# FROM BIG TO SMART DATA

The misconception of collecting useful predictive and matchable skills from big data.

JANZZ.technology, Stefan Winzenried, Founder & CEO September 19, 2019



### FROM BIG TO SMART DATA PROMISE...

With today's presentation, we will demonstrate the following:

- 1) Why the collection of skills and other occupation data from big data does not result in creating high quality smart data
- 2 Why implicit information is at least as important as the explicit, readily available information
- 3 How skill sets can be accurately modeled for both emerging and advanced employment markets
- 4) Why mainly experts and not algorithms will continue to be responsible for such data modeling in the future

It has always been a priority of JANZZ.technology to present in-depth and sound views on the topics that are most hyped these days – including **big data** and **skills prediction**, **re-skilling** and **upskilling**, **(X)AI** and **matching**, as well as **labor markets 3.0**. Globally, there is extensive discussion about **the future of work and the associated challenges** and whilst much of what is written sounds very inspiring and even reasonably plausible. Our research indicates that the reality is much more challenging.

In particular, job and skills matching is a significant topic with far-reaching implications for many new technologies in this field. It seems, however, that thus far only JANZZ.technology has been able to prove the ability to provide explainable AI (XAI) for accurate skill and profile matching. Many applications, including those by major platforms and providers of ATS and HR software, can rarely meet even modest expectations and appear to have made little progress in recent years.

### FROM BIG TO SMART DATA TYPICAL EXAMPLES OF DAILY MISMATCHES ...

#### **Executive Creative Director**

#### Film & Video Producer

### Graduate in Human Rights and International Relations

- Senior Director / Head of Health Economics
- Director Financial Planning and Analysis & M&A, EMEA
- Applications Supporter / System Specialist
- Senior Monitoring Architect
- Head Information Management Foundations
- Head of Procurement
- Corporate Director Business Development Strategic Products
- ...

- General practitioner for own medical office
- Associate Director for a fast-growing industry
- Senior Community Manager, chemicals industry
- Food Safety, Quality & Purchase Manager
- Senior Underwriter Mining & Energy Onshore
- Document Control Coordinator
- EHS Specialist
- ...

- Church Officer
- Meat Industry Expert
- Travel Agent
- Social Media Guru
- Tech Evangelist
- Medical Officer Specialized in TB
- Supply Chain Management Expert
- . . . .

Source: LinkedIn

There is much marketing hype/abracadabra in this space and we need to explore the underlying data set in detail to understand whether performance is actually delivered In real life application.

Within the scope of this article and its presentation today we offer high-level insight into this important subject. We do not claim to examine this topic in its entirety and remain available for deeper discussion on all topics, in particular, skills matching and knowledge representation outside of this paper/presentation.

In the course of our research we came across a remarkable publication by Prasanna Tambe, Peter Cappelli, Valery Yakubovich (Wharton School, University of Pennsylvania and ESSEC Business School in France) from August 2, 2019, from which we would like to quote the following:

### ARTIFICIAL INTELLIGENCE IN HUMAN RESOURCES MANAGEMENT: CHALLENGES AND A PATH FORWARD

Prasanna Tambe, Peter Cappelli, Valery Yakubovich First Published August 2, 2019

"When you talk to HR practitioners who see their colleagues in finance and marketing using these technologies with so much success, part of the question they ask is, why does it seem so hard for us? I think part of the point we wanted to make is that there are systemic and structural differences for HR that do make it harder, when you are building an Al-based system, you need to know what the right answer is, what a good employee looks like, and defining that in HR is already a difficult thing to do."

"As much as I would like to have people think that we're like wizards with this stuff, that we go into the computer, we take this data and then magic comes out, it's an awful lot of just plugging away."

The two quotes highlight the specific challenges faced by both HR and Public Employment Services.

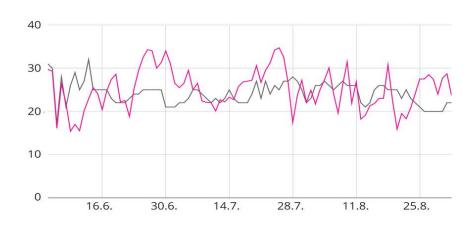
Big data and AI still have such a hard time today to find traction in the field of HR & employment because the quality and explanatory power of methods and results in these fields are limited and lack resilience.

