DEMB Working Paper Series

N. 119

WPTIP 1993-2017: a longitudinal analysis of issues, connections and impact of the discussion on technology and innovation policy at Oecd

Margherita Russo*, Pasquale Pavone**

December 2017

* University of Modena and Reggio Emilia CAPP, Centre for the Analysis of Public Policies Address: Viale Berengario 51, 41121, Modena, Italy
Email: margherita.russo@unimore.it

** CAPP, Centre for the Analysis of Public Policies
Address: Viale Berengario 51, 41121, Modena, Italy
Email: pasquale.pavone@unimore.it

ISSN: 2281-440X online

DEMB Working Paper Series 119

WPTIP 1993-2017: a longitudinal analysis of issues, connections and impact of the discussion on technology and innovation policy at Oecd

10th December 2017

by

Margherita Russo* and Pasquale Pavone**

* Dipartimento di Economia Marco Biagi and Centro di Analisi delle Politiche Pubbliche, Università di Modena e Reggio Emilia, Italy, margherita.russo@unimore.it

** Centro di Analisi delle Politiche Pubbliche, Università di Modena e Reggio Emilia, Italy, pasquale.pavone@unimore.it

Abstract
The occasion of the 50th plenary meeting of the OECD Working Party on Innovation and Technology Policy (TIP) is being taken to elaborate a reflexive analysis of the WPTIP activity to ground a discussion for future activities, particularly in view of the discussion of the next budget plan and for a future mandate after that of 2014-2019. This analysis aims at providing an overview of the theories and analyses grounding the recommendations proposed by WPTIP on technology and innovation policy, in a period in which 'innovation' has become the catchword for any discussion on policies to support economic development and social inclusion. To place this analysis in a historical perspective, this paper adopts a systematic automatic text analysis to organize figures and facts over 25 years. The analysis considers text documents on WPTIP events (such as conferences, workshops, plenary meetings) and activities (such as reports on specific topics).

Keywords: analysis of innovation policy, Oecd, WPTIP, organizational issues, textual analysis, similarity analysis, IRaMuTeQ, Taltac2

JEL classification: 03, 038, Z13

Acknowledgments
The authors wish to thank: Mario Cervantes (senior economist at OECD-DSTI) for his support in discussing our research idea with other TIP delegates; Caroline Paunov (senior economist at OECD) who coordinated the OECD staff to support the creation of the documental archive used in this paper; Dominique Guellec (Head of Division, OECD Directorate for Science, Technology and Innovation) for his invaluable support in opening up the analysis of Working Party on Technology and Innovation Policy (WPTIP) documents to the attention of the CSTP; Dr. Andrés Barreneche, who made available the Oecd Glossary on innovation. Dirk Meissner (professor at HSE, Moscow) for sharing the initial ideas and for the many discussions on methods and results with his team at HSE, Moscow. Alina Deniau (Information Management Officer, OECD-DSTI) for her support in ensuring the files' delivery. Discussions within the enlarged research team have been facilitated by Florence Hourtouat (Assistant, OECD STI/STP) who organized the remote meetings for the whole team. Luca Angelino (BA student on an internship at CAPP-Unimore) has provided invaluable support in the manual extraction of the files to create the data warehouse.
This paper has been prepared for the discussion at the conference WPTIP Conference on "TIP@50: What have we learnt? Where is innovation policy heading?" (Paris, 11th December 2017), and the WPTIP workshop on methods and results of text analysis, to be held in Paris, back-to-back with the Oecd CSTP meeting, in March 2018.

Table of contents

1. Objective and methods

> "The OECD Working Party on Innovation and Technology Policy (TIP) provides evidence-based advice to countries to foster STI policies that enhance productivity and foster sustainable and inclusive knowledge-driven economic growth; strengthen public research institutions and their contributions to innovation and economic performance and foster the creation of science and technology-based firms. Established in 1993, the TIP is composed of government officials and country experts in the field of science, technology and innovation (STI) policy and meets twice per year."
>
> *Excerpted from the presentation in the Innovation Policy Platform*
> https://innovationpolicyplatform.org/oecd-working-party-innovation-and-technology-policy-tip

Created in 1993 as a working group of the Oecd Committee on Science and Technology Policy (CSTP), the Working Party of Innovation and Technology Policy (WPTIP) is approaching its 50th plenary meeting, in December 2017[1].

This occasion is being taken to elaborate a reflexive analysis of the WPTIP activity to ground a discussion for future activities, particularly in view of the discussion of the next budget plan and for a future mandate after that of 2014-2019[2]. This analysis aims at providing an overview of the theories and analyses grounding the recommendations proposed by WPTIP on technology and innovation policy, in a period in which 'innovation' has become the catchword for any discussion on policies to support economic development and social inclusion.

To place this analysis in a historical perspective, this paper adopts a systematic automatic text analysis to organize figures and facts over 25 years. The analysis will consider text documents on WPTIP events (such as conferences, workshops, plenary meetings) and activities (such as country reviews and reports on specific topics).

Automatic text analysis is nowadays recognized as an effective tool when a large digital corpus is under analysis, and expert reading and annotation would not be possible, but also when the perspective is that of a comprehensive and systematic multidimensional analysis. The Automatic Text Analysis Techniques provide a perspective of qualitative and quantitative analysis of the contents, properties and characteristics of text documents, allowing you not to read the text but to analyse and interpret the representations of the information contained in it. These techniques represent the natural context for the application of Text Mining tools, which aim at extracting information written in natural language. Text Mining is a multidisciplinary research area that combines, with equal importance, instruments of Computational Linguistics, Information Retrieval and Statistics.

Approaching the 25 years of WPTIP activity from an outsider's perspective will not mean that no other knowledge than text mining will support the analysis and interpretation of results. On the contrary, results depend on a continuous dialogue between the two authors complementing their fields of knowledge in innovation theories and innovation policies and in text analysis. Further steps and tools in the analysis will be grounded on discussions of the results and on interviews with key informants (staff and delegates, but also experts involved over the 25 years) and the other research teams involved on the same topic. What makes this specific analysis even more intriguing is the attention that the proposal of a text analysis received by WPTIP and by

[1] In 1993 there was only one meeting. Regular schedule on June and December each year was started some years later.

[2] Margherita Russo and Dirk Meissner, who are, respectively, the delegate of Italy and Russian Federation at WPTIP, submitted a preliminary proposal of the project to the Oecd WPTIP Secretariat in June 2017.

two other research teams that are based, respectively, at the Higher School of Economics in Moscow[3], and at IFRIS-ESIEE Paris (Université Paris-Est)[4]. This will give rise to a dialogue in the interpretation of results with the WPTIP delegates (planned in the 50th plenary meeting, in December 2017) and in a dialogue among the research teams to confront their methods and results on the same corpus (as will occur in the international workshop planned back-to-back with the Oecd CSTP meeting in March 2018).

Together with reports (more structured documents summarizing the state of the art on specific topics and often the outcome of specific projects), the events organized by the WPTIP have the goal of introducing the delegates to ideas of experts, policy makers, entrepreneurs. Cluster analyses of the contents will contribute to interpreting their impact in the formation of the framework adopted/created by the WPTIP to discuss/to analyse/to orient innovation policy in the Oecd countries. The present analysis does not cover the country reviews on innovation policy, which would require an ad hoc analysis, since it is a very specific set of issues related to more than 20 countries all over the world.

The structure of this paper is the following. In Section 2 the archive is presented with regard (a) to the original sources provided by the Oecd Secretariat and (b) to the categories of documents adopted for implementing the text analysis. Section 3 presents the tools and methods for text analysis, a preliminary description of the corpus and the decisions on how to analyse the various entities (events and activities). Section 4 presents the main results on the analysis of three sub-corpora of texts, respectively of plenary meetings, conferences and workshops, and reports. Section 5 concludes with an overview of topics dealt by the WPTIP over the past 25 years and an indication of furthers steps in the analysis. Annexes (1-14) present detailed information, with tables and charts summarizing some detailed features of the corpora and graphs with a lower threshold for linkages between words.

2. Data: How to structure the archive and why

The original Oecd files for the text analysis of the WPTIP documents

The files received by the Oecd Secretariat on STIP are described in Annex 1, which presents our process of creation of our data warehouse.

The files received by Oecd refer to documents whose public use has been approved by delegates (they have been "declassified"), having the Oecd code "DSTI/STP/TIP" (Directorate on Science Technology and Innovation, Science and Technology Policy; Technology and Innovation Policy). This criterion of selection implies that they refer to activities involving exclusively, or promoted by, the WPTIP, i.e. it does not contain activities embedding other Oecd directorates, committees and working groups. This archive includes 934 files, in English, in French and in various versions (draft, revisions, …). The specific list of 307 files to be analysed - provided by the Secretariat - guided the extraction of documents to be analysed. Some additional files were specifically requested to complete the series of documents on the TIP plenary meetings[5]. Other files are supposed to be missing, as for several events there are associated only to an agendas with no minutes or summary reports. Moreover, a limitation in the analysis is due to

[3] Dirk Meissner with Pavel D. Bakhtin and Anton Timofeev.

[4] Philippe Laredo and Antoine Schoen.

[5] We singed out some missing files referring to workshops or reports that, except one, were received on request. One document is still missing, in its English version: "Defining subsidies in R&D and industrial innovation [cote: DSTI/STP/TIP(94)14, date 12/22/1994]. The title of French version (available in the original Oecd archive) is:" Definition des subventions à la recherche-developpement et à l'innovation industrielle". Other files are supposed to be missing, as for several events there are only the agendas and no minutes or summary reports. Moreover, the most recent events and projects are not included in the set of files received for the present analysis.

the lack of documents on the most recent events and projects, not included in the set of files received for the present analysis.

A further selection of Oecd documents should include the ones quoted in the agendas of the WPTIP plenary meetings. Those documents could provide indications about the range of interactions across the various Oecd directorates, committees and working groups, highlighting the specific topics that over time have featured in those interactions. These files are not included in the analysis presented in this paper.

In addition to the TIP files, the Oecd vocabulary of innovation topics was made available to the team. It is the vocabulary used by Oecd "to tag and semantically link documents and other types of content in the Innovation Policy Platform. [It] includes the main topic name along with synonyms [and the] URI for each topic"[6]. This file does not associate the tags to each of the documents received (the use of this resource is discussed below, in Section 3).

Categories of documents for text analysis in the Unimore data warehouse

The data warehouse organized by Unimore is made up of 305 documents, including the 31 country reviews. These documents will be examined as a separate corpus, but are included in the same process of pre-treatment of texts. Although mentioned in the Oecd list, 12 files have been excluded from the corpus because a more recent version is available in the archive (8 files) or because they are in French (4) and the corresponding English version has been considered.

Together with several features collected to control the process of classification (the database schema is listed in Annex 2), the Unimore data warehouse of the WPTIP's files has four main typologies to be used in text analysis: types of document, types of event, ID codes to classify activities (country reviews and reports) and events, year (see Annex 3). For each typology, information has been associated to each of the documents, according to their content.

Main differences in the categories: Oecd original archive vs. Unimore data warehouse

Three main differences between the two archives deserve attention.

a. *Almost the same total number of files but a different composition*. The Unimore data warehouse has 305 files: only two less than the ones in the list of the original Oecd archive. The result comes from changes due to the exclusion of files and the inclusion of files not previously available in the original Oecd archive.
b. *Similar categories, but different activities and events*. Categorization of documents has been refined by creating (a) a typology of events (conferences, workshops of different characteristics, plenary meetings) and (b) by disambiguating reports and summaries: 5 reports were categorized among summaries; while 11 summaries of events had been classified as reports. Details are presented in Table 1. In addition, Oecd categorization of agendas with codes "A" or "AH" was checked against the presence of the word "agenda" in the information gathered on the selected files. The result, returned in Table 2, suggests that eight cases were actually agendas, even though they did not have the code for agenda.
c. *The date of the document and the date of the event.* Years are not a clear-cut partition among concepts and ideas, but starting with a temporal classification according to the year in which the event occurs helps in clustering contents in a less arbitrary way.

[6] From Andrés Barreneche's email on 1st November 2017. See Annex 1 for details.

Table 1 - Oecd source vs. Unimore classification of types of documents

	Type of documents as classified by Unimore				
Source	agenda	summary	report	country review	**Total**
FolderOecd_TIP-Publications				31	**31**
not in the initial OECD TIP-list | 3.11.2018	1	1			**2**
OECD_TIP-list_Agenda_summ. reports f. analys.	80	67	5		**152**
OECD_TIP-list_Agenda_summ. reports f. analys. | 3.11.2017	2	4			**6**
OECD_TIP-list_Reports for analysis		6	91		**97**
OECD_TIP-list_Reports for analysis | 3.11.2017		5	5		**10**
other files from_TIP-Official-Documents	3	4			**7**
Total	**86**	**87**	**101**	**31**	**305**

Table 2 - **Oecd source vs Unimore classification of agendas**

	Type of documents as classified by Unimore				
is it an agenda?	agenda	summary	report	country review	Total
COTE...A	49				**49**
COTE..AH	29			6	**35**
other COTE..	8	31	101	81	**221**
Total	**86**	**31**	**101**	**87**	**305**

Will the specific criteria we used in cleaning the archive and creating the data warehouse have an impact on the result of text analysis? In principle, in any analysis the quality of data matters. In the present analysis, it has been of particular importance to have the clearest ideas of basic facts: which kind of activities are dealt with in the documents? Which format is the one adopted in the documents? An agenda has a different structure, and even register, from a summary, a report or a country review. Cleaning the database has allowed the creation of a list of conferences (Annex 7), workshops (Annex 8), terms used to refer to workshops (Annex 9). This will help to highlight missing information on those categories of events. When collected, a new release of text analysis will be realized.

There is a more specific need for categorizing documents. In the text analysis we propose in this paper, we do not work on glossaries extracted from a corpus, counting the relative importance of the various terms (applying indexes like the term frequency–inverse document frequency, TFIDF). We elaborate information grounded on the 'active forms' relevant in specific contexts: nouns, adjective and verbs, multi-words. This is why sub-corpora as units of analysis become relevant, since they can reveal polarizations - temporal and conceptual - that are context specific.

3. Text analysis: methodology, pre-treatment and corpus description

In this section we present methods and tools for text treatment and content analysis, a preliminary description of the corpus and the decisions on how to analyse the various entities (events and activities).

Methodology [7]

For the exploration of the corpus and the presentation of the results, the analysis was carried out by integrating a number of software tools: Taltac2[8] (Bolasco, 2010) to identify lexemes (Lexicon Analysis Units), both as simple and multi-word forms, occurring in the corpus; Spad,

[7] This section is excerpted from Pavone and Russo (2017, pp. 9-11).

[8] TALTAC - *Trattamento Automatico Lessicale e Testuale per l'Analisi del Contenuto* (Automatic Lessical and Textual Treatment for Content Analysis) http://www.taltac.it/it/index.shtml

for Multidimensional Analysis and Cluster Analysis; Iramuteq[9], for the analysis of co-occurrences of words and similarity analysis; Gephi[10], to display the graph resulting from the similarity analysis.

The analysis is characterized by three stages: pre-treatment and lexical processing; multi-dimensional analysis and network of co-occurrences.

Pre-Treatment of texts and lexical analysis

In the present analysis, texts are the ones described in Section 2. Pre-Treatment of texts consists in the acquisition of text by numerical indexing of words (their tokenization) for the recognition of the units of analysis, and in the creation of the document database.

The lexical analysis was carried out using Taltac2 software. This analysis provides: (a) the identification of lexical analysis units (the lexemes, "words"), intended both as single words and multi-words; (b) the attribution of grammatical meta-information to the "words" constituting the Vocabulary; (c) the selection of content words (nouns, verbs, adjectives) as the keywords to be analysed; (d) the probability-based analysis of the meaningful terms characterizing the specific language used in describing the activities and events under analysis.

Text mining: specific language to single out activities and products

A specific language representation was elaborated on a factorial plan[11], which graphically presents the combinations of a matrix of *Lexical Units* × *Characteristic* or a matrix of *Documents*[12] × *Lexical Units*. The position of words on the factorial plan is a function of the association of their occurrences in the sub-texts, thus expressing their similarity or diversity: two words are close because they are present in the same sub-texts. At the centre of the factorial plan are the most common terms among the different languages of the various sub-texts.

Through a correspondence analysis (CA), the row and column elements of the matrix are mathematically formalized as vectors, and the above profiles are represented by points in a multidimensional space. The distances between the lexical profiles are measured using a weighted Euclidean metric (chi-square metric). The complex multidimensional space of the variables [words, in our case] is then reduced to a few key factors that can represent, on dimensions named "factorial axes", the relationships between the elements of the data matrix. CA produces the best simultaneous representation of row profiles vs. column profiles in each factorial plan, and on each of its axes (Bolasco, 2013). To do this, we consider a matrix *Documents* × *Keywords* ($n \times p$). The matrix profiles the different documents to be classified with the occurrences of the keywords contained therein.

