Banco Interamericano de Desarrollo

A Skills Taxonomy for LAC: Lessons Learned and a Roadmap for Future Users

Labor Markets Division

Álvaro Altamirano
Nicole Amaral

NOTA TÉCNICA Nº
IDB-TN-02072

November 2020

A Skills Taxonomy for LAC: Lessons Learned and a Roadmap for Future Users

Álvaro Altamirano
Nicole Amaral

November 2020

Banco Interamericano de Desarrollo

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library
Altamirano, Álvaro J.
Skills taxonomy for LAC: lessons learned and a roadmap for future users / Álvaro Altamirano, Nicole Amaral.
p. cm. — (IDB Technical Note ; 2072)
Includes bibliographic references.
1. Vocational qualifications-Latin America-Classification. 2. Vocational qualifications-Caribbean Area-Classification. 3. Occupations-Latin America-Classification. 4. Occupations-Caribbean Area-Classification. I. Amaral, Nicole. II. Inter-American Development Bank. Labor Markets Division. III. Title. IV. Series.
IDB-TN-2072

JEL Codes: J01; J20; J23
Keywords: Skills, Occupations, Labor Markets Taxonomies, Job Vacancy Data; Machine Learning

http://www.iadb.org

A Skills Taxonomy for LAC: Lessons Learned and a Roadmap for Future Users

Alvaro Altamirano and Nicole Amaral

November 2020[1]

I. Introduction

Over the last several years, the Inter-American Development Bank has carried out a set of projects, initiatives, and partnerships to better understand how the demand for different skills and occupations is changing across Latin America and the Caribbean (LAC). The efforts to identify, track, and measure both occupations and skills have led us to a central challenge that the Inter-American Development Bank, and other organizations, have encountered: the lack of a common language and set of standards for comparing and analyzing skills across the region.

Occupational and skills taxonomies are structures developed to help provide this common language and set of standards for understanding and comparing skills. In its very simplest terms, a taxonomy is a classification scheme used to name and group related things based on discrete sets. Taxonomies often, but not exclusively, represent hierarchical relationships. A Skills Taxonomy, thus, provides a way to name and classify the variety concepts that we refer to when we say "skills" in standardized way that also expresses their relationship to each other. Such taxonomies also provide a base for analyzing which skills are foundational or transversal across many occupations, those currently most in demand for specific occupations, and how these skills needs are evolving over time. This analysis, in turn, can facilitate better career management, better coordination and articulation of training and recruitment needs (WEF), and empower more informed educational and career transitions.

The objective of this note is to bring together lessons from the IDB's and other institutions' efforts to adapt a skills taxonomy for Latin America and the Caribbean countries. These efforts have focused primarily on the ability to gather and make use of labor market information on skills demand from non-traditional data sources like online job vacancies. Most of these efforts have used the European Skills, Competences, Qualifications and Occupations (ESCO) taxonomy to underpin the identification and classification of skills.

This note is intended to be a starting point and set of considerations for policymakers who may be considering, or already embarking on, similar efforts to use ESCO or other taxonomical structures to help better analyze, understand and use skills-level information for decision making. It also seeks to motivate the need for additional classification systems that help governments take stock of its citizen's skills in increasingly complex and rapidly changing labor markets. This note is also companion to a set of earlier notes that focused in greater detail on understanding both occupational and skill taxonomies and ontologies, the classification of occupations and its trends in Latin America and the Caribbean using United States' Occupational Information Network (O*NET) (Amaral et al., 2018; Ospino, 2018; Altamirano et al., 2019a; Altamirano et al., 2019b).

---

[1] Alvaro Altamirano (alvaroalt@iadb.org) and Nicole Amaral (nicole5287@gmail.com). This note was developed within the regional technical cooperation RG-E1612. The authors appreciate valuable comments and suggestions from Juan Moreno, Diego Richards, Graciana Rucci, and Adolfo Unanue. All errors are the responsibility of the authors.

The lessons learned in this document are derived from several use cases and focus primarily on the lessons learned from the *adaptation* of ESCO in each. By adaptation, we mean the different methods used to make the ESCO classification work in the context of Latin American labor markets. While not the central focus of this note, we will also reflect on some most important considerations for a country looking to pursue a wide-scale *adoption* of a skills taxonomy at large, as for example the United States has done with O*NET and the European Union has done with ESCO, or for an institution to embed an ESCO-based system or tool on a smaller scale. These include the broader governance, infrastructure, legal and resource —both human and monetary— aspects that will allow these tools and taxonomies to remain relevant and useful over time.

This note is structured as follows: Section II provides a general overview of occupational and skills taxonomies, and their usefulness for both policy analysis and programmatic applications. Section III “zooms in”, using as case study the IDB and other institutions experiences in analyzing online job vacancies, illustrating opportunities and challenges of *adapting* the ESCO taxonomy for use in LAC. Section IV then “zooms out”, providing a set of considerations for policymakers for the broader *adoption* of a regional occupations-skills taxonomy, drawing on lessons learned from the IDB’s experience, but also the growing body of research and recommendations from other institutions that have embarked on similar processes.

II. **Context and Motivation: A Brief Introduction to Skills Taxonomies**

A. The development of occupational taxonomies[2]

In its most simple definition, a classification or taxonomy is a systematic array of objects placed into groups or categories according to an established criterion (European Commission, 2017). However, to understand skills taxonomies it is helpful to keep in mind the historical development of their precursors: occupations taxonomies and classification systems. Going as far back as the 1920s, the League of Nations and, subsequently, the International Labor Organization under the United Nations recognized the need to develop international standards and guidelines to help countries improve their labor administration as well as the quality and reliability of their labor statistics, and to improve international comparability of statistical data (ILO 1993). To these ends, an international standard classification of occupations, known as ISCO, was developed, and revised several times over the decades to its present iteration – ISCO 2008.

However, ISCO is intended to help provide alignment and comparability across countries, rather than replace individual countries’ efforts to classify occupations with the granularity and local specificity that reflect the reality of national labor markets. National-level classifications of occupations are used in local contexts for the collection and dissemination of statistics from sources such as population censuses, labor force surveys and other household surveys, employer surveys and other sources. They are also used by governments and companies in activities such as matching jobseekers with job vacancies, educational planning, reporting of industrial accidents, administration of workers’ compensation, and the management of employment-related migration (ILO ISCO-08 Methodological document, 2012).

The ideal, is that countries develop country-specific classification systems with ways to translate or align with international efforts like ISCO. Many countries have, in fact, developed their own local classifications systems independently. In LAC these tend to be bigger countries, examples

[2] For more detailed a conceptual discussion on taxonomies and ontologies, see Ospino 2018.

of which are Argentina's CNO, Brazil's CBO, and Mexico's CMO and SINCO[3]. Others have used ISCO as a starting point. The United States, for example, developed the Standard Occupational Classification (SOC) beginning in 1966, building subsequent "cross-walks" to ISCO to help better align the two systems for international compatibility[4]. Whereas Chile, for example, has the *Clasificador Chileno de Ocupaciones (CCO),* an occupational classification system that builds directly from ISCO, and is adapted to the Chilean context (See Ospino (2018) for other systems based on ISCO).