We recommend this article to everyone who is interested in AI, big data and HR technology and even more so to everyone who is concerned with predictions about skills, employment markets and the future of work in general.

In general, it should be noted that even in 2019, all types of predictive analysis based on big data and determined by a large number of variables are very rarely accurate.

The longer the horizon and the more variables included, the less likely it is that such predictions will be completely or partially true. There are some spectacular examples of this, such as the long-term weather forecasts for this summer in Switzerland, or the attempt to predict the future football world champion in 2018, and many more. That's why very big question marks should be placed behind such forecasts and only operated with the utmost caution.

Figure 1: Long-term weather forecast Switzerland summer 2019 using comprehensive big data-based models from AccuWeather compared to actual measured temperatures.



- Forecast AccuWeather
- Measured real temperatures MeteoSchweiz

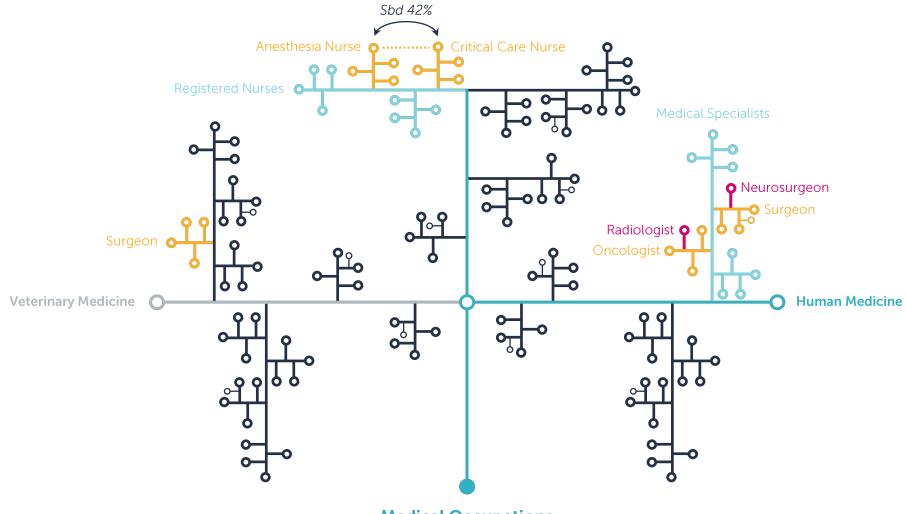
### FROM BIG TO SMART DATA THE KEY IS KNOWLEDGE REPRESENTATION

At JANZZ.technology we have been dealing exclusively with big occupation data since 2008. For over ten years we have been systematically modeling and curating the largest knowledge representation of occupation data worldwide, including skills and competences, soft skills and qualifications, education and training, certifications and authorisations, in over 40 languages. From the very beginning, the aim was to form big data into smart data using a structured and fully semantic ontological approach. We are convinced and have repeatedly proven that only this approach can work reliably in the area of skill matching, as well as in skill anticipation and prediction.

It can already be stated at this point that whoever works with big Occupation data evaluations, matching and skills predictions without a multilingual and sophisticated knowledge representation with all the necessary semantic dimensions, represents outdated methods and repeatedly disproved results with little to no significance.

As Ron Weinstein once so aptly put it: "A fool with a tool is still a fool."

# FROM BIG TO SMART DATA KNOWLEDGE REPRESENTATION VIA GRAPH STRUCTURES



### FROM BIG TO SMART DATA THE KEY IS KNOWLEDGE REPRESENTATION

Typical examples for this are, approaches that compile and evaluate all available job advertisements from all available sources of a market over a period of years. These aggregated data is used to make recommendations and to forecast the future employability and required skills of job seekers and other market participants.

In the course of this paper, we will analyse this methodology and this example again in detail and point out the shortcomings of the results produced by it.

What do we mean by the terms: skills, abilities and competences?

Definitions diverge greatly depending on who you ask. For some, **skills are** what **soft skills** are for others. **Competencies** are another topic all together. For example, at O\*net you can find **knowledge and technology skills** and tools, which for them refer only to the directly job-related or transferable knowledge.

