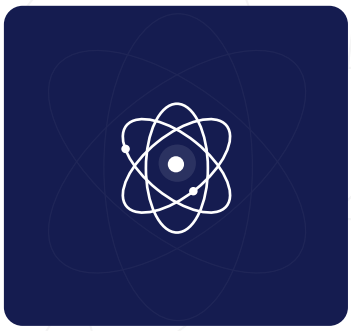
Jobs & skills ontology

DATA SOURCES, AI & ACCESS

SUMMARY

INTRODUCTION

For quite some time now, companies have seen corporate learning and development, talent management and HR technologies in general as an expense rather than as an investment in their own growth. However, under the impulse of a few software companies and thanks to data science and R&D in the HR field, this stand has evolved significantly.



Public authorities have worked tremendously over the past 15 years on workers' employability and on jobs evolution, in order to anticipate the impact of big trends (technological, societal, etc.) on the job market.

One of the main initiatives undertaken both at a country and regional levels has been the development of jobs and skills ontologies. It is the science of linking skills to jobs so as to be able to work on several fundamental HR aspects such as career path, learning and development, recruitment, strategic workforce planning, etc.

American and European initiatives, under the names of O*NET and ESCO, have been undertaken to list all existing jobs and match them with skills and job descriptions.

Skills and jobs data allow companies to gain visibility into the supply and demand of skills within their organization. However, collecting that data is far from being simple. The field is still emerging and lacks measurement, alignment and standardization.

IN THIS PAPER, WE WILL AIM AT ANSWERING 5 SIMPLE QUESTIONS :

1. **What is a jobs and skills ontology?**
2. **Why do the HR function and HR software vendors need to leverage ontologies?**
3. **How do we build jobs and skills repositories and what type of data do we find behind?**
4. **What data science techniques are used to exploit ontologies?**
5. **How can companies or HR software vendors access jobs and skills ontologies?**

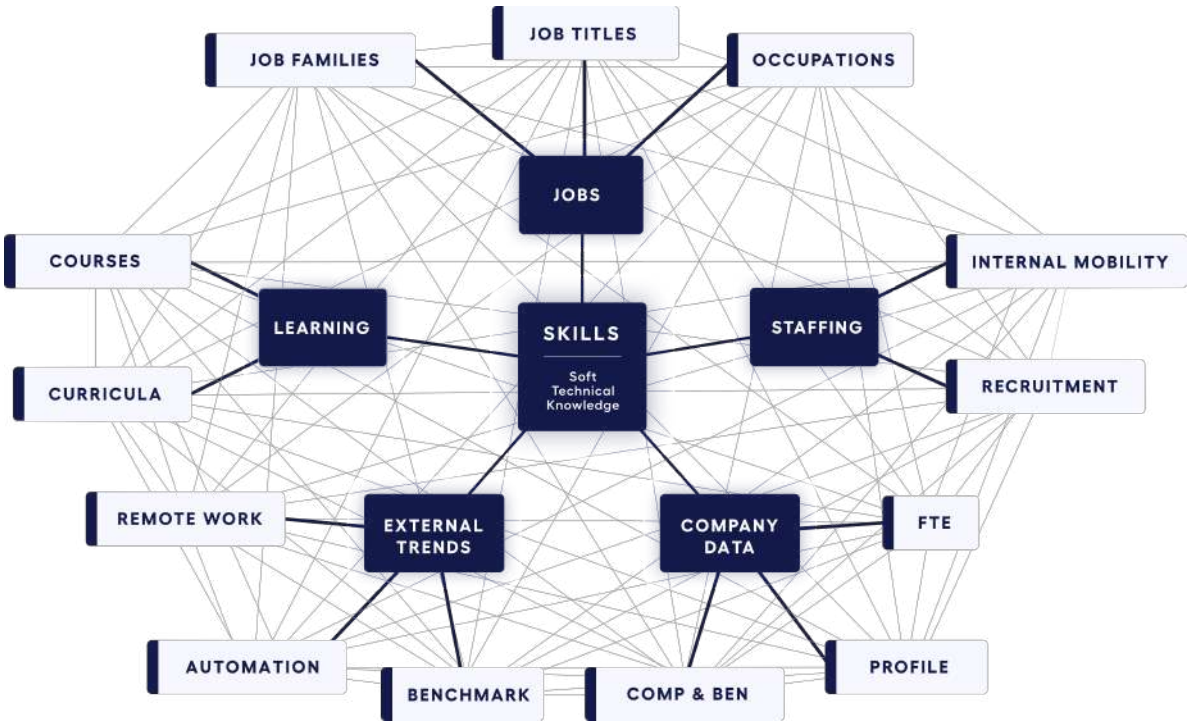
1.

