Compare different Web labour markets [IT, UK and DE here] to perform Skill Gap Analysis using the taxonomy as a baseline.

Question: “starting from a given occupation of the ISCO taxonomy for the ITA labour market, what is the occupation in the UK labour market whose requested skills with a better fit?”

Skills associated to “Web Technician” in ITA are more similar to a “Web and Multimedia Developer” in UK

Compare Different Countries (Web Technician)

OJV COLLECTED FROM UK, IT AND DE IN 2018

Nazione	Top 10 skill	<i>tf-idf</i>		
		IT	UK	DE
IT	web programming	100.0	38.0	100.0
IT	analyse software specifications	95.18	36.84	21.72
IT	computer programming	89.17	48.82	59.35
UK	Marketing	39.69	100.0	56.56
UK	online analytical processing	0	94.39	0
UK	adapt to change	69.21	87.21	79.47
DE	web programming	100.0	38.0	100.0
DE	use markup languages	80.47	34.03	82.97
DE	adapt to change	69.21	87.21	79.47

Some other ongoing research activities...

H/S Skills as factor for Job Replacement: Is there a correlation between the request for hard/soft skill and the probability for a job to be replaced by computerisation?

[Colombo, Mezzanzanica & Mercorio. AI Meets Labor Market: Exploring the Link Between Automation and Skills. Information Economics and Policy, 2019.]

Explainable LMI: improve the believability of the analyses provided by explaining the behaviour of black-box AI algorithms in LMI

[Under review]

GraphLMI: Use data from Web vacancies to enrich ISCO/ESCO taxonomies with analytics (job/skills relevance) and mentions from the Market

[Under review]

Thank you!!!

Some references to our research

(take a picture on QRcode to get the paper
Notice. PDFs are for personal use only)



1. Colombo, Mezzanzanica & Mercorio. [AI Meets Labor Market: Exploring the Link Between Automation and Skills](#). Information Economics and Policy, 2019.
2. Mezzanzanica & Mercorio: [Big Data enables Labour Market Intelligence](#). In Encyclopedia of Big Data Technologies. Springer, 2019.
3. Mezzanzanica et al. [WoLMIS: a labor market intelligence system for classifying web job vacancies](#). Journal of Intelligent Information Systems, 2018.
4. Mezzanzanica, Mercorio et al. [Using Machine Learning for Labour Market Intelligence](#). In ECML PKDD 2017, LNCS. Springer, 2017
5. Mezzanzanica, Mercorio et al. [A Language Modelling Approach for Discovering Novel Labour Market Occupations from the Web](#). In 2017 IEEE/WIC/ACM International Conference on Web Intelligence, 2017.