By treating this matrix in the classic sequence of multidimensional statistical analysis (simple match analysis and cluster analysis), you will get a partition in K document groups, homogeneous within them and heterogeneous among them: cluster analysis consists of grouping sub-texts according to their maximum lexical similarity, and thus characterizing them in terms of activities/events. The semantic field of expression of this homogeneity can be visualized by the proximity of the corresponding terms on the factorial plan. "The resulting K groups represent disaggregated classes and the word lists associated with each group define the theme or activity of the group" (Bolasco, 2013, Bolasco and Pavone, 2008).

[9] Iramuteq - *Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires* (R interface for Multidimensional Analysis of Texts and Questionnaires) http://www.iramuteq.org/

[10] Gephy (https://gephi.org/) is open-source and free.

[11] Factorial analysis was done with Spad.

[12] Event or activity.

Text mining: semantic network of co-occurrences to single out specializations

The last phase of our exploratory analysis consists in a word similarity analysis (Flament, 1962; Flament, 1981; Marchand and Ratinaud, 2012). The purpose of this analysis is to study the proximity and relationship between the elements of a set. In the textual case, therefore, it allows us to clarify the relationships between words, based on their co-occurrences within the fragments of texts. By co-occurrence we mean the existence of a pair of terms within fragments. Through the Iramuteq software, it is possible to perform this analysis based on an algorithm that measures how many times words are two-by-two "next" in the fragments of texts. As a result, a co-occurrence matrix is obtained, whose generic term contains the number of co-occurrences between each pair of words. A graph analysis of this matrix produces some synthesis of the relationships between the lexical units considered. In particular, "semantic communities" are generated through the modularity algorithm[13], which identifies sub-networks of keywords within the corpus and outlines the structure of the co-occurrence network. The graphical representation of co-occurrence networks is elaborated with Gephi (Bastian et al., 2009). In the co-occurrence matrix there are keywords with more than 10 occurrences, but the graph represents those above a specified threshold. The optimization algorithm for visualizing the graph is the Fruchterman-Reingold one. To each semantic sub-network is associated a colour. The vertices' size is proportional to the word's frequency and the edges' width indicate the word's co-occurrence strength.

Corpus and its parts

Before starting text analysis, a crucial step is defining the units of analysis, sub-corpora and fragments of texts that will help in a focusing on differences within relatively homogenous data. After a preliminary description of the entire corpus, criteria for specific pre-treatment and identification of sub-corpora will be outlined.

Preliminary description

The selected pdf files were converted into txt format (ANSI). Some basic figures on the four main categories of activities are returned in Table 3, with regard to the number of documents, by type and number of graphic forms (boxplot in Figure 1)

As expected, country reviews and reports are greatly different from the other categories of documents while workshops are made up of documents with a more heterogeneous distribution than that associated with plenary meetings.

Table 3 - Corpus main statistics: number of document and occurrences of graphic forms (quantile distribution, mean and total) by type of event/activity

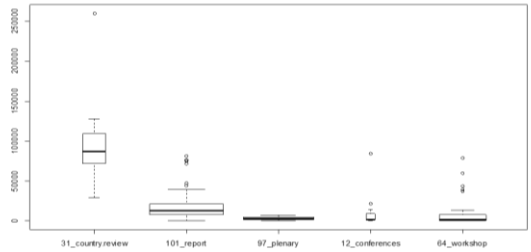
Type of event/activity	n. of doc.s	occurrences						
		min	25%	50%	75%	max	mean	Sum
country.review	31	29,208	71,776	86,374	109,546	259,731	92,489	2,867,156
report	101	362	8,071	13,384	21,535	81,174	17,607	1,778,257
conference	12	526	1,305	2,007	6,797	84,355	11,336	136,028
workshop	64	456	1,071	2,267	7,773	78,696	8,097	518,218
plenary	97	142	1,846	3,117	5,114	7,252	3,537	343,122

[13] See Fortunato and Hric (2016), for details on community detection algorithms and modularity.

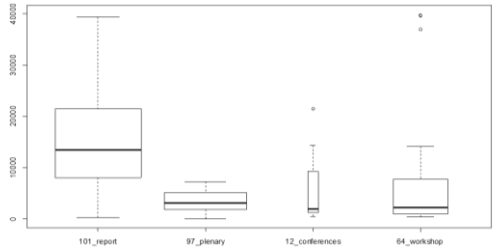
Figure 1 - Corpus: box plots of occurrences distribution of graphic forms, by type of event/activity

Boxplots' width: proportional to the number of documents in each category of documents (type of activity). Data elaborated with R

All types of documents



Only reports, plenary meeting, workshops, conferences
vertical axis truncated at 4000



Pre-treatment, semantic categorization, lemmatization

In order to identify the units of text analysis, a preliminary treatment is lemmatization of the graphic forms in the vocabulary, grouping together words in their singular and plural forms. A critical issue emerging in the pre-treatment of the Oecd documents is that Taltac2 has a limit in dealing with words of more than 99 characters (including dots and dashes). In an ordinary document this limit is never reached, but the documents under analysis have URLs that are not recognized as special entities but as words, often beyond the size that can be treated. This limitation has been dealt with by creating a mirror version of the corpus in which all the URLs, 3955, (both in the 'http' format and in the 'www' format) have been deleted. A complete version is available for further analysis dedicated to internet links[14].

Semantic tagging of documents has been carried out also by using the glossary of "terms on innovation and their synonyms" elaborated by the Oecd in 2015, which includes 1,065 keywords and 335 synonyms. The Oecd vocabulary has been useful to identify acronyms (listed in Annex 10).

To improve the semantic analysis, a second treatment is the application of a lexical-textual model to single out multi-words. A further treatment in semantic tagging to select the words for text analysis is the extraction of nouns, adjectives, verbs and adverbs, to implement a series of tools for the analysis of the graphic forms with content. Hereinafter, the words in this selected vocabulary are named as “keywords”.

Names of countries have been identified by using a dedicated resource and are specifically elaborated as supplementary words.

The goal of semantic tagging of documents is the grouping of documents according to their keywords in order to highlight specific language by categories of documents or the peculiar language of each document. A great heterogeneity across instances is a noise in the analysis that is reduced by excluding keywords occurring below a given threshold (set at 5). With regard to hapax, ad hoc analyses are explored.

Selection criteria in creating sub- corpora

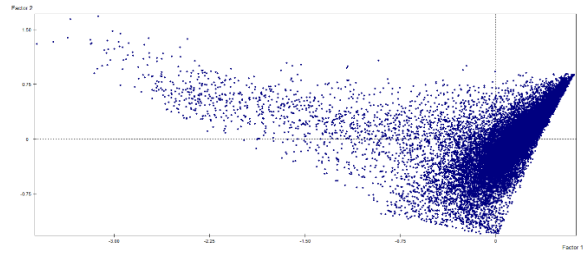
Having discussed the broad differences among the types of documents, and before starting the text analysis, it is useful to do a preliminary investigation on the lexicon of the entire corpus in order to select sub-corpora for the analysis. A first step is an analysis of the lexicon in the entire corpus to outline its main characteristics. This analysis is implemented by creating a matrix

[14] Such an analysis could be explored to reveal the set of underlying documents recalled by the links, the Oecd directorate or an external source that elaborated them, and to broaden the analysis to interlinking across themes, organizations and research groups.

Keywords × *Characteristics* (18,607×32) [15] analysed with a Factorial Analysis (Figure 2). The bulk of keywords is concentrated on the right side of the plan, where there are country reviews and reports. On the left side we find a more dispersed cloud, with plenary meeting at its far left. Hence, the type of activity characterizes Factor 1, underlining that documents of plenary meetings are very different from all the others in terms of contents and size. Factor 2 shows the polarization between agendas (very similar to plenary documents) and summaries (more similar to the documents characterizing workshops and conferences)[16]. The analysis referring to the entire corpus is not useful to investigate the changes of contents over time: the corpus is unbalanced by the 31 country reviews, mainly occurring in the last nine years.

Figure 2 - Factorial plan *f1-f2 Keywords* × *Characteristics*, all the documents

Lexicon



Categorical variables: type of activity, type of document, year

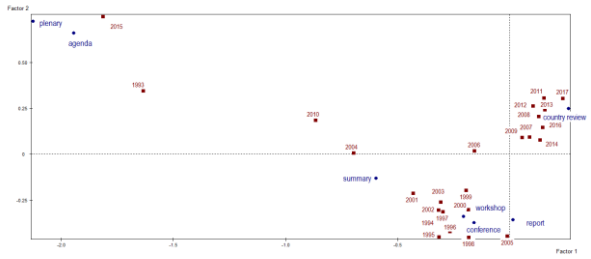
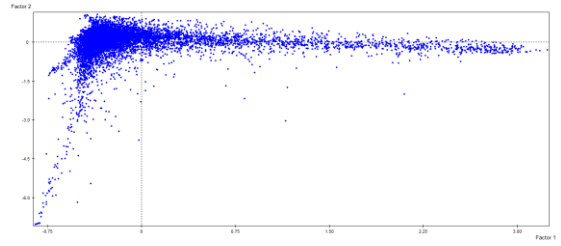
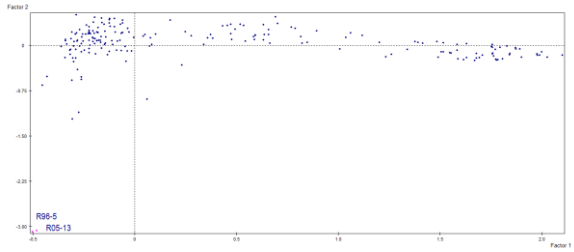


Figure 3 - Factorial plan *f1-f2 Events&activities* × *Keywords*: reports, plenary meetings, conferences and workshops

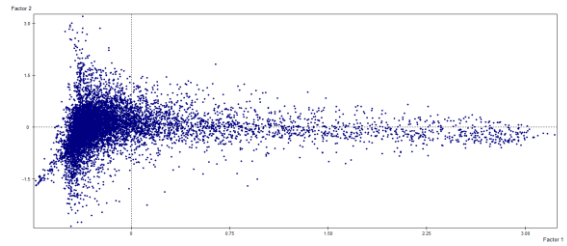
Lexicon



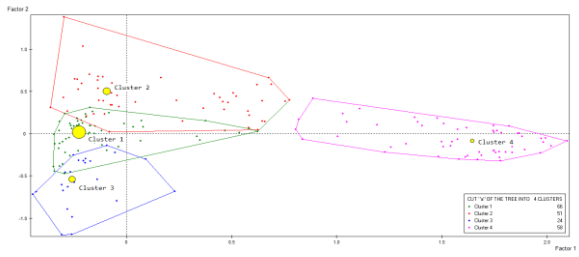
documents



Lexicon without the outliers



Clusters of documents without the outliers



Excluding country reviews from the analysis produces a sharp clustering among the documents. Hereinafter, documents are aggregated by event (for each plenary, or conference, or workshop) and by report (as a unit of activity). Figure 3 shows the results of the factorial analysis and cluster analysis on the matrix *Events&activities* × *Keywords* (201 × 9,510) for the remaining groups of documents: 97 reports, 49 plenary meetings, 47 workshops, 8 conferences. The analysis presented in the image on the left of Figure 3 shows a long tail on the left, due to two large documents (R96-05 and R05-13 that are two very large reports with an original lexicon). Excluding those outliers from the factorial and cluster analyses, we observe a clear parti-

[15] The set of characteristics under analysis are the years (25), the type of document (agenda, summary) and the type of event/activity (plenary meetings, conferences, workshops, reports, country reviews).

[16] See '/0 Corpus Completo AF/Tabulati AF.xlsx' with details of the lexicon and of the categorical variables characterizing F1, F2 and F3.

tion of the documents under analysis. Factor 1 still has a polarization between plenary meetings, on the right side, and all the other entities, on the left side. Factor 2 now shows the polarization between reports (at the top) and conferences and workshops at the bottom[17].

Table 4 returns both the number of entities (reports, conferences, workshops and conferences or plenary meetings) by cluster and the first ten entities characterizing the four clusters. They are characterized, respectively, by reports (cl #1), mainly by reports, but containing also workshops (cl #2), mainly by workshops and some reports (cl #3), exclusively by plenary meetings (cl #4).

Table 4 – Total number of entities (reports, conferences, workshops, plenary meetings) characterizing the four clusters of documents. For the first ten entities: distance to cluster's center and case identifier, by cluster

Cluster 1: 66 events		**Cluster 2: 51 events**		**Cluster 3: 24 events**		**Cluster 4: 58 events**	
Distance to cluster's center	Case identifier	Distance to cluster's center	Case identifier	Distance to cluster's center	Case identifier	Distance to cluster's center	Case identifier
0,00076	R05-15	0,01850	R07-7	0,00823	R02-15	0,00611	Plenary 28
0,00134	R03-11	0,02114	w-10-4	0,02903	w-95-1	0,01103	Plenary 20
0,00293	R11-2	0,02913	R14-12	0,03636	w-95-5	0,01242	Plenary 21
0,00309	R06-12	0,03147	R16-16	0,04431	R96-6	0,01326	Plenary 23
0,00632	R96-4	0,03451	R09-5	0,05234	R07-20	0,01384	Plenary 18
0,00680	R98-6	0,03928	w-94-1	0,05245	w-00-2	0,02274	Plenary 19
0,00787	RRD94-3	0,04556	R08-11	0,06817	R14-9	0,02394	Plenary 37
0,00843	R99-1	0,04664	R16-9	0,08452	C-97-Seoul	0,02427	Plenary 12
0,01230	R06-13	0,04911	R98-2	0,10840	w-98-1	0,02477	Plenary 25
0,01243	R96-2	0,04913	R13-3	0,14493	R95-2	0,02510	Plenary 14

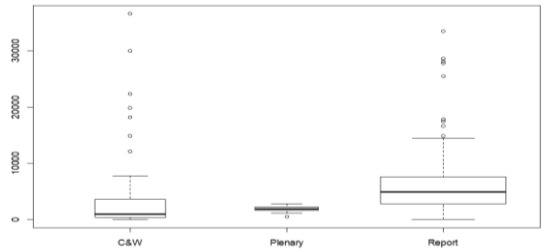
WPTIP events and reports are different with regard to their format and type of associated documents. Table 5 summarises the main statistics for events and reports, with regard to the occurrences of keywords identified in our analysis. Boxplots in Figure 4 highlight the large difference in the variety of lexicon of the three corpora.

Table 5 - Vocabulary and keywords in the three components of the original corpus

	n. of doc.s	**n. of events**	**Vocabulary [62,220]**		**Keywords [8,855]**	
			items	**Occorrences**	**items**	**Occorrences**
Plenary meetings	97	49	14,420	319,528	2,253	97,595
Workshops and conferences	64+12	47+8	25,683	626,457	4,436	232,671
Reports	101	97	44,119	1,677,141	7,842	634,029

Figure 4 - Corpus: box plots of occurrences distribution of graphic forms, by type of activity

Boxplots' width: proportional to the number of events/activity in each category of documents (type of activity).



Each plenary meeting is associated with two documents: the agenda and the minutes[18], both of which are schematic, with a number of keywords ranging from 543 to 2877 (mean 1992).

[17] See \Risultati AT\1 Analisi Corpus Report Plenary WorkConf\03 Risultati Cluster Analysis.xlsx for details on the lexicon and the documents characterizing each cluster.

[18] For the last plenary meeting, the 49th held in June 2017, the minute is not yet available, being prepared for the approval by delegates attending the 50th meeting on December 2017.

Reports generally have a table of contents and are relatively longer texts, with a greater variety of keywords than that of workshops and conferences.

Workshops and conferences have a more similar format, differing mainly for the range of topics (an issue that will be discussed in detail in the next section). In general, this group of events is associated with more than one document: an agenda, as a separate document or included in the minutes, minutes prepared by rapporteurs, but also concept notes issued in preparing the workshops and extended summaries prepared after the workshop[19]. This last feature explains why some of these events appear in the same cluster when also reports are considered in the same corpus. The overall number of keywords in workshops and conferences is 232,671, with a more dispersed distribution (see boxplots in Figure 4).

The results of this preliminary analysis support the decision to analyse, as separate corpora, the documents referring, respectively, to plenary meetings, to reports and to workshops and conferences (together). Hence, we created three sub-corpora - respectively with documents on plenary meetings, reports, workshops and conferences - contents of which are examined separately in the next section.

4. Results

For each of the three corpora mentioned above, in this section we present a set of descriptive statistics, the results of the factorial analysis and cluster analysis, the pattern of contents emerging from these analyses with regard to the specific lexicon of each cluster.

Plenary meetings

The corpus with documents of the 49 plenary meetings is the first one that we analysed to frame the 25 years of WPTIP. This corpus is composed of 2.253 keywords. First, a cluster analysis is implemented to single out a pattern of contents; secondly, a content analysis is carried out in order to frame the content; thirdly, the specific topics are presented. Subsequently, there is focus on the geographical space underlying the Oecd plenary meetings. Finally, a summing up on the plenary meetings is presented.