Why do the individual terms and expressions overlap significantly? If you want to think about the skills which will be required in the future and make forecasts about how they will change (which skills will gain in importance and which skills will disappear) or if you want to perform target-oriented skill matching, you **first have to be able to correctly recognize**, **understand**, **assign and classify today's skills**.

Most approaches and big data evaluations fail miserably at this and hence they cannot be used for matching.

Thus, the same representations and forecasts are produced repeatedly, which, if they are then retroactively checked for their forecast quality, can usually not even fulfill very modest expectations.

Figure 2: Typical forecasts of general "top skills" as published regularly by LinkedIn or in the case shown by the WEF.

In 2020		In 2015	
01	Complex Problem Solving	01	Complex Problem Solving
02	Critical Thinking	02	Coordinating with Others
03	Creativity	03	People Management
04	People Management	04	Critical Thinking
05	Coordinating with Others	05	Negotiation
06	Emotional Intelligence	06	Quality Control
07	Judgment and Decision Making	07	Service Orientation
80	Service Orientation	80	Judgment and Decision Making
09	Negotiation	09	Active Listening
10	Cognitive Flexibility	10	Creativity

Source: Future of Jobs Report, World Economic Forum

In 2020		In 2015	
01	Complex Problem Solving	01	Complex Problem Solving
02	Critical Thinking	02	Coordinating with Others
03	Creativity	03	People Management
04	People Management	04	Critical Thinking
05	Coordinating with Others	05	Negotiation
06	Emotional Intelligence	06	Quality Control
07	Judgment and Decision Making	07	Service Orientation
80	Service Orientation	80	Judgment and Decision Making
09	Negotiation	09	Active Listening
10	Cognitive Flexibility	10	Creativity

Source: Future of Jobs Report, World Economic Forum

The figure above does not indicate the survey method, nor the sources, nor the number of samples used in the forecast, nor does it name the occupations and activities for which these skills should be relevant. Moreover, the skills often overlap or are to be understood very differently depending on the context, industry or activity. The skills mentioned above would be much too generic/general and hardly usable for matching, as they are barely relevant for many occupations.

Figure 3: Typical, current example from the skilling, up-skilling and training area, which shows further dimensions of skill definitions and their mapping/classification.

Business Skills Collection Help employees develop the right hard and soft skills they need

Skillsoft's Business Skills collection contains Skillsoft's most rigorous and comprehensive content offerings, including state-of-the-art content such as critical and adaptive thinking, virtual collaboration, cross-cultural competency, and new media literacy. Combining that with industry-leading instructional technique and design and highly effective application resources, this content collection will drive their personal development while moving their organizations forward.

On the platform Percipio, the Business Skills collection contains over 65 curated channels that have been mapped to core competencies across popular topics such as:

- Business Operations
- Customer Service
- Professional Improvement

- Management & Leadership
- Project Management
- Sales & Marketing

The previous description shows the typical, mostly generic and unspecific structuring and disordered classification of skills and soft skills into completely irrelevant categories such as "Sales & Marketing" or "Project Management".

Additionally, they sometimes behave differently, sometimes even contradictorily, in terms of their meaning in different languages. Plus, there are the difficulties of a reliable leveling: what does good or very good knowledge mean and what distinguishes an expert in a certain skill? Is it theoretically acquired knowledge or Is it already applicable in a real, professional environment? In contrast to other big data, there are no binding scales and validations. There is a pattern of conceptual ambiguity, unspecificity and inconcise form throughout (see Figures 2 and 3).

Even though NLP (Natural Language Processing) has made and will continue to make great progress in the next years, the binding processing of language remains extremely difficult for the computer to manage. However, other clear big data categories and criteria such as numbers, prices, GIS data and coordinates, temperatures, clicks and likes, products, elements of computer vision, pictures, etc. are not applicable.

NLP is and remains the second decisive component in the structuring and processing of data in the big occupation data area besides knowledge representation.

## A BRIEF EXAMINATION OF THE IMPORTANCE NLP CHALLENGES...

Highly experienced in **Software Development Tools**. **SKILL** is a plus.

→ Software Development Tools

→ Level 4

→ A PLUS

Bachelor in Food Research EDU , would be nice to have.

→ Bachelor Food Research

→ Bachelor

EXP

→ Nice to have

Minimum 3–4 years in a fast-paced kitchen, high-volume restaurant.

