



Re-search Alps

Research laboratories in the Alpine Area

INEA CEF-TELECOM Project

AGREEMENT No INEA/CEF/ICT/A2016/1296967

D 1.1 – Modeling and estimating the economic and social impact of the results of the project

Code	D 1.1
Date	31/10/2017
Type	Deliverable
Responsible	Margherita Russo (UNIMORE)
Participants	UNIMORE (leader) Data Publica, UMIL, MESRI, MIUR, PAs
Authors	Margherita Russo, Francesco Guerra, Francesco Pagliacci, Matteo Paganelli, Loïc Petit, Frederic Olland, Emmanuel Weisenburger, Emanuela Zilio
Corresponding Authors	Margherita Russo



Modeling and estimating the economic and social impact of the results of the project.

1 Introduction

The idea behind the Re-search Alps project has been conceived inside within the EUSALP Action Group 1 - “to develop an effective research and innovation ecosystem” (AG1). EUSALP is the EU-Strategy for the Alpine Region, which is composed of seven countries: Austria, France, Germany, Italy Liechtenstein, Slovenia and Switzerland. The strategy aims at ensuring mutually beneficial interactions between the mountain regions at its core and the surrounding lowlands and urban areas.

The goal of the Re-search Alps project is the publication on the web of an open dataset describing the private and public laboratories, research and innovation centers (hereinafter, referred as “labs”, in short) existing in the seven aforementioned countries, with particular reference to the 48 Regions constituting the Alpine Area.

The main goal of Task 1 of this project is retrieving and analysing available lists of data about those labs. Data have to refer to the countries involved in the EUSALP area but available lists of data could exist also at sub-national level. In some other cases, relevant information could be provided at international level as well. All these pieces of retrieved information will eventually feed the web-dataset that represents the main input for the outcome of this project. The web-dataset will reach at least 3-stars on Tim Berners-Lee's 5 star schema (making it available on the web under an open licence in a non-proprietary, structured format) and it will support a number of services on it for different types of users: for instance, enterprises; researchers; policymakers and citizens. To reach this goal we have to describe the information currently available on labs for each country of the EUSALP area, and their conditions of usage.

In our analysis we have to consider that the seven countries considered differ under several socio-economic dimensions (such as size of population, number of employees, industrial specialisation); EU membership (involved countries are both EU and non-EU members, as in the case of Switzerland and Liechtenstein); involvement in the European Strategy for the Alpine area as the entire country (as in the case of Austria, Lichtenstein, Slovenia and Switzerland) or only with some regions (as in the case of France, Germany and Italy) (Figure 1).

This area is composed of territories with different demographic, social and economic size and trends. The Alpine area also shows a great cultural and linguistic diversity, as four languages are actually used in the EUSALP area (French, German, Italian and Slovenian). This diversity goes along with a great variety of governance systems and traditions, which comes from specific and diverse historical national experiences. According to this specific set of features, the Alpine region can be considered as a unique territory, with important potentials for dynamism, but facing major challenges:

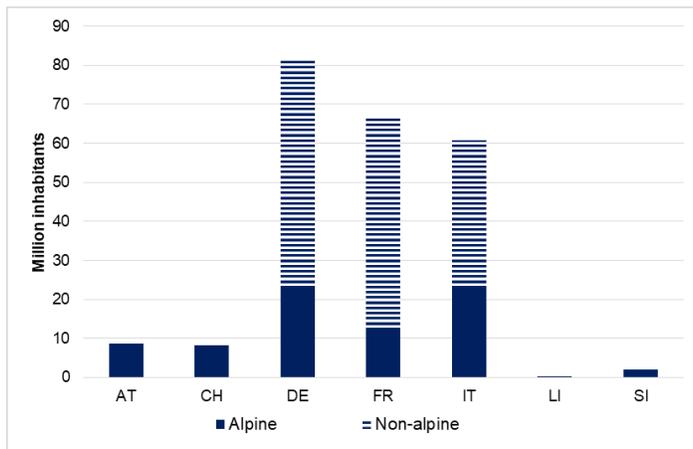
- economic globalisation that requires the territory to distinguish itself as competitive and innovative by developing the knowledge and information society;
- demographic trends, combining ageing and new migration models;
- climate change and its effects on the environment, biodiversity and on the living conditions of its inhabitants;
- the energy challenge at the EU and worldwide scales;



- its specific geographical position in Europe, as a transit region but also as an area with unique geographical and natural features with set the frame for all future developments.

Given the existence of all these challenges, both the common specificities of the Alpine area and its variety and diversity call for cooperation, and in particular for cross-border cooperation as well as for identifying common goals and implementing them more effectively. Better cooperation between the regions and States is needed to tackle those challenges and to fuel specific programmes of local and regional development.

Figure 1 – Population by country and EUSALP area (2015)



Source: Eurostat (2017): http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_r_d2jan&lang=en

Sources that we have explored are of many different types. Some international items describing publications and research institutions (e.g., ORCID, Google Scholar...) have to be evaluated as well, as a possible source of details about EUSALP countries' labs activities. Other sources, providing data about companies and their research activities, will be investigated (e.g. INPI, namely the French *Institut national de la propriété industrielle*; the European Patent Office, Bureau van Dijk; Orb Intelligence, i.e. a US data provider).

Deliverable D.1 aims to return a broad state-of-the-art analysis about information already existing on labs in the seven countries, its condition of use for different types of users, conditions of the data update, coverage (in terms of number of instances). This is preliminary condition to feed a model to assess the economic and social impact of the results of the project (metadata, dataset and software application).

The strategy adopted to collect and analyse information is as follows.

Firstly, we move from the number of firms that sell research and technology services to other companies. This is the case of the professional, scientific and technical activities, as included into the NACE rev. 2 classification under the section M. In order to retrieve data on them that are useful for this project, business registers have to be considered. Across EUSALP countries, national heterogeneity is a relevant issue: business registers are instituted at national level but with different rules of access across countries.

Secondly, even if accessible, the business registers suffer from a limitation for our analysis. They do not cover public research institutions (such as universities and public research centres), with the only



exception of France (where SIRENE covers all legal entities, including public ones, but not research labs as they are not legal entities). Thus, other sources of information have been explored such as: national and regional archives from the ministries; private companies' archives (e.g. Kompass); and international organisations' archives (e.g. ORCID).

Thirdly, an additional limitation, related to business registers, has to be overcome. As a matter of fact, the set of companies under the section M cannot be actually considered as fully exhaustive, for the purposes of our project. Indeed, other companies or institutions, even though their main aim is not selling research and technological services, can be relevant as well (such as Renault or FCA). To identify them, we have to consider also those firms that have filed a patent, the ones that have participated in research consortia of different types or whose employers have published scientific contributions. Information on these activities can be retrieved by exploring several kinds of source, both at international and national level.

According to this strategy, this deliverable is structured as follows. Section 2 describes the geographical area under study in the project in terms of active enterprises classified under section M of the NACE Rev. 2 Classification. Section 3 focuses on the role and characteristics of business registers across EUSALP countries. Section 4 describes the main characteristics of items (namely, datasets and websites) that we have collected in order to detect information on labs across the EUSALP countries. In particular, a group of 157 items have been considered as potential feeders of the Re-search Alps web-dataset.

Grounded on this information, we can feed in a model structured on three aspects: two of them refer on pre-post comparison; the third one on changes within the institutions that support innovation policies to enhance territorial eco-systems. The first pre-post comparison refers to the information on the enterprises belonging to NACE Rev. 2 Section M and to public and public-private labs (i.e. the target information of our project), that are available now and that will become accessible in our web-dataset. The second pre-post comparison refers to the more general availability of data on labs, and to the conditions of their usage, across the seven EUSALP countries. The third aspect of our model aimed at assessing the economic and social impact of the results of the project concerns the process of change within institutional organisations that make accessible information (e.g. national and regional ministries). Section 5 returns a discussion of these three aspects.

2 Professional, scientific and technical activities: an overview from NACE rev.2

2.1 Statistical classification of economic activities

The general classification of economic activities, which applies to business and other private companies, is useful in order to disentangle those businesses that sell research services to other corporations in the whole set of corporations. They are supposed to be private research organisations and private laboratories that are directly involved in supporting scientific and technical activities.

Moving from this idea, we adopt the statistical classification of economic activities that is currently adopted at EU level, i.e. the NACE Rev. 2¹. In particular, such a classification can be adopted in order

¹ *Nomenclature statistique des activités économiques dans la Communauté européenne.*



to single out the total number of enterprises that perform professional and scientific activities: in principle, these organisations should include possible private research organisations and laboratories.

To this respect, Section M of the NACE Rev. 2 identifies "Professional, scientific and technical activities". This section includes activities that tend to require a high degree of training, and make specialised knowledge and skills available to users (Eurostat RAMON, 2017).

Nevertheless, this section is not particularly homogeneous. Indeed, it includes seven different divisions (i.e. 2-digit classification of the economic activity), which can be defined as follows (Eurostat RAMON, 2017):

- **69 - Legal and accounting activities:** legal representation of one party's interest against another party (e.g. advice and representation in civil cases, advice and representation in criminal actions, advice and representation in connection with labour disputes); preparation of legal documents; other activities of notaries, accounting, bookkeeping services;
- **70 - Activities of head offices; management consultancy activities:** provision of advice and assistance to businesses and other organisations on management issues (strategic and organisational planning; financial planning and budgeting; marketing objectives and policies; human resource policies, practices, and planning; production scheduling; and control planning); activities of head offices;
- **71 - Architectural and engineering activities; technical testing and analysis:** provision of architectural services, engineering services, drafting services, building inspection services and surveying and mapping services, as well as performance of physical, chemical, and other analytical testing services;
- **72 - Scientific research and development:** including three types of research and development: 1) basic research; 2) applied research; 3) experimental development. Research and experimental development activities in this division are subdivided into two categories: natural sciences and engineering; social sciences and the humanities;
- **73 - Advertising and market research:** creation of advertising campaigns and placement of such advertising in periodicals, newspapers, television...;
- **74 - Other professional, scientific and technical activities:** provision of professional scientific and technical services (except legal and accounting activities; architecture and engineering activities; technical testing and analysis; management and management consultancy activities; research and development and advertising activities);
- **75 - Veterinary activities:** provision of animal health care and control activities for farm/pet animals.

According to the aforementioned definitions, Section M clearly includes a large number of pretty heterogeneous activities. In particular, both "legal and accounting activities" (division 69) and "other professional, scientific and technical activities" (division 74) include, by definition, a variety of services. They all differ from veterinary activities, as included into division 75. Nevertheless, all these typologies of enterprises could represent the ultimate target of this project, concerning business organisations, as they could represent an important part of the labs in the area.

2.2 Number of enterprises, at EU level and in the EUSALP countries

As shown in Table 1, according to Eurostat data, in 2015 there were about 4.3 million enterprises classified under Section M in the 28 Member States of the European Union. As far as the EUSALP



countries are considered², enterprises classified under Section M are about 1.8 million, representing about 40% of the total number of enterprises in the EU-28.

The biggest division of the section M in terms of enterprises is division 69 (519 thousands across the EUSALP countries, equal to 29% of the total), followed by divisions 71, 70 and 74 (respectively accounting for 26%, 20% and 17% of the total number of Section M enterprises in the EUSALP countries) (Figure 2).