Clusters of documents revealing temporal patterns

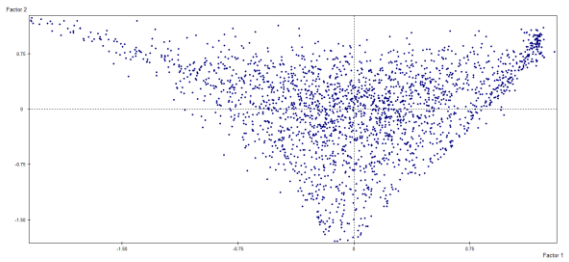
Based on the matrix *Events* × *Keywords* (49 × 2,253), Figure 5 returns the factorial plan *f1-f2* of the lexicon (left image) and the clusters of events (right image). The five clusters show a clear temporal pattern: 1993-1996, 1996-2001, 2001-2007, 2007-2014, 2015-2017. In Table 6 the periods are listed with the associated series of plenary meetings. In four clusters, plenary meetings of the same year belong to two different clusters. The cluster of the last period (cl #4) is explained by the specificity of the relative events captured by factor 3, as can be observed in the image of the factorial plan *f1-f3*, on the right in Figure 5.

Having identified the five clusters of plenary meetings showing a marked temporal pattern, we can analyse the content dealt with in those five periods. By focusing on the specific lexicon of each cluster, we observe that there is a mix of specific keywords not necessarily referring to topics related to innovation policy. This is due to our choice of creating the vocabulary from the corpus and, in the case of plenary meetings, texts refer to more than just the innovation topics. To identify which other topics could be singled out from text analysis, if any, we implement a "Similarity Analysis", within the whole set of keywords in each cluster.

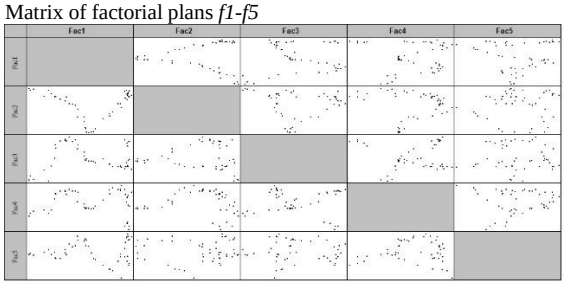
[19] Details on the number of documents by type of workshop is summarised in Annex 8.

Figure 5 - Plenary meetings: lexicon and clusters of documents

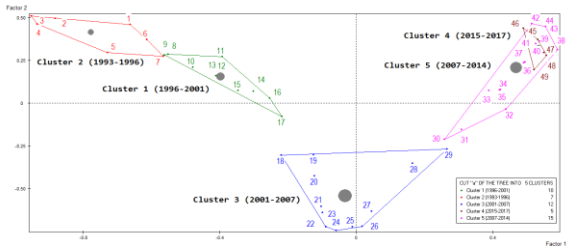
Lexicon in the factorial plan *f1-f2*



Matrix of factorial plans *f1-f5*



Clusters of plenary meetings #, in the factorial plan f1 f2



Clusters of plenary meetings #, in the factorial plan *f1-f3*

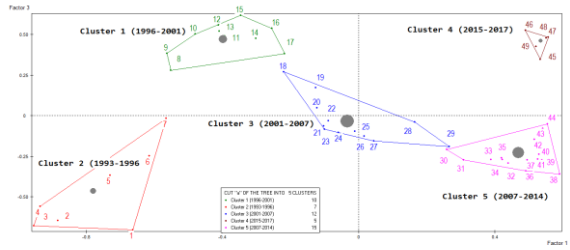


Table 6 - Clusters of plenary meetings according to the lexicon characterizing the documents (agenda and summary) associated with each plenary meeting

Cluster	period	plenary #	Number of meetings
2	1993-1996	1-7	7
1	1996-2001	8-17	10
3	2001-2007	18-29	12
5	2007-2014	30-44	15
4	2015-2017	45-49	5

Organizational discourse vs topics: pivotal words in the five periods

To implement the Similarity Analysis, all documents of each cluster are taken together in order to analyse all their fragments of texts as if each cluster were a single document. Similarity Analysis aims at identifying communities of words based on their co-occurrences in the fragment of texts. For each cluster, the modularity algorithm returns pivotal words as the ones linked to other words in the fragments of text. Communities aggregate words co-occurring with the pivotal word.

From this analysis on the plenary meetings, we obtain a clear perspective on the entire corpus of texts that appear to be made of two main parts: one referring to the organization of the Working Party on Innovation and Technology Policy, the other to the topics characterizing the discussion during the plenary meetings. This perspective emerges by analysing the five graphs in Figure 6. They represent, for each period, communities of words, identified by Iramuteq with the modularity algorithm and elaborated graphically in Gephy. In Figure 6, colours highlight different communities of words (only words with occurrences greater than a threshold, ranging from 12 to 30 according to the size of the cluster, are represented; the corresponding graphs with threshold greater than 4 are presented in Annex 12). Size of words is proportional to their occurrence in the corpus; width of the edge is proportional to the number of co-occurrences within fragments in the corpus.

In the first two periods, three words are pivotal in the discourse on organizational issues: *working group*, *secretariat*, *meet*. The pivotal words characterizing the topics of these meetings are: *technology*, *science*, *innovation*.

In the second period, *delegate* is the new word of the organizational domain (linking *secretariat* to *meet*), together with *action*, *report* and the verb "to present".

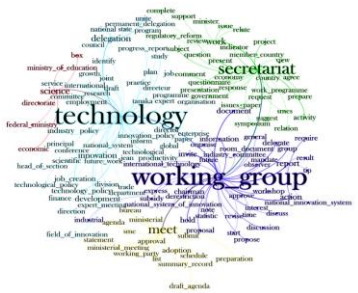
In the periods, 2001-2007 and 2007-2014, *delegates* are at the core of the organizational discourse, and *actions* are linked to them, and no longer to the secretariat, as in the first two periods. Another change occurs in the topics: *innovation* is now clustering issues discussed in the meetings as well as *technology*, and it is directly linked in the discourse mentioning the delegates. *Science* is still a subtopic of *technology*, but also *industry* assumes a more relevant part of the discourse.

In the last period, 2015-2017, the documents are very limited in number but significantly different in terms of their content specificity. *Policy* emerges as a pivotal word, clustering topics linked to the *secretariat*. In this period, the *secretariat* is again at the centre of the organizational discourse, strongly marked by its activity in the meetings: *to present*. Topics are clustered around *innovation*, while *science* has a relatively lesser importance.

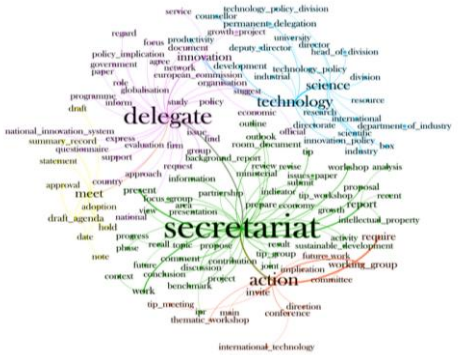
Figure 6 - Pivotal words and communities of words in the five sub-periods of plenary meetings (1993-1996, 1996-2001, 2001-2007, 2007-2014, 2015-2017)

Communities aggregate words co-occurring with the pivotal word. Colours highlight word communities. Size of words is proportional to their leverage degree in the corpus, width of the edges is proportional to the number of co-occurrences within fragments in the corpus. Modularity is implemented with Iramuteq. Graphs are represented with Gephy

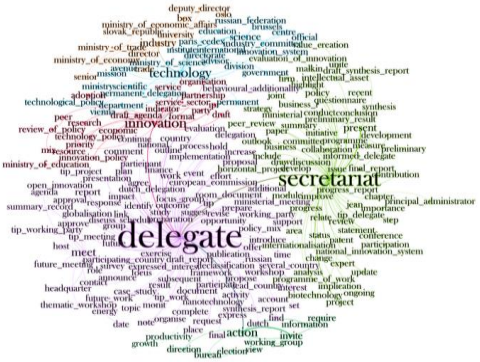
1993-1996: plenary meetings 1-7
[keywords with occurrences >15]



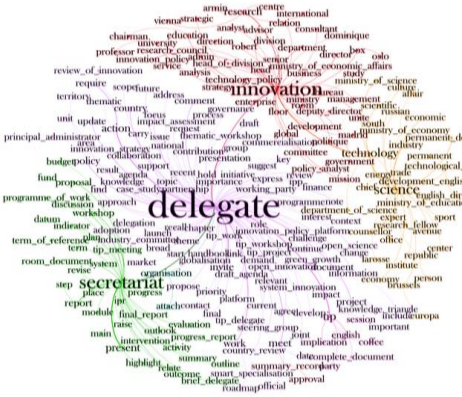
1996-2001: plenary meetings 8-17
[keywords with occurrences >30]



2001-2007: plenary meetings 18-29
[keywords with occurrences >25]



2007-2014: plenary meetings 30-44
[keywords with occurrences >25]



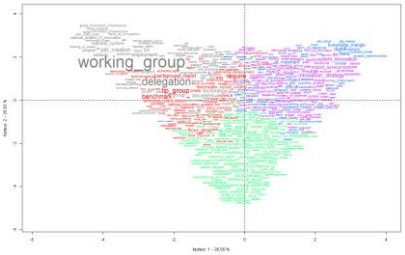
2015-2017: plenary meetings 45-49
[keywords with occurrences >12]

Issues on organization and on technology and innovation policy

The topics characterizing these periods emerge from the analysis of the specific keywords of each cluster. Figure 7 returns the same factorial analysis above discussed. Here the matrix *keyword* × *clusters* returns the factorial plan *f1-f2*, where keywords have different colours according to the cluster for which they are specific (size of words is proportional to their test-value (only words with the highest test value, 100% occurrence within the cluster and probability lower than 0.003 are represented).

Figure 7 - Factorial plan of the matrix *keywords* × *cluster*

Legenda: grey: 1993-1996; red: 1996-2001; green: 2001-2007; pink: 2007-2014; blue 2015-2017



A selection of keywords with the highest test-value and a relatively high number of occurrences in the cluster is returned in Annex 13[20]. For each period, we highlight first the organizational features and then the topics on technology and innovation policy[21].

_Issues on TIP organization

With regard to the organization, the following three main features emerge: the components of the working party itself, specific types of meetings, the tools adopted.

In the period 1993-1996, the *Working group, delegation, member country* are referred to in the plenary with other organization components: *chairman, observer* (countries), *rapporteur*. The activity is structured around a *mandate*, a *work programme*, a *work plan*. On which "the *working group agreed*". *Meeting of expert,* but also *informal meeting, ministerial meeting, summit* are the activities organized around *pilot case study, national case study, issues paper*.

In the period 1996-2001, the word "*TIP*" starts to be used, as in "*TIP group*". *Technological Working group* is also used. The *renewal* of the CSTP *mandate* is under discussion. A *group of national experts* starts to be organized to support TIP work. Beyond the *issues papers*, a range of new tools start to be mentioned: *reports (analytical/background), summary records, institutional mapping* (related to NIS work/activity/focus group). In creating a shared view, *draft* versions begin to circulate.

In the period 2001-2007, the structure of the organization is made of *TIP Working party/delegate /work/contribute* to reach a common goal, with *participating country/volunteer country* providing *volunteer contribution. There* seems to be a marked need for engagement. No

[20] The test-value returns the statistical measure of the specificity of the keyword within the cluster. For each keyword, other statistics are returned: the probability that the keyword does not belong to the cluster (it is the statistical measure of the position of the keyword in the tail over the probability distribution); its number of occurrences in the whole corpus and its share of the occurrences within the cluster. Words with a higher number of occurrences create links across more documents in the cluster and across clusters. We adopt the same rule of thumb proposed by Bolasco and Pavone (2007) and use a threshold of p-value, to discriminate keywords when computing their TFDF.

[21] A critical aspect in this analysis is the need to disambiguate some keywords that assume different relevance in the specific contexts in the same document. The agendas and the minutes of the plenary meetings contain information on addresses of people involved in the meeting (from the secretariat staff to keynote speakers). In order to keep the corpus as informative as possible, we decided not to delete this type of information because in principle an analysis of those names and affiliations could be of some interest in analysing the links across the meetings within the Oecd Directorates and outside. This decision left some noise in the corpus and in several cases the specific keywords had to be checked by controlling for concordances in the original text to disambiguate their relevant specificity.

specific type of meeting are introduced in this period, while new important tools are set in the TIP agenda: projects. *MONIT project* (proposed in 2002[22]), *monit group*, *tip case study* (on kibs, on evaluation of innovation policy, in 2006), *country study* (see list in Annex 14), *peer review*, *chapter/report/summary reports/synthesis*, and their various versions*: draft/revised/final publication*

In the period 2007-2014, the *innovation policy platform (ipp)* enters the stage (in 2012, and has become fully implemented in 2016). The *secretariat statement/oral* play an important part in the plenary presentations. In this period *staff changes* occur. It is only ten years ago that *TIP Steering group* entered into action more systematically (despite being active also in the previous periods (in the 8th,14th and 17th plenary meetings). Again, *revised mandate* is under discussion. What new types of meetings characterize this period? *tip policy roundtable* (on STI governance), but also *expert workshop*, *global forum* (on the knowledge economy, on green growth and sustainable development). There is a wide range of tools: *handbook* (STI policy platform, formerly 'handbook', 2010-11), *strategic P/PPs*, *strategic innovation areas/programme*[23], *tip project* (digital and open innovation, system innovation, knowledge triangle), *secretariat summary*, *revised term of reference* (project on open science), *piloting* (on the IPP), *review of innovation*, *expert review*, *technology study*, *scoping paper*, *annotated agenda.*

The last period of plenary meetings, 2015-2017, seems not specifically marked in terms of the working group structure, but rather by the opportunity of intervention of countries' delegates to express their interest in a specific issue/topic (*"country expressed interest in*" is the recurring expression). The focal tool of this period is the *blue sky conference* (held in Ghent, in September 2016). Other tools of this period are: *peer reviewer* (by other countries on country reviews), *keynote presentation*, *mapping* (of the governance of national policy research / of the policy mix / knowledge flow in innovation processes), but also road mapping, (as in "....approaching a transition to a low carbon future using road mapping"), *scoping* (the TOR on assessing the impact of policy mix / for the work on digitalization / scoping and framing), *focusing* (as in the expression *'countries suggested focusing on'*). *Powerpoint*, mentioned since 2014, becomes a standard for describing the format for presentations. Two terms are specific to this period: *important* and *different.* Those terms occur specifically in the cluster (grouping more than 40% of total occurrences of the entire corpus of plenary meetings). This specificity points to the fact that these documents are strongly marked by the linguistic style of the person in charge of drafting the summary. Although it is a non-secondary consideration when we analyse text, it will not be investigated in this paper.

_Topics on technology and innovation policy

Clusters are characterized by specific topics that we list below, in decreasing order of test value, by adopting a threshold on probability 0.0, % of cluster in frequency >80%. The selection excludes the terms that we have classified as "organizational", commented in the previous paragraphs.

1993-1996, **national system of innovation**

[22] " iii) proposal for monit (Monitoring and assessing DSTI/STP/TIP (2002) horizontal innovation policy) project to monitor implementation of NIS policy the norwegian delegation (m. remoe) presented a revised proposal for a new activity to be conduct under the aegis of the tip that would examine the implementation of horizontal innovation policy in OECD country. the proposal reflect the result of meeting and workshop held in April, august, and october 2002 that convene researcher from country that plan to participate in the project. tip delegate supported the proposal to proceed with the work under the aegis of the tip. several additional country expressed interest in participating." [excerpt from the summary record of the 22nd session] [post-treated text, all capital letters are removed]

[23] In many cases, 'strategic' refers to the contact of a person in a 'strategic division', these instances are not considered in the analysis.

national system of innovation, fiscal measure, financing innovation, subsidy, information infrastructure, global information infrastructure, technology: technology policy, technological strategy, technology research, technology flow, national technology programme, field of innovation, field of biotechnology, impact of technology, effectiveness of technology, government foresight exercise, intellectual property policy, intellectual property protection, international science, international technology, service enterprise, environmental protection, climate, ocean, employment, job creation, productivity, venture capital, information society, information technology, creativity, organisational change, regulatory reform.

1996-2001, **labour market**

national innovation system, growth project, labour market, technology labour market, growth performance, mobility of human resource, policy implication, technology foresight, policy evaluation, technology diffusion, Korean conference on international technology co-operation, globalisation of industrial research, development of innovation, technology development, role of competition, industrial competitiveness, collaborative R&D, commercialisation of research, technology incubator, spin-offs, innovative firm, technology programme, technology foresight exercise, innovative network, management of intellectual property, intellectual property, science system, globalisation of research, sustainable development, environmental technology, productivity growth, venture capital, mature industry.

2001-2007 **intellectual assets, behavioural additionality**

globalisation, monit project, behavioural additionality, review of policy, innovation module, policy mix, business R&D, innovation system review, evaluation of innovation, innovation performance, structural adjustment, economic performance, internationalisation, fuel cell, nanotechnology, biotechnology, pharmaceutical biotechnology, business environment, social policy, treatment of investment, financing of business, promotion of innovation, evaluation of innovation, health care service, open innovation, innovative entrepreneurship, service sector, IPR, patent regime, licensing practice, patented invention, patented knowledge, european patent office, exploitation of intellectual property, horizontal project, effect of government, development of science, globalisation indicator.