→ Hospitality/Gastronomy Industry

→ Amount: Min. 3–4

→ Type: years

→ Is mandatory

### A BRIEF EXAMINATION OF THE IMPORTANCE NLP CHALLENGES...

#### Help wanted

Job Overview: Assist with the organization, preparation and service on the line as directed. Ensure the quality preparation of all menu items to established specifications and the proper handling /storage of all food items in accordance with standards. Contribute to the fluid operation of the cooking line. Inform the General Manager or Supervisor of any ingredients that need to be ordered for service to ensure preparation of a fresh, high quality product. Work at assigned station and ensure consistent presentation and high quality of all prepared plates. Reports To: Restaurant General Manager, Assistant General Manager, Manager Key Relationships: Internal: Restaurant Staff, Bar Staff, Kitchen Staff External: Restaurant Guests and Food Vendors Qualifications/Skills: Requirements are representative of minimum levels of knowledge, skills and /or abilities. To perform this job successfully, the employee will possess the abilities or aptitudes to perform each duty proficiently with or without reasonable accommodation. Essential: 1. Ability to effectively communicate in English with co-workers and management to their understanding. 2. Ability to perform assigned duties with attention to detail, speed, accuracy, follow-through, courtesy, cooperativeness and work with a minimum of supervision. 3. Ability to prioritize work assignments. 4. Ability to solve practical problems and deal with a variety of concrete variables in situations where only limited standardization exists. 5. Ability to work well under pressure of organizing and attaining production schedules. 6. Ability to ensure quality standards are met. 7. Ability to operate, clean and maintain all equipment required in job functions. 8. Three years experience in a similar position. 9. Ability to work all stations on the line for designated meal period and assist in the training of new culinary staff members. 10. Punctuality and reliable attendance. 11. Ability to follow directions and create respectable menu items. 12. Ability to quarantee execution and presentation excellence of all menu items. 13. Interpersonal skills and the ability to work well with others and the public. Desirable: 1. Culinary training, 2. Ability to communicate in a second language 3. Serv-Safe certification and training Essential Physical Abilities: ? Endure various physical movements throughout the work areas, such as reaching, bending and stooping. ? Ability to stand and/or walk continuously throughout the shift to perform essential job functions. ? Sufficient manual dexterity to be able to grasp, load, and carry tools. ? Ability to endure repetitive motions for extended periods of time. ? Ability to maintain good coordination while serving orders quickly. \* Ability to grasp, bend, and stoop; push or pull heavy loads weighing up to 50 lbs.; and lift and/or carry or otherwise move received goods and boxes. Essential Job Functions: \* Ensure that all preparation and closing duties are completed to standard before signing out. \* Oversee and help with the supervision and the maintenance of the cooking and prep lines. \* Prepare menu items following recipes guidelines. \* Work on line during service and assist wherever needed. \* Inform Management of shortages, excesses and abnormalities in kitchen. \* Practice correct and safe usage and care of all machinery in the kitchen operation. \* Assist Management in ongoing training of new Cooks. \* Thorough knowledge of all menu items produced by assigned station \* Assist in overseeing plates prepared before they are sent out to dining room or bar areas \* Keep work area clean and organized. ? Work in a clean and orderly fashion. ? Ensure that all prepared plates going out to dining room are in order with allergy requests. \* Promote a work environment conducive to teamwork. Secondary Job Functions: ? Attend meetings, training sessions and tastings as scheduled.