Table 1 – NACE Rev.2 section M and its divisions: number of active enterprises by EUSALP country* and total EU-28, year 2015

	M	69	70	71	72	73	74	75
	Professional, scientific and technical activities	Legal and accounting activities	Activities of head offices; management consultancy activities	Architectural and engineering activities; technical testing and analysis	Scientific research and development	Advertising and market research	Other professional, scientific and technical activities	Veterinary activities
Austria	65,393	13,383	15,593	16,851	1,037	9,679	6,966	1,884
France	463,476	89,303	159,919	91,543	5,769	26,205	82,284	8,453
Germany	478,857	118,979	107,932	128,545	6,823	33,498	72,935	10,146
Italy	714,934	283,827	56,686	202,122	8,796	19,879	130,870	12,754
Lichtenstein	n.a.	n.a	n.a	n.a	n.a	n.a	n.a	n.a
Slovenia	30,757	6,776	9,947	6,771	1,221	1,940	3,943	159
Switzerland	23,822	6,772	3,967	8,783	558	1,290	1,815	637
Total EUSALP countries	1,777,239	519,040	354,044	454,615	24,204	92,491	298,813	34,033
Total EU-28	4,383,000	1,218,282	984,487	1,028,946	62,759	306,219	704,667	77,640

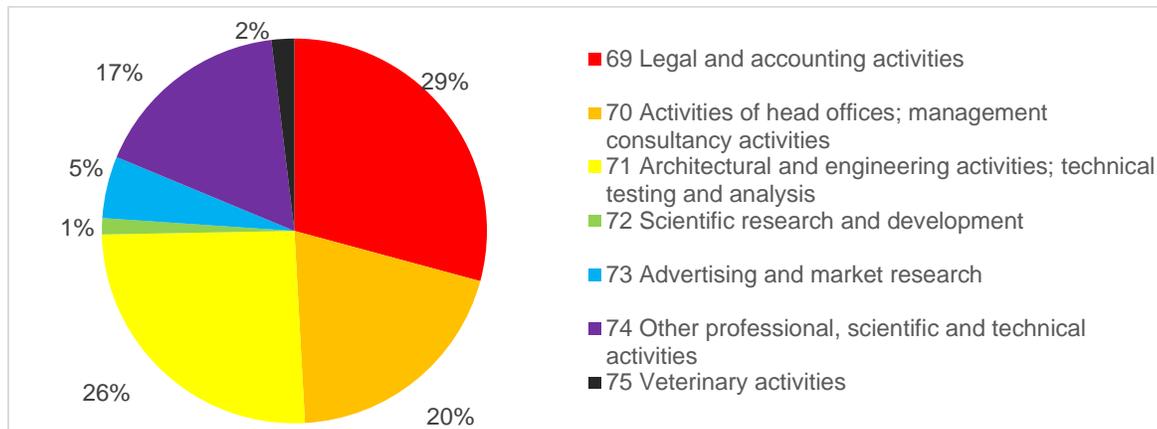
* Data about Lichtenstein are not available.

Source: Eurostat (2017): http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=sbs_r_nuts06_r2&lang=en

² Data about Lichtenstein are not included here, as they are not available in the Eurostat (2017) source adopted in this analysis.



Figure 2 - NACE Rev.2 section M divisions: percentage share of active enterprises in the EUSALP countries*, year 2015



* Data about Lichtenstein are not available.

Source: Eurostat (2017): http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=sbs_r_nuts06_r2&lang=en

2.3 Number of enterprises at regional level in the EUSALP countries

When we consider EUSALP countries, unless Lichtenstein and Slovenia (which participate to the strategy as national governments), regional data should be considered by referring to the NUTS (Nomenclature of territorial units for statistics) classification adopted by the EU. As a matter of fact, we observe that different NUTS level administrations³ are involved in the EUSALP. In particular, NUTS 1, NUTS 2 and NUTS3 levels (which are defined according to the size of the population) appear to be involved in the seven countries, as follows:

- NUTS 1
 - Germany (Länder: Baden-Württemberg and Bavaria);
- NUTS 2:
 - France (Régions: Franche-Comté; Rhône-Alpes; Provence-Alpes-Côte d'Azur);
 - Italy (Regioni and Province Autonome: Piemonte; Valle d'Aosta/Vallée d'Aoste; Liguria; Lombardia; Provincia Autonoma di Bolzano/Bozen; Provincia Autonoma di Trento; Veneto; Friuli-Venezia Giulia),
 - Austria (all Länder);
- NUTS 3:
 - Switzerland (all Cantons).

As far as the sub-national levels are considered, we observe that data about single divisions on local units⁴ follow similar patterns. In each of the aforementioned territorial areas (i.e. EU-27, countries

³ The NUTS classification (Nomenclature of territorial units for statistics) is a hierarchical system for dividing up the economic territory of the EU for the collection, development and harmonisation of European regional statistic. In particular, the NUTS regulation defines minimum and maximum population thresholds for NUTS regions' size. In particular, it is possible to define (<http://ec.europa.eu/eurostat/web/nuts>): NUTS 1: major socio-economic regions (approximately 3 to 7 million inhabitants); NUTS 2: basic regions for the application of regional policies (approximately 0.8 to 3 million inhabitants); NUTS 3: small regions for specific diagnoses (0.15 to 0.8 million inhabitants). However, for practical reasons the NUTS classification mirrors the territorial administrative division of each Member States.

⁴ Please note that Eurostat regional data are available only for local units and not for enterprises. Moreover, no data about Switzerland and Lichtenstein are available. Data about local units might not necessarily coincide with data on active enterprises.



affected by the EUSALP Strategy, regions directly covered by the Strategy), the relevance of each division is fairly similar: both local units classified as division 69 and local units classified as division 71 represent about 29% of Section M local units, each. On the opposite, "scientific research and development activities" (division 72) represent less than 1.5% of total local units in Section M (Table 2).

Table 2 – Section M and its divisions: number of local units within the EUSALP area, year 2011

	M Professional, scientific and technical activities	69 Legal and accounting activities	70 Activities of head of offices; management consultancy activities	71 Architect ural and engineering activities; technical testing and analysis	72 Scientific research and developm ent	73 Advertisin g and market research	74 Other profession al, scientific and technical activities	75 Veterina ry activitie s
Germany	393,739	109,350	75,040	114,340	5,668	29,120	51,016	9,206
Baden- Württemberg	55,401	13,303	13,673	17,030	1,043	3,584	5,930	839
Bayern	70,881	19,192	13,234	20,587	969	5,279	9,896	1,723
France	510,488	103,534	172,643	111,430	5,352	34,178	72,521	10,830
Franche-Comté	4,721	932	1,154	1,422	52	230	722	209
Rhône-Alpes	47,554	8,336	15,394	12,112	579	2,882	7,269	982
Provence-Alpes- Côte d'Azur	47,783	10,041	14,706	11,018	531	3,220	7,356	911
Italy	732,414	281,236	55,577	219,118	10,024	22,748	132,357	11,354
Piemonte	54,981	15,216	7,061	18,321	619	1,602	10,950	1,212
Valle d'Aosta	2,108	561	97	975	45	33	358	39
Liguria	20,633	7,884	1,179	6,806	292	483	3,653	336
Lombardia	146,769	47,429	15,937	38,981	1,785	6,452	33,801	2,384
Provincia Autonoma di Bolzano	5,880	1,690	577	1,869	62	201	1,376	105
Provincia Autonoma di Trento	6,609	1,616	451	2,727	110	186	1,418	101
Veneto	61,099	20,424	4,897	19,762	603	2,229	12,252	932
Friuli-Venezia Giulia	14,150	4,343	1,030	4,355	262	424	3,501	235
Austria	63,847	12,521	15,180	17,168	1,076	9,614	6,386	1,902
Slovenia	23,922	6,141	6,621	5,568	1,013	1,690	2,765	124
Switzerland	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Lichtenstein	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Total Eusalp Countries	1,724,410	512,782	325,061	467,624	23,133	97,350	265,045	33,416
Total Eusalp Regions	626,338	169,629	111,191	178,701	9,041	38,109	107,633	12,034

Source: Eurostat data (2017)

Indeed, a single enterprise could include more than one local units not necessarily in the same region. This is why local units must be considered when analysing employment data as sub-national level.



2.4 Comparing Nace and Isic

In order to identify those businesses that sell research services to other corporations, beyond the EU NACE classification other sources could refer to the International Standard Industrial Classification of all economic activities (ISIC), adopted by the United Nations. NACE and ISIC classifications are linked by structure, so that ISIC codes can be easily converted into NACE ones, that are the ones that we use in this project (Eurostat, 2008).

3 The characteristics of business registers in the EUSALP countries.

At national level, as far as private business companies are concerned, it is possible to retrieve specific information on their activities, by querying the business registers. In the seven EUSALP countries, business registers are in charge by law of registering and maintaining detailed information on firms. Despite this common goal, in Europe, the national business registers offer a range of services, which actually vary from one Member State to another. With regard to the goals of the Research-Alps project, the access to information is quite different and overall particularly poor.

The following analysis has been carried out in order to highlight pieces of information provided by these registers that could be usefully adopted according to the Re-search Alps goal of the creation of a unified dataset of the labs. Here, a country-by-country analysis is shortly returned, while Table 3 summarizes all these characteristics.

Information on national business registers are available on European e-justice portal (2016), for all EUSALP countries but Switzerland.

3.1 Austria⁵

The Austrian commercial register (*Firmenbuch*, also known as the ‘principal register’ or *Hauptbuch*) contains information on all registered Austrian companies. In this case, the access to the Austrian commercial register is chargeable. In other words, every piece of information collected by the Austrian commercial register is accessible only for a fee.

It is observed that the Austrian commercial register is not the only repository of institutional data about Austrian firms and business companies. Also other national institutions provide similar information. However, in all the aforementioned cases, information on business corporations is available only for a fee.

3.2 France⁶

Insee (the French *Institut national de la statistique et des études économiques*) has set up the distribution of the database Sirene®, by enriching the directory of companies and establishments, improving its quality and offering support to users. The website presents the data from the Sirene® database and the consultation and list building tools. In particular, it represents a unique and

⁵ “European E-Justice Portal - Business Registers - Austria.” 2016. https://e-justice.europa.eu/content_business_registers_in_member_states-106-at-en.do?member=1. Last accessed on: October, 17, 2017.

⁶ European E-Justice Portal - Business Registers- France.” 2016. https://e-justice.europa.eu/content_business_registers_in_member_states-106-fr-en.do?member=1. Last accessed on: October, 17, 2017.



comprehensive database of companies and their establishments, which is characterised by more than 10,000 daily updates and covering all the sectors of economic activity. In terms of coverage, it brings together economic and legal information about 10 million French establishments. The search for a can be done both by name of the company and by the SIREN code (*Système d'identification du répertoire des entreprises*) and/or the SIRET code of the company (given to their single establishments and facilities). Additional search can be by referring to the company's address and by municipality name or postal code.

Among the main economic and legal pieces of information, the following are returned as open data:

- Company address;
- Company's name;
- Company's address;
- SIREN/SIRET code;
- Legal organization form of the company;
- Business activities classified by NAF code;
- Total number of employees.

3.3 Germany⁷

The portal of the German Commercial Register (*Handelsregister*) is run by the Ministry of Justice of the federal state of North Rhine-Westphalia on behalf of other German *Länder*. This register provides centralised access to all federal state registers of companies, cooperatives and partnerships and to announcements for the register. The register can be consulted through two alternative ways of search. The normal search allows users to find out German companies by their official name; and advanced search also allows users to find out companies by the name of the federal state in which they operate as well as by the company's legal status. However, the only freely accessible piece of information is the address of the company (through the UT – entity link). Further details about registered companies are available only for a fee. This is the case of the AD – Current hard copy; CD – Chronological hard copy; HD – Historic hard copy; DK – Document view and SI – Structured register content.