Most examples of occupational taxonomies mentioned above can be said to have include two minimum components (ILO ISCO 2008):

1) a *dictionary* of occupations. Dictionaries *describe* of the common tasks and duties that make up an occupation. Occupations are essentially composed of similar jobs, with a job defined as a set of tasks and responsibilities performed by a person, for an employer or for oneself (International Labor Organization (ILO) 2012). While an occupation is understood as a set of jobs that are carried out, with slight differences, in multiple establishments, and not necessarily within the same industry (Ospino 2018). The range and types of tasks that make up each occupation can, and does differ, across different national occupational classifications.
2) *criteria* for classification. These are the rules or guidelines that decide how to classify jobs into occupations. Again, these can differ slightly across different national contexts, depending on the dynamics of local labor markets.

B. Skills as the building blocks

Over time, however, occupational taxonomies have evolved to better describe what makes up an occupation, not only in terms of the tasks and duties that characterize an occupation, but also to include the *skills* required for their execution. O*NET and ESCO are two most well-known examples skills taxonomies, which classify and map both occupations and the requisite skills for each. Increasingly, both taxonomies have become "ontologies" which not only help classify occupations and skills, but how these also relate to each other, as well as to other classification systems, such as qualifications frameworks.[5]

However, there is no universally accepted definition of "skill", and there is significantly more dispersion in the use of the term, than that of occupation. Skill is a term that is used interchangeable with other terms like competency, or as a catch-all that is used to describe a range of concepts that include those that are learned or acquired—whether on the job or through training and education-- those that are more frequently referred to as innate personality traits, or even to indicate the mastery of certain tools or technologies (such listing mastery of the Python programming language as a skill).

Each taxonomy has approached the variety of skill-related terms in different ways. O*NET, for example, differentiates between skills and abilities. Skills are defined as competencies developed

[3] This does not mean that smaller countries have not developed their own occupational classifications. Examples of this are Bolivia (COB), Costa Rica (COCR), or Paraguay (CPO). The main difference is that smaller countries' taxonomies are generally more aligned with ISCO, at least for higher aggregation levels; and for that matter provide clearer paths for what is known as taxonomy crosswalks.

[4] See: https://www.bls.gov/soc/2000/frn-feb-28-1995.pdf

[5] An ontology defines each element of a knowledge base and how they relate to each other. In the context of occupations, an ontology allows one taxonomy of occupations to be related to other taxonomies (Ospino 2018).

through training or experience, while abilities are defined as relatively stable attributes for individuals' capacity to perform a particular set of tasks (Fleisher and Tsacoumis 2012a, 2012b). Abilities are grouped into four categories: cognitive, psychomotor, physical, and sensory. For their part, skills are grouped into seven categories: content, process, social, complex problem solving, technical, systems, and resource management. Information regarding abilities and skills is collected through continuously surveying occupational analysts, since they understand the abilities and skills constructs better than those performing the job (Reeder and Tsacoumis 2017a, 2017b).

ESCO, on the other hand, contains 13,485 concepts that distinguished between two “skill-types”: i) skill/competence concepts ii) knowledge concepts. These are further broken down in a hierarchy which contains four sub-classifications which include 1) Knowledge, 2) Skills, 3) Attitudes and values, and Language skills and knowledge (See box 1 for more detail on ESCO).

Box 1. A spotlight on the European Classification of Occupations and Skills (ESCO).

What is ESCO?

ESCO is the European classification of skills, competencies, qualifications, and occupations. Although many countries in European maintain their own national-level classification systems, ESCO was developed to provide a common understanding of the most common occupations and skills that characterize the broader European Union Labor Market. ESCO aims to meet multiple objectives among which are:

- Improve communication between the training sector and the European labor market;
- Support labor mobility in Europe, make data transparent and readily available to different stakeholders, such as public employment services, educational institutions, and statistical agencies;
- Facilitate the sharing of information between employers, education providers, and job seekers regardless of language.

To do this, it is structured around three pillars: 1) occupations, 2) skills, knowledge, and abilities, and 3) qualifications. The occupations pillar is based on the international ISCO-08 classification, but the level of detail is much higher; while ISCO has 426 occupations at its most detailed level, ESCO has about 3,000 detailed occupations (with the ability to aggregate occupations at levels comparable to ISCO-08, thus allowing for international comparison). The skills, knowledge, and abilities pillar contain a comprehensive list of relevant European labor market skills. The v1 version of ESCO has 13,485 objects in this pillar between skills, knowledge, and abilities. The ESCO qualifications pillar collects information from two sources: (1) The qualifications databases of member countries and (2) the qualifications provided directly to ESCO by those conferring these qualifications.

In ESCO, the first four occupational taxonomy levels correspond exactly to the ISCO-08 classification, while levels 5 and up are ESCO-specific. There are advantages in being able to aggregate each occupation listed in ESCO at the most disaggregate ISCO level, therefore ensuring cross-country comparison. On the other hand, specificity at the local level is not sacrificed, since each occupation is broken down into relevant categories for which there is detailed information on qualifications, skills, and abilities, according to the needs of each country.

In each occupation description, ESCO defines two types of skills/competences[6]: essential and optional. The first refers to the competences that are generally relevant to an occupation, regardless of context, employer, or country. While the latter are those that do depend on the context, employer, or country. These are considered of vital importance for job matching in the labor market as they reflect the variety of jobs within the same occupation (European Commission 2017).

Sources: Adapted from ESCO; Ospino (2018).

[6] ESCO uses the terms skills and competences interchangeably.

Both ESCO and ONET, however, have occupational-skill taxonomies that are *expert-driven*. This means that they have been developed largely through a top down process, in which experts with knowledge of each industry and occupation define those tasks, duties and required skills that characterize an occupation. The benefit of expert-driven approach is that the categories are highly meaningful, standardized and with clear criteria. The downside, however, is that both developing such taxonomies and keeping them updated are resource and time-intensive activities that few countries have been able to undertake. The development and upkeep of such an initiative demands high levels of institutional coordination and requires specialized knowledge on the part of the stakeholders involved. Therefore, processes can be lengthy, and it is common for updates to take decades (Ospino 2018).

More recently, however, the advent of large scale, non-traditional data sources combined with a range of data science techniques, including Machine Learning (ML), have made it possible to develop a "ground-up" approach to skills taxonomies, that take significantly less time and resource to develop and maintain. Companies like LinkedIn, ESMI or other similar initiatives use actual jobs and skills data to determine the mathematical relationships between skills and occupations,[7] and let a classification system arise from the data itself, rather than classifying occupations and skills based on pre-defined categories and criteria. While this approach allows for a more flexible taxonomy that can be developed faster and more easily adapts to change, it continues to present challenges of comparability across initiatives and analysis as the categories from these "organic" taxonomies are not standardized, they may change as the underlying data changes, and the criteria for classification arising from complex algorithms may not always be transparent or meaningful to external audiences.