2007-2014 **knowledge network, social challenge**

system innovation, innovation strategy, impact assessment, smart specialisation, open science, governance, country review, form of innovation, ppps, knowledge network, global challenge, knowledge market, priority Setting, policy thinking, open access, social challenge, knowledge economy, collaborative mechanism, financial crisis, skill, technology evaluation, policy message, service innovation, green technology, green growth strategy, green innovation, marketing, helix, changing nature of innovation, innovation research, incentive, Frascati Manual, public procurement, research performance, source of growth, internationalization strategy, social innovation, research policy, tip policy, taxation, specialisation, industrial innovation.

2015-2017 **digital economy, system transformation**

heis, digital, knowledge triangle, circular, policy instrument, system transformation project, assessing, knowledge triangle project, system transformation, sustainability, business model, outcome, Oslo Manual, smart regulation, higher education system, HEI, knowledge transfer, megatrend, system transition, innovation policy mix, digital economy, technology trend, university system, research centre, metrics, scientific discipline, competitive funding, cluster policy, smart, student, Assessing, leadership.

The geographical space underlying the Oecd plenary meetings

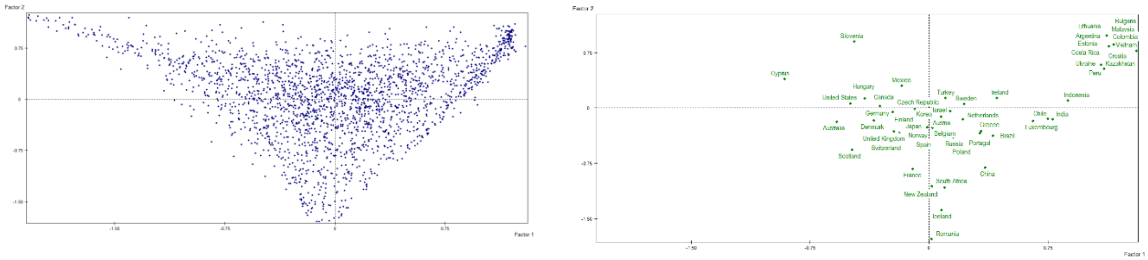
To conclude the text analysis of this corpus, we now consider the geography of countries emerging from the analysis of documents associated with the WPTIP plenary meetings. The complete list of countries named in the five periods, with the number of occurrences of their citations of countries, is in Annex 14[24]. Of the 53 countries cited in the plenary meetings, 27 are

[24] Two non-countries entries were detected by the automatic identification of names of countries: *Uruguay*, mentioned in connection to the Uruguay round (in 1994) and *England*, when referring to the Bank of England. Four other cases (*Georgia, Montenegro, Morocco, Panama*) were excluded because they referred to a person's name, an address or part of the name of an institution.

cited in all periods, some are cited only in one period[25]. Here we comment on the temporal pattern of citations by representing them in the factorial plan. Although excluded from the set of keywords used for clustering documents, the names of countries can be represented in the factorial plan of the lexicon as supplementary items in the documents. The position of each country in the *f1-f2* plan highlights their relative occurrence in the entire set of documents (see Figure 8).

Figure 8 - Plenary meetings: lexicon and name of countries in the factorial plan *f1-f2*

Lexicon (as in **Errore. L'origine riferimento non è stata trovata.**) Names of countries



Three main groups of countries are observed, and their temporal pattern is quite marked, in the same way as for the plenary meetings. In the period 2001-2007 (at the bottom of factor 2) we find: Romania, Iceland, New Zealand, South Africa, China; in the last ten years (on the right of factor 1), we find: Luxembourg, Chile, India, Indonesia, Peru, Ukraine, Kazakhstan, Croatia, Costa Rica, Vietnam, Estonia, Argentina, Colombia, Malaysia, Lithuania, Bulgaria. As said above (with details in footnote 25), the majority of countries are mentioned in various plenary meetings, and we can see that they occupy a position in the central part of the factorial plan.

From this analysis, we cannot associate countries to a specific activity, such as the reference to the country review on innovation policy or the intervention of a country delegate or her/his nomination in the board. If considered relevant for WPTIP, such details could be explored in a further stage of the analysis.

Summing up on plenary meetings

It is a clear fact that the structure of the twin sets of documents of the plenary meetings (an agenda and a summary, approved in the next plenary) are linked in an almost continuous thread from one to the next. Notwithstanding this chain and the need to keep delegates informed about the just-ended, the ongoing and the future events/topics/activities, we can identify five periods within the 25 years, for which we can provide an overview of WPTIP activities and topics.

WPTIP's activities are outlined within the mandates that express the goals of the TIP as a working group of CSTP. 'Working on what?' and 'how is it organized to work?' are the two questions around which the interpretation of text analysis has been developed in this section. This organizational perspective on contents of plenary meetings has been possible because our text analysis is grounded on a general set of keywords (identified by using a the process of semantic tagging described in Section 3), beyond the ones that could be specifically associated to innovation policy issues.

[25] The 27 countries named in the first period and also in all the following ones are: Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Italy, Japan, Korea, Mexico, Netherlands, Norway, Poland, Portugal, Russia& Russian Federation, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States, while Australia, Iceland, New Zealand and Slovenia are cited discontinuously in the following periods. Brazil, Cyprus, Luxembourg, Scotland, South Africa are first cited in the period 1996-2001; China, Estonia, India, Indonesia, Romania in the period 2001-2007; Argentina, Colombia, Costa Rica, Croatia, Lithuania, Malaysia, Peru, Ukraine, Vietnam in the period 2007-2014, and Bulgaria and Kazakhstan in the last period.

The structure and organization of the WPTIP has different characteristics over the five periods: in the first period the *working group* is at the core of the organizational issues while in the others, the *delegates* are at the centre of the scene of the action. Delegates' *actions* refer to decisions on future workshops, conferences and meetings, on WPTIP budget and projects, on the election of the board, on the approval of documents to be declassified, and terms of reference for the WPTIP's projects. *Presentations* made by the *secretariat* become an increasing part of the plenary meetings: they provide the linkages across the various directorates and the CSTP. A complementary organizational tool should be considered: the board made up of delegates and the chair of the WPTIP, selected among the proponent delegates. When was the board first nominated and convened? What was its role in defining the agenda and the activities to be proposed to the delegates? From agendas and minutes of the plenary meetings there are references to the 'Vice-chairs' and the role of members of the board[26] only in some of the years[27].

Over the 25 years, the organizational framework of the WPTIP changes: from the kick-off phase, to an interim transformation that starts in the second period and configures the structure observed over the third and fourth period. The last period is another turn in the organization, as if the secretariat regains the centre of coordination and action.

Although it may seem obvious, the interplay between countries' delegates and the Oecd Secretariat is the core of the organization but its effective denouement is not a given: it depends on the specific sets of competences within the Secretariat and on the engagement of delegates. A complementary analysis of people involved on both sides (delegates and secretariat) and the team in the board would ground this hypothesis, but in fact, documents of plenary meetings register what was planned, in the agendas, and what actually happened, in the minutes.

A complementary interpretation of organizational structure is the conduct dimension, i.e. the topics addressed by the WPTIP: they are not independent from the structure, moreover, their increasing complexity demands for a more structured organization (at least, this is what can be observed). A more general view would be provided by analysing the TIP's mandates, expressing the strategic perspective in which both the organizational structure and its conduct should be analysed.

Can we highlight continuity and change in the plenary meetings?

Over the five periods, the topics at the core of the plenary meetings move smoothly from *national systems of innovation* to *system transition*, passing through *patents* and *taxation incentives to foster R&D, environment, information society, social innovation, open innovation, incubator, megatrends, knowledge triangle, digital transformation.*

Notwithstanding their recognized relevance, some topics lose centre stage in the succeeding plenary meetings. Let us consider the case of behavioural additionality, mentioned in the meetings in the years 2003-2007. Its entry in the debate is introduced by Luke Georghiou's presentation (in the 21st plenary meeting, held in 2003). His seminal work on behavioural additional (Georghiou, 1997) refers to government financing business R&D. A topic at the core of a cross-country analysis, discussed in the Vienna workshop (2004) and in a pilot study with a focus on methodology. Behavioural additionality is last mentioned in 2007 (28th meeting)[28]. No further

[26] There is no reference to the term 'board' associated to vice-chairs: none of the 26 concordances of the term "board" refers to the board of TIP. In six fragments, it is part of the expression "to take on board" a specific issue/topic, in the remaining 20 is the board of other institutions.

[27] Meetings mentioning the nomination of Vice-chairs are the 9th, 30th, 34th, 37th, 38th, 42nd, 44th.

[28] " i) conclusion of the Seoul conference on evaluation the secretariat (Mr. Dirk Pilat) presented a short summary of the outcome of the Seoul conference on evaluation. among the finding were the need to improve the capacity for systematic evaluation; the international comparison of evaluation system; the need to deepen work on behavioural additionality and to examine the role of ex ante evaluation in priority setting . the secretariat will prepare a short summary of the conference for the next tip meeting".

mention is found in the following meetings. We shall return on this case as a general issue of evaluation in the next subsection on reports and workshops.

Structure and conduct, explored in this section focusing on plenary meetings, pave the way for an analysis of "performance" of a think tank like Oecd-WPTIP. This is the essential core of the entire analysis, which can be formulated in the following terms: In which directions, in which way (i.ee through which channels) and with what impact has WPTIP shaped the technology and innovation policy of OECD countries, and beyond? Which theories, which policy practices (of which countries) have oriented the performance of those countries?

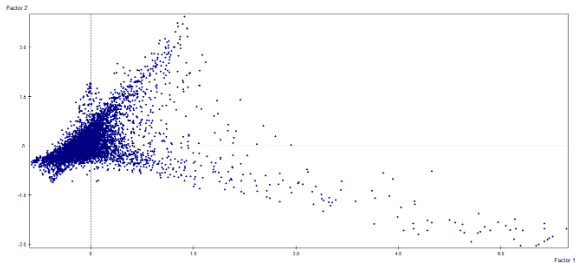
To answer these questions, our focus shifts first to reports and subsequently to conferences and workshops, arenas for a broad dialogue on innovation theories and policies.

Reports

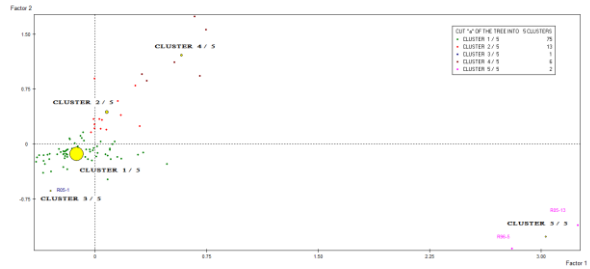
This corpus is made up of 101 documents. Since they are parts of the same reports, eight of the 101 reports have been joined, two-by-two. Hence, the total number of reports is 97. In analyzing the overall corpus, the lexicon is represented in the factorial plan *f1-f2* in Figure 9, where five clusters have been identified. Reports have a large number of unique documents. The factorial plan shows a polarization on factor 1 between a cl #3 on the left and cl #5 on the right: these clusters have respectively one and two documents. They are outliers by size (in terms of number of graphic forms and keywords). We first examine the specific topics of those clusters and then we discuss how to analyse the remaining reports. A summing up ends the presentation of reports that are the locus in which topics are presented in some detail.

Figure 9 - Factorial Analysis matrix *Reports* × *Keywords* *(97 × 7842)*

Lexicon *f1-f2* plan



Clusters of documents in the *f1-f2* plan



Clusters of reports that are outliers in the analysis

In cl #3 there is only one report published in 2005 [R05-1][29] on **energy technologies**, with a specific focus on innovation in **hydrogen fuel cells**. A selection of specific topics are list below, by adopting a threshold on probability below 0.001, % of cluster in frequency >40% and keywords listed in decreasing order of test value [48 out of 300 specific keywords are listed]):

> *hydrogen fuel cell, fuel cell innovation, combustion, fuel cell, hydrogen, oil, fuel cell technology, energy technology, gas, fuel, energy security, automotive, stationary, turbine, fossil, temperature, electrolyte, deployment, power generation, organizational network, energy sector, ballard, phosphoric, acid, environmental benefit, portable, gas technology, oxide, statoil, energy innovation, oil industry, membrane, entrenched, greenhouse gas emission, durability, carbon dioxide, ford, environmental quality, energy supply, gases, fossil fuel, international energy agency, shell, powered, lewis, role played, proprietary knowledge, automobile industry,*

The two reports in cl #5 refer to **taxes as a leverage of R&D** and were published in 1996 and 2005. The one produced in 1996 is specifically on "R&D tax treatment in Oecd member

[29] INNOVATION IN ENERGY TECHNOLOGY. SYNTHESIS REPORT.txt

countries" [R96-5][30], the other one, in 2005, is an international comparison on "tax treatment of business investments in intellectual assets" [R05-13][31]. A selection of specific topics are list below, by adopting a threshold on probability below 0.000, % of cluster in frequency >80% and keywords listed in decreasing order of test value [29 out of 320 specific keywords are listed]

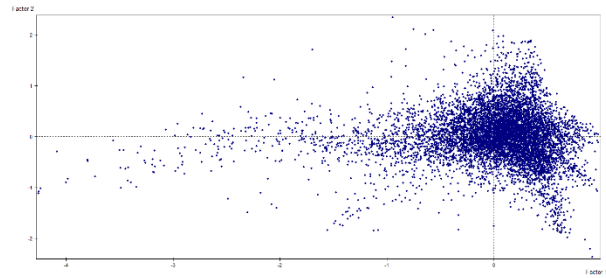
> *deductible, tax subsidy, tax treatment, depreciation, corporate income tax, deduction, allowance, depreciated, taxable income, tax system, income tax, useful life, current expense, investment allowance, organizational expense, present value of depreciation, capital asset, depreciation allowance, additional allowance, income tax rate, tax purpose, Calculating, tax provision, depreciable, generosity, incremental tax credit, capitalized, deductibility,*

The other clusters

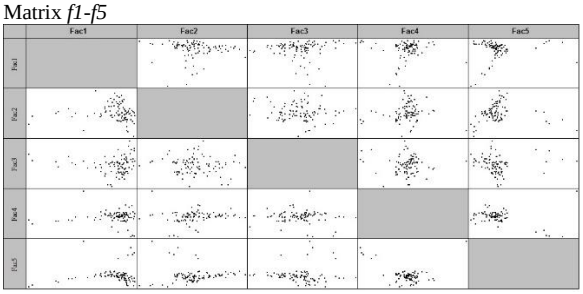
Given the peculiar characteristics in terms of size and content of those reports, to analyse the remaining corpus it has been necessary not only to exclude them from analysis, but also to remove their specific lexicon from the corpus.

Figure 10 - Factorial plan with *Reports* × *Keywords* (*94 × 7732*): reports

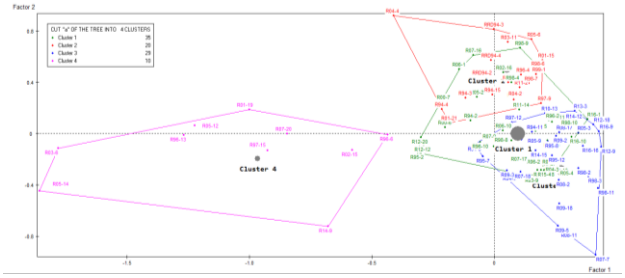
Lexicon without the outliers [in cl #3 and #5 above]



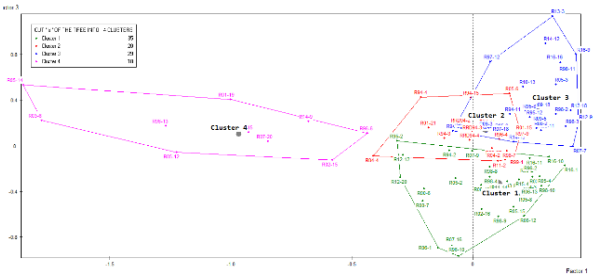
Matrix *f1-f5*



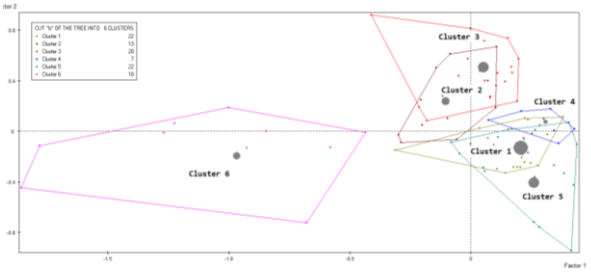
Clusters of reports without outliers [in the cluster 5 above], in the factorial *f1-f2* plan



Clusters in the factorial *f1-f3* plan: 4 clusters



Clusters in the factorial *f1-f3* plan: 6 clusters



The sub-corpus with 94 reports has been analysed by considering only keywords with more than 5 occurrences, a selection not necessary for any other of the sub-corpora of Oecd documents. With this selection, 7,732 keywords are under analysis and the lexicon in the matrix *94 reports by 7732 keywords* is summarized in Figure 10. The clusters of reports generated on the factorial analysis returns four clusters, balanced in terms of number of documents and size of

[30] This report is composed of two files that have been joined: R&D TAX TREATMENT IN OECD MEMBER COUNTRIES ..txt and R&D TAX TREATMENT IN OECD MEMBER COUNTRIES. A COMPARISON ..txt

[31] TAX TREATMENT OF BUSINESS INVESTMENTS IN INTELLECTUAL ASSETS. AN INTERNATIONAL COMPARISON.txt

reports. To provide a better perspective on those clusters, the matrix of one-by-one combinations *f1-f5* of factorial plans is presented in Figure 10. Let us call the new "cl_scr#", where "scr" reminds us that we are referring to the cluster elaborated on the sub-corpus of reports without the three outliers examined above. We observe that in the factorial plan *f1-f2* the polarization of reports is between the ones in cl_scr#4, on the left side of the f1 axis, and the cl_scr#1 on the right. Another polarization emerges along the f2 axis between cl_scr#2, at the top, and cl_scr#3, at the bottom.