Another source of data that can be partially used to search for institutional (i.e. official) and up-to-date structural and economic details about German companies is represented by the portal of the German Companies Register (available at the following link: <https://www.unternehmensregister.de/ureg/>). Through this web-portal, it is possible to have direct access to some annual financial statements, made available by German companies. However, the consultation of these documents cannot be made automatic: actually, captcha tests are used. In addition to this critical issue, it can be observed that these documents show a large variety when dealing with the completeness of the contents that are provided to the users. For instance, as far as the smallest companies are concerned, the financial statements only return some basic budget data. Conversely, the largest (most important) companies also return some further structural details, such as:

- a wider description of the company;
- a list of affiliated companies, if any;

⁷ "European E-Justice Portal - Business Registers - Germany." 2016. https://e-justice.europa.eu/content_business_registers_in_member_states-106-de-en.do?member=1. Last accessed on: October, 17, 2017.



- the business model characterising them;
- products and services sold;
- the industrial sectors which are covered by the business.

3.4 Italy⁸

In Italy, *Registro delle Imprese* is run by the Chambers of Commerce, with a support from Unioncamere, under the supervision of the Ministry of Economic Development. The ICT infrastructure is run and maintained by Infocamere, a consortium of the Chambers of Commerce in a public limited company form. The register provides detailed pieces of information about active businesses, such as:

- complete name of companies;
- VAT number;
- legal form of the company;
- governing bodies;
- capital;
- legal representatives;

It also provides access to public documents concerning companies, including full financial statements; and the list of the shareholders.

The access to the business register and to a limited amount of details (such as name and address of company) is free of charge. On-line access to the complete set of information is available only upon request and only for a fee.

In particular, companies search can be performed by referring to either their name or their main activity, which have to be selected from a closed list of possible alternatives. Although these alternatives do not directly refer to the NACE classification, they actually refer to the same classification⁹.

As far as the freely available economic details are concerned, they just include:

- the address of the company;
- its legal organization form;
- the business activities that are carried out by each of them.

It should be noticed that since April 2016, the web portal Atoka provides access to information as a search engine for Italian companies (<https://atoka.io/>). The service is available for a fee.

3.5 Liechtenstein¹⁰

A business register (*Handelsregister*) is held for the entire country of Liechtenstein. It is a public register, whose main aim is to ensure the legal certainty of commerce by disclosing legal relationships governed by private law, in particular the situations of natural and legal persons engaging in business when it

⁸ “European E-Justice Portal - Business Registers - Italy.” 2017. https://e-justice.europa.eu/content_business_registers_in_member_states-106-it-en.do?member=1. Last accessed on: October, 17, 2017.

⁹ As an example, all the activities identified by codes 69 to 75 in the NACE Rev. 2 classification, as discussed in the second section are actually included into the list of alternatives that are suggested by the register for the search.

¹⁰ https://e-justice.europa.eu/content_business_registers_in_member_states-106-ii-en.do?member=1



comes to liability and authority to act. The business register for Liechtenstein is currently available at the following link: http://www.oera.li/hrweb/ger/firmensuche_afj.htm.

When considering the search option, companies can be found by selecting the field 'name' and the field 'legal organization form'. Nevertheless, the only significant piece of information that is provided by this register is company's address.

3.6 Slovenia¹¹

The Slovenian Business Register (PRS) is managed by the Agency of the Republic of Slovenia for Public Legal Records and Related Services (*Agencija Republike Slovenije za javnopravne evidence in storitve* – AJPES). It is accessible only by the ePRS application at the following link (<https://www.ajpes.si/prs/Default.asp?>). The Slovenian Business Register is the central public database providing details about all businesses and private enterprises that have a registered office in Slovenia and engaged in gainful or not-for-profit activities. Moreover, additional details on subsidiaries and other divisions of those businesses are provided as well.

The registration to this application as well as the access to companies' profile is completely free. Moreover, this application also provides, with respect to other registers, useful search filters and a wide variety of economic details. For example, it allows users to search companies directly for the NACE code of the business.

Although the completeness of economic details among different companies cannot be considered as fully comparable, some useful pieces of information can be extracted from this register, such as:

- company address;
- VAT number;
- contacts (sometimes also web site);
- business activities by NACE classification;
- legal organization form.

3.7 Switzerland

As far as the online Swiss business register (*UID-Register@BSF*) is concerned, the companies search requires to specify the company name or the company IDI (namely, a unique company identification number). Once this parameter is specified, it is possible to add more selective filters such as: the geographical location of the business (e.g., its canton or its municipality); the legal organization form of the business.

Available details about private companies that can be retrieved actually include:

- company address;
- legal organization form;
- VAT number.

¹¹ "European E-Justice Portal - Business Registers - Slovenia." 2017. https://e-justice.europa.eu/content_business_registers_in_member_states-106-si-en.do?member=1. Last accessed on: October, 17, 2017.



3.8 *Summing up*

The short analysis performed so far clearly suggests that business registers cannot be insightfully adopted in order to retrieve (and re-use) detailed information about EU businesses and private enterprises, unless for France. In most cases, these details are available only in principle and can be retrieved by single users for their private purposes. As a matter of fact, in most cases, these details cannot be re-used or published on a brand new online dataset.

Moreover, business registers only cover the set of private research organisation, while no information is returned for public research centres or public laboratories.

Given these open issues, a different methodology has been followed in order to retrieve details that are useful for the successful outcome of this project.



Table 3 – Company information available in business registers, by EUSALP country

Country	URL	Access	Search	Available data	Terms of use for the Re-search project's purposes
Austria	http://www.justiz.gv.at/firmenbuch	Restricted	-	-	It is possible to submit an application for re-using data. A reply is due in 4 weeks.
France	https://www.sirene.fr/sirene/public/accueil	Free	<i>normal search:</i> <ul style="list-style-type: none"> • by name; • by SIREN - SIRET code. Additional search by address / municipality name or postal code	<ul style="list-style-type: none"> • Company address; • Company's name; • Company's address; • SIREN/SIRET code; • Legal organization form of the company; • Business activities classified by NAF code; • Total number of employees 	Open data
Germany	http://www.handelsregister.de/	Free	<i>normal search:</i> <ul style="list-style-type: none"> • by name; • by legal form. <i>advanced search:</i> <ul style="list-style-type: none"> • by federal state. 	<ul style="list-style-type: none"> • company address; • financial statements. 	A further publication or further uses of the available data is not allowed.
Italy	http://www.registroimpresa.it/	Free	<ul style="list-style-type: none"> • by name; • by NACE code. 	<ul style="list-style-type: none"> • company address; • legal form; • business activities. 	It is forbidden to carry out data distribution and/or sale activities, and to attempt to access data in some ways other than the ones allowed. In particular, it is forbidden to extract data automatically and massively in order to speed up activities or create self-contained databases.
Liechtenstein	http://www.oera.li/hrweb/ger/firmensuche_afj.htm	Free	<ul style="list-style-type: none"> • by name; • by legal form 	<ul style="list-style-type: none"> • company address. 	Terms of use are not openly stated in the website.
Slovenia	via the ePRS application (https://www.ajpes.si/prs/Default.asp?)	Free	<i>normal search:</i> by name <i>advanced search:</i> <ul style="list-style-type: none"> • by legal form; • NACE code; • Location. 	<ul style="list-style-type: none"> • company address; • contacts; • VAT number 	The use of significant parts (both qualitative and quantitative) of the databases (e.g. the reproduction of the structure of the databases, of some parts of its structure, the creation of private and public databases, etc.) is not allowed.
Switzerland	https://www.uid.admin.ch/Search.aspx	Free	<i>normal search:</i> <ul style="list-style-type: none"> • by name; • by IDI. <i>advanced search:</i> <ul style="list-style-type: none"> • by canton, municipality, place; • by legal form (normal search parameters are mandatory). 	<ul style="list-style-type: none"> • company address; • VAT number. 	No clear information have been retrieved on data use so far.

Source: authors' elaboration



4 A comprehensive dataset to describe research organisations and laboratories

4.1 From significant rough information to a comprehensive dataset: the adopted strategy

In order to overcome the lack of available data provided by business registers in Europe, but also in order to overcome the fact that the set of companies under the section M of the NACE Rev. 2 cannot be considered as fully exhaustive, the following empirical strategy has been adopted for the purposes of this project.

Firstly, a form has been created to be submitted online for being filled by the regional and national representatives of the EUSALP AG1. The form contained a set of information to describe datasets existing at national or regional level, as described in the left column of the Table 4.

Given the short time span over the holiday season (deadline was early September 2017), we were able to collect only a few replies. Thus, the team from the University of Milan started an exploration of the available data sources, across the EUSALP area. This exploration produced, by mid-September, a list of 157 items, including datasets, relevant websites, PDF files. As shown in the right column of Table 4, the database schema provided by UMIL added to the previous one some fields, namely "Body/Institution" and "Link".

Table 4 – Database schema in the online form and in the dataset elaborated by UMIL

Database schema of the online form (with no replies)	UMIL database schema
	Body/institution
Title of the suggested dataset	Title of the suggested dataset
Type	
Short description of dataset: content and regions involved	Short description of dataset: content and regions involved
Reference Country	Reference Country
Reference person for the dataset (First name, Last name, email)	Reference person for the dataset (First name, Last name, email)
Dataset update frequency/last update, if known	Dataset update frequency/last update, if known
Comments / Suggestions	Comments / Suggestions
Person filling the form (First name, Last name, email)	
Email address	
	Link

Source: authors' elaboration

Provided that the final list contained the link of each of the 157 items, we were able to retrieve detailed information that are essential to the successful outcome of the project. All the details gathered together from the single items have been reorganized within a comprehensive framework, described in the following sub-sections.

Sub-section 4.1 focuses on the database schema, describing its main characteristics. Then, following sub-sections focus on each of the most important parts of this schema: source of the item (sub-section 4.3), data included and policies for their use (sub-section 4.4), and characteristics of the instances occurring in each item (sub-section 4.5).

4.2 The database schema to build a comprehensive dataset of items

The 157 items identified refer to a broad and heterogeneous set of information. In order to properly describe the main characteristics of each of the original items, we have created a comprehensive



dataset. This task makes possible a proper description of the most significant pieces of information that are currently available and that could usefully describe the state-of-the-art about the knowledge on labs in the EUSALP area. Such an exploration is essential to single out what sources can be used for the goals of the Re-search Alps project, and at which conditions they can be used.

Our comprehensive dataset has been filled according to the database schema provided in Annex 1 (Table A.1) that describes the details of its structure. In Table 5, after a short description of the item, information is organised in three macro-categories: source of the item, data included and characteristics of the instances. For each of the fields in our comprehensive dataset, the number of available, missing and non-relevant information is returned. We can observe that the fields describing the sources are the ones showing the largest number of available information across the 157 items. On the opposite, critical aspects refer to the presence of missing values with regard to ID codes and updates of information (Table 5).