### A BRIEF EXAMINATION OF THE IMPORTANCE NLP CHALLENGES...

#### Help wanted

'Skills': [Ensure the quality preparation of all menu items to established specifications,' Contribute to the fluid operation of the cooking line,' Ensure preparation of a fresh, high quality product,' Work at assigned station,' Ensure consistent presentation and high quality of all prepared plates,' Ensure quality standards are met,' Operate, clean and maintain all equipment,' Assist in the training of new culinary staff members,' Create respectable menu items', 'Guarantee execution and presentation excellence of all menu items', 'Culinary training,' Ensure that all preparation and closing duties are completed to standard before signing out,' Oversee and help with the supervision and the maintenance of the cooking and prep lines', 'Prepare menu items following recipes guidelines', 'Work on line during service', 'Inform Management of shortages, excesses and abnormalities in kitchen', 'Practice correct and safe usage and care of all machinery in the kitchen operation', 'Assist Management in ongoing training of new Cooks', 'Thorough knowledge of all menu items produced by assigned station', 'Assist in overseeing plates prepared before they are sent out to dining room or bar areas', 'Keep work area clean and organized', 'Work in a clean and orderly fashion', 'Ensure that all prepared plates going out to dining room are in order with allergy requests', 'Attend meetings, training sessions and tastings'], 'Softskills': ['Attention to detail', 'Speed', 'Accuracy', 'Follow - through', 'Courtesy', 'Cooperativeness', 'Work with a minimum of supervision', 'Ability to prioritize work assignments', 'Solve practical problems', 'Deal with a variety of concrete variables in situations', 'Work well under pressure of organizing and attaining production schedules', 'Punctuality', 'Reliable attendance', 'Follow directions', 'Interpersonal skills', 'Ability to work well with others and the public', 'Maintain good coordination', 'Serving orders quickly', 'Promote a work environment conducive to teamwork'], 'Educations': [], 'Languag

'Experiences': ['Three years experience in a similar position'],

'Supervisor': ['Restaurant General Manager', 'Assistant General Manager'], 'Authorizations': [], 'Salary': [], 'Working\_conditions':

['Endure various physical movements throughout the work areas', 'Reaching, bending and stooping', 'Stand and / or walk continuously throughout the shift', 'Sufficient manual dexterity to be able to grasp, load, and carry tools', 'Endure repetitive motions for extended periods of time', 'Ability to grasp, bend, and stoop', 'Push or pull heavy loads weighing up to 50 lbs', 'Lift and / or carry or otherwise move received goods and boxes']}

Some previously mentioned, frequent and typical examples from predictive evaluations and recommendations:

#### **Digital Skills**

What exactly does this mean? Does it include operating digital devices such as smartphones or computers, posting on Instagram, knowledge of complex BIM applications in real estate drafting and planning, or dealing with the Internet, Google Analytics, Facebook or social media? The definition of the skill is completely imprecise.

#### Marketing or Digital Marketing

Same situation. The term has thousands of different meanings and is far too broad and too unspecific. Therefore, it is completely meaningless for almost all applications, especially for matching.

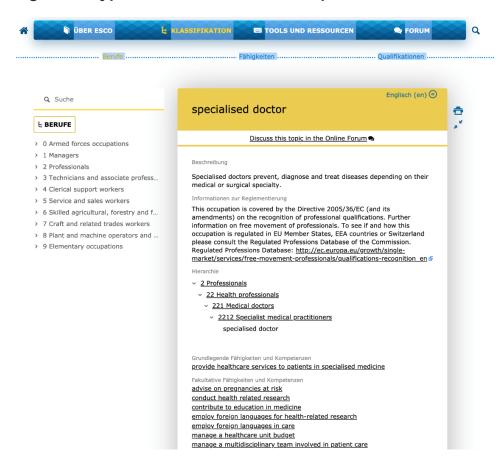
#### **Project Management Skills**

In this case as well, the term is almost completely useless. Almost everyone has project management skills on some level, but this knowledge cannot be compared or categorized. For example, solid project management knowledge of a foreman on a large tunnel construction site, the pm knowledge of a project manager of a large IT application, the pm knowledge of a campaign manager of a public authority or that of a process engineer or an event manager are very specific and thus obviously have very different meanings. It is nonsensical to comprise them all into a single "matchable" skill.

Skills are usually presented and processed out of context in an incomprehensible way, mostly without any relevant semantic context. Despite this being the game changer, which should go beyond what can be extracted linguistically and can be elucidated e.g. via textmining/NLP. Many applications are still based on keywords and strings, but not on context and multidimensional semantic concepts.

Moreover, in almost all current methods and standard classifications, the skills are often assigned to superordinate occupational groups. This, too, is a systemic error that urgently needs to be corrected. In the future, meaningful and helpful matchable skill sets should be defined and, if necessary, specifically promoted.

Figure 4: Typical relation of skills and professions in ESCO or O\*Net.