As a result, other initiatives like The Open Skills Project[8] led by the University of Chicago have combined the two approaches: they are focusing on providing a dynamic, up-to-date, locally-relevant, and normalized taxonomy of skills and jobs that builds on O*NET and data science techniques along a variety of data sources to expand and update O*NET. Moreover, initiatives like the Open Skills project are also open source[9], providing their different tools either as API interfaces and/or as public code repositories for developing and analyzing skills and competencies from unstructured text.

For Latin America and the Caribbean, several studies and initiatives have taken advantage of O*NET and ESCO/ISCO for analyses related to *occupations* (Ospino, 2018; Altamirano et al., 2019a; Altamirano et al., 2019b). However, there is not yet a standard for a regional occupations-skill taxonomy that countries can adopt that would guarantee consistency with each of the local occupational classification systems, but also generate comparable regional statistics on the skills and abilities across the region.

A skills classification system for LAC could help underpin analysis both within countries, as well as across them, to help guide job seekers through information on job opportunities and vacancy requirements; and assisting those who wish to be trained for entry into the labor market or make a career change. It can also help policymakers develop programs and policies that focus on skills

[7] See Amaral et al (2018) for an example.
[8] See http://dataatwork.org/data/
[9] Open source denotes software for which the original source code is made freely available and may be redistributed and modified (given specific licensing criteria).

as the dynamic building block of jobs, helping to better transition people from one job to another, or to upskill or reskill in a more efficient manner, when needed.[10]

While O*NET and ESCO and other taxonomies provide a starting point on which to build, these were developed in other contexts – based on the United States' and European labor markets--and provide a number of challenges and necessary adaptations to be useful in Latin American contexts. To better understand and illustrate these challenges this note provides a set of lessons learned from analyzing online job vacancies data as a potential source of labor market information, leveraging ESCO. These case studies help illustrate and motivate the need for tools that support the greater standardization of skills level data—and potentially the development of a skills classification system that could ground future initiatives in LAC.

The subsequent sections will discuss lessons learned from the Inter-American Development Bank's and other institutions' experiences with gathering, processing, and analyzing occupations and skills level data and information from online job vacancies, leveraging ESCO to support this effort.

III. Adapting ESCO: Challenges and Solutions in Classifying Skills and Occupations from Online Job Vacancies (use case)

A. Learning about occupations and skills from online job vacancies

Online job boards have grown all over the world in recent decades and have become an increasingly important source of labor market data (Kurekova et al, 2015; Kuhn & Masour, 2011; Stefanik, 2012). These portals provide massive, real-time information on job-specific skill demands, with adequate regional and occupational granularity (Cedefop, 2018). Additionally, most websites or portals have well-structured and detailed indicators for each job vacancy advertisement.

Many countries and both public and private entities have recognized the potential of online job vacancies as a source of labor market information. In the developed world, two initiatives stand out, the European Centre for the Development of Vocational Training's (CEDEFOP) Online Vacancy Analysis Tool for Europe[11], and Burning Glass-based research for the USA. In Latin America, a few national projects using job portals data include Colombia's Public Employment Service's (PES) vacancies registry, Chile's Sistema de *Análisis de Bolsas de Empleo (*SABE) project, Ecuador's INEC explorations -in collaboration with the World Bank- (Benitez et al., 2018), and México city Labor Bureau's DiCoDe project (Sierra et al., 2020).

One of the central challenges, however, once the data is obtained, is identifying, and classifying both occupations and skills from the text of thousands, or even millions of online job postings. Between 2019-2020, the IDB undertook an effort to cull information from online job boards in LAC and to employ Natural Language Processing (NLP) and Unsupervised Machine Learning (UML) algorithms to build a model capable of identifying and classifying occupations and skills from online job postings. The IDB's model leveraged ESCO to enrich the occupations and skills data.

---

[10] The Open Skills Project provides a good overview of the different potential use cases of skills classification systems and tools. See Crockett et al (2018). Skills-ML: An Open Source Python Library for Developing and Analyzing Skills and Competencies from Unstructured Text. http://dataatwork.org/skills-ml/SkillsMLWhitepaper.pdf

[11] https://www.cedefop.europa.eu/en/data-visualisations/skills-online-vacancies/online-job-advertisements-providers

Several other initiatives in the region were, simultaneously, conducting similar exercises that also used ESCO (See Box 2 for a description of the techniques employed in each exercise).

From the IDB's own experience, as well as interviews conducted with the teams leading these initiatives, we gathered a set of lessons learned, challenges, and potential solutions described below. It's important to note, however, that all of these initiatives are a work in progress, and the aim of this note is to help share knowledge and facilitate discussion and exchange of experiences as different countries and institutions across the region embark on similar efforts. This note does not seek to evaluate these initiatives and their outputs, but rather to make more transparent the types of questions encountered and decisions to be made in the process.

Box 2. Understanding Occupations and Skills Data from Online Job Vacancies: On-going efforts across Latin America and the Caribbean.

Inter-American Development Bank: The Inter-American Development initiated a web scraping exercise to cull information from several of the principal online job boards across five countries: Mexico, Colombia, Argentina, Chile, and Peru. This exercise resulted in 2.9 million observations from job vacancy ads in those countries. The goal of the exercise was to explore how online job vacancies could be used to identify and extract data on jobs and skills to analyze changes in demand, as well as the skill composition of different occupations, in a manner comparable across several countries. To do so, the IDB collaborated with Quantil, an applied mathematics firm, to develop an unsupervised machine learning model capable of ingesting the job vacancy information and classifying each vacancy into a corresponding set of ESCO occupations and skills categories. The final product of this project is an open-source Python repository that allows users to input raw vacancy data (Title and Description text strings), process it, and return a classified dataset using ESCO occupations and skills taxonomies. More information about this project's methods and challenges is detailed over the rest of the note. In parallel, the IDB also consulted with and interviewed other teams in the region undertaking similar initiatives to exchange experiences, challenges, and potential solutions.

Colombia: As of 2015, the Colombian Public Employment Service has been collecting information from online job vacancies reported by employers to the Public Employment Service (PES). The PES contracted researchers from the Universidad Javeriana to conduct a skills text analysis of the vacancy description and applicant resumes received by the PES with the end goal of measuring the skill mismatch by calculating the minimum distance between the skills, competencies and capacities required in vacancies against those that workers claim in their resumes. The project matched skills found in vacancies and resumes with the ESCO ontology using manual matching, combined with *fuzzy matching,* a technique that calculates correspondences between segments of text that allows for non-exact matches of the target skill. The database was created with all the matched skills and were subsequently classified according to the ESCO network into Cross-cutting skills and Occupation-specific skills. The complete registry of skills was combined to build an ontology of skills and occupations for the Colombian Public Employment Service based on the data they had collected from job vacancy texts and resumes. Construction of this ontology later facilitated a set of calculations regarding skill mismatch and gender imbalances in the Colombian labor market (See Diaz & Salas, 2020).