Table 5 – The database schema (source, data and instances) and their availability

Macro category	Name of the field	Available values	Missing values	Non relevant
Source	Short description of the item	157	0	0
	Source name	157	0	0
	Source location	157	0	0
	Title of the item	157	0	0
	E-format of the item	157	0	0
	Type of organization publishing the item	157	0	0
	Reference: URL	157	0	0
	Reference: person name	93	64	0
	Reference: e-mail contact	91	66	0
Data	Policies for the use for the project	157	0	0
	Policies for the use - details	144	13	
	Data retrieval	124	0	33
	Language	157	0	0
	Territorial coverage	157	0	0
	Geographic information of instances	157	0	0
	ID codes of instances	71	86	0
	Update: Frequency	52	105	
	Update: Date of the last update	32	125	
	Update: additions and changes	25	132	
	Used in scanR	157	0	0
	Time series	157	0	0
Instances	Instances by type	153		4
	Instances: number	128	25	4
	Instances type 'multiple': comments	36	0	121

Source: authors' elaboration; details in Annex 1.



4.3 Sources

The sources of the 157 items about the labs in the EUSALP area are characterised by means of eight different fields: name, location, title, e-format, type of organization publishing the item, references. In this sub-section, we discuss the information regarding to:

- **e-format of the item** (namely datasets, PDF files, websites, websites with an underlying dataset): this information is essential to populate our web-dataset¹²;
- **source location** (the country and/or the region of the source publishing the item), to highlight the relative importance of the retrieved items published across the EUSALP area;
- **type of organisation** that publishes the item (namely, business organisations; public institutions; not-for-profit organisations and projects): this helps in summarising the kind of contacts that Re-search Alps has to undertake for having access to data.

The analysis of the sources by **e-format** and country represents an important feature to single out the territorial heterogeneity that affects the amount and type of available information on labs in the EUSALP area.

As far as e-format of the item is concerned, according to Table 6, items are in most cases websites (133 items out of 157); conversely datasets are only 20 (plus one additional website that is linked to an underlying dataset). Lastly, three objects just return a PDF file with information on labs.

Moving from e-format heterogeneity, in Table 6 we observe that also territorial heterogeneity occurs. Most of information on **source location** is gathered on a national basis, but different countries of the EUSALP area are characterised by a different number of items, as highlighted in Figure 3¹³. According to it, Austria is the country with the largest number of items to detect information on labs (37): in this case, only one dataset occurs among them. Similar patterns are observed in Germany (25 items), Switzerland (22 items), Italy (21 items), Slovenia (16 items) and Lichtenstein (6 items). In each of these countries, items mostly comprise websites. As far as French items are concerned, five out of 14 are datasets, accessible to users.

In addition to these 141 items, 16 of them are defined as "Worldwide". Such a definition refers to those items that are provided by sources not located in the EUSALP area but that cover even labs located within its boundaries. A possible example of this type of sources is represented by the dataset provided by ORCID, namely an international organisation publishing information on researchers and their affiliations, hence a source to detect labs, even in the EUSALP area¹⁴.

¹² To this respect, dealing with a PDF file has very different implications in comparison with dealing with a website that can be crawled but even in comparison with a structured dataset.

¹³ Compared to Table 6, Figure 3 ignores the fact that some sources are identified on a regional basis. This occurs both in Austria (where 9 items are specifically produced at regional level) and in Italy (6 regional items).

¹⁴ This is just an exploratory analysis. In the next steps, these "worldwide" sources will be disentangled at national level, when appropriate.

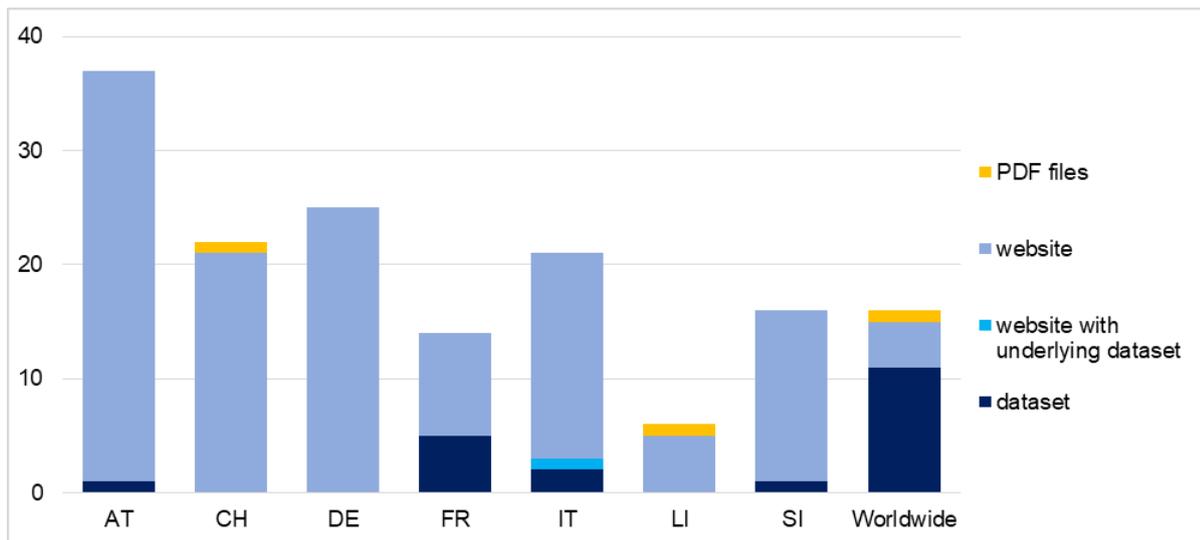


Table 6 – Items to detect information on labs, by e-format and by geographical location

Geographical location of the source	datasets	website + underlying datasets	websites	PDF files	Total
Austria	1		27		28
Austria_Carinthia			1		1
Austria_Salzburg			1		1
Austria_Styria			1		1
Austria_Tirol			1		1
Austria_Upper Austria			1		1
Austria_Vienna, Lower Austria and Burgenland			3		3
Austria_Vorarlberg			1		1
France	5		9		14
Germany			25		25
Italy	2		13		15
Italy_Friuli Venezia Giulia			1		1
Italy_Lombardy		1			1
Italy_Piedmont			2		2
Italy_Veneto			2		2
Liechtenstein			5	1	6
Slovenia	1		15		16
Switzerland			21	1	22
Worldwide	11		4	1	16
Total	20	1	133	3	157

Source: authors' elaboration

Figure 3 – Items to detect information on labs, by e-format and by country



Source: authors' elaboration

When considering the country of the sources of the item, also the **type of the organisation** matters. To this respect, four different types have been disentangled, with the idea that each of them might be characterised by a very different strategy of action for the Re-search Alps project, with regard to the kind of contact to be undertaken in order to have access to data. To this regard, the four identified types of sources are:

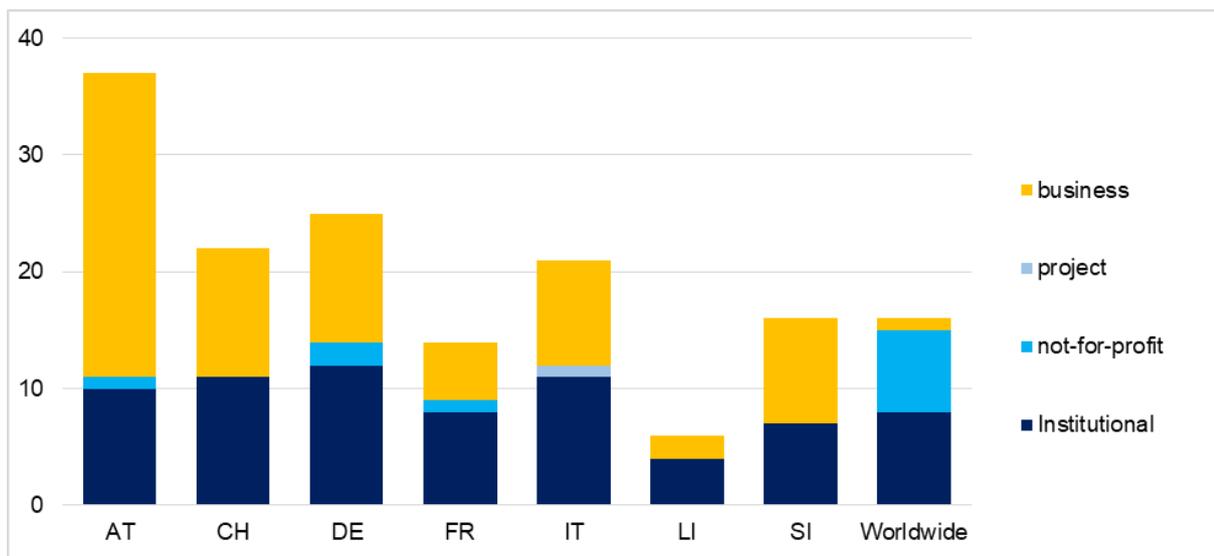
- business organisations: both for-profit companies and public-private companies (such as Fraunhofer);
- public institutions, publishing information on labs as part of their own duties;



- not-for-profit organisations: namely, those organisations that are active in the research field and publishing information for other scientific purposes;
- projects (e.g. the ones financed by EU programmes or funds).

Items whose sources come from both public institutions and business organisations are widely represented in the EUSALP area. However, their relative importance differs country by country. In Austria, the largest part of the available items is published by business sources, while in Germany and in Italy items from institutional sources are more numerous than business ones. Also in France, despite a lower total number of items, institutional sources are the prevailing ones. Lastly, not-for profit sources represent a minority of sources, with one only exception: worldwide sources. Lastly, only one project-related source is observed, in Italy (Figure 4).

Figure 4 – Items to detect information on labs, by type of organisation of the source and country



Source: authors' elaboration

4.4 Data and policies for their use

Items also differ in terms of the features of the data they return. As far as data are concerned, the following features are important for the Re-search Alps project: the policies for using the data they publish; the way of retrieving data; the adopted language; the territorial coverage; the presence of geographic information; the presence of an ID code; their update and the use in scanR; time series.

In this sub-section, adopting the perspective of the potential use of the data in the Re-search Alps project, we discuss the information regarding to:

- the **policies for using the data**, namely: free, restricted, no, Not usable_no data, Not usable_no structured data, unknown;
- the **way of retrieving data**, namely: raw data, raw data (CSV; XLS; XML), Both. Api & csv files, Both. Api & JSON, Api, api_OAI PMH, Crawling, pdf, to be checked.
- the **language** adopted for describing the pieces of information. Three different criteria are implemented, according to the type of e-format of the item: for PDF files, the language in which the file is written; for websites, the original language(s) in which the website is published; for datasets, the language(s) through which queries are possible.
- **territorial coverage** of the pieces of information published, namely: regional; national; international.



- **availability of geographic information** for the instances published in the item, namely: no; geographic coordinates available, Only Address field; Only City & country; Only country; yes (but not specified); yes_(c)swisstopo; yes_geobasis DE; yes_Google Maps view; yes_Openstreetmap.
- the presence of a **unique ID code**, which indicates whether instances are identified through a unique identification code, such as the VAT code or other internal ID numbers.
- **frequency of the update**, namely: updated; realtime crowdsourcing; daily; weekly; monthly; every 2 to 4 months; every 4 months; yearly; Not regular.
- the **use in scanR** project (and available via <https://scanr.enseignementsup-recherche.gouv.fr/api/swagger-ui.html>), namely: yes; yes_only for French ones; yes_can be extended; soon; no.

The **policy for the use** of data largely differs across the 157 items: 94 of them return free data¹⁵; 30 of them return restricted data¹⁶, 11 items return data with no permission to be used, 22 items return either no structured data ("not usable_no structured data": 18 cases) or data that are not usable at all ("not usable_no data": 4 cases) (Table 7). With regard to this classification, it has to be noticed that those items whose data are classified as "restricted" could be included into the final Re-search Alps web-dataset¹⁷.

