STI 2018 Leiden

23rd International Conference on Science and Technology Indicators
"Science, Technology and Innovation Indicators in Transition"

STI 2018 Conference Proceedings

Proceedings of the 23rd International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Chair of the Conference

Paul Wouters

Scientific Editors

Rodrigo Costas
Thomas Franssen
Alfredo Yegros-Yegros

Layout

Andrea Reyes Elizondo
Suze van der Luijt-Jansen

The articles of this collection can be accessed at https://hdl.handle.net/1887/64521

ISBN: 978-90-9031204-0

23rd International Conference on Science and Technology Indicators (STI 2018)
"Science, Technology and Innovation indicators in transition"
12 - 14 September 2018 | Leiden, The Netherlands
#STI18LDN

Mapping the educational content of the science of science, technology, and innovation policy: an international comparison[1]

Asako Okamura[*], Shinano Hayashi[**], Hitoshi Koshiba[***] and Yui Nishimura[****]

**a-okamura@grips.a.cjp*
SciREX Center, National Graduate Institute for Policy Studies, 7-22-1 Roppongi, Minato-ku, Tokyo 106-8677, Japan

***shinano.hayashi@jst.go.jp*
Center for Research and Development Strategy, Japan Science and Technology Agency, K's Gobancho, 7 Gobancho, Chiyoda-ku, Tokyo 102-0076, Japan

*** *hitoshi.koshiba@nistep.go.jp*
National Institute of Science and Technology Policy, 16th Floor, Central Government Building No.7 East Wing, 3-2-2, Kasumigaseki, Chiyoda-ku, Tokyo 100-0013, Japan

****
School of Science for Open and Environmental Systems, Graduate School of Science and Technology, Keio University, 301401 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan

Abstract
It has been seven years since the launch of Japan's Science for RE-designing Science, Technology and Innovation Policy (SciREX) programme. It is therefore an opportune time to take stock, not only of the Japanese programme, but also of leading science, technology, and innovation (STI) research and education programmes in other nations. This paper updates our work (Okamura et.al. 2017), applying a quantitative approach to syllabi analysis by using topic model methodology to identify frequent topics on the syllabi of various STI educational programmes. The number of education programmes analysed has increased from 12 to 24 Japanese and overseas programmes. Although this is still on-going study, we have found that worldwide STI education programmes can be characterised using twelve topics. In 2016, the SciREX programme in Japan held discussions with STI stakeholders at multiple workshops to share key terminologies and science questions relating to key chapters of core parts of the curricula used by six participating Japanese universities. The aim of both activities (the quantitative analysis of syllabi and the development of core curricula) was to identify research and educational agendas that could respond to societal needs and changes.

Background

The need for informed science, technology, and innovation (STI) policies is increasing amidst fast-growing social and economic problems in various areas. Many countries have sought to

[1] The research team would like to thank the individuals and organisations who generously shared their time, experience, and materials for this project. The work has been supported by the Science for RE-designing Science, Technology, and Innovation Policy (SciREX) Programme in Japan. The views and opinions expressed here are solely those of the authors and do not represent in any way those of the institutions to which they are affiliated.

design more effective and efficient STI policies that encourage innovation by mobilising multi-disciplinary scientific fields. In response to the growing demand for evidence-based infrastructure in STI policymaking, the Japanese government implemented the Science for RE-designing Science, Technology, and Innovation Policy (SciREX) programme in 2011. This programme consists of competitive research funding, data infrastructure building, research projects centred on prioritised policy areas (mission-oriented research) and graduate education programmes. In the United States, where a former science advisor to the president, Dr John H. Marburger III, addressed the need to establish a Science of Science Policy in 2005, the Science of Science and Innovation Policy (SciSIP) programme of the National Science Foundation has been producing research and practice in this interdisciplinary field for more than a decade.