By increasing the number of factors in analysis we obtain a more detailed clustering in six clusters, represented in the last image in Figure 10. This further clustering allows the identification of sub-groups of documents. To keep the description of topics more compact, we present the topics referring to the four clusters and to their sub-groups. Clusters are characterized by specific topics that we list below, in decreasing order of test value, by adopting a threshold on probability 0.0, % of cluster in frequency >80%.

Cl_scr#1, **education & research** [two sub clusters]

35 reports [41 of the 814 out of 1241 specific keywords are listed]

university, centre, student, million, heis, entrepreneurship, fund, education, expenditure, funding, research, spanish, higher education, total, researcher, research centre, knowledge triangle, icelandic, loan, SMEs, venture capital, programme, entrepreneurship education, projet, R&D, initiative, graduate, average, spin-offs, school, entrepreneurial, biotechnology, gdp, business R&D, partnership, regional, years, usd, par, competence centre, R&D intensity, financing, mission, science, chilean, pharmaceutical industry, young, pharmaceutical, increase, guarantee, academic, support, teacher, aaaa aaaa aaaa, BERD, company, knowledge transfer, participation, excellence, technological centre, dpi, commercialization

22 reports on education: HEIs, knowledge triangle, entrepreneurship education, entrepreneurial university
13 reports on specific research domains: pharmaceutical sector, biotechnology innovation system,

Cl_scr #2, **kibs & networks**

20 reports [41 of the 123 out of 969 specific keywords are listed]

intensive service, software industry, focus group, KIBS, leisure, service firm, software firm, *national system of innovation, service provider, manufacturing firm, service activities, distribution power, cluster analysis, isic, mile, forest industry, knowledge distribution, cluster study, market service, mining technology service, mobility rate, codified knowledge, health care service, retail trade, organisational innovation, innovative firm network, external service, stock of knowledge, informal network, service sector firm, bunt, service company, market transaction, kind of knowledge, innovative density, institutional mapping, software company, knowledge system, city district, external provider, pilot case study,*

Cl_scr #3, **system innovation & assessment & climate issues** [two sub clusters]

29 reports [40 of the 289 out of 1230 specific keywords are listed]

system innovation, impact assessment, transition, foresight, environmental technology, expert review, smart specialisation, priority setting, reviewer, evaluator, waste, evaluation system, environmental policy, social impact, emission, behavioural additionality, specific evaluation, technology evaluation, evaluation team, evaluation process, climate change, programme evaluation, incumbent, energy efficiency, evaluation of national, evaluation result, enquiry, impact of research, cleaner production, impact assessment exercise, demonstration project, policy cycle, well-being, socioeconomic, policy intervention, programme manager, recycling, transition management, meissner, programme management, policy intelligence, evaluation committee,

7 reports on: system innovation, system transformation, smart city, digital platform
22 reports on impact assessment, social impact, cleaner production

Cl_scr #4, **open science vs. patent**

10 reports [40 of the 156 out of 921 specific keywords are listed]

open access, exemption, open science, valuation, court, patent holder, patent law, research exemption, filing, patented invention, infringement, litigation, research tool, licensee, patent protection, patent pool, patent system, patentee, trip agreement, Options paper, statutory, foreground, patent licensing, patent application, dataset, subject matter, crown, jurisdictions, refusal, crown procurement contract, patent valuation, patent regime, data centre, gold, doctrine,

patent act, license income, patent portfolio, copy, patent infringement, economic asset, settlement, copyright law,

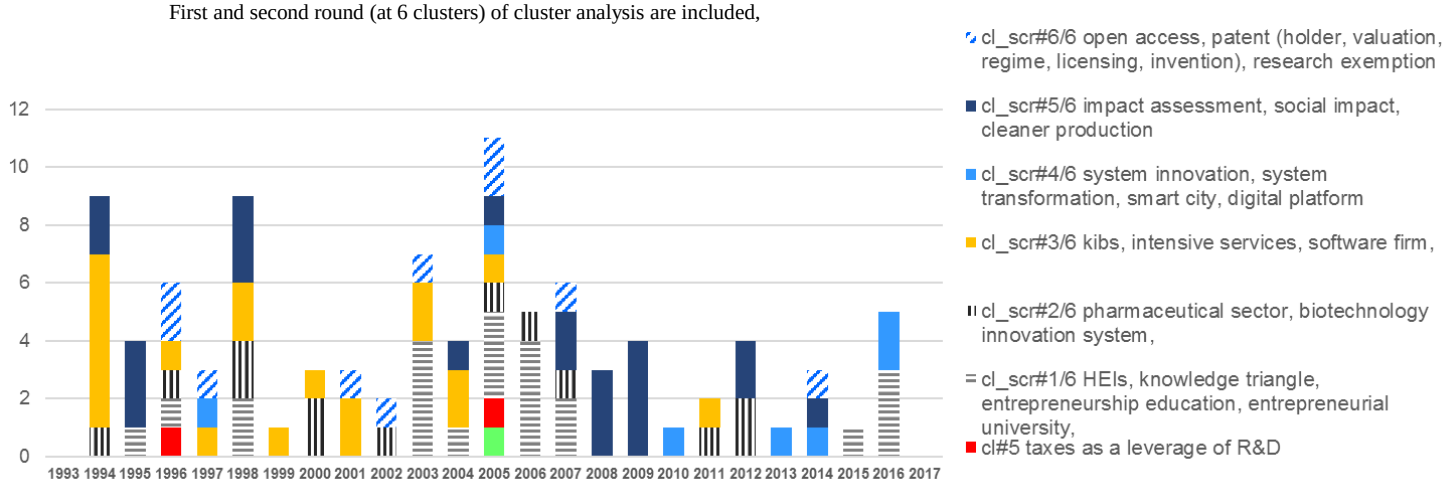
Summing up on the topics of reports

The topics represent a broad synthesis of the 97 reports produced in 25 years of activity, published by the WPTIP in the period 1994-2016: *energy technologies, taxes as a leverage of R&D, education & research, kibs & networks, system innovation & assessment & climate issues, open science vs. patent*.

Although a temporal dimension has no relevance in clusters' identification, we can observe a temporal pattern in the mix of topics addressed by the reports over the 22 years in which the 97 WPTIP's reports have been published. Figure 11 returns, by years, the six topics identified by the automatic analysis. In this sub-group of documents, year 2006, in which all the six topics are dealt with in the 11 reports published that year, is a divide: before that year the mix of topics was exploratory in all the topics, while after 2006 the key topics around which the analysis is developed are mainly system innovation & assessment & climate issues and education & research. A more detailed analysis could single out the differences within the contents in the various reports over the period under analysis.

Figure 11 -Number of reports by year and cluster (2+6) identified by automatic text analysis

First and second round (at 6 clusters) of cluster analysis are included,



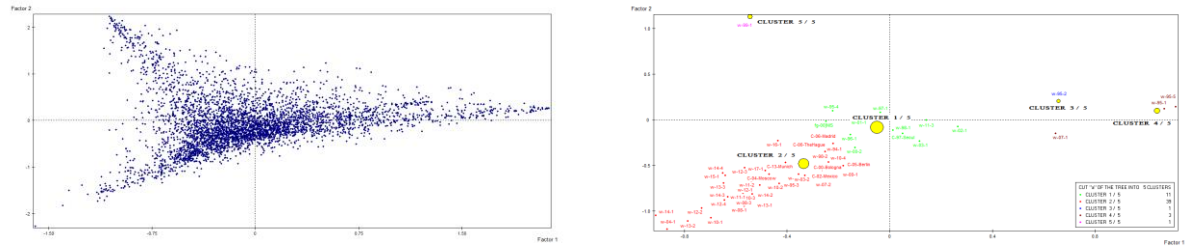
Workshops and conferences

An event-based categorization of the 76 documents associated with workshops, focus groups, seminars and conferences returns 55 events: 47workshops and 8 conferences[32]. The factorial plan in Figure 12, with the distribution of the lexicon and the clusters, singles out some outliers forming clusters with only one or few events. The distribution of lexicon shows a strong polarization on factor 1 between the largest part of events on the left (cl #2, with 39 events) and two small clusters of events on the right (cl #3, with 1 event, and cl#4, with three events). The other tail on the left top of the cloud characterizes the polarization along factor 2 with another cluster (cl#5) made up of just one event. A set of topics generally shared by many events characterises a cluster with 11 events (cl #1). In order to explore the contents of this corpus, we first analyse the three clusters of outlier events, then we re-elaborate the cluster analysis without those outliers. We conclude the analysis of this corpus by highlighting some critical issues emerging in the analysis.

[32] Conferences list is in Annex 7, workshops list is in Annex 8. The 47 workshops by type are summarized in Annex 9.

Figure 12 - Factorial plan *Events × Keywords (55 × 4436)*: conferences and workshops

Lexicon documents



Clusters of workshops that are outliers in the analysis

In order to summarize the specific topics of clusters #3, #4 and #5, we refer to the frequency of characteristic keywords of each cluster, by adopting a threshold on probability equal to zero, % of cluster in frequency >90%. Keywords are listed in decreasing order of test value.

cl_#3, **venture capital**, 1995

1 event: **w-95-2** [the resulting 42 out of 353 specific keywords are listed]

informal venture capital, business angel, venture capital firm, mason, business angel network, venture capital fund, venture capital market, venture capitalist, informal investor, private equity, capital gain, participation company, investment opportunity, pension fund, venture capital company, european venture capital, risk capital, informal investment, technology rating, institutional venture capital, private investor, growth company, appraisal, unquoted company, nasdaq, business introduction service, private individual, formal venture capital, gaston, investee company, potential investor, oakey, private equity investment, bygrave, succession, Venturing, riding, friend, backing, investee, chapman, unquoted,

cl_#4, **fiscal measures**, in 1995 and 2007

3 events: w-95-1, w-95-5, w-07-1 [the resulting 64 out of 509 specific keywords are listed]

tax credit, tax policy, rebate, deduction, fiscal measure, tax concession, calendar year, concession, proprietary technology, tax withholding organisation, refund, tax authority, leyden, tax liability, corporate tax, guilder, outlay, atp, tax scheme, technology element, application form, national insurance contribution, incremental tax credit, generic technology research, increased deduction, BERD, depreciation, wage costs, fiscal unit, volume mechanism, maximum rebate, main project, tax incentive scheme, development spending, tax credit system, generic research, provisional reduction, research joint venture, incremental mechanism, auditor, tax subsidy, calendar quarter, incremental system, tax credit mechanism, eisner, accelerated depreciation, neutrality, bauer, depreciated, increment, paperwork, reclassification, previous years, tax measure, expensed, submitting, taxable, gauge, enclosed, calibration, omnibus, stoneman, refundable, stone,

cl_#5, **HEIs**, in 1999

1 event: w-99-1 [the resulting 65 out of 445 specific keywords are listed]

PhD, occupation, field of study, qualification, earnings, unemployed, university graduate, KIBS, gender, PhD graduate, postgraduate, fellowship, immigration, occupational, post-doctoral, university council, tenure, enrolment, baccalaureate, vocational, engineering graduate, main indicator, technology personnel, tomlinson, technology labour market, job shift, teacher, graduate student, find employment, labour market performance, university degree, Extended, tertiary, inactive, cohort, subgroup, salaried, isco, graduate survey, insecure, population census, technology worker, medical science, university education, skill shortage, secondary education, enrolled, field of information, staffing, chemist, diploma, equivalence, engineering field, engineering degree, intake, physicist, upskilling, ISCED 6, grad, jari, hired, earning, predicting, categorisation, corsi,

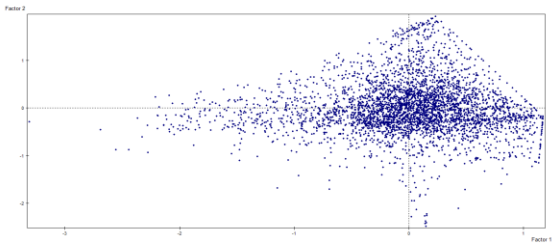
The other clusters

The factorial analysis and the cluster analysis on the remaining events, presented in Figure 13, produce a more informative perspective on the topics of the other 50 events, grouped in four clusters. Let us call the new "cl_scw#", where "scw" reminds us that we are referring to the cluster elaborated on the sub-corpus of events without the three outliers examined above. On factor 1, there is a polarization of the lexicon between cl_scw#5 (made of 25 events), on the left,

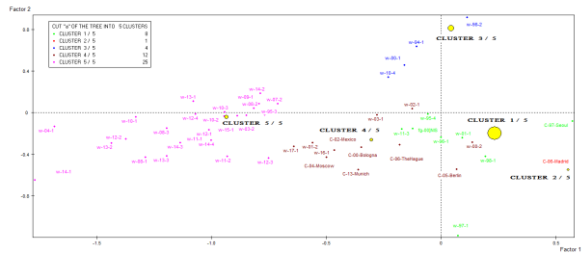
and cl_scw#2 on the right, with only one event, the Madrid conference in 2006. Two smaller clusters characterize the polarization along factor 2: cl_scw#3 (4 events), at the top, and cl_scw#4 (12 events) at the bottom. The lexicon of cl_scw#1 (8 events) occupies the central part of the plan.

Figure 13 - Factorial plan *50 events* × *keywords*: conferences and workshops

Lexicon without the outliers



Clusters of documents without the outliers [in the clusters cl 3-4-5 above]



The five clusters are characterized by specific topics that we list below, in decreasing order of test value, by adopting, as above, a threshold on probability 0.0, % of cluster in frequency >90%.

cl_#scw1[33], **incubator, networking, science park, technology diffusion programme**

8 events: [45 of the 145 out of 353 specific keywords are listed]

incubator, technology incubator, technology diffusion, foreign, trip, networking, korean, business incubator, technology diffusion programme, mobility rate, wto, employee, science park, logistics, framework programme, export, international technology, computer, migration, fixed asset, korean firm, knowledge market, tenant firm, aaaa aaaa aaaa, venture, labour market, samsung, harmonization, intellectual property protection, enforcement, technology park, foreign company, dispute, package, flexible, foreign firm, technical assistance, innovation centre, liberalisation, located, cluster policy, innovative firm, foreign investment, import, incubation,

cl_#scw2, **research exemption**

1 event: **C-06-Madrid**[34] [the resulting 30 out of 575 specific keywords are listed]

research exemption, research tool, csic, exemption, patented invention, spanish, patented, madrid, statutory, law school, research purpose, hitotsubashi, research input, commons, subject matter, amplification, rosenthal, gene, bird, asher, exclusivity, prize, clinical, disclosed, sociology, instituto, merck, visiting, derivative, homogeneous,

cl_#scw3, **environmental issues**

4 events: **w-94-1, w-98-2, w-00-1, w-10-4** [the resulting 40 of the 95 out of 575 specific keywords are listed]

environmental, foresight, technology foresight, sustainable development, environmental technology, water, waste, carbon, emission, foresight exercise, realisation, thesis, forecast, carbon dioxide, thematic field, critical technology, environmental problems, sustainable technology, cleaner, forecasting, solar, pollution, Cleaner, recycling, environmental issues, efficient technology, environmental policy, renewable energy, separation, wind, energy technology, soil, environmental impact, prediction, environmental performance, foresight activities, environmental sustainability, environmental challenge, atmosphere, ecological,

cl_#scw4, **open innovation, evaluation, econometric studies, behavioural additionality**

12 events, of which, 6 workshops: w-00-2, w-01-2, w-02-1, w-03-1, w-16-1, w-17-1; and 6 conferences C-00-Bologna, C-02-Mexico, C-04-Moscow, C-05-Berlin, C-06-TheHague, C-13-Munich [the resulting 14 out of 575 specific keywords are listed]

open innovation, business R&D, behavioural additionality, additionality, valuation, econometric, econometric study, tax incentive, leverage effect, pilot study, knowledge sharing, fier, boris, silo,

[33] In this cluster, we observe the most distant lexicon: on average specific terms of w-97-1 are distant almost 3 times from the centroid of the cluster.

[34] Conference on "Research Use of Patented Knowledge", held in Madrid, Spain, on 18-19/05/2006.