With regard to policy for the use of data by geographical location of the sources, the percentage of items whose data are freely available is the highest in the set of worldwide and French sources (87.5% of the total and 78.6% of the total, respectively). In other EUSALP countries, the percentage of free-use items is lower, ranging from 50 to 60%, on average. Two negative exceptions occur: in both Lichtenstein and Slovenia, less than one third of items publish free data (Table 7).

With regard to policy for the use of data by type of organisation publishing the item, business sources are confirmed to be the type of sources returning the most restricted data: the 11 items that return data with no permission to be used are published by business sources. On the opposite, when considering the 11 items published by not-for-profit sources, 9 of them are characterised by data whose use is free (Figure 5).

¹⁵ This type also includes the items regulated by creative commons (CC) license (which enables the free distribution of an otherwise copyrighted work, giving other users the right to share, use, and build upon a previous work), and by the Etalab Licence (which is an open license within the French system). Nevertheless, for the purposes of the Re-search Alps project, when websites are considered, if no policies are clearly stated, we have assumed that information has to be considered as freely available.

¹⁶ Access is considered as restricted also when a registration/login is explicitly required. Moreover, in some cases, it is not possible to use data for massive searches or for creating other datasets; in other cases all contents such as texts, images, movies, trademarks, logos, banners, headlines, audio-visuals, etc. are protected under copyright, patents and intellectual property laws.

¹⁷ As a matter of fact, restrictions refer to the obligation to acknowledge the original source or to the limitation to non-commercial uses.

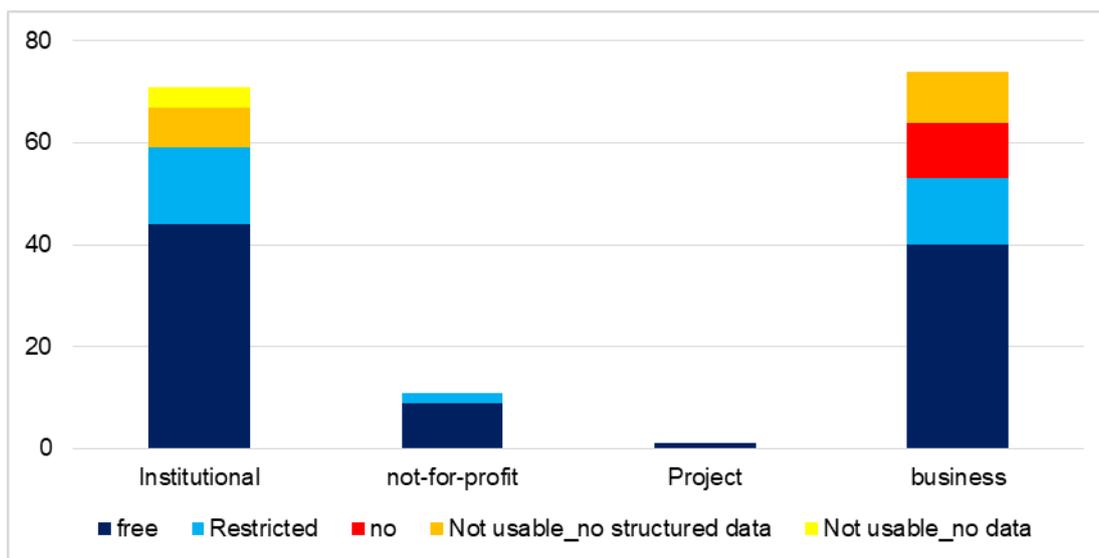


Table 7 – Items to detect information on labs, by and by geographical location

Geographical location of the source	free	Restricted	no	Not usable_no structured data	Not usable_no data	Total
Austria	18	2	2	6		28
Austria_Carinthia				1		1
Austria_Salzburg		1				1
Austria_Styria	1					1
Austria_Tirol				1		1
Austria_Upper Austria				1		1
Austria_Vienna, Lower						
Austria and Burgenland	3					3
Austria_Vorarlberg				1		1
France	11	2	1			14
Germany	15	8	2			25
Italy	8	4	3			15
Italy_Friuli Venezia Giulia	1					1
Italy_Lombardy		1				1
Italy_Piedmont	1			1		2
Italy_Veneto	2					2
Liechtenstein	2			3	1	6
Slovenia	5	5	1	2	3	16
Switzerland	13	6	2	1		22
Worldwide	14	1		1		16
Total	94	30	11	18	4	157

Source: authors' elaboration

Figure 5 – Items to detect information on labs, by policy for the use of data and by organisation publishing them



Source: authors' elaboration

In general terms, 124 items (namely, 94 free and 30 restricted items) out of 157 represent the relevant set of items to be considered for **data retrieval**, the ultimate goal for feeding the Re-search Alps web-dataset¹⁸. However, the importance of this set of relevant items must be considered under

¹⁸ According to the analysis on the policies for the use of data, this information is non-relevant for 33 items, namely the 11 items that return data with no permission to be used, the 22 items that return either no structured data or data that are not usable at all.



the light of their actual importance in terms of number of instances they return, which may largely vary item by item (as we will discuss in sub-section 4.5).

With regard to data retrieval, different ways are admitted. While in Table 8 each of these ways is returned singularly, Figure 6 refers only to a few macro-categories, by country. In particular, the following are considered to: raw data (i.e. the easiest way to access data, when the download of a file containing the original information is admitted); Api (when an application programming interface is provided to retrieve data); crawling (which is possible, especially when the item is a website); PDF files (when information is just published that way) and to be checked (which occurs when policy for use of the data is restricted).

Accordingly, in the largest number of cases, data retrieval is possible only through crawling (84 cases out of 124): such a technique is almost the only one admitted across Austrian, German, Italian and Swiss items. French items admit Api in a larger number of cases (7 out of 13 cases). Conversely, the worldwide sources are the ones that provide raw data in a larger number of observations (Figure 6).

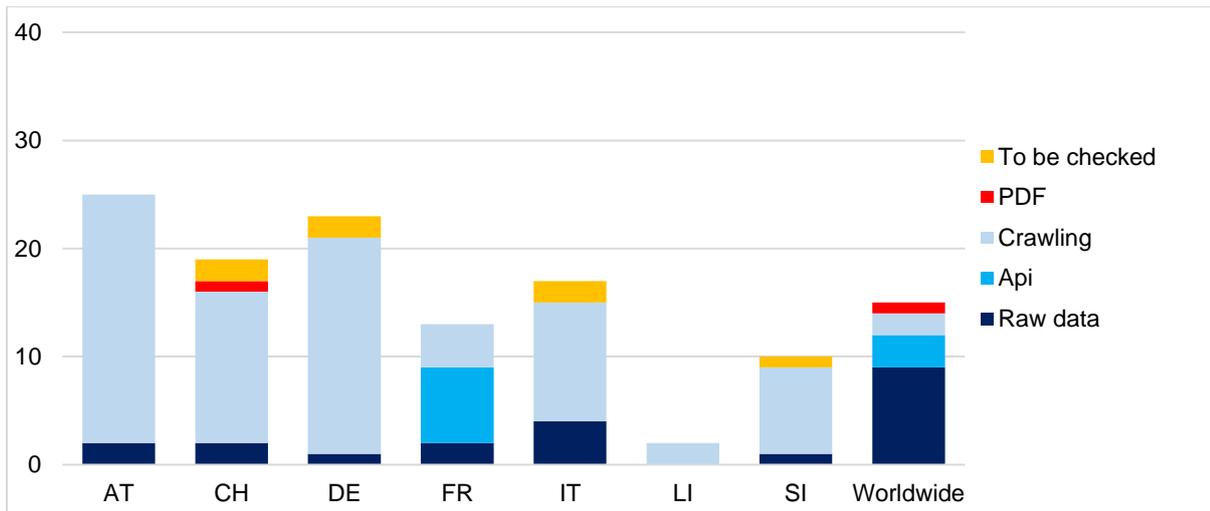
Table 8 – Ways of retrieving data of items to detect information on labs, by geographical location of the source

Geographical location of the source	raw data	raw data (CSV)	raw data (XLS)	raw data (XML)	Both. Api & JSON	Both. & csv files	api	api_OA I PMH	crawling	pdf	To be checked	Total
Austria		2							18			20
Austria_Carinthia												
Austria_Salzburg									1			1
Austria_Styria									1			1
Austria_Tirol												
Austria_Upper Austria												
Austria_Vienna, Lower Austria and Burgenland									3			3
Austria_Vorarlberg												
France		1		1			5	2	4			13
Germany		1							20		2	23
Italy	1	2	1						7		1	12
Italy_Friuli Venezia Giulia									1			1
Italy_Lombardy											1	1
Italy_Piedmont									1			1
Italy_Veneto									2			2
Liechtenstein									2			2
Slovenia		1							8		1	10
Switzerland		2							14	1	2	19
Worldwide	3	1	2	1	1	1	3		2	1		15
Total	4	10	3	2	1	1	8	2	84	2	7	124

Source: authors' elaboration



Figure 6 – Items to detect information on labs, by way of retrieving data and by country



Source: authors' elaboration

Language is an additional critical issue in the implementation of a project such as Research-Alps, since four different languages are used in the EUSALP area. This is the reason why a multilingual website will be created by Re-search Alps.

As a matter of fact, the 157 items gathered together so far are characterised by the presence of different (and multiple) languages: each of the official languages spoken within the EUSALP boundaries but Slovenian may occur: items are published (for PDF files/website) or queries can be performed (for datasets) in French, German and Italian. In addition to the use of national languages, some items are also in English (national languages plus English) while others are published only in English. As a matter of fact, data suggest that English is the most used language: in 107 cases out of 157, English is the only adopted language; in 23 cases items are translated (both national language and English). Moreover, three more cases are about items published by worldwide sources, which show all the possible language. On the opposite, in 24 out of 157 cases, information is available only in the national language. This is the case of two Austrian and four German items, six French items; nine Italian items and 3 Lichtenstein items. In all these cases, some typical issues related to the adoption of multiple languages may arise, especially when a comprehensive dataset is implemented (Table 9 and Figure 7).