If we really want to look ahead by using big data in order to finally make progress in the matching of complex and multi-layered profiles and skill data, we must make sure to methodically include the entire picture and as much context as possible.

How many professions will disappear in the future, and which ones? What jobs will be affected by robotics and automation? And when? What share do these occupations have in the overall market? What are the relevant skills that will ensure the professional advancement of future generations? What are the most promising fields of study and vocational training for a secure future? Which skills and competences should be part of the much-discussed lifelong learning and can they even be taught to a large number of people?

All these issues are integral parts of the question whether we can use big data to anticipate and match skills.

# ILLUSTRATIVE EXAMPLES DATA COLLECTION AND ENTITY RECOGNITION

Goal:



### ILLUSTRATIVE EXAMPLES DATA COLLECTION AND ENTITY RECOGNITION

With these examples we will Illustrate specifically the quality and quantity challenge of extracting data from job specifications and adverts.

#### Figure 4: Various examples of job advertisements.

### **Curator – Botanist**

Royal Botanic Gardens Kew – Kew Gardens Apply On Company Site Kew Gardens

In addition to this, the role includes facilitating access for our many visitors to the collections from around the world, identifying African & Madagascan specimens and providing research support to the team. Our successful candidate will have a botanical background with experience of working with Herbarium / natural history collections. You'll also need to have excellent interpersonal skills and enjoy both working in a team and independently.

The salary will be £21,527 - £23,282 per annum, depending on skills and experience.

We offer a fantastic range of benefits including a broad range of Learning and Development opportunities, with access to the Civil Service training curriculum, generous annual leave entitlement for new starters, family friendly policies, a choice of competitive pensions and flexible benefits scheme.

If you are interested in this position, please submit your application through the online portal, by clicking "Apply for this job". We are committed to equality of opportunity and welcome applications from all sections of the community. We guarantee to interview all disabled applicants who meet the essential criteria for the post.

No Agencies please.

### Carpenter

New York University 972 reviews – New York, NY Position Summary SHIFT: Monday - Friday 8:00AM – 4:30PM

Primarily responsible for building, installing, repairing and refinishing building fixtures (e.g., doors, floors, stairs, ceilings, walls, windows, cabinets, etc.) of assorted materials (wood, ceramic tiles, concrete, stone, glass, plexiglass formica, metal, etc.).

#### Qualifications

#### Required Education:

High School diploma or equivalent

#### Preferred Education:

Trade school or other post-elementary education.

#### Required Experience:

Three years of hands-on carpentry work experience for general building maintenance.

#### Preferred Experience:

Experience in higher education, health care, telecom or other sensitive facilities where high levels of performance and reliability are required.

#### Required Skills, Knowledge and Abilities:

The ability to use all power and hand tools. The ability to read mechanical sketches and shop drawings. Basic arithmetical and measuring skills. Common knowledge of materials. Good communication skills with the ability to follow verbal and written instructions. Positive attitude, willingness to collaborate and team with others to accomplish work efficiently and to high quality standards. Proficient in English (e.g., reading directions and instructions written in English). The ability to lift and carry up to 50 pounds, climb a ladder, work in small, cramped spaces, and in outdoor environments and elements (combined 40% – 50%). MUST BE AVAILABLE TO WORK NIGHTS AND/OR WEEKENDS.

#### Preferred Skills, Knowledge and Abilities:

Knowledge of University and department policies, processes, and quality standards.

#### **Additional Information**

EOE/AA/Minorities/Females/Vet/Disabled/Sexual Orientation/Gender Identity

### Möbelschreiner / Monteur (m/w)

Job3000 – Zürich, ZH Weiter zur Bewerbung Zürich. ZH

#### Ihr Profil:

- Ausbildung als Schreiner EFZ oder erfahrener Möbelmonteur
- Umzugserfahrung
- Körperlich top fit
- Gute Deutschkenntnisse von Vorteil
- Führerausweis Kat. B, B/E von grossem Vorteil
- Flexibel und belastbar

#### Ihre Aufgaben sind:

- Transport und Montage von teuren Designermöbel
- Ein- und ausladen
- Montagearbeiten
- Mithilfe in der Entsorgung
- Ein und Auspacken

Fühlen Sie sich angesprochen? Wir freuen uns auf Ihre vollständigen Bewerbungsunterlagen (Lebenslauf mit Foto, Zeugnisse, Diplome), welche sie an die folgende E-Mail Adresse schicken können:

### Carpintero autónomo

AVERALIA - Madrid, Madrid provincia

#### Solicitar

Madrid, Madrid provincia

24.000€ - 36.000€ al año

Se necesitan operarios autónomos con amplia disponibilidad y que pueda hacer una o dos guardias al mes los fines de semana. Somos una empresa que trabaja para asistencias de seguros.