It is therefore important to assess leading initiatives in the science of science and innovation policy, not only to look back at what has been accomplished so far, but also to establish forward-looking research agendas, examine the focus of relevant research, and identify any existing or potential bottlenecks. In recent years, researchers have carried out similar studies on innovation (Martin, 2016), science policy and innovation (Martin, 2012), and science and technology (Martin, Nightingale, and Yegros-Yegros, 2012). In addition, the OECD's decennial Blue Sky III Forum in 2016 discussed future research agendas, based on previous accomplishments in this field.

To identify the most important research questions, researchers must draw on their own expertise, insight, and conjectures. Most studies have relied on qualitative approaches such as literature reviews and discussions in workshops and interviews (e.g. Alexander, Hart, and Hill, 2016; the National Science and Technology Council [NSTC]; and the Office of Science and Technology Policy [OSTP], 2008). Quantitative approaches have rarely been used; Martin (2012), one of few exceptions, adopted a bibliometric approach. However, statistical technologies and text-mining approaches that categorise sentences and identify clusters of words have existed for years. The significant body of currently available data has expanded the possibilities for using these technologies. The present study therefore adopts a mixed-methods approach to identify the core questions of the science of science and innovation policy research. Nevertheless, there remain challenges associated with the effective synthesis of qualitative and quantitative approaches.

Objectives

Currently, six universities in Japan are participating in the SciREX programme. Each has its own educational syllabus, which is not synchronised with those of other universities. To develop human resources in the field of science policy and innovation studies, we are leading ongoing discussions to determine whether neighbouring sciences share a core interdisciplinary STI knowledge base. This study-in-progress explores the distribution of educational elements in the syllabi of various educational institutes to define the structure of educational programmes. Our present and future findings will shed light on the likely relationship between input (educational programmes) and output (the social impact of programme graduates).

As the science of science and technology innovation policy is an interdisciplinary field, can we find common ground?

Since starting the SciREX programme, we have realised that the 'science of STI policy' covers a wide range of scientific fields, not just the natural sciences, such as medicine and biology, but also the social sciences and humanities. This makes it difficult to identify the core of this interdisciplinary area of study. Martin (2012) has observed that the 'field of science policy and innovation studies (SPIS) *is still a more fragmented and heterogeneous field than established social science disciplines*'. It is significant step to define the 'core' of 'science of STI policy' shared among various neighbouring sciences. Each scientific field is its own discipline; this study asks whether the science of STI policy as interdisciplinary field is associated with a unique discipline or theory of its own. Carrying out this research has involved extensive discussions about the 'science of STI policy' as a field, along with efforts to define its range and core areas of knowledge. This work is based on a recognition that a 'science of STI' policy will be indispensable for the future development of the field and necessary to provide benchmarks.

Two approaches have been used to analyse the structure and range of the 'science of STI policy': (i) A qualitative approach to define 'core content' was based on science questions that were shared among stakeholders in the SciREX community and the consensus obtained in workshops; (ii) A consensus on the kind of courses (curricula) to include in 'science of STI policy' educational programmes and the field internationally was used as a proxy for the structure of the science. The present study has used a quantitative approach to objectively and systematically assess the overall trends in educational programmes, including those at Japanese and other universities.

The first approach (i) used a procedure based on discussions developed in 2016; specialists (mainly researchers and practitioners in the SciREX community) discussed and later wrote core science content for a science of STI policy. The final draft will be released on the web in late 2018. To conduct the quantitative analysis (ii), syllabi collected through the Internet and collaborators were analysed using the topic model.

Our 2017 study ran a trial, analysing a quantitative analysis of syllabi. The present study has updated and revised the data, analytical method, and analyses. In addition, some commentators responded to presentations of the earlier analysis by pointing out the value of focusing, not only on research topics within the 'science of STI policy', but also on its longer-term educational effects. For instance, one way to understand the educational effect of a programme is to track graduate jobs and careers; this can help to assess the extent to which the programme design has successfully developed human resources to support STI policy. In future research, it will be important to interpret and synthesise the results of both the (i) and (ii) approaches.