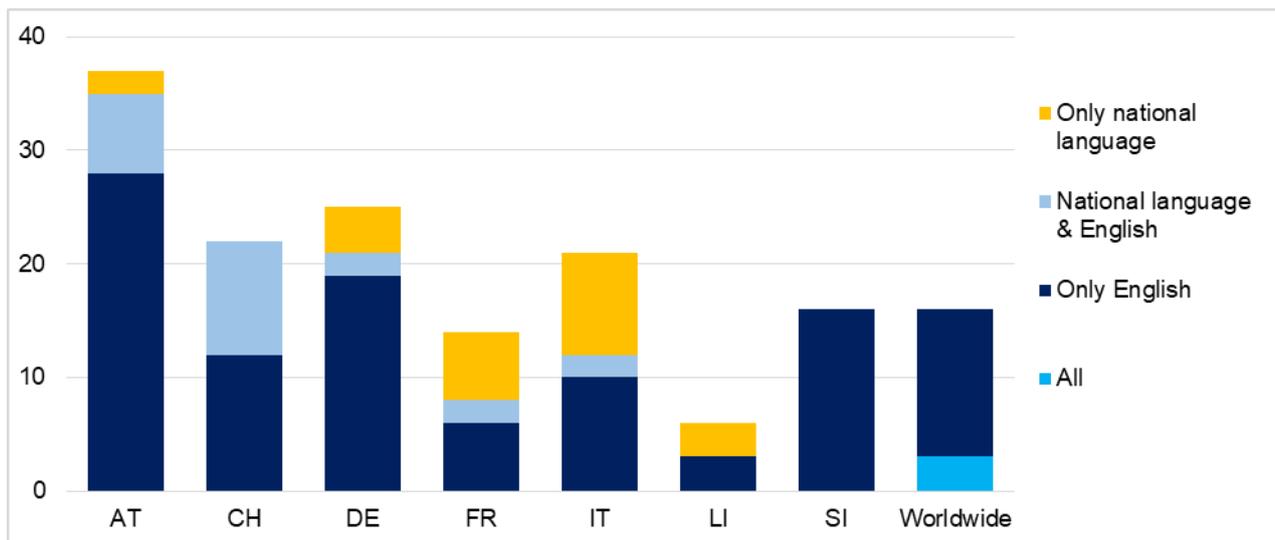


Table 9 – Language of the retrieved objects, by geographical location of the source

Geographical location of the source	DE	DE / EN	EN / FR / DE	EN / FR / DE / IT	FR	FR / EN	IT	IT / EN	EN	All	Total
Austria	2	6							20		28
Austria_Carinthia									1		1
Austria_Salzburg		1									1
Austria_Styria									1		1
Austria_Tirol									1		1
Austria_Upper Austria									1		1
Austria_Vienna, Lower Austria and Burgenland									3		3
Austria_Vorarlberg									1		1
France					6	2			6		14
Germany	4	2							19		25
Italy							5		10		15
Italy_Friuli Venezia Giulia								1			1
Italy_Lombardy							1				1
Italy_Piedmont							1	1			2
Italy_Veneto							2				2
Liechtenstein	3								3		6
Slovenia									16		16
Switzerland			6	4					12		22
Worldwide									13	3	16
Totale complessivo	9	9	6	4	6	2	9	2	107	3	157

Source: authors' elaboration

Figure 7 – Items to detect information on labs, by language and by country



Source: authors' elaboration

In addition to the aforementioned issues, which are critical in the implementation of a comprehensive dataset, two additional issues play a role: geographical coverage of original data and availability of geographical information about the labs described by each of the items.

As far as **geographical coverage** is concerned, this has to do with the territorial area that each of the item covers in terms of labs. Accordingly, items can have a regional, national or international relevance. Table 10 returns the geographical coverage of the items to detect information on labs, by



the geographical location of the sources providing them. According to these figures, most of items show only a national or regional coverage (124 and 12 cases out of 157, respectively). In this context, regional coverage always refer to a region within the EUSALP area. Conversely, only 21 items out of 157 cover labs across international areas: among them, worldwide institutions are the vast majority (Table 10).

In general, such a finding supports the conclusion that most of the information about labs within the EUSALP area does not refer to international level. Accordingly, additional efforts are required, in order to provide homogeneous information on labs throughout the area.

Table 10 – Geographical coverage of the items to detect information on labs, by geographical location of the source

Geographical location of the source	International	National	Regional	Total
Austria		28		28
Austria_Carinthia	1			1
Austria_Salzburg		1		1
Austria_Styria		1		1
Austria_Tirol			1	1
Austria_Upper Austria			1	1
Austria_Vienna, Lower Austria and Burgenland			3	3
Austria_Vorarlberg			1	1
France	1	13		14
Germany	2	23		25
Italy		15		15
Italy_Friuli Venezia Giulia			1	1
Italy_Lombardy			1	1
Italy_Piedmont			2	2
Italy_Veneto			2	2
Liechtenstein		6		6
Slovenia		16		16
Switzerland	1	21		22
Worldwide	16			16
Total	21	124	12	157

Source: authors' elaboration

In addition to territorial coverage, geography has also to do with the availability of **geographic information** about labs, in the original dataset. This piece of information is particularly insightful for the end users, which may be interested in knowing the exact location of labs within a given area of interest. Unfortunately, the vast majority of items (95 out of 157) does not return geographic location of labs. Even if available, in 1 case, location is referred to just as the name of the country; in 7 cases just the name of the city is available. Conversely, 25 items return the complete address of the lab: through a software for geocoding, addresses can be converted into geographic coordinates. In other cases geographic information is returned in a much more detailed way, such as by means of the presence of geographic coordinates, or by geocoded information (to this regard, adopted maps refer to Google maps view or Openstreetmap, in most cases (

Table 11).

Figure 8 returns the same information, by summing up it by country and by only returning four different types of geographic information availability: yes; address of the lab; city&country or only country; no. The two countries returning the largest number of items with at least the complete



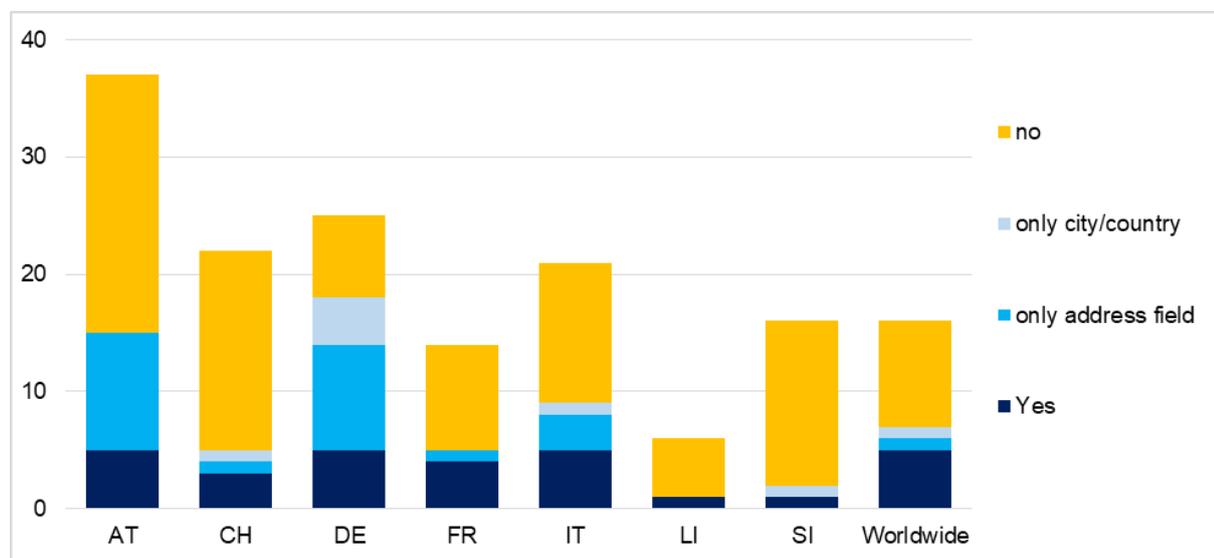
address field or a more complete set of geographic information is Germany and Austria. Conversely, Switzerland, France and Slovenia share the lowest number of observations returning geographic information (Figure 8 and Table 11).

Table 11 – Availability of geographical location of labs and others, by typology of the source

Geographical location of the source	no	Only country	Only city & country	Only address field	geographic coordinates available	yes	yes_(c) swisstopo	yes_geobasis DE	yes_Google Maps view	yes_OpenStreetMap	Total
Austria	16			10		1				1	28
Austria_Carinthia	1										1
Austria_Salzburg	1										1
Austria_Styria	1										1
Austria_Tirol	1										1
Austria_Upper Austria	1										1
Austria_Vienna, Lower Austria and Burgenland									3		3
Austria_Vorarlberg	1										1
France	9			1		2				2	14
Germany	7		4	9				2	2	1	25
Italy	7		1	2		3			1	1	15
Italy_Friuli Venezia Giulia				1							1
Italy_Lombardy	1										1
Italy_Piedmont	2										2
Italy_Veneto	2										2
Liechtenstein	5					1					6
Slovenia	14		1							1	16
Switzerland	17		1	1			2			1	22
Worldwide	9	1		1	2	2			1		16
sTotal	95	1	7	25	2	9	2	2	7	7	157

Source: authors' elaboration

Figure 8 – Items to detect information on labs, by availability of geographic information and by country



Source: authors' elaboration



Additional issue, when gathering together information that will feed a comprehensive dataset about research organisations and laboratories in the EUSALP area, has to do with the presence of a **unique ID code**. Unfortunately, in 86 cases the information is missing; 50 items explicitly lack this piece of information; 9 items return internal ID codes, codes & registration numbers. In the remaining cases, 8 items adopt a unique VAT code while, in 4 French cases, the RNSR (Répertoire National des Structures de Recherche) ID code for public labs and the SIRET/SIREN code¹⁹ for other institutions is returned.

Even the **frequency of the update** of the information published by items is an important characteristic for the Re-search Alps project. In order to regularly update information in the final comprehensive dataset on labs, it is essential that even original datasets are updated with a given regularity. Unfortunately, this field is returned only in 52 items out of 157. Even the quality of this piece of information is rather poor: In 17 out of 52 items, data are "generically" updated, while in 3 cases, the update is generated by a realtime crowdsourcing and in 8 cases data are updated daily. However, for 17 items the update is only once per year (Table 12).

Table 12 – Availability of geographical location of labs and others, by typology of the source

Geographical location of the source	Updated	Realtime crowdsourcing	daily	weekly	Monthly	every 2 to 4 months	every 4 months	yearly	Not regular	Total
Austria	2							3		5
Austria_Carinthia										
Austria_Salzburg										
Austria_Styria										
Austria_Tirol										
Austria_Upper Austria										
Austria_Vienna, Lower Austria and Burgenland										
Austria_Vorarlberg										
France		2	3					5		10
Germany	11							2		13
Italy	1		1				1	2	2	7
Italy_Friuli Venezia Giulia										
Italy_Lombardy	1									1
Italy_Piedmont									1	1
Italy_Veneto										
Liechtenstein										
Slovenia	1							2		3
Switzerland	1							1		2
Worldwide		1	4	1	1	1		2		10
Total	17	3	8	1	1	1	1	17	3	52

Source: authors' elaboration

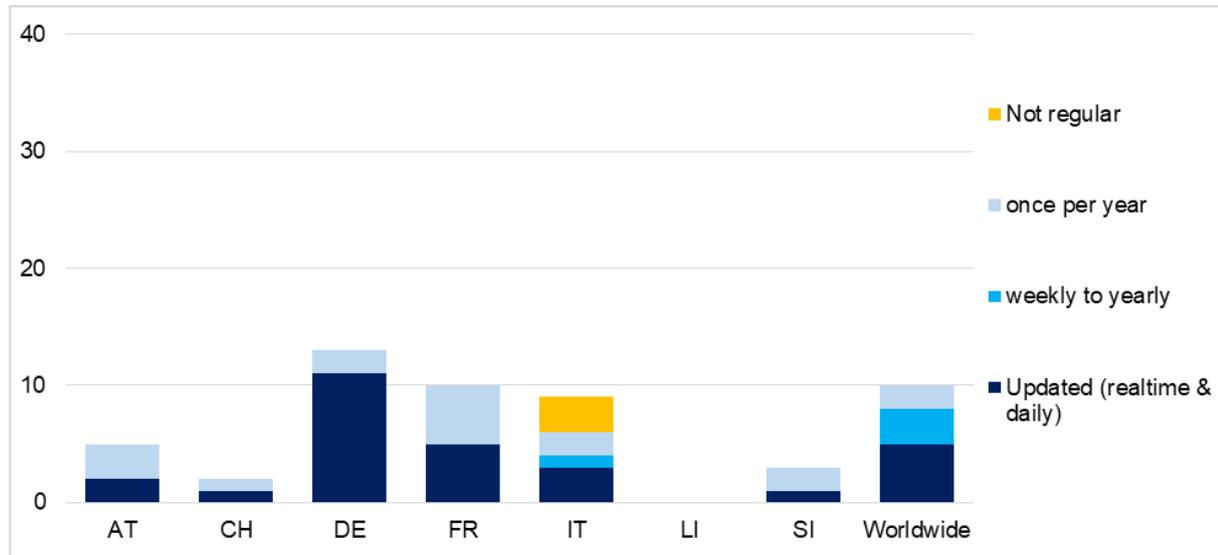
With regard to the country of location of the source, the largest number of items whose update frequency is known is located in Germany and France. In Italy, it is possible to notice a largest share of items whose frequency of the update is not regular (Figure 9 and Table 12). Such a feature could

¹⁹ SIREN (*Système d'identification du répertoire des entreprises*) codes are given to all French businesses and non-profit associations; SIRET codes are given to their establishments and facilities. Both codes are maintained by INSEE.



actually introduce some limitations, when trying to regularly update the final dataset to be collected and published.