What is a jobs and skills ontology?

One of the most trendy concepts in the HR Tech space today is 'ontology'. At a high level, the purpose of a jobs and skills ontology is to create a common « language » which links seemingly distinct areas of HR. For example, it could be used to describe the details of a job posting – or a CV for a job seeker, list the skills attached to a job and be able to link trainings to those skills.

From a technical point of view – an ontology is a system which describes a number of concepts such as jobs, skills, courses or candidates, and additionally specifies the relationships among them. While there is a need to organize jobs (or skills) in a hierarchical classification scheme (usually known as « taxonomy »), an ontology extends beyond the hierarchy to connect entities in separate areas.

For example, a job could be conceptually viewed as a collection of skills, but the relationship can be also turned around to represent a skill via the jobs which require it. There is no inherent hierarchical relation between the job and the skill concepts, and instead each concept can become central depending on the application at hand: a professional training organization might focus on skills, while a recruitment company would put more emphasis on jobs.



Artificial Intelligence has made it easier to map out the links between the various building blocks of an ontology. For example, a job title extracted from a CV – say SEO (Search Engine Optimization) specialist would be recognized as a marketing job and be matched to a Digital Marketing class provided by a training company.

Similarly, jobs and skills can be connected to broad industry trends such as automation or remote work capability, or be connected to a more narrow framework present in one company.

By leveraging modern techniques such as vectoring and AI (artificial neural networks, natural language processing and multivariate algorithms), ontologies have now become more agile and have developed the capacity to become self-learning systems. The more content they scan and analyze in terms of job titles, functions, skills, etc., the more correlations they can see and the smarter they get.

Note. It is important to point out the distinction between an ontology and the data which populates it. Strictly speaking, the ontology is simply the conceptual representation of concepts along with the relationships amongst them.

For example, as soon as job and skill data are uploaded into an ontology we create what is usually referred to as a « jobs and skills repository ».

Thus, a repository is simply a linked data set which specifies and inter-connects all jobs and skills present within the perimeter considered.

IN A NUTSHELL

What is a job and skills ontology?

It is a categorization of jobs and skills that allows it to build a common language, defining the aspects of a specific job and its skills rather than relying on blanket terms and vague descriptions.

2. Why do the HR function and HR software vendors need to leverage ontologies?

2.1. Why do HR and HR software companies need it?

Why is it critical for companies but also for HR software companies (ATS, HRMS, LMS...) to invest in a jobs and skills ontology? Because it can help companies build an internal database of the jobs and skills in its workforce.

This is foundational as it serves various purposes such as:

- **Recruitment:** improving recruiting by better understanding the skills gaps that exist in the workforce, the tension areas and the 'hot skills' on the market.

- **Learning & Development (L&D):** proactively identify training needs and recommendations, improve the user learning experience by improving the training to trainee match, assess upskilling and reskilling needs.

- **Strategic Workforce Planning (SWP):** designing a process within the company to anticipate current and future staffing needs by analyzing the current employee skills and highlighting the resource gaps to reach the company's objectives.

A jobs and skills ontology is always evolving: new jobs appear, news skills arise, and as they do, the underlying ontology is used to categorize them.

The system is thus continually updated and can also be enriched (impact of automation on jobs and skills, capacity to work remotely, emerging jobs or jobs in demand...) to provide companies and HR software editors useful information for their job announcements, their HRMS or their LMS.

The various purposes of a jobs and skills ontology

2.2. Why not build it on their own?

For companies or HR software companies willing to leverage a jobs and skills ontology, a major question is: should we build it ourselves or should we buy it ? To answer that question, it is essential to understand what it takes to build a jobs and skills ontology.