Chile: The Chilean Labor Market Observatory launched the construction of a system to analyze job vacancy data from online job boards in Chile (known as *Sistema de Análisis de Bolsas de Empleo (SABE)*, in Spanish). SABE is one of several instruments in a battery of analytical tools being developed to generate labor market information about skills gaps, and changes in the demand and supply of jobs and skills in Chile at both national and subnational levels. SABE is intended to complement traditional sources like household and employer surveys providing a way to monitor the changing demand for

occupations (as classified by the Chilean occupations classification CIUO-08-CL) and skills (using ESCO's skills taxonomy as reference) expressed by companies in the principal online job boards used in Chile. The SABE team uses a Supervised Machine Learning (SML) classification approach, taking the initial data set of online job vacancies provided through data sharing agreements with several of the key public and private job boards across Chile, and manually tagging 5000 vacancy announcements with occupational titles found in CIUO-08-CL, and skills using the ESCO taxonomy. The labeled sample of five thousand job ads will then be used to train occupations and skills classification models, which in turn will classify unlabeled vacancies. The project is in still in its initial stages, having nearly completed occupational tagging, but not yet having initiated the tagging of skills using ESCO's taxonomy.

Source: Authors based on interviews.

B. Choosing a Reference Taxonomy

In the absence of a skill taxonomy in LAC that links skills with occupations, choosing an existing occupational-skills taxonomy to serve as the basis of a classification exercise provides a shortcut, with the important acknowledgement and caveat that taxonomies developed to reflect other labor markets may require significant adaptation to be useful locally. Frequently, the choice is cited as being between O*NET and ESCO. For the IDB exercise, as well as several of the others interviewed for this note, ESCO was selected as the best available taxonomy to adapt and leverage.

One of the primary reasons it is frequently selected in LAC over O*NET is language availability. The set of terms and descriptions for both occupations and skills available in the ESCO ontology is available in 26 European Languages as well as Arabic[12]. This includes the 4 primary European languages spoken in the LAC region. O*NET, in contrast, only has a Spanish translation[13] of its SOC-O*NET taxonomy.

Second, occupation codes in ESCO are also mapped with ISCO-08, which (as noted in earlier in this note) most countries in the LAC region already have mapped to, or use as the basis for, their national-level occupational classification systems. This facilitates a faster integration of additional occupational information, as well as mapping ESCO's skills level taxonomical structure to local occupational classification systems—providing a way to link occupations with their requisite skill set.

Third, ESCO recently updated the structure of its skills pillar (see Box 1) to provide additional hierarchy, rather than just a dictionary of skill terms. Prior to this 2019 update, an advantage of O*NET was that its skill terms were organized with a meaningful hierarchy, allowing for aggregation/disaggregation and association of skill terms across its levels and related occupations, whereas ESCO provided a dictionary of skill terms without a hierarchical structure. In fact, to remedy this weakness, ESCO drew on and adapted both O*NET's skill taxonomy as well as some of the groups of the Canadian glossary of skills to develop the first and second layers of the ESCO skills hierarchy (Figure 1).

[12] In Europe, ESCO's language availability has permitted to address the specific training needs of different waves of migrant workers.
[13] https://www.onetcenter.org/spanish.html

Figure 1. Example of ESCO's Occupation and Skills Hierarchy in Spanish

a. Occupations tree

- 0 - Ocupaciones militares
- 1 - Directores y gerentes
- 2 - Profesionales científicos e intelectuales
- 3 - Técnicos y profesionales de nivel medio
  - 31 - Profesionales de las ciencias y la ingeniería de nivel medio
    - 311 - Técnicos en ciencias físicas y en ingeniería
    - 312 - Supervisores en ingeniería de minas, de industrias manufactureras y de la construcción
    - 313 - Técnicos en control de procesos
    - 314 - Técnicos y profesionales de nivel medio en ciencias biológicas y afines
    - 315 - Técnicos y controladores en navegación marítima y aeronáutica
  - 32 - Profesionales de nivel medio de la salud
  - 33 - Profesionales de nivel medio en operaciones financieras y administrativas
  - 34 - Profesionales de nivel medio de servicios jurídicos, sociales, culturales y afines
  - 35 - Técnicos de la tecnología de la información y las comunicaciones
- 4 - Personal de apoyo administrativo
- 5 - Trabajadores de los servicios y vendedores de comercios y mercados
- 6 - Agricultores y trabajadores calificados agropecuarios, forestales y pesqueros
- 7 - Oficiales, operarios y artesanos de artes mecánicas y de otros oficios
- 8 - Operadores de instalaciones y máquinas y ensambladores
- 9 - Ocupaciones elementales

b. Skills tree

- S1 - comunicación, colaboración y creatividad
- S2 - competencias en materia de información
- S3 - prestar asistencia y cuidados
  - S3.0 - prestar asistencia y cuidados
  - S3.1 - ofrecer orientación
  - S3.2 - prestar atención sanitaria o administrar tratamientos médicos
  - S3.3 - proteger y velar por el cumplimiento
  - S3.4 - facilitar información y apoyo al público y a los clientes
  - S3.5 - preparar y servir alimentos y bebidas
  - S3.6 - prestar cuidados personales en general
- S4 - capacidades de gestión
- S5 - trabajar con ordenadores
- S6 - manipular y mover
- S7 - construir
- S8 - trabajar con maquinaria y equipo especializado

Source: Authors based on ESCO.

Fourth, ESCO is also linked to the skills and qualifications content of European occupations[14]. While several countries in LAC have embarked on building their own qualifications frameworks, ESCO's linkage to European qualifications can serve as a starting point in the construction of qualifications frameworks in LAC countries, or help to reinforce and update those qualifications frameworks already under construction. The mapping to qualifications is one of the critical elements that connect the changing skill needs of the labor market to the education and training institutions that help people build these skills. The ESCO governing body is currently undergoing a pilot project for testing the use of an *automated* approach for linking learning outcomes of qualifications to ESCO skills that could also provide lessons learned for countries in the LAC region.[15]

Finally, ESCO is open-source and can be directly linked to national classification systems through its API. ESCO is not unique in this regard, but it nonetheless allows countries to legally leverage, modify and adapt ESCO's taxonomy for their own needs. The API facilitates the incorporation of updates to ESCO in near real-time, as well as to build public-facing applications that use the

[14] https://ec.europa.eu/esco/portal/qualification

[15] See https://ec.europa.eu/esco/resources/escopedia/20200320_150541/bad1c9dd-57ff-4e20-9f94-63dd97983054ESCO_MSWG_11-3_Initial_results_from_the_pilot_linking_Learning_Outcomes_of_qualifications_with_ESCO_skills.pdf

ESCO taxonomy.[16] The main limitation of ESCO, however, is that it is of recent creation, while O*NET has been in place for years. For this reason, O*NET can continue to shed light on how updates should be carried out and on good practices for collecting the information that nourishes the ontology thereof (Ospino, 2018).