Figure 9 – Items to detect information on labs, by frequency of the update and by country



Source: authors' elaboration

Lastly, also the field **Used in scanR** is of relevance. Being the French scanR experience the most relevant one for the purposes of the Re-search Alps projects, it has been analysed whether any of the original 157 items is also used by scanR or not. Actually, only 11 items are currently used by it, and one additional item will be soon integrated. As expected, items used in scanR are only French items and worldwide items.

4.5 Type of Instances and their number

In section 4.3, we have already discussed the number of items collected at country level, highlighting the heterogeneity that occurs across the area. Indeed, number of items is definitely higher for some countries (e.g. Austria) than for others (e.g. France). As the two mentioned examples suggest, this heterogeneity not only depends on the size of the country or the number of labs the currently exist at country level.

However, for the purposes of Re-search Alps project, the number of selected items is just part of the story: as we discussed above, items also differ in terms of the sources providing them and in terms of data and policies for their use. However, also information on instances matter: instances by type; number of instances; comments for instances of the type 'multiple'.

In this sub-section, we discuss the information regarding to:

- **instances by type**, namely: Academies of science and humanities; Business associations and chamber of commerce; Clinical Trials; companies; ministries; multiple; multiple_organisations; National Centres of Competence in Research (NCCR) ; patents; people_H2020 Expert Evaluators (49 lists); people_researchers; people_university staff; Phd dissertations; projects; Publications; regional offices; research laboratories; research organisations; research organisations_medical; research organisations_networks; scientific institutes; to be explored; universities. Here, the value "multiple" refers to the chance of having collected more than one typology of instances (e.g. organisations, people and



projects). The value "multiple_organisations" is adopted when only different typologies of organisations are returned.

- **Instances: number**; namely the number of the instances returned. If type of instance is multiple or multiple_organisations, then the sum of the instances of all types is returned, here.
- **Instances type 'multiple': comments**: when type of instance is multiple or multiple_organisations, then the single types of instances are returned together with the number of each of them.

Information about instances' types is available in 153 out of 157 items, as the field is non-relevant for the four items whose policies for the use is not usable_no data. According to the available typologies, four broader groups of instances can be detected (Table 13): organisations (90 items), people (3 items); projects (19 items) and multiple (36 items). The latter group of instances can be eventually disentangled into: i) multiple_organisation (28 cases), when only instances belonging to the first group (such as, research organisation, universities...) occur; and ii) multiple (8 cases), when instances from organisations, people and project are jointly published. To this respect, a typical example is the ORCID dataset, which contains data about researches (people), their publications (projects) but also about researchers' affiliation (e.g., universities or other research organisations).

Table 13 – Items to detect information on labs, by type of instances

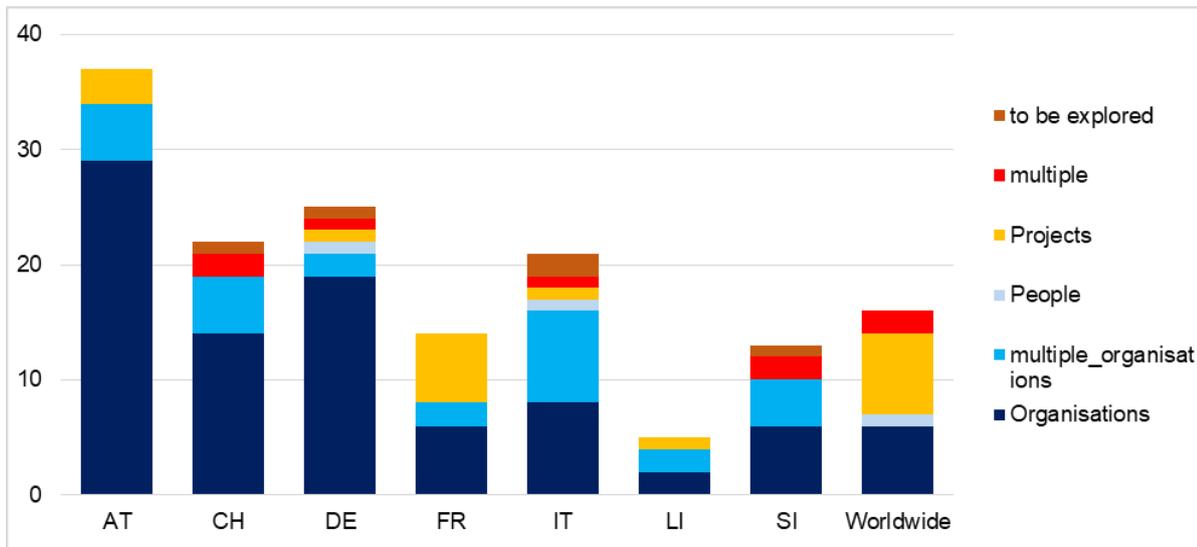
Instance type		Number
Organisations	Academies of science and humanities	1
	business associations and chamber of commerce	1
	companies	5
	ministries	1
	National Centres of Competence in Research (NCCR)	1
	regional offices	1
	research laboratories	6
	research organisations	53
	research organisations_medical	1
	research organisations_networks	3
	scientific institutes	1
	universities	16
People	people_H2020 Expert Evaluators (49 lists)	1
	people_researchers	1
	people_university staff	1
Projects	Clinical Trials	1
	patents	2
	Phd dissertations	1
	projects	8
Multiple	Publications	7
	multiple_organisations	28
	multiple	8
to be explored	to be explored	5

Source: authors' elaboration

Figure 10 returns the number of items by broad category of types of instances and by country. As expected, the wide majority of items returns organisations or multiple organisations. However, both among French items and worldwide items, a larger number of items is about projects. Lastly, a number of items either only return multiple instances (namely, organisation, people and projects) or they are still to be explored (this type mostly occur when policies for the use are restricted, i.e. a login is required).



Figure 10 – Items to detect information on labs, by type of instances (broad categories) and by country



Source: authors' elaboration

In addition to this analysis, which has to do with the number of items, also number of instances that is returned is important. In order to comment comparable results, here we just focus on those items that refer to both organisations and multiple_organisations, ignoring items about people, projects or items returning multiple instances. Thus, the final set of items comprises 118 items, which cover all EUSALP countries. Table 14 returns the number of instances by type and by geographical location of the source.

Data by location of the source publishing items are incredibly heterogeneous. Indeed, France, although it was characterised by a smaller number of items compared to other countries (only 8) is actually the country with the largest number of instances (more than 9 million, mostly companies). A large amount of instances (although it is not comparable to the French one) is also available for Germany (53 thousand instances, collected by 21 items). Even in the German case, the largest amount of information has to do with companies. Among large countries, Italy is the one with the lowest amount of instances (a bit more than 12 thousand), when including both nation-level and regional level items). Moreover, in the Italian cases, the most represented type is the multiple_organisation one. Lastly, as expected, even worldwide items are characterised by a large number of instances (Table 14 and Figure 11).



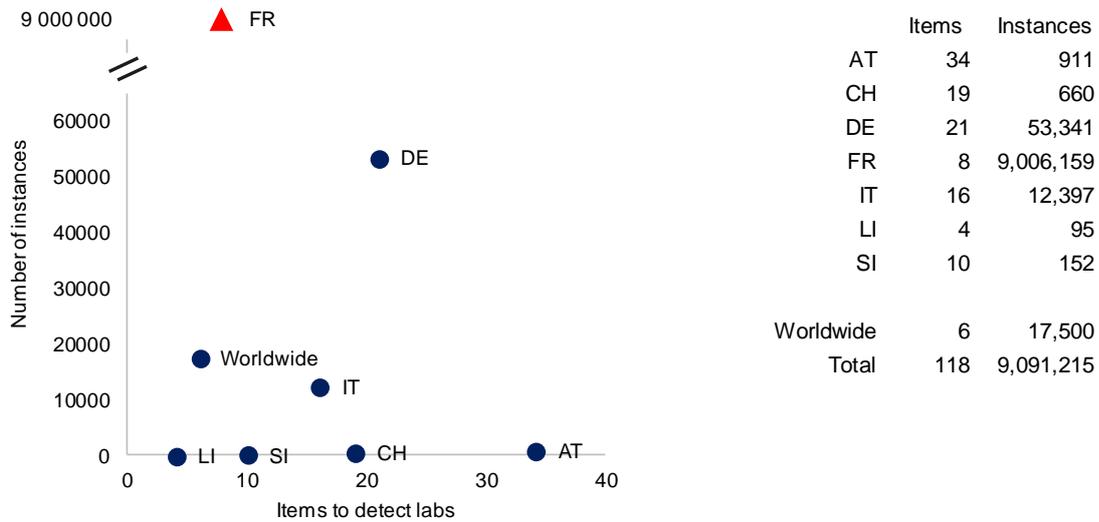
Table 14 – Number of instances in each item referring to organisations, by location of the source

Geographical location of the source	Academies of science and humanities	business associations and chamber of commerce	companies	ministries	National Centres of Competence in Research (NCCR)	regional offices	research laboratories	research organisations	research organisations - medical	research organisations - networks	scientific institutes	universities	multiple_organisations	Total
Austria			70	2			21	330				63	88	574
Austria_Carinthia													56	56
Austria_Styria			200											200
Austria_Tirol													11	11
Austria_Upper Austria								8						8
Austria_Vienna, Lower Austria and Burgenland												29		29
Austria_Vorarlberg								27						27
France			9,000,000				1,514	4,220				14	411	9,006,159
Germany	8		50,000				307	1,478		571		679	298	53,341
Italy							544	1,006				18	8,796	10,364
Italy_Friuli Venezia Giulia														29
Italy_Lombardy													1,508	1,508
Italy_Piedmont													496	496
Liechtenstein		6						5					84	95
Slovenia							1	26			13		112	152
Switzerland					37	10	42	268	12	30		106	155	660
Worldwide								17,500						17,500
Total	8	6	9,050,270	2	37	10	2,429	24,903	12	601	13	909	12,015	9,091,215

Source: authors' elaboration



Figure 11 – Items referring to organisations and relative number of instances, by country



* In the case of France, instances are 9,006,159, of which 9 million are companies.
Source: authors' elaboration

According to these figures, it is possible to compare them with some kind of benchmark, as represented by the total number of enterprises that in each country are classified under the Section M or the division 72 of the NACE Rev. 2 (as returned in the Section 2 of this document). As shown in Table 15, France is definitely overrepresented in the items gathered, as the number of instances is well above the total number of enterprises active under Section M of the NACE Rev. 2. Conversely, as far as Germany and Italy, underrepresentation of the identified instances seems to occur. Indeed, when comparing data on instances with data on companies, we have to bear in mind that instances also include public labs, such as universities or other public research organisations.