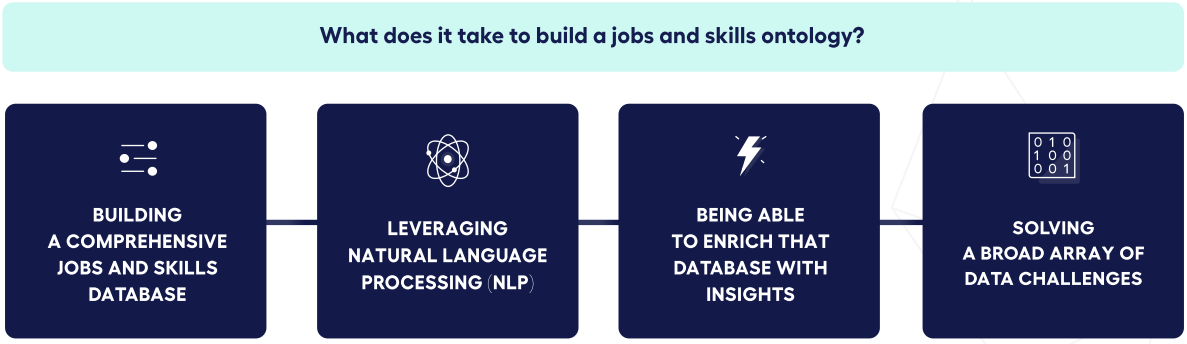
We can think of 4 main factors here:

Building a comprehensive jobs and skills database – the objective here is to build an enhanced and curated dataset on top of existing databases (ESCO, O*Net, ROME, industry specific datasets...)

Leveraging Natural Language Processing (**NLP**), multivariate algorithms and AI to exploit that database – the purpose here is to be in capacity to use several tools that are needed to exploit a jobs and skills ontology, such as job and skill title matching, career path projection, text-to-skills and text-to-jobs matching, artificial neural networks...

Being able to enrich that database with insights – the idea here is to enrich the jobs and skills ontology with insights to make the database richer and more insightful for the HR function: soft skills to job matching, capacity to work remotely (for each job, for each skill), impact of automation on jobs and skills, insights on senior employment, benchmarking capabilities with industry comps...

Solving a broad array of data challenges to fit a jobs and skills ontology brick into your technology – scoping and defining the needs is quintessential as APIs and data solutions must be designed to meet specific needs. Testing and iterating are compulsory so as to ensure proper delivery of the solution.



When answering the buy or build question, a critical decision criterion is the minimum amount of time it takes to develop a fully functional jobs and skills ontology. When looking at the 4 factors mentioned above, we can easily add 9-12 months to build the database, 12 months to develop the algorithms & AI to exploit it, 16 months to enrich the database with a minimum set of insights and 6-9 months for the delivery part. Even when parallelizing some of these 4 steps, it thus takes a minimum 3 years to get it up and running.

To build and operate a jobs and skills ontology, Boostrs has relied on a team of 6 experienced data scientists, composed of 5 PhDs and 1 MS – with more than 50 years of combined data science experience. As for the product development team needed to develop the back-end infrastructure, we can count on a team of 6 developers (1 CPO, 3 Full Stack developers, 1UX/UI designer and 1 QA tester).

IN A NUTSHELL

Why do HR and HR software editors need a jobs and skills ontology?

Because it its foundational for many tasks inherent to HR: recruitment, learning and development, strategic workforce planning... It serves as a common language to harmonize and operate the full HR stack.

Should they build it or buy it?

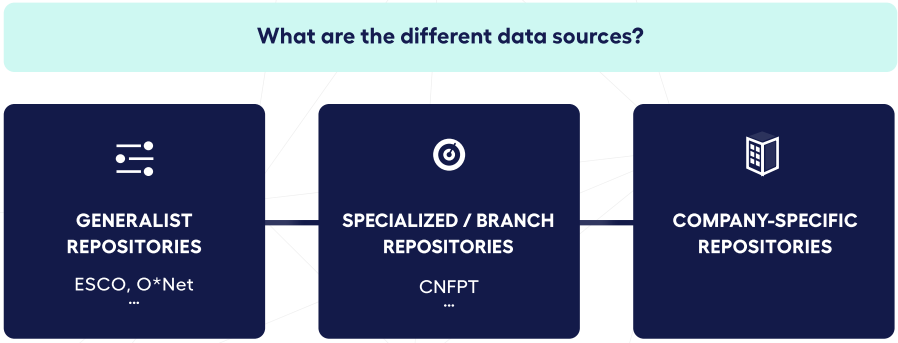
Building a jobs and skills ontology takes a minimum 3 years with a team of 6 seasoned data scientists and the same number of software developers. 4 steps have to be undertaken to build it: building the database itself, developing the algorithms and AI to exploit it, enriching it with complementary insights and ensuring proper delivery. To sum up, building your own ontology will take you 3 years and cost you around $3M.

3.