Data and Methodology

Data

The present study has expanded the number of targeted education programmes from 12 to 24. All of the targeted programmes were graduate level (Master's or doctoral courses) but non-degree programmes were also included: 6 programmes at 6 Japanese universities (JPN), 8 US programmes (USA), 2 programmes each from the UK (GBR) and Turkey (TUR), and one programme each from the Netherlands (NLD), South Africa (ZAF), Canada (CAN), Belgium (BEL), Italy (ITA), and the Russian Federation (RUS). The programme names are listed in the Appendix. Table 1 below shows the number of courses offered in each programme.

Apart from five programmes for which information was collected through an exchange of emails after last year's STI 2017 presentation, information was gathered from programme websites. The targeted programmes were selected through a website survey that assessed each programme's science policy experts and researchers; for US programmes, we viewed the 'Engaging Scientists & Engineers in Policy (ESEP) Coalition' for each programme. Some programmes were excluded from the analysis (for example, specialised programmes on the Management of Technology) because they did not offer any classes on policy. Programmes involving few lectures (such as Ghent, Bocconi, and MIT) were excluded from interpretation parts as the results were strongly biased to specific topics. The Japanese and French syllabi were translated into English.

Table 1: The number of courses offered by each programme

Name	Country	Number of Courses
Ghent University	BEL	1
Polytechnique Montréal, Grad. (POLY)	CAN	8
The University of Manchester	GBR	7
University of Sussex	GBR	15
Bocconi University	ITA	2
The National Graduate Institute for Policy Studies (GRIPS)	JPN	17
Hitotsubashi University	JPN	8
Kyoto University	JPN	25
Kyushu University	JPN	8
Osaka University	JPN	21
University of Tokyo (U Tokyo)	JPN	34
UNU-MERIT	NLD	31
Higher School of Economics (HSE)	RUS	18
Middle East Technical University,MS. (METU)	TUR	38
Middle East Technical University,Ph.D. (METU)	TUR	38
Arizona State University (ASU)	USA	6
Carnegie Mellon University (CMU)	USA	12
Duke University	USA	32
Georgia Institute of Technology (Georgia Tech)	USA	10
The George Washington University (GWU)	USA	21
University of Michigan	USA	74
Massachusetts Institute of Technology (MIT)	USA	2
Princeton University	USA	25
Stellenbosch University	ZAF	13
Total	—	466

Note: The acronyms used are as follows: JPN for Japan; USA for the United States; GBR for the United Kingdom, NLD for the Netherlands, ZAF for the South Africa, CAN for Canada, TUR for Turkey, BEL for Belgium, ITA for Italy and RUS for the Russian Federation.

The syllabus-related information included title text and the outlines of courses. Words with no discriminatory power, such as unspecified nouns and general verbs were excluded as stop words[2] . In addition, several frequently used words that were not necessarily related to

[2] See http://xpo6.com/list-of-english-stop-words

educational content, such as 'description' and 'week' were excluded as stop words. Conjugations were also treated, using the WordNet and Python NLTK package. For example, 'practised' and 'practice' were considered the same word, in our analysis.

Methodology

The recent progress made by Natural Language Processing (NLP) has been remarkable: multiple methods, the topic model (e.g. Blei, 2013), and Word2Vec (e.g. Mikolov, 2003) have been developed and used widely in practice to understand the implied structure and categories of large amounts of textual information, including news articles and SNS content.

Although these methods may be relevant to an analysis of educational programmes, we found that few syllabus analyses used the topic model. To examine the applicability of NLP to the content of education programmes, we therefore applied a topic model to extract common elements from syllabi mechanically and to classify them as topics without relying too much on our own subjectivity[3].

Topic model analysis is a technique in the field of NLP, which has been developed mainly to identify potentially implied categories within large amounts of textual data. Topics are calculated using a generative probabilistic model. In principle, topics are calculated using a given set of documents (sets of words), on the condition that word probabilities in a given document are generated independently among topics. Multiple topic model methods have been proposed, including Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (pLSA), and Latent Dirichlet Allocation (LDA), among others. LDA is one of the multi-topic models that can be used to relate plural topics to one document. This is a preferred method when the amount of text data analysed is relatively small, as it avoids over-learning, which decreases the generalisability of results.