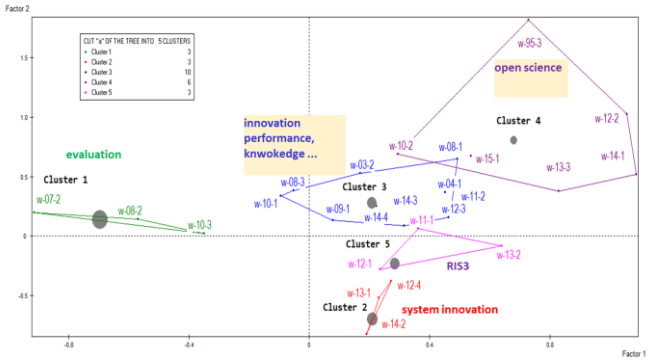
cl_#scw5, **system innovation, smart specialisation, priority setting, open science, global challenge**

12 events: w-95-3, w-03-2, w-04-1, w-07-2, w-08-1, w-08-2, w-08-3, w-09-1, w-10-1, w-10-2, w-10-3, w-11-1, w-11-2, w-12-1, w-12-2, w-12-3, w-12-4, w-13-1, w-13-2, w-13-3, w-14-1, w-14-2, w-14-3, w-14-4, w-15-1 [the resulting 33 out of 575 specific keywords are listed]

system innovation, smart specialisation, priority setting, peer review, ipp, handbook, conference centre, prioritisation, governance arrangement, outlook division, tour, smart specialisation approach, governance mechanism, keenan, marshall Building, keynote presentation, principal administrator, leuven, enquiry, expert presentation, head of division, reviewer, presenter, Setting, innovation policy platform, federal ministry of science, country case study, hein, monit, taxonomy, vlaanderen, interdepartmental, Enabling,

We observe that cl_scw#5 is made of 25 events within a more disperse cloud. Information on these events are presented in detail in Figure 14. Actually, five main topics characterize this cluster, but in two of them, comprising 16 out of 25 events[35], information refers only to the agendas. The remaining three, made of three events each, are respectively about :*evaluation*; *system innovation, transition, green growth, technological system, enabling technology; smart specialisation (strategy/approach), regional development*. Such level of detail confirms the general topic assigned to identify the overall cl_scw#5, but underlines a critical feature of consistency of this corpus that requires a more complete set of documents, not only because some events are missing in our database, but also because in many cases there are only agenda of the events.

Figure 14 - Factorial plan *25 events* × *keywords*: conferences and workshops in cl_scw#5_



Summing up on the topics of workshops and conferences

The topics addressed in the 55 events (47 are workshops and 8 conferences) organized by WPTIP since 1994 cover a wide range of topics: venture capital; fiscal measures; HEIs; incubator, networking science park, technology diffusion programme; research exemption; environmental issues; evaluation, econometric studies, behavioural additionality; system innovation, smart specialisation, priority setting, open science, global challenge.

Also in the case of workshops a temporal dimension has no relevance in clusters identification, but also in this case we can observe a temporal pattern in the mix of topics addressed by the 55 events.

Figure 15 returns, by years, the eight topics identified by the automatic analysis. Year 2007 is a divide: before, topics cover more aspects than after 2007, but this may be a shift in the paradigm that becomes inclusive of more topics under the same general label: "system innovation, smart specialisation, priority setting, open science, global challenge". The evaluation issue, and in particular the topic of behavioural additionality, is centre stage in the years 2000-2006, be-

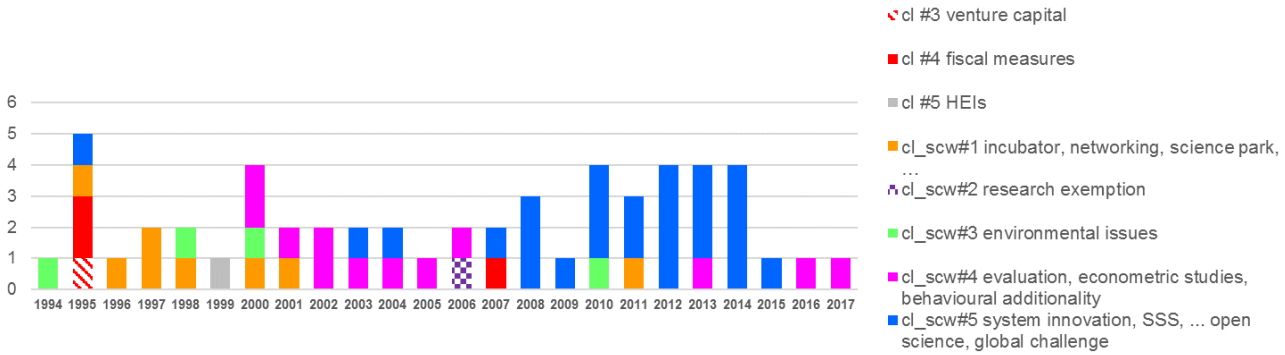
[35] The topics are "innovation performance, knowledge triangle, knowledge network, knowledge transfer, monit" (10 events) and "open science, open access, ipp" (6 ecents).

coming again a key topic in the last two years. HEIs seem to have been a minor focus, but this is due to a lack of documents of the most recent events, one of them being the high-level meeting, held in Paris in September 2016, on the role of HEIs in enhancing innovation.

In general, incompleteness of this corpus is critical in terms of missing information on events, also because in many cases there are only agenda of the events, thus affecting the heterogeneity within the corpus.

Figure 15 - Number of events (workshops and conferences) by year and cluster identified by automatic text analysis

First and second round of cluster analysis are included



5. Topics in a temporal perspective

The ongoing mandate of WPTIP expires in 2019. Our contribution, looking backward over the past 25 years, is a way to discuss what could characterize a mandate after 2019.

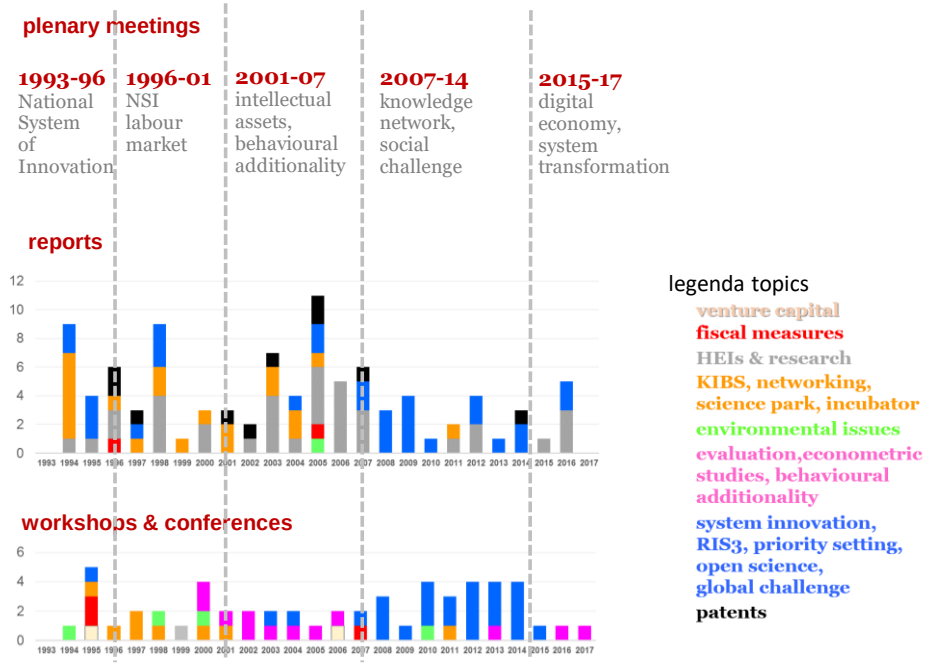
Text analysis has provided an overview of the main issues addressed in 25 years of activities, at different level of granularity. The five periods emerging from the analysis of the plenary meeting are punctuated by analyses presented in reports and debated and disseminated through a wide array of events (from seminars, workshops to conferences). At a greater granularity, we can single out sub topics, as shown in Figure 16.

We were unable to single out the documents containing the mandates of WPTIP. Those documents would provide an opportunity to compare the plan of activities/goals with the themes and actions undertaken. Consistency is generally expected, but the divergence would help in understanding what actions prevail and why.

The year 2007 seems to be a divide in the main discourse developed in 25 years. We interpret this in terms of a shift in the perspective on innovation policies, which begin looking at networks (across PRIs, business companies, users, civil society, public institutions) as a source of creation of knowledge, but also to systems transformation, with many and different topics related to it.

A further step in the analysis will focus on a systematic review of the evaluation methods promoted so far by WPTIP (as they emerge from the TIP's documents), in order to outline which are the objectives and tools proposed in the different domains of policies. If system transformation is the new perspective that frames the last period of TIP's activities, it needs methods that are more complex to assess systemic impacts of innovation policies, whose roots are not necessarily within the ones so fare proposed.

Figure 16 - Temporal pattern of WPTIP' topics by periods of plenary meetings and topics of reports and events by year



References

Bastian, M., Heymann, S., Jacomy, M., others, 2009. Gephi: an open source software for exploring and manipulating networks. Icwsm 8, 361–362.

Benzécri, J.-P., 1992. Correspondence analysis handbook, Statistics, textbooks and monographs. Marcel Dekker, New York.

Bolasco, S., 2010. TaLTaC 2.10: sviluppi, esperienze ed elementi essenziali di analisi automatica dei testi. LED, Milano.

Bolasco, S., 2013. L'analisi automatica dei testi: fare ricerca con il text mining. Carocci, Roma.

Bolasco, S., Pavone, P., 2007. Automatic dictionary and rule-based systems for extracting information from text, in Data Analysis and Classification. p. 189-198. BERLIN-HEIDELBERG: Springer

Bolasco, S., Pavone, P., 2008. Multi-class categorization based on cluster analysis and TFIDF. JADT2008, vol 1, 209-218.

Flament, C., 1981. L'Analyse de Similitude, une Technique pour les Recherches sur les Représentations Sociales. Cah. Psychol. Cogn. 375–395.

Flament, C., 1962. L'analyse de similitude. Cah. Cent. Rech. Opérationnelle 63–97.

Fortunato, S., and Hric, D., 2016. 'Community Detection in Networks: A User Guide'. Physics Reports 659 (2016): 1–44. https://doi.org/10.1016/j.physrep.2016.09.002.

Georghiou, L., 1997. Issues in the evaluation of innovation and technology policy, in: Policy evaluation in innovation and technology: towards best practice, chapter 3, OECD, Paris.

Marchand, P., Ratinaud, P., 2012. L'analyse de similitude appliquée aux corpus textuels: les primaires socialistes pour l'élection présidentielle française (septembre-octobre 2011). Actes 11eme Journ. Int. D'Analyse Stat. Données Textuelles JADT 2012, 687–699.

Pavone, P., 2010. Sintagmazione del testo: una scelta per disambiguare la terminologia e ridurre le variabili di un'analisi del contenuto di un corpus, in: Bolasco, S., Chiari, Isabella, Giuliano, Luca (Eds.), Statistical Analysis of Textual Data. Proceedings of 10th International Conference Journées d'Analyse Statistique Des Données Textuelles 9-11 June 2010 - Sapienza University of Rome. Led on Line, Roma, pp. 131–140.

Pavone, P. Russo, M., 2017. Clusters of specializations in the automotive supply chain in Italy.

An empirical analysis using text mining, DEMB Working Paper Series 116, http://merlino.unimo.it/campusone/web_dep/wpdemb/0116.pdf

Annexes

Annex 1 - Documents received from OECD

Three sets of documents have been received from the Oecd STIP Directorate.

16th October 2017

The first set of documents was received on 16th October 2017. This archive of 944 files is organized in 47 folders and subfolders with the following ramification.

Two main folders: TIP-List and TIP-PDF.

- **The folder "TIP-List"** contains only one file "`DSTI-STP-TIPdocument list_05Oct 2017_Text_Mining_Selection.xlsx`" with two sheets:
  - 'Reports for analysis', with 116 record, and the following fields: (order number), COTE; FINAL REPORTS; Summary; Security; Date
  - 'Agenda_summ. reports f. analys', with 160 records; and the following fields: (order number), COTE, AGENDA / SUMMARY REPORT, Summary, Security, Date
- **The folder "`TIP-PDF`"** contains all the files extracted by the Secretariat but only those listed in the file "`DSTI-STP-TIPdocument-list_05Oct2017_Text_Mining_Selection.xlsx`" have to be considered for text analysis. It is made up of two main -folders:

TIP-Official-Documents:

- with 24 folders, one for each of the years 1994-2017
- with a folder "DSTI-STP-TIP-AH", with 16 sub folders, one for each of the years 1998-2001 and 2004-2015
  - TIP-Publications: with 31 files

Further details on this Oecd archive were sent by Caroline Paunov on 17th October

- **Regarding acronyms**, the important ones to look out for are A = agenda, M = summaries of agendas, AH/A or AH = workshop agendas (note, however, that we stopped using AH and have some workshop agendas just with simple codes). All other documents (basically the reports) will be labelled either REPORT (used in the past) or just simply numbers (i.e. the TIP(yyyy)X) where numbers simply start at 1 to the last document generated, following generally the order they appear on agendas).
- You can ignore the **security** mentions, all documents we will look at are declassified i.e. can be made public as they are generally approved versions. .
- As to **associated metadata**, we generated the information by going through the documents. I am afraid we may not have more metadata available, but I will check and get back should there be good news.
- Regarding the **tagging and vocabulary**, I leave it to Michael to get back on this.
- As to **participant information**, we would suggest not to focus on this topic at this point in time.

1st November 2017

The file with the Oecd vocabulary of innovation topics ("IPPvocabulary terms and synonyms 19-Nov-2015") was received on 1st November 2017, sent by Andrés Barreneche, Policy Analyst at the Oecd Directorate for Science, Technology and Innovation.

The vocabulary has been used by Oecd "to tag and semantically link documents and other types of content in the Innovation Policy Platform (IPP). The table includes the main topic name (Display Label column) along with synonyms (Alternative Label column). The table also includes an URI for each topic (Concept ID column)".

Although each Oecd document in the IPP has been tagged, the Oecd documents received for text analysis have no such tags.

3rd November 2017

The extraction from the pdf directories of documents to be analysed was made manually because the list provided by Oecd, in the excel file contained the title of the document - in the column 'Agenda & summary' and in the column 'Final Report' - not the name of the file. In many cases they differ significantly (moreover, for half of the documents, names are duplicated in that cell).

By checking the file in the subfolder by year and in the subfolder by AH and year, and having in the name of the file some indication of related to the content, we were able to make the extraction. The overall checked list is made up of 258 documents (from the two lists "agenda&summary" + "Final reports").

From the excel original Oecd list, we have excluded the following 5 files because we chose the most recent version available in the directory (and in the excel list)

DSTI/STP/TIP(96)10
DSTI/STP/TIP(96)10/REV1
DSTI/STP/TIP(2004)4/FINAL
DSTI/STP/TIP(97)15/REV2
DSTI/STP/TIP(96)13/REV1

But, not all the files mentioned in the excel original Oecd list are available in the folder of pdf files. In particular we singled out 20 + 1 files that should be added in the corpus

- the following 9 are missing
  - DSTI/STP/TIP/RD(94)4
  - DSTI/STP/TIP/RD(94)3
  - DSTI/STP/TIP(94)17
  - DSTI/STP/TIP(95)4/REV2
  - DSTI/STP/TIP(96)3
  - DSTI/STP/TIP(2002)16/REV1
  - DSTI/STP/TIP(2006)12/FINAL
  - DSTI/STP/TIP/SFRI/AH(2008)1/REV1
  - DSTI/STP/TIP/RIHR/A(2012)1/REV1
- the following 9 are available in the French version and the English version would be needed
  - DSTI/STP/TIP(97)1
  - DSTI/STP/TIP/M(96)2
  - DSTI/STP/TIP(94)14
  - DSTI/STP/TIP(96)7
  - DSTI/STP/TIP(96)7/REV1
  - DSTI/STP/TIP(97)4/REV1
  - DSTI/STP/TIP(96)2
  - DSTI/STP/TIP/M(98)2
  - DSTI/STP/TIP/M(96)1
- the following is available in French, in a previous version: the English version is needed
  - DSTI/STP/TIP(96)3/ADD1
- the following is the cote of the first WPTIP meeting in 1993. It is mentioned in the agenda of the second meeting, but not available in the list or the folder of pdf files DSTI/STP/TIP/M(93)1
- in addition to those 20 files, possibly, the first agenda should be found and made available

The additional files were received on 3rd November; 18 of them have been included in the Corpus.