What are some of the data that lie behind jobs and skills repositories?

Jobs and skills repositories can be classified into three main groups: generalist repositories, specialized / branch repositories and company-specific repositories.

Generalist repositories include jobs and skills databases like ESCO (European initiative – European Skills/Competences, Qualifications and Occupations) and O*NET (the Occupational Information Network – the American counterpart of ESCO). While O*NET covers around 1,100 jobs, the ESCO repository covers over 3,000 jobs and offers the most in-depth job and industry coverage. Other public sources exist like the World Economic Forum (good coverage of soft skills) and the International Labor Organization (particularly useful to organize and classify jobs into groups and subgroups). While ESCO and O*NET constitute the largest and most well-known continental repositories, country level initiatives also exist like the ROME repository in France for instance. ROME is much smaller than ESCO and is organized on 3 levels: 14 grand domains, 110 domains and 532 occupations.



Specialized / branch repositories are industry specific, like the CNFPT repository for public sector jobs in France for instance. By nature, they are much smaller than ESCO or even ROME. They often use a language that is specific to their own industry / branch (the French public sector in the case of the CNFPT) but lack several key aspects that generalist repositories have: they often have too many skills to be truly useful (skipping the notion of essential vs. optional skills), their skills description are generally very long and wide, making them somewhat fuzzy or even impossible to understand and their career paths are not fact-based from an analytical standpoint, meaning they don't use skills to derive career paths. When adopting a language that seems at first sight more colorful from an industry standpoint, those specialized repositories make the choice of customization vs. standardization. Yet standardization enables to compare a job in two different industries (e.g., a data scientist in the French public sector and a data scientist in the business services space) and allows to achieve scale.

Last but not least, there are **company-specific repositories**, using a verbiage specific to the company (product names, processes, ways to describe a specific skill or job...). While having a 'flavor' specific to the company is understandable, and in some cases even beneficial, one should not forget that the ultimate capacity to use a common language is paramount to success – else the company-specific repository will be fully customized and adopt the company's own language but will not be fit to exchange with other repositories, thus completely losing sight of one of the main purposes of having a jobs and skills ontology: ensuring the employee's employability in the outside world.

That does not mean that handling languages specific to the company is not possible: at Boostrs, if clients request it, we can take into account their specific wording in the names and descriptions of their jobs and skills. For example, for some clients in the public sector, the word 'company' was replaced by 'public organization', the words 'client' and 'customer' by 'user', etc.

So, what makes one ontology better than the others? At Boostrs, while we cannot claim to bring a definitive answer to that question, we know that we tick at least two boxes:

☑ Client-testing ☑ Exhaustivity

Our clients in the public sector, pharmaceutical sector, transportation, etc. have pushed us to extend the reach of our repository beyond the standard job coverage offered in generalist repositories.

Testing with clients also means we removed a certain fraction of job-to-skills associations, we added alternate titles found in our clients' specific repositories, we added new jobs we previously did not have covered, we modified job descriptions to reflect the reality of the jobs…

About exhaustivity, we brought in additional data, mostly proprietary: soft skills (through the analysis of 1M+ job descriptions to understand which soft skills are most often associated with which jobs), automation (impact of automation and robotization on the different jobs and skills), remote work capability, etc.

Furthermore, and this goes along with the data we have, we have developed algorithms to make this data talk – that is, to use our data in order to link concepts across the entire HR spectrum: from L&D to recruitment to mobility… So, our data & algorithms approach gives us an edge to tackle a whole variety of HR challenges: strategic workforce planning, career paths, learning and development…

IN A NUTSHELL

- **Jobs and skills repositories can be classified into three main groups: generalist, branch- or industry-specific and company-specific.**
- **Client-testing and exhaustivity are two important criteria to take into account when looking at ontologies.**
- **Having data is great, making them talk is even better: data is worthless without a profound layer of algorithms to link jobs, job titles, occupations, hard skills, soft skills and trainings together.**

4.

What are some examples of data science techniques behind it and why is it hard?

We use several data science techniques such as Natural Language Processing (NLP), Artificial Intelligence (AI), topic modelling, clustering techniques.

Some of the NLP techniques we use include word embedding and sentence embedding models such as BERT (Bidirectional Encoder Representations from Transformers, developed by Google) and USE (Universal Sentence Encoder) – which convert a piece of text into a numerical vector. Following this, we use the associated algorithms to compute distances between vectors (cosine similarity, Euclidian distances...).