As the present study has included information drawn from 466 syllabi in total, the data set was relatively small; we therefore applied LDA. For sampling, we used the collapsed Gibbs method. For the calculation, we used GibbsLDA++ developed by C++. When using LDA, the number of topics (k) and parameters (hyper parameters α and β) should be set *a priori*. We followed some findings from Griffiths and Steyvers (2004); that $\alpha = 50/k$ and $\beta = 0.1$ are considered to be appropriate.

Defining topics

There are no clear-cut criteria for defining the number of topics (k). We have therefore analysed the results of topic numbers from 5 to 12. As a result, we reviewed the number of topics: although Okamura et al. (2017) classified only 8, the present study classified 12 topics, based on an easier comprehension of results. The topic names were labelled (it is not possible to eliminate subjectivity completely when assigning labels).

Data challenges

The data are still far from complete, as we only have access to a limited number of STI education programmes around the world. Although we have access to syllabi, the amount of information publicly available differs significantly among programmes. As there appears to be no universal standard for writing syllabi, the available information may be affected by forms of bias associated with university-,program-, and teacher-specific issues. Other forms

[3] The 2017 presentation (Okamura et al., 2017) used topic model and correspondence analysis. Although correspondence analysis is a convenient method, it was difficult to apply directly to achieve our purpose. As the results could have been somewhat difficult to interpret, the present study uses only the topic model.

of bias may result from the process of translating syllabi from their original languages to English.

Analysis and observation

Tables 2, 3 and 4 show the top 20 words and the probability distribution of words to topics. Table 5 provides examples of topic composition by course. Explanations for each topic are as follows.

(T_1) *Science and society*:

The words with highest probabilities are science (14.1%) and technology (13.6%) in this topic. Those with the second highest probability level include social (3.6%), society (2.9%), and communication (1.7%). Most courses attached to this topic involve Science and Technology Studies (STS). However, some management courses also have high probabilities within this topic.

(T_2) *Law and ethics*:

Words that relate to justice studies, including law (2.7%), legal (1.4%), conflict (1.1%), and regulatory (0.9%) are attached to this topic, as are issue-specific words such as environment (2.5%), international (2.3%), and space (1.3%); the latter may represent specific legal areas. Most of the higher probability courses involve justice studies, but some cover standards and rules from a management perspective. The ratios attached to words are relatively small, with the highest below 3%.

(T_3) *Technology adoption*:

Words with higher probabilities in this topic include market (3.4%), technological (2.7%), and work (1.8%). This topic appears to cover issues such as the economic and social impact of technologies on the market, work, and organizations. Courses with higher probabilities include “Technology and work organization” (METU, 61.9%). However, the ratios attached to words are relatively small overall, with the highest probability word just over 3%.

(T_4) *Technology and management*:

Words frequently used in management studies, including management (5.7%), business (4.1%), strategy (2.8%), and company (1.4%) are attached to this topic. The highest probability courses are “Technology and Corporate Strategy” (43.2%, METU) and “Business Model Innovation” (38.5%, HSE).

(T_5) *Intellectual assets*:

This topic specifically relates to intellectual property assets and its rights. Words with higher probabilities include application (3.4%), design (2.7%), property (2.7%), intellectual (2.7%), right (2.6%), and trademark (1.7%). Courses with higher probabilities include “Intellectual property rights and regulation” (87.5%, METU), and “Intellectual Property Management” (32.5%, HSE).

(T_6) *Energy and environment*:

Words with higher probabilities in this topic include environment (4.5%), change (3.1%), energy (3.0%), and climate (2.0%). Courses such as “Environmental psychology for public policy” (39.4%, Michigan) and “Sustainable Energy Systems” (37.6%, Michigan) have higher probabilities.

(T_7) *Development and STI*:

Words with higher probabilities in this topic include development (5.5%), economic (4.1%), global (1.9%), and regional (1.5%). Courses with higher probabilities include "Financing Social Protection" (53.2%, UNU-MERIT) and "The Global Challenge: Beyond Poverty and Inequality" (36.0%, UNU-MERIT).