Annex 2 - Database schema of the Oecd WPTI documents created by Unimore

ID_unimore	*progressive number*		
source	*values*		count ID_unimore
	FolderOecd_TIP-Publications		31
	not in the initial OECD TIP-list_received_3.11.2018	2	
	OECD_TIP-list_Agenda_summ. reports f. analys.	152	
	OECD_TIP-list_Agenda_summ. reports f. analys._received 3.11.2017	6	
	OECD_TIP-list_Reports for analysis		97
	OECD_TIP-list_Reports for analysis_received 3.11.2017	10	
	other files from_TIP-Official-Documents		7
n.ord. OECD original	*progressive number in the original source from Oecd*		
year_fromCote	*year extracted from the COTE*		
YEAR of the event	*In the case of 'country reviews' and 'reports', 'year' is the one of the document as it is defined, respectively, in the publication and in the Cote. In the case of 'events' it refers to the year in which the event occurred. It could be different from the one in which the documents linked to the events were published (eg., for events occurring at the beginning of the year, the 'agenda' could be issued in the year before the event; for events at the end of the year, the 'summary' could be is-sued in the year after the event)*		
file_name	*file name converted in txt (note that in one case it has been truncated manually, because it was too long)*		
type of the event	*values*		count ID_unimore
	conference		12
	country review		31
	plenary		97
	report		101
	workshop		64
IDcode#	*values*	different IDcodes	count ID_unimore
	for countries: it is a code consisting of yy_Country	31	31
	for reports: Pasquale will assign a code to each report	98	101
	for the plenary meetings: it is a progressive number from 1 to 49	49	97
	for conferences: "C_yy_country"	8	12
	for workshops: w_yy_#"	45	61
	for the focus group: "FG_yy)topic"	1	3
yy)# form cote	*year and number extracted from the COTE*		
Type of document	*values*		count ID_unimore
	agenda		86
	country review		31
	report		101
	summary		82
	summary&agenda		5
"type of workshop"	*values*		count ID_unimore
	ad hoc meeting		10
	conference_Berlin		1
	conference_Bologna		1
	conference_Madrid		2

	conference_Mexico	1
	conference_Moscow	1
	conference_Munich	1
	conference_Seoul	3
	conference_The Hague	2
	CSTP	2
	CSTP/TIP	2
	focus group	3
	High level meeting	1
	seminar	1
	thematic workshop	18
	working meeting	1
	workshop	26
Workshop_Title_long	*entire title from the document*	
Workshop_Title_short	*only the title*	
Workshop_date	*date drwan from the document*	
Workshop_place	*place drawn from the document*	
agenda summary & report	*from the original file DSTI-STP-TIPdocument list_05Oct 2017_Text_Mining_Selection.xlsx*	
Summary	*from the original file DSTI-STP-TIPdocument list_05Oct 2017_Text_Mining_Selection.xlsx*	
COTE	*from the original file DSTI-STP-TIPdocument list_05Oct 2017_Text_Mining_Selection.xlsx*	
Date	*from the original file DSTI-STP-TIPdocument list_05Oct 2017_Text_Mining_Selection.xlsx*	
Missing/other version/other language_details		
notes (or different COTE, or other comments on workshops)		
description of the document excerpted form the document first page		
Agenda_from Cote[A or AH]	control: is it an agenda?	

Annex 3 - Categories of documents adopted in the text analysis

Type of document: country review, report, agenda, summary (details are listed in Annex 4 and Annex 5).[36]

- *Country review* OECD Reviews of Innovation Policy: 31 documents on 25 countries, produced in the period 2006-2017; six countries have two reviews).
- *Report:* 98 reports with 101 documents (3 reports have 2 documents: in two reports there are part 1 and part 2 separate files; in one document there is the main document and its annex)
- *Agenda:* only the agenda;
- *Summary*: concept note, issue, summary, proceedings; the agenda may be included.

Type of the event. The following types have been adopted in our classification:

- *plenary meetings*;
- *conference*s;
- *workshops*. This type of events includes: ad hoc meetings; focus groups; high level meetings; seminars, thematic workshops, workshops.

- **ID**: a code has been assigned in order to create a short label for documents or to group documents referring to the same activity or event (see details in Annex 2).
  - *Country_review*: the 31 country reviews have been coded to create appropriate attributes, for each document, with year and name of the country (yy_Country)[37]
  - *Report*: we have identified 101 documents referring to 97 reports (three reports are split in 'part one' and 'part two'; one report consisted of two files: a main document and a draft set of recommendations)
  - *Plenary*: we have identified the 49 meetings and assigned them a code from 1 to 49, for each meeting, but for the last one, there are two documents (respectively: agenda and summary report)
  - *conference/workshop/focus group/seminar*: we have identified 54 events with a total of 76 documents. Each event has been identified as conferences or workshop or focus group, with the following IDs:
    - for conferences: "C_yy_country" (8 events, 12 documents),
    - for workshops: w_yy_#" (45 events, 61 documents);
    - for the focus group: "FG_yy)topic" (one event, 3 documents)
- **Year**: 25 years from 1993 to 2017. In the case of 'country reviews' and 'reports', 'year' is that of the document as it is defined, respectively, in the publication and in the Cote. In the case of 'events' it refers to the year in which the event took place. It could be different from the year in which the documents linked to the event were published (eg., for event occurring at the beginning of the year, the 'agenda' could be issued in the year before the event; for events at the end of the year, the 'summary' could be issued in the year after the event).

[36] In order to classify documents in this category as "country reviews", we have adopted the Oecd classification, as emerged from the original Oecd archive. For all the other documents, categories have been assigned by browsing all the documents referring in their title to the words 'report', 'conference', and 'workshop' 'meeting', 'seminar', 'focus group', 'agenda': the corresponding file has been opened to read the text on the front page and first pages.

[37] A table setting the country reviews in the scenario of Oecd members and key partners is presented in Annex 11.

Annex 4 Number of documents by type of document and by activity

Type of document	Country reviews	Reports	Events			Total
			plenary meetings	conferences	workshops	
country review	31					**31**
report		101				**101**
agenda			49	4	33	**86**
summary			48	8	26	**82**
summary&agenda*					5	**5**
Total	**31**	**101**	**97**	**12**	**64**	**305**

* This category has been created to allow a specific retrieval of information from the agendas

Annex 5 - Number of documents by year and type of activity

Year*	Country reviews	Reports	Events			Total
			plenary meetings	conferences	workshops	
1994			9	5	2	16
1995			5	4	10	19
1996			7	4	2	13
1997			3	5	3	2	13
1998			9	4	3	16
1999			1	4	2	7
2000			3	3	1	5	12
2001			3	4	2	9
2002			3	4	1	1	9
2003			7	4	2	13
2004			4	5	1	1	11
2005			12	3	1	16
2006		1	5	4	4	14
2007		5	6	4	3	18
2008		3	3	5	3	14
2009		2	4	4	1	11
2010			1	3	4	8
2011		2	2	4	4	12
2012		2	3	4	6	15
2013		3	2	4	1	3	13
2014		5	3	4	5	17
2015				4	1	5
2016		4	6	4	1	15
2017		4		2	1	7
Total		**31**	**101**	**97**	**12**	**64**	**305**

* For country reviews and reports: it is the year of the document; for the events: it is the year in which the event took place

Annex 6 - ID# of country reviews, plenary meetings, conferences and workshops/focus group/seminar

Country reviews: 31 events and documents

ID (yy_Country)	n. of documents
06_Switzerland	1
07_Chile	1
07_China	1
07_Luxembourg	1
07_New Zealand	1
07_South Africa	1
08_China	1
08_Hungary	1
08_Norway	1
09_Korea	1
09_Mexico	1
11_Peru	1
11_Russian Federation	1
12_Slovenia	1
12_Sweden	1
13_Croatia	1
13_Mexico	1
13_Southeast Asia	1
14_Colombia	1
14_France	1
14_Korea	1
14_Netherlands	1
14_Viet Nam	1
16_Lithuania	1
16_Luxembourg	1
16_Malaysia	1
16_Sweden	1
17_Costa Rica	1
17_Finland	1
17_Kazakhstan	1
17_Norway	1

Plenary meetings: 49 events, 97 documents

ID (#)	n. of documents	ID (#)	n. of documents	ID (#)	n. of documents
1	1	2	31	2	
2	2	2	32	2	
3	2	2	33	2	
4	2	2	34	2	
5	2	2	35	2	
6	2	2	36	2	
7	2	2	37	2	
8	2	2	38	2	
9	2	2	39	2	
10	2	2	40	2	
11	2	2	41	2	
12	2	2	42	2	
13	2	2	43	2	
14	2	2	44	2	
15	2	2	45	2	
16	2	2	46	2	
17	2	2	47	2	
18	2	2	48	2	
19	2	2	49	1	
20	2				
21	2				
22	2				
23	2				
24	2				
25	2				
26	2				
27	2				
28	2				
29	2				
30	2				

Conferences: 8 events, 12 documents

ID event	n. of documents
C_00_Bologna	1
C_02_Mexico	1
C_04_Moscow	1
C_05_Berlin	1
C_06_Madrid	2
C_06_TheHague	2
C_13_Munich	1
C_97_Seoul	3

Workshops*: 47 events, 64 documents

Year** ID event	**n. of documents**
1994	**2**
w_94_1	2
1995	**10**
w_95_1	3
w_95_2	4
w_95_3	2
w_95_4	1
1996	**2**
w_96_1	2
1997	**2**
w_97_1	2
1998	**3**
w_98_1	1
w_98_2	2
1999	**2**
w_99_1	2
2000	**5**
fg_00)NIS	3
w_00_1	1
w_00_2	1
2001	**2**
w_01_1	1
w_01_2	1
2002	**1**
w_02_1	1
2003	**2**
w_03_1	1
w_03_2	1
2004	**1**
w_04_1	1
2007	**3**
w_07_1	2
w_07_2	1
2008	**3**
w_08_1	1
w_08_2	1
w_08_3	1
2009	**1**
w_09_1	1

Year ID event	**n. of documents**
2010	**4**
w_10_1	1
w_10_2	1
w_10_3	1
w_10_4	1
2011	**4**
w_11_1	1
w_11_2	1
w_11_3	2
2012	**6**
w_12_1	2
w_12_2	1
w_12_3	1
w_12_4	1
w_13_2	1
2013	**3**
w_13_1	2
w_13_3	1
2014	**5**
w_14_1	1
w_14_2	2
w_14_3	1
w_14_4	1
2015	**1**
w_15_1	1
2016	**1**
w_16_1	1
2017	**1**
w_17_1	1

* It includes the following types of workshops:

type of workshop	**Number of events**	**Number of documents**
ad hoc meeting	5	10
CSTP workshop	2	2
CSTP/TIP workshop	2	2
focus group	1	3
high level meeting	1	1
seminar	1	1
thematic workshop	13	18
working meeting	1	1
workshop	21	26
total	47	64

** It is the year in which the event took place

Annex 7 - List of conferences organized by the WPTIP

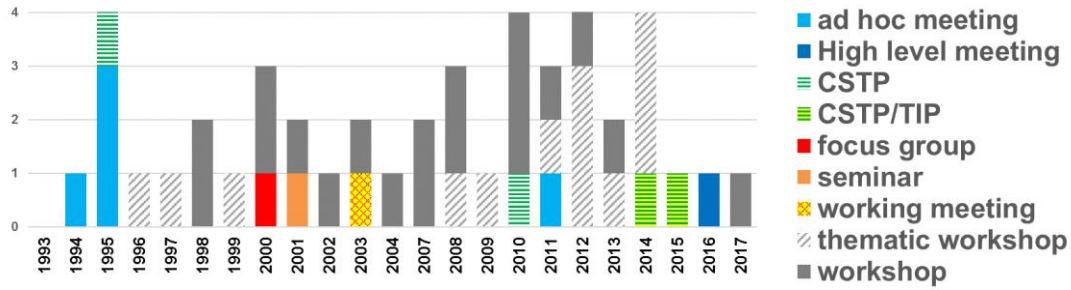
ID_unimore	YEAR	Title	Date	Place
59, 62, 63	1997	Facilitating International Technology Co-operation in a Globalised Knowledge-based Economy	13-14/10/1997	Seoul, Korea
81	2000	Enhancing the Competitiveness of SMEs in the Global Economy: Strategies and Policies	14-15/06/2000	Bologna, Italy
105	2002	Public/private partnerships for innovation	Autumn 2002	Puerto Vallarta, Mexico
132	2004	Public-private partnerships for innovation	16-17/12/2004	Moscow, Russia
147	2005	Intellectual Property as an Economic Asset: Key Issues in Valuation and Exploitation	30/06/-01/07/2005	Berlin, Germany
148,166	2006	Research use of Patented Knowledge	18-19/05/2006	Madrid, Spain
161, 179	2006	Globalisation and Open Innovation	06/12/2006	The Hague, netherland
241	2013	Creating Markets from Public Research Results	06-07/05/2013	Munich, Germany

Annex 8 List of the 47 workshops organized by the WPTIP, 1993-2017

"type of workshop"	Title	date	place
ad hoc meeting	GOVERNMENT FORESIGHT EXERCISES	14/09/1994	Paris
ad hoc meeting	FISCAL MEASURES TO PROMOTE R&D AND INNOVATION	19/01/1995	Paris
ad hoc meeting	VENTURE CAPITAL	29/11/1995	Paris
CSTP workshop	Points Arising for Action or Information of the Working Group on Innovation and Technology Policy	20-22/03/1995	missing
workshop	TECHNOLOGY, ECONOMIC GROWTH AND EMPLOYMENT IN THE SERVICE SECTOR	03-05/05/1995	Washington, D.C.
ad hoc meeting	FISCAL MEASURES TO PROMOTE R&D AND INNOVATION	missing	missing
thematic workshop	TECHNOLOGY DIFFUSION: A TYPOLOGY OF PROGRAMMES	13/12/1996	missing
thematic workshop	TECHNOLOGY INCUBATORS	25/06/1997	Paris
workshop	COMMERCIALISATION OF GOVERNMENT-FUNDED RESEARCH	25/11/1998	Canberra
workshop	TECHNOLOGY FORESIGHT FOR SUSTAINABLE DEVELOPMENT	11/12/1998	Budapest
thematic workshop	SCIENCE AND TECHNOLOGY LABOUR MARKETS	17/05/1999	Paris
focus group	NATIONAL INNOVATION SYSTEMS: Knowledge Markets and Innovation Systems: Nurturing the Institutions of Innovation	missing	missing
workshop	INNOVATION AND THE ENVIRONMENT	19/06/2000	missing
workshop	THE MANAGEMENT OF INTELLECTUAL PROPERTY RIGHTS FROM PUBLIC RESEARCH	11/12/2000	Paris
seminar	Innovation and the Valorisation of Science and Technology in Russia	1-2/03/2001	Helsinki
workshop	PUBLIC/PRIVATE PARTNERSHIPS FOR INNOVATION	10/12/2001	Paris
workshop	CHANGING STRATEGIES FOR BUSINESS R&D AND THEIR IMPLICATIONS FOR SCIENCE AND TECHNOLOGY POLICY	10-12/06/2002	Paris
working meeting	THE EFFECTIVENESS OF GOVERNMENT POLICIES FOR INFLUENCING BUSINESS R&D STRATEGIES	04/02/2003	Brussels
workshop	ASSESSING NATIONAL INNOVATION PERFORMANCE	24/03/2003	London
workshop	GOVERNANCE OF INNOVATION POLICY: PRELIMINARY FINDINGS OF THE MONIT PROJECT	06/12/2004	Paris
workshop	R&D TAX TREATMENT IN OECD COUNTRIES: COMPARISONS AND EVALUATIONS	10/12/2007	Paris

workshop	Rethinking Evaluation in Science and Technology	11-12/12/2007	missing
workshop	ASSESSING THE SOCIO-ECONOMIC IMPACTS OF PUBLIC R&D INVESTMENT	11/06/2008	Paris
workshop	ENHANCING RESEARCH PERFORMANCE THROUGH EVALUATION AND PRIORITY SETTING	15-16/09/2008	Paris
thematic workshop	THE INNOVATION STRATEGY -- DRAWING ON THE WORK OF TIP AND THE COUNTRY REVIEWS OF INNOVATION	03/12/2008	Paris
thematic workshop	FUTURE ORIENTATIONS FOR SCIENCE, TECHNOLOGY AND INNOVATION POLICY	14/12/2009	Paris
workshop	GOVERNANCE OF STI POLICY MAKING AT NATIONAL LEVEL	18-19/02/2010	Vienna
workshop	THE OECD INNOVATION HANDBOOK	23/06/2010	Paris
workshop	NATIONAL STI GOVERNANCE	25/06/2010	Vienna
CSTP workshop	Green technology and innovation policies	13-15/12/2010	Paris
ad hoc meeting	COMPARATIVE ADVANTAGE THROUGH SMART SPECIAL-ISATION	05/05/2011	Leuven-Heverlee
workshop	KNOWLEDGE NETWORKS AND MARKETS	15/06/2011	Paris
thematic workshop	FINANCING R&D AND INNOVATION IN THE CURRENT MACROECONOMIC CONTEXT	07/12/2011	Paris
workshop	SMART SPECIALISATION STRATEGIES FOR INNOVATION-DRIVEN GROWTH	10-11/05/2012	Paris
thematic workshop	THE INNOVATION POLICY PLATFORM (IPP)	25/06/2012	Paris
thematic workshop	KNOWLEDGE TRANSFER, EXPLOITATION AND COMMER-CIALISATION	05/10/2012	Paris
thematic workshop	SYSTEMS INNOVATION	12/12/2012	Paris
workshop	SMART SPECIALISATION	03-04/04/2013	Paris
workshop	SYSTEMS INNOVATION	28/03/2013	Paris
thematic workshop	OPEN SCIENCE AND OPEN DATA	12/12/2013	Paris
thematic workshop	OPEN SCIENCE AND OPEN DATA	10/03/2014	Warsaw
thematic workshop	SYSTEM INNOVATION: POLICY RELEVANT FINDINGS FROM THE BERLIN WORKSHOP	28-29/04/2014	Berlin
thematic workshop	STRATEGIC PUBLIC-PRIVATE PARTNERSHIPS IN STI	18/06/2014	Paris
CSTP/TIP workshop	THE KNOWLEDGE TRIANGLE PROJECT 2015-2016	15-16/12/2014	Paris
CSTP/TIP workshop	THE OECD SCIENCE, TECHNOLOGY AND INDUSTRY OUT-LOOK	17/06/2015	Paris
High level meeting	HIGH LEVEL MEETING ON THE KNOWLEDGE TRIANGLE	15-16/09/2016	missing
workshop	Innovation and the Digital Economy: What Role for Innovation Pol-icies	14/06/2017	Paris