Some of the initiatives we interviewed also considered proprietary and/or local variations of occupation-skills taxonomies; the Colombian team, for example, found proprietary taxonomies built by private companies they consulted to be adequate but very costly, as well as potentially unsustainable as the long-term foundation of publicly available services and analysis desired by the Colombian public employment service. Similarly, in Colombia, the National Training Institute -SENA- maintains the Colombian National Occupational Classification System, which includes some associated skills; however, the researchers found that these were not sufficiently mapped to each occupation and were not organized as a skills classification system to be useful for the planned exercise. As a result, ESCO was also selected as the most adequate starting point for the classification of occupations and skills from the Public Employment Services vacancy and resume datasets. An exemption is Mexico's city Ministry of Labor and Employment Promotion (STyFE, in Spanish) vacancies project, which used one of the job portal's (Bumeran) occupation taxonomy to train their classification models. They did not use Mexico's official classification system (SINCO) because they considered "its classification system and vocabulary/jargon differs from the one used by" online jobs postings (Sierra et al., 2020).

C. Identifying and Classifying Occupations and Skills

One of the first choices faced in the IDB exercise, as well as others examined, is whether to pursue a supervised or an unsupervised machine learning approach[17]. In simple terms, a supervised learning classification model uses a labeled[18] dataset to train an algorithm that is later used to classify the rest of the data. A common example is the classification of tweets or movie reviews into positive or negative sentiments, which under a SML approach relies on previously annotated tweets or movie reviews that help the chosen algorithm create numerical vector representations for the association of words used for positive or for negative human expressions in those contexts.

An unsupervised machine learning classification model, in contrast, does not need previously labeled data. It relies on the extracting of features and patterns from co-occurrence statistics of words in a semantic space (Figure 2). In other words, it aggregates statistics of the contextual use of words in a given corpus, both by representing those words in a shared mathematical space and by reducing the dimensional representation of their contextual usage.

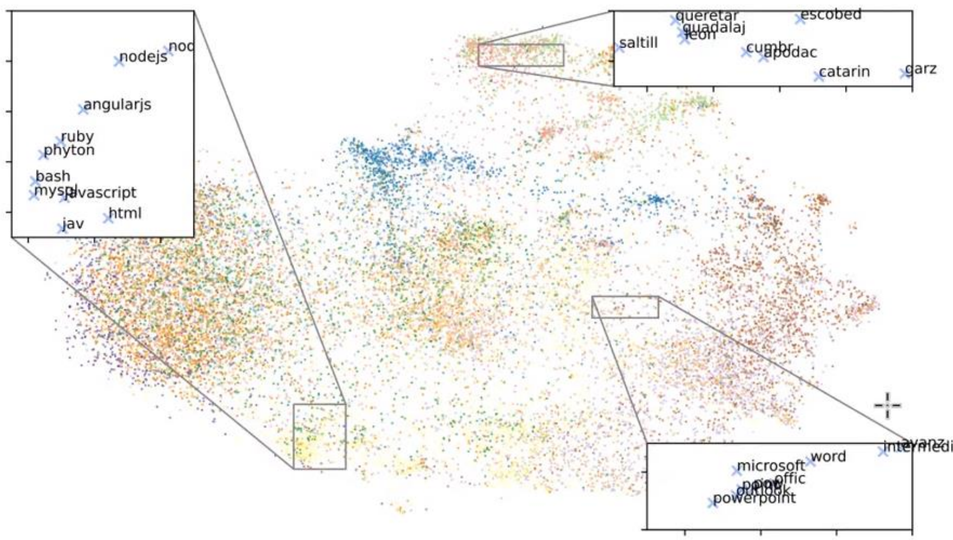
In the context of classifying labor occupations, for example, it first creates a dictionary from job postings' texts composed by a factorization of the numeric vector representations of the co-occurrences of their words in a shared space. A second step uses ESCO occupation titles and descriptions to create another similar dictionary, a condensed or reduced matrix representation

[16] See, for example, Amsterdam's *House of Skills* project. https://www.houseofskillsregioamsterdam.nl/about-house-of-skills/

[17] Machine learning models can be used to classify (eg. True or False, Poor or Non poor) or to estimate mean predicted values (eg. Sales, Income). This note focuses both on *classification* and *natural language* experiences represented by the need to classify human generated texts into standardized occupation and skills. Specifically, it details initiatives classifying job boards data.
[18] Ideally classified by human subject-experts.

containing most of the information ESCO labels provide. A final step applies ESCO's labels to each vacancy based on estimates of the similarity of their texts using their cosine distance as matching criteria.

Figure 2. Representation of Words in a Mathematical Space



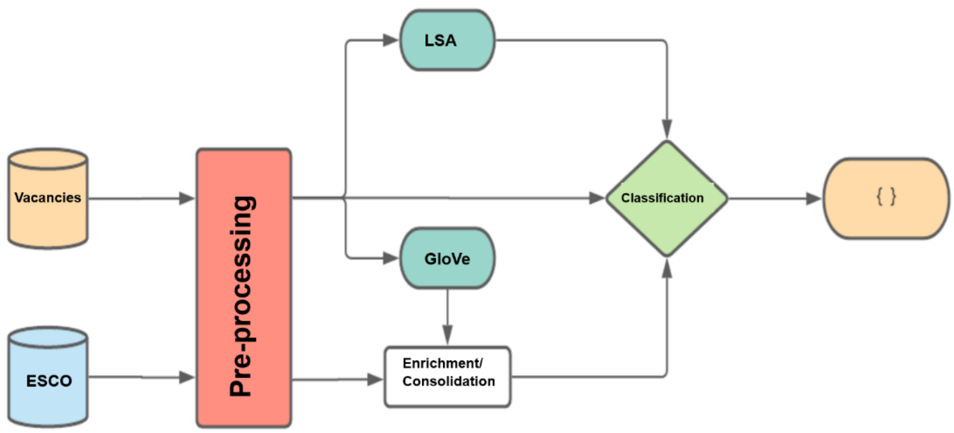
Source: IADB-Quantil (2020).

Each approach offers pros and cons: The supervised approach requires the manual tagging of thousands of vacancies announcements collected to create a set of labeled data. This step can be expensive in human resources and in time, especially given the volume of data (2.9 million vacancies). Ideally, hundreds of thousands of labelled data would be needed to accurately classify occupations and skills at their outmost detailed ESCO level. Assuming a project has sufficient resources to label (in-house or outsourced) a vast amount of observations, this approach potentially offers more accuracy for a specified classification level (e.g. ESCO at its 3-digits or even 4-digits disaggregation). For example, in Chile the SABE team currently has about 10 experts tagging of a set of 5,000 vacancies using the labels from the CIUO-CL_08 occupation titles (and is subsequently planning to tag skills in vacancies using ESCO skill terms). This team has been working over the course of 5 months nearly full time. Each vacancy text is tagged at least twice by different people to ensure consistency, resulting in 10-15k occupation tags. However, two main challenges arise: the first, is the need to dedicate a team of sufficient people, with the requisite knowledge and expertise to make informed decisions about how to tag the vacancy information accordingly, which incurs a high cost up front. A second challenge is determining the quantity of vacancies that need to be tagged to both train and validate the model. There is no pre-determined threshold, but rather is dependent on the end target to be identified and classified, and the precision desired. At a higher level of aggregation for occupational groupings (ESCO at 1 or 2 digits), a smaller training set may be feasible to tag manually. At the level of skills, which in ESCO for example there are thousands, consulted ML experts told us more than a 100,000 vacancy texts tagged may be needed to train and test a model. Labeling that quantity of vacancy texts may be unfeasible where resources are limited.