Tipo de puesto: Jornada completa

Salario: 24.000,00€ a 36.000,00€ /año

Imagine a technology of a big data company collects data of all available job advertisements from the English-speaking area over a period of 10 years. It extracts all possible entities, above all, skills.

Statistically most relevant are ads from the US, as they are the highest in quantity. Nevertheless, all geographical areas are modeled individually for separate evaluation.

Also relevant is the fact that ads from the US are unique in their detailedness, which makes their comparison with other data sets almost impossible. Thus, when parsing ads from the US one can find a lot of information that is only implicit or entirely unvailable in comparable data from the UK and other English-speaking regions.

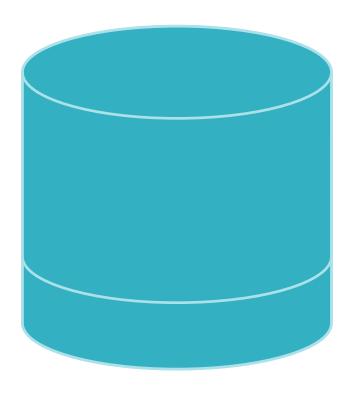
### ILLUSTRATIVE EXAMPLES BIAS AND MISREPRESENTATION

#### The Challenge of these Examples

These simple examples show impressively that the information density and relevance varies massively. They also show that a large part of the information is often implicit (approximately 30–40%), as it is contained in education/training, qualifications and experience.

If such information is not represented in an accurate and semantically meaningful manner, the collected data is quickly distorted: pseudo-correlations and statistical significances occur when in fact they should not exist at all. In addition, the following deviations are to be expected when using such a survey method:

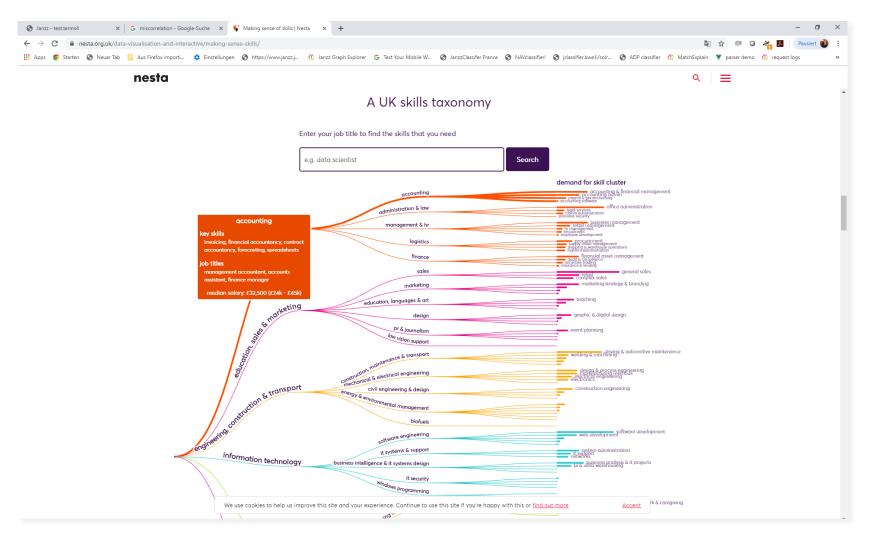
### ILLUSTRATIVE EXAMPLES BIAS AND MISREPRESENTATION



30%
Non published jobs/missed jobs

- Large companies or those with a duty to publish are statistically overrepresented
- technical/ICT-related occupations provide more specific skills, as in this domain it is more common to mention all the necessary tools and technologies
- segments and industries that suffer from a shortage of skilled workers tend to advertise jobs on several platforms; in addition, there is a high number of advertisements, which are not aimed at hiring people with the skills you are looking for, but are only for employer branding or self-marketing, which makes them hardly distinguishable from real advertisements
- implicit information is only recorded to a limited extent and is statistically underrepresented despite its high relevance
- the extracted data is often very imprecise and incomplete, job titles cannot be assigned or mapped correctly due to cryptic formulations ...