Annex 9 - The 47 WPTIP workshops, by type, 1993-2017



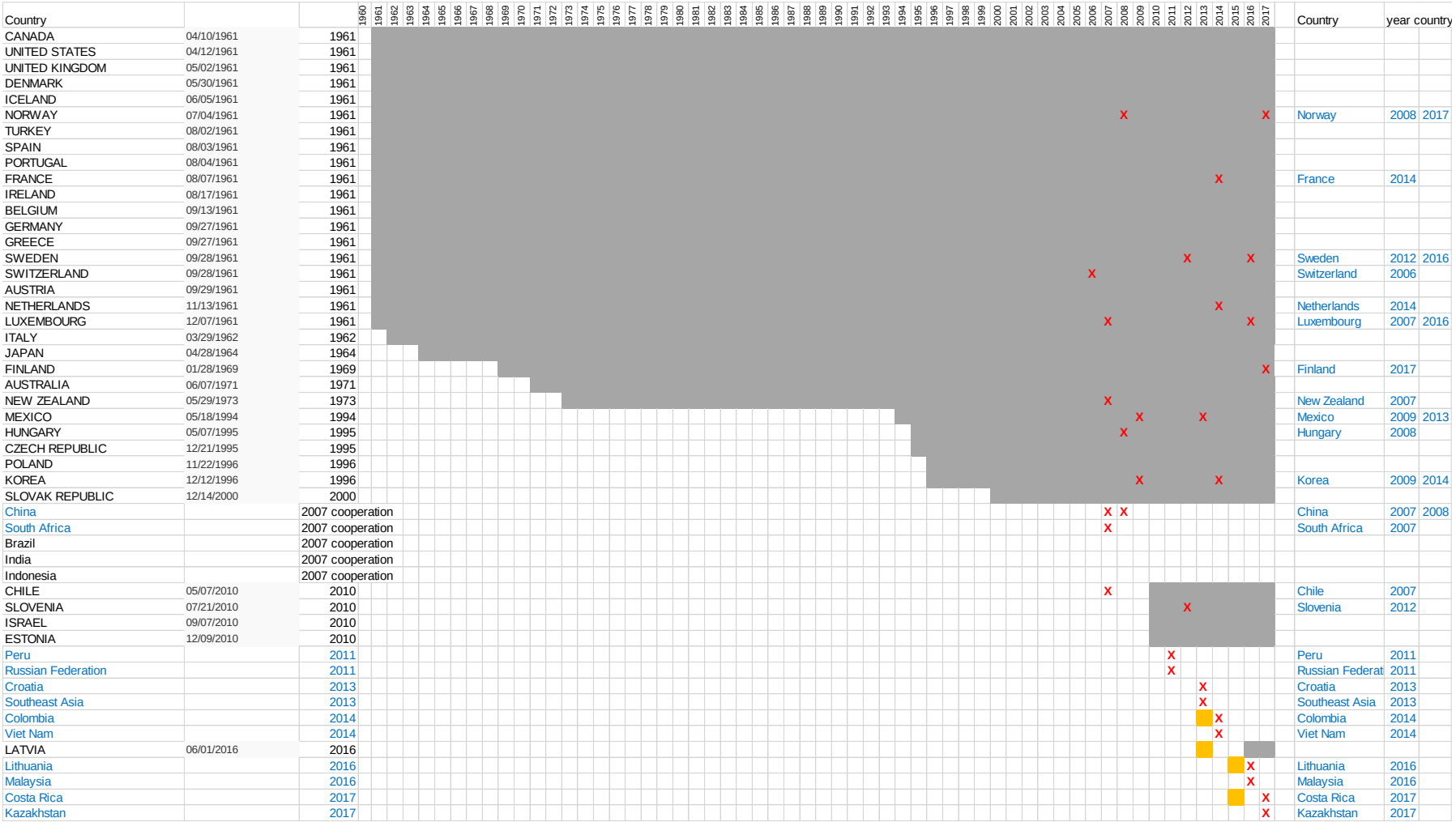
Annex 10 - List of acronyms

The selection refers to documents of plenary meetings, workshops and conferences, reports
Source: our elaboration using the Oecd vocabulary on innovation (2015)

BERD
CIS
EGT
FDI
GBAORD
GERD
GOVERD
H index
HDI
HEI
HERD
HRST
ICT
IP
IPR
ISCED 5
ISCED 6
ISO
KAM
KBC
KIBS
MNC
MNE
NGO
NIE
NPM
NTBF
OEM
PCT
PIAAC
PISA
PLC
PMR index
PPP
PhD
R+D
R&D
R+D funding
R+D intensity
R+D labs
RCA
RDA
RTA
SBIR
SEZ
SME
SMEs
SOE
TFP
TNC
TRIPS
TTO
VET

Annex 11 - Oecd country members and observers, by year, and country reviews

Country reviews, by country and year, are marked with a red cross.
Years in which a country has been invited to open formal OECD accession talks are marked in orange, for the corresponding countries.



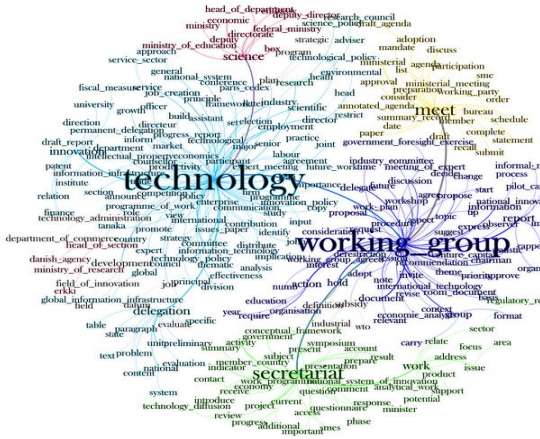
Country			Country	year	country
CANADA	04/10/1961	1961			
UNITED STATES	04/12/1961	1961			
UNITED KINGDOM	05/02/1961	1961			
DENMARK	05/30/1961	1961			
ICELAND	06/05/1961	1961			
NORWAY	07/04/1961	1961	Norway	2008	2017
TURKEY	08/02/1961	1961			
SPAIN	08/03/1961	1961			
PORTUGAL	08/04/1961	1961			
FRANCE	08/07/1961	1961	France	2014	
IRELAND	08/17/1961	1961			
BELGIUM	09/13/1961	1961			
GERMANY	09/27/1961	1961			
GREECE	09/27/1961	1961			
SWEDEN	09/28/1961	1961	Sweden	2012	2016
SWITZERLAND	09/28/1961	1961	Switzerland	2006	
AUSTRIA	09/29/1961	1961			
NETHERLANDS	11/13/1961	1961	Netherlands	2014	
LUXEMBOURG	12/07/1961	1961	Luxembourg	2007	2016
ITALY	03/29/1962	1962			
JAPAN	04/28/1964	1964			
FINLAND	01/28/1969	1969	Finland	2017	
AUSTRALIA	06/07/1971	1971			
NEW ZEALAND	05/29/1973	1973	New Zealand	2007	
MEXICO	05/18/1994	1994	Mexico	2009	2013
HUNGARY	05/07/1995	1995	Hungary	2008	
CZECH REPUBLIC	12/21/1995	1995			
POLAND	11/22/1996	1996			
KOREA	12/12/1996	1996	Korea	2009	2014
SLOVAK REPUBLIC	12/14/2000	2000			
China		2007 cooperation	China	2007	2008
South Africa		2007 cooperation	South Africa	2007	
Brazil		2007 cooperation			
India		2007 cooperation			
Indonesia		2007 cooperation			
CHILE	05/07/2010	2010	Chile	2007	
SLOVENIA	07/21/2010	2010	Slovenia	2012	
ISRAEL	09/07/2010	2010			
ESTONIA	12/09/2010	2010			
Peru		2011	Peru	2011	
Russian Federation		2011	Russian Federat	2011	
Croatia		2013	Croatia	2013	
Southeast Asia		2013	Southeast Asia	2013	
Colombia		2014	Colombia	2014	
Viet Nam		2014	Viet Nam	2014	
LATVIA	06/01/2016	2016			
Lithuania		2016	Lithuania	2016	
Malaysia		2016	Malaysia	2016	
Costa Rica		2017	Costa Rica	2017	
Kazakhstan		2017	Kazakhstan	2017	

Source: our elaboration on http://www.oecd.org/about/membersandpartners/ and from list of country reviews in the Oecd archive

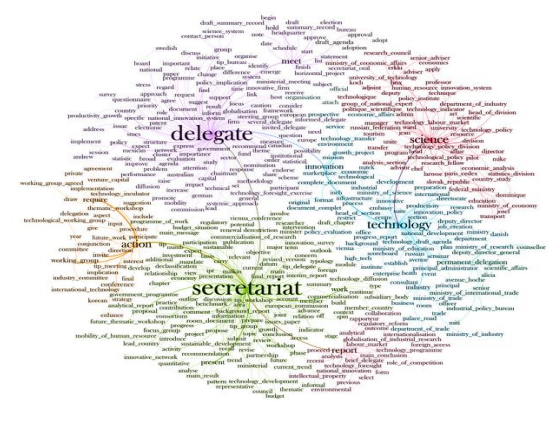
Annex 12 Plenary meetings: web of semantic clusters

Communities aggregate words co-occurring with the pivotal word (occurrences >4). Colours highlight word communities. Size of the words is proportional to their leverage degree in the corpus, width of the edges is proportional to the number of co-occurrence within fragments in the corpus. Modularity is implemented with Iramuteq. Graphs are represented with Gephy

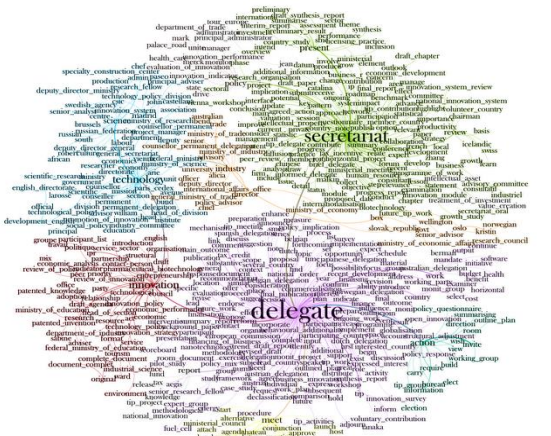
1993-1996: plenary meetings 1-7



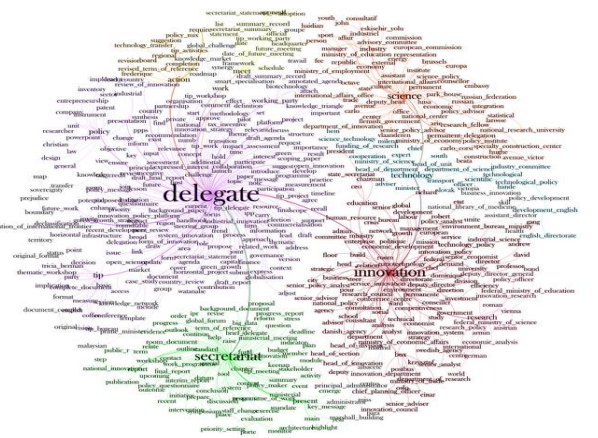
1996-2001: plenary meetings 8-17



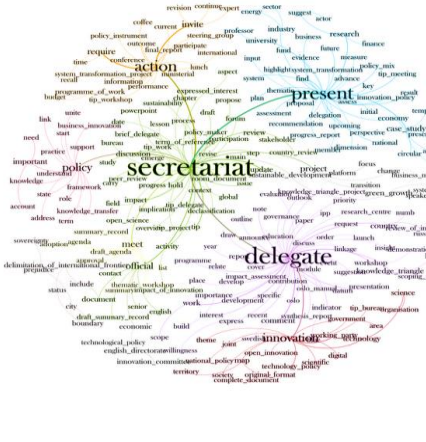
2001-2007: plenary meetings 18-29



2007-2014: plenary meetings 30-44



2015-2017: plenary meetings 45-49



Annex 13 - Plenary meetings: a selection of specific keywords on organizational aspects, by cluster

	1993-1996	1996-2001	2001-2007	2007-2014	2015-2017
WP structure	**Working group,** delegation, member country. Specifically, the organizations has: chairman, observer (countries), rapporteur). The activity is structured around a **mandate, a work programme, a work plan. ...** Working group **agreed.**	**TIP group /** technological Working group **Working group agreed** the **renewal** (of the CSTP mandate)	**TIP Working party/delegate** /work/contribution **participating country** **volunteer country** **volunteer contribution**	**innovation policy platform (ipp)** **secretariat statement**/oral **staff change** **TIP Steering group** **Contact information of** **revised mandate**	
specific types of meetings	Meeting of **expert,** but also **informal** meeting, **ministerial** meeting, **summit**	**group of national expert**		**tip policy roundtable (**on STI governance) **expert workshop** **final conference** **global forum** (on the knoledge economy, on green growth and sustainable development)	**blue sky conference**/indicators
tools	**pilot** case study **national** case study **issues paper**	**report** (analytical/ background) **draft summary record** issues paper **institutional mapping** (related to NIS work/activity/focus group)	**MONIT project/group** **tip case study** (on kibs, on evaluation of innovation policy, in 2006) **country study** (see list) **peer review** chapter/report/summary reports/synthesis **Various versions:** draft/revised/final publication	**handbook** (STI policy platform (formrely handbook, 2010-11) **strategic:** P/PPs, innovation areas/programme; but also in many cases refernce contact of persons in strategic divisions) secretariat summary **tip project** (digiatl and open innovation, system innovation, knowldge triangle) **revised term of reference** (project on open science) **piloting (on the IPP)** review of innovation expert review technology study scoping paper annotated agenda timeline	**peer reviewer** (by other countries on country reviews) **keynote presentation** **mapping** (of the governance of national policy research / of the policy mix / knowledge flow in innovation processes / "....approaching a transition to a low carbon future using **road mapping**) **scoping** (the TOR on assessing the imoact of policy mix / for the work on digitalization / scaping and framing **focusing: "...**countries suggested focucing on..." **powerpoint**

Annex 14 - Countries named in the plenary meetings, by period

Country	1993-1996	1996-2001	2001-2007	2007-2014	2015-2017
Argentina	0	0	0	1	0
Australia	31	41	57	16	0
Austria	16	20	67	33	26
Belgium	14	14	63	27	17
Brazil	0	1	3	3	0
Bulgaria	0	0	0	0	1
Canada	36	33	57	32	7
Chile	0	0	12	4	9
China	0	0	61	12	15
Colombia	0	0	0	15	3
Costa Rica	0	0	0	10	18
Croatia	0	0	0	8	2
Cyprus	0	1	0	0	0
Czech Republic	6	7	18	6	8
Denmark	11	6	17	8	2
Estonia	0	0	1	8	3
Finland	32	39	72	47	24
France	16	20	160	35	3
Germany	25	28	53	15	20
Greece	1	2	7	2	3
Hungary	23	15	24	7	11
Iceland	1	0	25	1	0
India	0	0	2	1	0
Indonesia	0	0	1	1	0
Ireland	6	1	16	15	17
Israel	10	3	21	15	4
Italy	10	17	36	13	5
Japan	36	34	83	46	19
Kazakhstan	0	0	0	0	7
Korea	11	29	50	25	13
Lithuania	0	0	0	1	5
Luxembourg	0	1	24	30	3
Malaysia	0	0	0	14	11
Mexico	25	24	29	30	14
Netherlands	15	11	88	52	25
New Zealand	1	3	45	6	0
Norway	18	24	70	41	13
Peru	0	0	0	4	0
Poland	5	4	45	9	10
Portugal	2	5	36	5	24
Romania	0	0	3	0	0
Russia	6	21	67	33	14
Scotland	0	1	2	0	0
Slovenia	9	0	1	8	0
South Africa	0	2	39	4	0
Spain	12	9	55	12	7
Sweden	9	17	37	21	30
Switzerland	13	6	37	10	3
Turkey	10	8	13	20	5
Ukraine	0	0	0	1	0
United Kingdom	12	6	32	11	7
United States	42	11	47	22	8
Vietnam	0	0	0	1	0