Due to both resource and time limitations, the IDB opted for an unsupervised approach (see Box 2) that combined multiple natural language processing techniques to overcome the lack of a pre-

labeled set of data. This combination of machine learning techniques in essence created its own training data set by matching vacancies' texts with ESCO taxonomic trees. For the classification of occupations, the model's default configuration creates vacancies vectors that give a weight of 77% to job titles and 23% to job descriptions. The classification of skills only used the job posting's description strings, it did not include their titles. On ESCO's end, the words embedding process used the concatenation of their text string at all levels (1 to 4-digit tree-maps). Besides a robust battery of text pre-processing tools, the exercise relied mostly on the semantic consolidation and enrichment of vacancy data and ESCO's occupations and skills classification trees through Latent Semantic Analysis (LSA) and Global Vectors for Word Representation (Glove). These NLP methods construct semantic dictionaries which are finally used as training data over a cosine distance-based similarity classifier (Figure 3).

Figure 3. Model's General Structure



Source: IADB-Quantil (2020).

However, this approach is also not exempt from challenges. The first is represented by the NLP code development needed to create the training set from the underlying corpus of text being used from both the vacancies themselves as well as from the multi-layered ESCO database of occupations and skills terms. Figure 4 illustrates two examples of the type of raw job vacancies fed into the model, composed by two main elements: its title (*Título*) and description (*Descripción*).

Figure 4. Two Examples of Job Vacancies

Titulo de la vacante: sastre palacio hierro huixquilucan estado mexico

Descripción: el palacio de hierro interlomas te invita a vivir la mejor experiencia de trabajo en mexico como: -i-i- sastre -i-i- requisitos: escolaridad: preparatoria o bachillerato concluido indispensable experiencia: preferentemente 1 aÑo en area solicitada. actividades a realizar: promover servicio de hechuras a la medida colaborando con el vendedor durante la labor de prendas a la medida tomar medidas de acuerdo a procesos por tipo de producto solicitar al proveedor el servicio o materiales dar seguimiento a produccion. horario de trabajo: lunes a domingo de 10:45 a 21:15 hrs. con un dia de descanso entre semana. ofrecemos: - atractivo sueldo base $9600 mensuales - prestaciones superiores a las de ley vales despensa servicio de comedor servicio medico transporte gratuito entre otras - contratacion via planta. - capacitacion. zona de trabajo : el palacio de hierro interlomas

Titulo de la vacante: auxiliar bodega montacarguista empresa

Descripción: importante empresa de logistica requiere auxiliar de bodega montacarguista bachiller con experiencia minima de un 1 año en bodega y manejo de montacargas electrico y de combustion realizando labores de recepcion almacenamiento alistamiento y despacho de mercancias con el uso correcto del montacargas. debe tener obligatoriamente curso de manejo de montacargas y certificado de alturas ambos vigentes . persona dinamica agil excelente trabajo en equipo y relaciones interpersonales horario: lunes a viernes de 7:30am a 5:30 y sabados medio diasalario: $ 1.000.000 + auxilio de transporte + prestaciones de ley. interesados en la vacante presentarse los dias 14 y 15 de agosto de 9:00 a.m. a 10:00 y de 2:00pm a 3:00pm en la calle 72 #10-70 torre a oficina 606 con hoja de vida fotocopia de cedula certificado montacargas y alturas documentos al dia.

Source: Authors.

In the same vein, the unsupervised approach also presented several methodological and computational challenges. As an unsupervised approach, the IDB and Quantil researched and tried several models to see which would produce the best results.[19] When recreating models such as lda2vec or doc2vec, a custom training for the vacancies database was done. In the case of lda2vec, this took 48 hours with 100% of the database. On the other hand, for doc2vec, the computational resources initially allocated were not sufficient, finally resulting in a 4-day turnaround. Although much larger quantities of data are often used in the data science industry, such work is usually carried out with pre-trained models. However, to generate a custom solution for the job vacancies database, these models were developed from scratch, which translates into longer development and execution times.

Lower metrics of accuracy are expected for classification tasks involving occupations and skills at higher levels of detail in terms of the ESCO hierarchy. Consequently, the model's default classification output was set to return the nearest (cosine distance) 3 occupations and 5 skills for each inputted vacancy, at their 2-digit[20] and 3-digit level, respectively. Hyperparameters' tuning allows to obtain either a deeper or a more aggregated classification level in terms of ESCO's hierarchy (e.g. 4, 3 or 1-digit ESCO's occupations and skills).

Finally, the challenge with most unsupervised models is evaluating their performance, and frequently expert human input or revision is needed to ensure the results are meaningful. A validation of the model's classification performance metrics in on course, and as an initial step the IDB hired a group of 4 experts to label a random sample of 3 thousand vacancies into ESCO's 2-digit level occupational classifications[21]. This labeled sample will then be used to estimate the model's performance metrics (precision, recall, F1) as they relate to occupational classifications[22].

D. Translating ESCO to local dialect and contexts

As mentioned, one of the advantages of using ESCO is the availability of language packs, including the languages most used across LAC: English Spanish, Portuguese, and French. However, differences in dialects across LAC mean that the Spain-based dialect on which ESCO is based creates a set of additional hurdles for adapting and using the taxonomy locally. The first of these challenges is that the entirety of the ESCO taxonomy has not yet been fully translated. For the most part, this translation includes all titles for both occupations and skills taxonomies, at all levels of their tree map disaggregation, but it rarely includes occupations and skills descriptions text[23]. For the IDB exercise, which relied on using the full set of ESCO descriptions, translations of terms were required and conducted using the Google Translate API[24]. The validation of the automated translation of these terms was not conducted, and the impact of any potential errors is unclear, although likely minor.

The second challenge is matching ESCO terms in European Spanish, with those in in local dialects that convey similar meaning with through a different term (i.e., synonyms), or even a different meaning with the same term (polysemy). Labor markets vary greatly across geographies

---

[19] These included LDA, word2vec, LDA2vec, Doc2vec, TF-IDF, BERT among others.
[20] Equivalent to ISCO's Sub Major Group.
[21] An ad hoc tool was developed for this task: https://anotador.herokuapp.com/
[22] The validation for skills is less parsimonious as it has a bigger number of classes. As such, we are currently discussing whether a supervised machine learning approach for the validation of skills classification would suffice.
[23] See https://ec.europa.eu/esco/portal/occupation
[24] https://py-googletrans.readthedocs.io/en/latest/

or even across different industries in the use of different terms to indicate specific occupations or skills. A common example that has risen in the process of building the validation sample discussed in the previous section is the case of "Executives" (Ejecutivos in Spanish). In Latin American labor market jargon, the word "Ejecutivo" is used in different contexts and represents distinct positions within firms. As such, a vacancy calling for *Ejecutivos* may refer to customer attendants, sales reps, or managers. On the other hand, in the standard ESCO/ISCO taxonomy, the descriptions for *Executive* occupations denote a higher hierarchical level that cannot truly be associated to, for example, sales reps or customer attendants.