One of the various tasks we need to achieve is to calculate the 'distance' between two jobs. To do that, as mentioned, we use sentence embeddings, which is a way to represent any phrase as a vector of numbers which captures its semantic meaning (i.e., the sentence meaning in context). From there, we can calculate the distance between two jobs by simply calculating the distance between their vector representations.

Another technique we use is artificial neural networks. The sentence embedding models we use rest on an advanced type of neural network architecture that has been developed to capture the meaning of texts and trained on all of Wikipedia's articles to do so. Neural networks offer several benefits:

1. They model well non-linear relations, which is highly adaptable to language processing.
2. They scale easily to accommodate training with very large amounts of data.
3. They mimic the architecture of the brain – which so far performs very well on language.

Other AI techniques we use include (non-exhaustive): feed forward neural networks, linear regression, multivariate algorithms, clustering techniques and topic modelling.

IN A NUTSHELL

- **Data science sits at the core of everything Boostrs does, from job title normalization to assigning the correct set of skills to everyone.**
- **In general, through the use of data science we have attained the most knowledge around our jobs and skills database so that we can apply it to any number of problems in the HR space.**

5.

How does Boostrs distribute its technology and what is the tech behind it?

One of the main questions for HR departments within companies or for HR software editors who wants to access a jobs and skills ontology is how they can integrate it into an IT system (whether it be an HRMS, an LMS, an ATS...).

There are two answers to that :

Long answer : it depends on the need at hand.

Short answer : through APIs (or Application Programming Interface). APIs are basically a software intermediary that allows two applications to talk to each other.

Most companies use a lot of various software tools when it comes to HR (in all categories: ATS, LMS, core HR, talent management...) because legacy systems have been inherited throughout the past of the company (different business units, mergers, etc.). Adding another layer of software is not necessarily what companies dream of.

The beauty of the API system we use at Boostrs is that our clients do not need to change their internal applications to implement our jobs and skills ontology. Instead, they can use directly the results of Boostrs algorithms and data to address their business needs. These needs could be, for example, to match people with jobs or with trainings. Long story short, our clients' applications remain unchanged, and we bring the insights directly in their applications via APIs.

So, in the end, our clients access our data in two ways :

1. They can browse through our data using the Skillmapper, which is a SaaS application containing the jobs and skills data of our clients.
2. They can access our data through APIs. We have Skillmapper-related APIs that will give them access to their customized data stored in the Skillmapper as well as other APIs that make use of our internal algorithms to search jobs, skills, extracted skills from a text, etc.

Accessing our data and algorithms is relatively quick. Once the Skillmapper account of a client has been loaded with the client's data – which takes typically 3 to 6 weeks from the kickoff meeting to the Skillmapper delivery, clients have immediate access to it with their own account.

If a client uses off the shelf's APIs, they also have immediate access to them after being granted authentication credentials. However, if an API has to be created or modified to suit a client need, additional time will be required – the length depending on the nature and magnitude of the changes that need to be done.

IN A NUTSHELL

- **Clients can access our data and algorithms through a SaaS or through APIs – which helps leave their applications unchanged while accessing our insights directly.**
- **Setup times are fast and generally take less than two months from kickoff to go-live.**

RESOURCES

- About Human Resources Management Ontology, the European Union
- HR Technology 2021 Now Published: Shattering Changes in the Market, Josh Bersin
- The War of the Skills Cloud (2020), Josh Bersin
- Workday Skills Cloud: A Big Idea with Much More to come, Josh Bersin
- The Ultimate Skill Data Handbook, Degreed
- How HR-ontologies map the working worlds, Filip de Geijter
- How a skills ontology framework keeps enterprises ahead of the curve, Adobe
- Forbes, taxonomies vs. ontologies (2019)

ABOUT BOOSTRS :

Boostrs is an API-first skills mapping engine.
By seamlessly connecting our data to HCM solutions, we help companies develop their employees in an ever-changing world.

We deal with 3 major challenges in the education and HR spaces:

- Managing a skills ontology
- Understanding how the Future of Work impacts jobs and skills
- Matching users their future career and learning paths

For more information, please visit :
www.boost.rs

Contact us :
hello@boost.rs

Our studies :

Remote work : an untapped potential ?
A skills-based measurement of the extent to which different jobs can be performed remotely.

The impact of automation on jobs and skills :
Who needs a skill boost?

The importance of soft-skills in the job market
An analysis of the soft skills demand by job, function and seniority level