Table 15 – Number of active enterprises and number of instances by country*

	M Professional, scientific and technical activities	72 Scientific research and development	Instances
AT	65,393	1,037	911
FR	463,476	5,769	9,006,159
DE	478,857	6,823	53,341
IT	714,934	8,796	12,397
LI	n.a.	n.a.	95
SI	30,757	1,221	152
CH	23,822	558	660
Worldwide	n.a.	n.a.	17,500

* Data on Lichtenstein and on worldwide locations are not available.
Source: authors' elaboration

5 Estimating the economic and social impact of the results of the project

The analysis presented in the above sections contributes in outlining the main characteristics of our modelling design of the potential impacts of the results of the Re-search Alps project.

With regard to the two pre-post comparisons about the information on the enterprises that belong to the section M of the NACE Rev. 2 and about the more general availability of data on labs in the



EUSALP area, Section 3 and 4 have provided a comprehensive overview of the state of the art, from which we can derive three main conclusions on the main impacts of this project.

Firstly, the creation of a unique access point for the available information on private and public labs would provide an effective opportunity for companies belonging to the different regions to access information on labs across the area.

Secondly, building a multilingual web-dataset would enlarge such an opportunity to users that belong to different linguistic areas, given the specific multilingual feature of the involved area. As a matter of fact, it is characterised by the co-existence of four languages, with the addition of English as a fifth language to be adopted for international and extra-EUSALP area retrieval of information.

Thirdly, there is an additional impact which is very specific of the Re-search Alps project: namely, the cumulative and comprehensive set of information that will be provided by the web-dataset as a unique opportunity for the seven countries. This occurs in a condition in which supporting innovation in the regional eco-systems, characterised by SMEs, strongly demands for easy and open access to information. To this extent, the delivery of geo-referenced information represents an important condition in order to properly select a provider of services.

Along this line of achievements, there would be a complementary line of interventions the Re-search Alps will promote across the seven countries: namely, the conditions to have an open access at EU level to the national registries of companies, as recently started by the French government (as discussed in Section 3).

With regard to the assessment of the impact of the project, we can refer to the experience of the scanR project. After one year of activity, it is difficult to know a priori who will be interested in the data produced by scanR. They have been gradually discovering uses and users of their own web-dataset. In Re-search Alps, we start detecting information on possible uses and users after the first year of the project, when the dataset will gather complete information on the French and the Italian cases. Then, our experience monitoring our uses and users will provide hints to the implementation of the web-dataset to other EUSALP countries, acquiring a better understanding of potential users and uses.

For example, if we consider the scanR experience, we can highlight at least seven types of activities/actors that have found interest in the development of the web-dataset:

- projects targeting innovative companies (see www.c-radar.com);
- public investment banks (also supporting companies) that need to profile their targets in order to support their own investment decisions (e.g. Bpifrance);
- companies that aims to concentrate all data on the agricultural sector in order to provide digital services to farmers and to the food industry (as in the case found in www.api-agro.fr);
- companies whose services consist in identifying experts and expertises (as in the case of en.expernova.com/?noredirect=en_GB or www.idexlab.com/en/);
- companies that monitor research activities in universities and public research organisations (such as the case of <http://university-analytics.com/>);
- technology transfer companies (for example www.sattse.com/);
- and all public research actors that can find relevant information made available by the website, such as universities and other public research bodies (such as OECD, EC), or foreign delegations (for a better understanding of the research in their host country, in order to find new partners and to develop international collaborations).

This example highlights a target of users and uses of the web-dataset that Re-search Alps will provide that goes well beyond the set of SMEs that only look for labs providing scientific and technical



services in their area, in order to support their own activities (i.e. the initial target we focused on when outlining the Re-search Alps proposal).

As far as the one-year scanR experience after its launch, we also know some specific uses of the web-platform, provided through Google Analytics: in one year, nearly 300,000 pages were viewed by more than 72,000 different visitors who spend an average of 7 minutes and 30 seconds on the site. In addition, all the open data sets, associated with scanR open data platform, have been downloaded almost 30,000 times over one year.

From these data, we should consider that even though 72,000 different visitors could appear by itself a significant impact, their uses lead to a more disruptive effect. Indeed, their uses reflect their keen interest for scanR. What is even better is that scanR has provoked a mobilization of all public research and higher education institutions. In addition, many companies asked to improve their image via scanR, and asking how scanR could better report on their activities. For example, scanR team answered them to systematically deposit their publication on HAL, i.e. the application of open archives. This, in turn, contributes to the improvement of sources on research and innovation.

In the case of Re-search Alps, we consider that this type of interaction with users will foster a virtuous circle through which building up an increasing quantity of data, of better quality on research and innovation. In addition, this will support a big shift in providing useful data to define and evaluate public policies.

Beyond the pre-post comparisons, this is then the third aspect of our model of analysis of the impact of Re-search Alps, which refers to process of change within institutional organisations that make accessible information (e.g. national and regional ministries). Again, in the experience of scanR, the French Ministry of Research improved its reputation and had the opportunity to enhance its activity. In the Re-search Alps projects, ministries as well as regional government, all the innovation agencies, innovation intermediaries organisations will be supported in being at the cutting edge of digital technologies, encouraging the construction of powerful and exhaustive databases.

A final achievement will be to provide grounded support to transparency, by encouraging public laboratories and companies to surpass themselves. By sharing information on their activities, both strengths and weaknesses will be made comparable. This is an additional value added of the Re-search Alps project.

References

European e-justice portal (2016), Business registers in Member States. Available at: https://e-justice.europa.eu/content_business_registers_in_member_states-106-en.do?clang=en. Last accessed on: October, 17, 2017.

Eurostat (2008), NACE Rev. 2. Statistical classification of economic activities in the European Community. Luxembourg: Office for Official Publications of the European Communities, available at: <http://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF>).

Eurostat RAMON (2017), Statistical Classification of Economic Activities in the European Community, Rev. 2 (2008). Available at http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=NACE_REV2&StrLanguageCode=IT&IntPcKey=&StrLayoutCode=HIERARCHIC

Istat (2011). 9° Censimento Generale dell'Industria e dei Servizi. <http://dati.istat.it>



Re-search Alps



Annex 1

Table A.1 – The database schema

Macro category	Name of the field	Description
	Short description of the item	Short description of the item (when available) and any other comment, which is useful for properly interpreting the item
Source		
	Source name	Source of the item. Name of the organization publishing the pieces of information.
	Source location	Location of the source of the item (i.e. the organization publishing the pieces of information). Values: <ul style="list-style-type: none"> • "worldwide" (for international organisations, operating even outside the EUSALP area); • name of the country (one of the EUSALP countries); • name of the country_region (when relevant).
	Title of the item	Title of the published item (dataset/website/piece of information), shortly explaining the content of the object.
	E-format of the item	Typology of the item that is published. Values: <ul style="list-style-type: none"> • dataset: when data are directly accessible, through queries; • websites: for those pieces of information that are not returned as a structured dataset; • websites + underlying dataset: mixed version of the two previous cases; • PDF: when just a PDF file is returned, containing relevant information.
	Type of organization publishing the item	Characteristic of the sources according to the characteristics of the organisation publishing them. Values: <ul style="list-style-type: none"> • business organisations (both for-profit companies and public-private companies, such as Fraunhofer); • public institutions (publishing information as part of their own duties); • not-for-profit organisations (active in the research field and publishing information for other scientific purposes); • projects (e.g. the ones funded by EU projects).
	Reference: URL	Web link to the item.
	Reference: person name	Name(s) of the contact person(s) to ask for information, if any.
	Reference: e-mail contact	E-mail address to contact the organisation publishing the item
Data		
	Policies for the use	This field returns the policy for the use of available information, as stated by the source publishing the item. Values: <ul style="list-style-type: none"> • free; • restricted; • no; • Not usable_no data; • Not usable_no structured data • Unknown.
	Policies for the use - details	Details on the policies for the use of data.
	Data retrieval	The way to retrieve relevant information from the item. If nothing is specified, crawling is always possible. API is specified when application programming interface (API) is provided in order to retrieve data. Raw data is specified when it is possible to download files containing information. According to this framework, possible values are: <ul style="list-style-type: none"> • raw data; • raw data (CSV); • raw data (XLS); • raw data (XML); • Both. Api & csv files;



		<ul style="list-style-type: none"> • Both. Api & JSON; • Api; • api_OAI PMH; • crawling; • pdf; • to be checked.
	Language	<p>Language adopted for describing the pieces of information. Three different criteria are implemented, according to the type of e-format of the item:</p> <ul style="list-style-type: none"> • for PDF files: the language in which the file is written; • for websites: the original language(s) in which the website is published; • for datasets: the language(s) through which queries are possible.
	Territorial coverage	<p>Territorial coverage of the pieces of information published. Values:</p> <ul style="list-style-type: none"> • regional; • national; • international. <p>Please note that the details that are returned in this field may differ from the one returned in the field "source location".</p>
	Geographic information of instances	<p>Availability of geographic information for the instances published in the item. Values:</p> <ul style="list-style-type: none"> • no; • geographic coordinates available (e.g. the ones returned by wikipedia); • Only Address field; • Only City & country; • Only country; • yes (but not specified); • yes_(c)swisstopo; • yes_geobasis DE • yes_Google Maps view; • yes_Openstreetmap
	ID codes of instances	<p>It indicates whether instances are identified through a unique identification code, such as the VAT code or other internal ID numbers.</p>
	Update: Frequency	<p>Frequency of the update. Values:</p> <ul style="list-style-type: none"> • updated; • realtime crowdsourcing; • daily; • weekly; • monthly; • every 2 to 4 months; • every 4 months; • yearly; • Not regular
	Update: Date of the last update	<p>Time of the latest update of the available pieces of information</p>
	Update: additions and changes	<p>This field specifies whether it is possible to add / change data or not., for instance by directly inserting data (user) or by asking the reference person(s). in some other cases it is clearly stated that it is not possible to modify / update data.</p>
	Used in scanR	<p>This information is about the use of this item in the ScanR project (and available via https://scanr.enseignementsup-recherche.gouv.fr/api/swagger-ui.html). Values:</p> <ul style="list-style-type: none"> • yes; • yes_only for French ones; • yes_can be extended; • soon; • no.
	Time series	<p>Dummy variable (yes/no) to highlight whether information are provided over time</p>
Instances		
	Instances by type	<p>Types of the instances returned in the object:</p> <ul style="list-style-type: none"> • Academies of science and humanities;



		<ul style="list-style-type: none"> • Business associations and chamber of commerce; • Clinical Trials; • companies; • ministries; • multiple; • multiple_organisations; • National Centres of Competence in Research (NCCR) ; • patents; • people_H2020 Expert Evaluators (49 lists) ; • people_researchers; • people_university staff; • Phd dissertations; • projects; • Publications; • regional offices; • research laboratories; • research organisations; • research organisations_medical; • research organisations_networks; • scientific institutes; • universities; • The value "multiple" refers to the chance of having collected more than one typology of instances (e.g. organisations, people and projects). The value "multiple_organisations" is adopted when only different typologies of organisations are returned.
	Instances: number	Number of the instances returned. If type of instance = <i>multiple</i> , the sum of the instances of all types is returned, here.
	Instances type 'multiple': comments	If type of instance = <i>multiple</i> , the single types of instances are returned together with the number of each of them.

Source: authors' elaboration