In Colombia, researchers from Universidad Javeriana addressed this issue using a technique called *fuzzy matching*. As a first step they matched occupations and skills with exact terms found in ESCO, and a second matched against terms that were not a 100% match using fuzzy matching, which calculates the similarity between the component parts of words. Fuzzy matching was used to match skills terms in the job vacancy texts to those in ESCO that had similarity of 90% and over, and those that had a similarity index of 60-89% were then manually reviewed and tagged. In Chile, however, the "translation" to Chilean Spanish and local usage of terms is done through manual coding by the experts employed in the tagging exercise, who make these judgement calls using their own knowledge of local labor markets.

The IDB's exercise, on the other hand, relies on NLP techniques, and enriching the vacancy data with ESCO's skills terms, descriptors and hierarchy to build a model that can identify occupations and skills from job vacancies across 5 different countries. A potential challenge, however, is that language used to describe different occupations and skills, even within the same sector, can vary across country. The ability of the model to learn these differences depends on having a sufficient sample of job vacancy texts from each country. Similarly, ESCO's descriptions may simply differ from the contextual description of occupations across Latin American countries, all of which increase the possibility the model commits classification errors. For the IDB team, the extent of these potential errors is not yet clear, as the final test stage of the model is planned for next year.

E. Representation challenges

A final set of challenges in adapting ESCO for use in Latin America is how to understand the representativeness of the ESCO taxonomy—at both the level of occupations and skills—for labor markets in LAC. As aforementioned, ESCO was developed through expert-consultants, and it represents those skills and occupations mostly commonly found across countries in the European Union. Labor markets in LAC, however, differ in significant ways from those in many EU countries—as evidenced in some of the exercises conducted by the IDB and other initiatives interviewed for this note. First, there are occupations and even economic sectors that are found in ESCO, but that are less common in LAC. However, the reverse is also true; in Colombia, for example, the *Javierana* team noted that the mining sector and the military are important sectors of the Colombian labor market, however these occupations are less represented in the ESCO taxonomy. This is in large part because ESCO is expert-driven, meaning that resources were allocated to develop occupational and skill profiles for those sectors that were most common across the EU. Although many EU countries have an important mining sector, this sector has been less well-developed in the taxonomy, and does not reflect the range of occupations in Colombia's mining sector.

For the IDB exercise, which seeks to provide a machine learning model that is capable of adapting to different countries in the region, the upcoming validation exercises will be critical for

determining how well the model will identify these differences, as well as how/whether it will be capable of identifying occupations that are currently not in the ISCO/ESCO occupations taxonomy –whether because these are emerging or are unique to the local labor markets. As yet, there hasn't been a comparison done with local "ISCOs"—those local occupational taxonomy adaptations that better reflect the particularities of each countries labor markets, including accommodations for the large informal economy that is absent in Europe. This kind of comparison will be a critical part of understanding the usefulness and accuracy of the IDB's model.

A similar challenge may occur at the level of skills, although the exercises conducted by the IDB and the SABE team in Chile are still in progress and are not yet able to evaluate this aspect. However, in Colombia the Javierana team noted that many skills in the job vacancy and resume data from the Public Employment Service did not have a closely-worded match in the ESCO taxonomy and were manually classified by the team of researchers. Additionally, a lingering question is the extent to which the skill composition of an occupation differs by geographic location—whether at the national or subnational level. For example, in Chaparro and Franco (2018) surveyed human resources professionals in Colombia using a survey comparable to that used by O*NET in the United States when collecting information on skills and abilities. In short, the authors used the O*NET methodology to gather information from the Colombian market. Their preliminary results for thirteen occupations in the civil engineering industry confirm that there are significant differences in ability intensity reported by occupational analysts. In Colombia, physical ability requirements would be higher in almost all occupations when compared to those reported in O*NET. For less-skilled occupations, there appears to be greater demand for social and cognitive abilities in the United States than in Colombia (Ospino 2018). For further adaptations and applications of ESCO, it will be critical for teams to consider how to measure and manage such differences in skills composition. It also begs the question of what level of skill disaggregation is needed to be useful for different end users and objectives. For example, because of technical and computational challenges, the IDB's model was able to incorporate ESCO taxonomies' third level of skills disaggregation. While this level of disaggregation may be useful for population level analyses, it's unclear whether further disaggregation is needed (4th level) to be useful for identifying skills with the granularity needed for educational institutes to update curricula or for jobseekers to know which skills they need to acquire to transition to a new job. The final section of this note will discuss some of these considerations in more detail.

IV. Conclusions: Considerations for Adopting ESCO or other Skills Taxonomies

The previous section discussed different ways in which ESCO has been *adapted* and leveraged to help underpin skills and occupations classification exercises. The result—or expected result—being a set of models that can ingest unstructured data on jobs and skills, extract them, and classify them. A subsequent question is whether these models could help provide the basis for an skills-occupations taxonomy for an institution like the IDB, for individual countries to *adopt* on a national level, and/or perhaps even at a regional level the way ESCO functions for the European Union. And, if so, what steps and considerations would adoption of this adapted taxonomy at one or more of these levels entail? Here we highlight a set of **governance considerations** for the adoption of skills taxonomies for LAC, drawing on experience with adoption of machine learning models more generally, as well from ESCO's ongoing adoption across Europe.

Defining the Objectives and Users

Building a taxonomy, and furthermore and ontology, is a complex process and requires adequate institutions both politically and technically. Nevertheless, these classification systems are public goods that enable the development of information systems around them to help inform decision-

makers in education and labor markets. For LAC, for example, a multilingual regional ontology might allow for better labor market coordination in different countries by enabling a comparison of occupations and the skills needed to perform them.

However, before embarking on taxonomy adaptation/adoption initiatives—including those utilizing online job vacancies such as those discussed in this note—it's critical to define the objective of the information it will produce and/or the decisions it will help make—whether at the level of the individual or for public policy. Projects like those of utilizing online job vacancy data to understand skills and jobs are often marketed as a panacea project –promising to be able to provide information for a range of objectives and users. While this is possible, there are important practical decisions and tradeoffs to consider when developing the machine learning models that underpin these. One of these, for example, is what level of disaggregation of occupations and skills is needed for the intended end-use. The IDB project encountered this tradeoff as getting to more granular levels of skills classifications required additional computational power and verification hurdles. The team opted to remain at the third level of skills classification and second level of ISCO occupations for this first iteration of the model, in large part because of the computational challenges arising for reaching a greater level of granularity. These levels of ISCO occupations and ESCO skills provide sufficient information for decisions at the population and policy level. However, they may be less useful for curricula development of use by individual job seekers. Ramachandran et al (2020) present a helpful case study, on the retention in care of HIV patients, of the benefits of considering models that work at the individual, rather than population level and population subgroups, illustrating the tradeoffs in resource allocation cost effectiveness of an individual level focus.