Combining this compiled data with other assumptions and criteria, such as assumptions about the upcoming automation of different professions and sectors (Attention: correlation vs. causality...), results in an avoidable, data-based forecast with a very high probability of error. The data are hardly suitable for matching anyway ...



Of course, this results in decision-making and orientation aids such as Nesta's SkillsMaps, which, on closer inspection, have hardly any significant information relevance. These are then conveyed with such summaries and lists. Decide for yourself what you can do with it and whether you can say anything about the future of skills and abilities with it. Or whether, for example, the skill Mental Health (if it is a single skill at all, it's much more e.g. treatment of mental health, mental health care... etc.) would be suitable for matching profiles and multidimensional data.

### FROM NESTA.ORG.UK WHICH 10 SKILLS ARE NOW MENTIONED MORE FREQUENTLY?

These skills have shown the fastest growth in the number of mentions between 2012-14 and 2014-16. Several factors can drive growth, including an increase in the number of vacancies for the job that requires this skill, or an increase in the range of jobs that use this skill. A number of these skills (looking at the top 100) relate to caring for others, such as patient care, mental health, and working with patients who have dementia. A second group of skills reflect the opportunities and threats that come from living in a more connected world. These include digital marketing, big data, social media, information security and firewalls.

- 1. Big data
- 2. Information technology industry experience
- Contract accountancy
- 4. Onboarding
- 5. Digital marketing

- 6. Information security
- 7. Transportation logistics
- 8. Front-end development
- 9. Patient care
- 10. Mental health

### FROM BIG DATA TO SMART DATA SUMMARY

### Why the collection of skills and other occupation data from big data does not result in creating high-quality smart data

We are convinced that a purely big-data-based prognosis of future skills will hardly be possible in a way that goes beyond the obvious. These forecasts will be similarly unspecific and mostly just as wrong as the current long-term weather forecasts. If, however, the quality and consistency of the data should improve significantly in the coming years and at the same time reduce the extremely high variability, we can certainly expect somewhat more reliable and meaningful predictions. The methodology for collecting and structuring data plays a decisive role here.

#### Why implicit information is at least as important as the explicit, readily available information

We have highlighted that information density and relevance varies massively. A large part of job information is often implicit and therefore hidden in education/training and qualifications, but also in experience. If these are not represented accurately, semantically and in a knowledge representation, the collected data is distorted. <a href="mailto:limportantly">Importantly</a>, geographic differences in the description and detailedness of jobs makes comparison with other data sets almost impossible.

### FROM BIG DATA TO SMART DATA SUMMARY

of candidates, it would be necessary to consistently check and record which criteria and skills were decisive and why a candidate was finally hired. Were these congruent with the specifications of the job ad? Were they different because a hired team member already possessed certain skills required for the new position and, hence, complementary criteria were weighing in on the decision? Are these criteria reviewed periodically? Which factors are responsible for the good or bad performance? Which skills seemed important at the time of recruitment or placement, but irrelevant for the actual job performance? What are the real skill gaps? Have there been significant changes in the framework over time that will influence the decision or assessment of similar jobs in the future?

All these criteria would have to be gathered in a data collection and evaluated in order to facilitate the development of significant and precise big data models and Al processes.

### FROM BIG DATA TO SMART DATA SUMMARY

How skill sets can be accurately modeled for both emerging and advanced employment markets

With regard to matching, an acceptable quality level with improved results can be achieved relatively quickly by means of a close integration of knowledge representation with big data applications. However, with big data and ML/AI approaches that lack in prior standardization, systematic structuring and semantic expansion this undertaking will not succeed.

Why mainly experts – and not algorithms – will continue to be responsible for such data modelling in the future.

### THANK YOU! CONTACT

#### Stefan Winzenried

Founder & CEO T +41 43 499 71 04 s.winzenried@janzz.technology

#### Diego Rico

VP Customer Integration & Deputy CEO T +41 43 499 71 04 d.rico@janzz.technology