Another consideration is that policies must help match individuals not only for jobs that fits their skillset, but with employment that also meets their personal goals. This cannot be answered by looking at vacancy data alone, and governments must first (or simultaneously) invest in analyzing résumés, as well as other sources of labor demand trends (Unamune, 2020 interview). Overall, its critical to ask what kinds of information are needed for decision making, and by whom, to ensure that models are providing the right kinds and levels of information.

Precision, accuracy, and ethics

A related point is the ethical considerations regarding the precision and accuracy of machine learning models that are the basis of these exercises. First, its critical to decide what level of accuracy is necessary for the decisions that will be made with the outputs of the model. Like the previous challenge mentioned about who the end users will be and what information they require, the accuracy of the model required similarity depends on its intended use and context. Decisions about accuracy and precision, thus, have ethical implications when humans will make decisions based on the information or predictions produced (See Ackermann et al (2018) for a detailed discussion).

There are also tradeoffs with greater levels of accuracy at the cost of feasibility. The Netflix Challenge illustrates this point: Netflix put out a challenge to improve the accuracy of its prediction algorithms. The most accurate model that won the prize, however, was highly costly and thus unfeasible to implement.

Finally, a model's accuracy may decrease over time, as underlying data shift. This means that the model needs to be consistently monitored, tested, and adjusted to ensure its usefulness over time—a capable team and governance model is needed for this maintenance and supervision to

be sustained over time. The example of using job vacancies is paramount, as changing labor markets will not be represented by a model based on past data alone.

In the case of skills and occupation, the accuracy of models intended to classify skills and occupations produced can have critical resource allocation implications at a policy of individual level—for example, where these models and the taxonomical structures they provide are used to make decisions about where to provide additional education or training resources for certain jobs of skills at the expense of others (Unamune, 2020 interview). Tradeoffs and decisions about accuracy are critical to consider during the development of machine learning models meant to identify and classify information on jobs and skills, but also in their adoption for different uses.

Implementation, Deployment and Maintenance

The governance structure and resource provisioning needed to sustain a machine learning project with public policy uses is complex. Without correct planning, projects of this nature tend to carry significant sink costs and low fruition or policy impact. They require very specialized technical knowledge and profit from inter-institutional collaboration. The next paragraphs detail some of these requirements and ideal features.

Today's reality is not only one of rapidly changing labor markets, but of constant paradigmatic changes in natural language processing and artificial intelligence technologies. The state-of-the-art in NLP and ML classification models vary from year to year. Therefore, a project of this sort must plan not only for the continual ingestion of new data, but also for the advent of new/better models, and the necessary finetuning of its test and production versions.

Assuring that models stay up to date with technological change is one component of several that prevent them turning obsolete. In the case of unsupervised machine learning models, teams need retrain data every six months or every year. Creating and maintaining data sharing agreements and/or collection algorithms can add bureaucracy and costs. As such, accounting future data availability is key, and plans for its compilation or acquisition should be part of the initial project's blueprints.

Regularly maintaining and updating the model also depends on having the team in place or outsourcing the model's management. In-house solutions have the advantage of tracking activities and advances more closely, but at higher costs than outsourcing. Some of the regular tasks include re-training the model with new data, finetuning the model's features when validation results become available, researching for new methodologies, creating test versions, creating and maintaining proper documentation and guides of use and applications. For example, according to consulted experts, unsupervised models used to generate labor market information using job vacancies data should be updated every 6 months or once every year. Following these updates, teams must conduct new validations to monitor the precision and general applicability of the model's predictions.

Governing Bodies and Agreements

Finally, machine learning is not just a model or a software, but the greater infrastructure and set of governance agreements and structures that ensure its usefulness and sustainability overtime. The same can be said of taxonomies in general. Although occupations and skills taxonomies were built before the advent of machine learning, there are many parallels with the governance structures required to develop and maintain these as useful tools in the public good. Moreover,

taxonomies like ESCO and O*NET are increasingly informed by machine learning models that make use of non-traditional data like that found in online job vacancies highlighted in the previous section of this note. Both required resources, a governing body capable of making high level decisions and allocation of resources, as well as the human capital with the sufficient technical skills and know-how to build, maintain and correct these systems over time. The key message is that building a taxonomy is a complex process and requires adequate institutions both politically and technically.

V. References

Altamirano, A., Azuara, O., González, S., Ospino, C., Sánchez, D., & Torres, J. (2019). Clasificación de ocupaciones en América Latina y el Caribe. Nota técnica. Banco Inter-Americano de Desarrollo.

Altamirano, A., Azuara, O., González, S., Ospino, C., Sánchez, D., & Torres, J. (2019). Tendencias de las Ocupaciones en América Latina y el Caribe 2000-2015. Nota técnica. Banco Inter-Americano de Desarrollo.

Amaral, N., Eng, N., Ospino, C., Pagés, C., Rucci, G., & Williams, N. (2018). Clasificación de ocupaciones en América Latina y el Caribe. Nota técnica. Banco Inter-Americano de Desarrollo.

Benítez, D., Lucero, S., Pazmiño, A. 2018. Elaboración de estadísticas de vacantes publicadas en internet. Una experiencia en Ecuador. Revista de Estadísticas y Metodologías. Número 4, abril 2018. Instituto Nacional de Estadística y Censos.

CEDEFOP. (2018). Mapping the landscape of online job vacancies. Background country report: Spain, http://www.cedefop.europa.eu/en/events-and-projects/projects/big-data-analysis-onlinevacancies/ publications.

CAF-DOCUMENTO DE TRABAJO # 2020/04 Julio 21, 2020. Brecha de habilidades de los jóvenes en el mercado laboral colombiano. Ana María Díaz1 | Luz Magdalena Salas.

Diaz, Ana Maria. Salas, Luz. (2020). Brechas de habilidades de los jóvenes en el mercado laboral colombiano.

ILO. (2012). International Standard Classification of Ocupations: ISCO-08. International Labour Office: Geneva.

IADB & Quantil. (2020). Technical support to develop and deploy machine learning algorithms to extract skills from online job vacancy texts. Mimeo.

Ospino, C. (2018). Ocupaciones laborales: Clasificaciones, taxonomías y ontologías para los mercados laborales del siglo XXI. Nota técnica, IDB-TN-1513. Banco Inter-Americano de Desarrollo.

Sierra, S., Bel-Enguix, G., Gómez-Adorno, H., Moreno, J. M. T., Hernández-García, T., Guadarrama-Olvera, J. V., ... & Martínez, S. A. (2020, May). Enhancing Job Searches in Mexico City with Language Technologies. In Proceedings of the 1st Workshop on Language Technologies for Government and Public Administration (LT4Gov) (pp. 15-21).