(T_8) *Innovation and policy*:

Words with higher probabilities in this topic include innovation (11.4%), knowledge (3.4%), technology (3.3%), policy (3.2%), and evaluation (2.2%). Courses with higher probabilities include "Knowledge and Technology Transfer in Innovation Systems" (72.2%, METU) and "Innovation Policy and Governance: Trends and Challenges" (50.3%, METU).

(T_9) *Policy science*:

Words with higher probabilities in this topic include policy-related words such as policy (18.5%), public (5.7%), political (1.8%), decision (1.3%), and intervention (1.1%). Health (5.1%) also has a high probability, possibly associated with courses on health policy. Courses with higher probabilities include health-policy-related courses such as "Epidemiology, Health Services & Policy" (46.6%, Michigan) and "Research Design in Policy Science" (30.1%, Georgia Tech).

(T_10) *Research design*:

Words with higher probabilities in this topic include research (9.3%), lecture (4.3%), learn (3.3%), and study (2.2%). Courses with higher probabilities are distributed across a variety of courses, including "Negotiation and agreement" (41.4%, U Tokyo) and "Applied Economics" (38.8%, Kyoto). This may indicate that the topic is poorly defined.

(T_11) *Quantitative methodology*:

Words with higher probabilities to this topic include analysis (6.7%), method (5.5%), theory (2.6%), data (2.3%), and model (2.1%). Courses with higher probabilities to this topic include "Introduction to Quantitative Methods" (44.8%, GRIPS), "Regression Analysis" (38.8%, UNU-MRIT).

(T_12) *Qualitative methodology*:

Words with higher probabilities in this topic include student (6.8%), class (3.1%), develop (2.5%), and seminar (1.7%). Courses with higher probabilities include "Advanced Science and Technology Policy" (43.7%, ASU) and "Introduction to Analysing Sociotechnical Systems" (37.6%, UNU-MRIT).

Table 2: Estimated Latent Topics using LDA (1)

ID	Science and society Topic 1 (T_1)		Law and ethics Topic 2 (T_2)		Technology adoption Topic 3 (T_3)		Technology and management Topic 4 (T_4)	
1	science	14.1%	issue	2.7%	market	3.4%	management	5.7%
2	technology	13.6%	law	2.7%	technological	2.7%	business	4.1%
3	social	3.6%	environment	2.5%	role	2.6%	risk	3.7%
4	society	2.9%	international	2.3%	introduction	1.9%	project	3.1%
5	communication	1.7%	include	2.3%	work	1.8%	strategy	2.8%
6	scientific	1.6%	current	1.9%	state	1.7%	issue	2.2%
7	relationship	1.2%	legal	1.4%	process	1.5%	process	1.5%
8	knowledge	1.1%	future	1.4%	case	1.4%	company	1.4%
9	think	1.1%	human	1.3%	change	1.4%	strategic	1.3%
10	addition	0.9%	space	1.3%	debate	1.3%	organization	1.3%
11	field	0.8%	global	1.2%	topic	1.2%	area	1.1%
12	require	0.7%	explore	1.2%	production	1.1%	focus	1.0%
13	question	0.7%	conflict	1.1%	issue	1.0%	idea	1.0%
14	various	0.7%	life	1.0%	development	1.0%	source	1.0%
15	critical	0.6%	concern	1.0%	approach	1.0%	industry	1.0%
16	module	0.6%	topic	1.0%	firm	0.9%	control	1.0%
17	study	0.6%	activity	1.0%	organization	0.9%	practice	0.9%
18	concrete	0.6%	regulatory	0.9%	network	0.9%	knowledge	0.9%
19	discipline	0.6%	non	0.9%	identify	0.8%	development	0.8%
20	scientist	0.5%	modern	0.9%	technical	0.8%	creativity	0.8%

Table 3: Estimated Latent Topics using LDA (2)

ID	Intellectual assets Topic 5 (T_5)		Energy and environment Topic 6 (T_6)		Development and STI Topic 7 (T_7)		Innovation and policy Topic 8 (T_8)	
1	application	3.4%	environmental	4.5%	development	5.5%	innovation	11.4%
2	design	2.7%	change	3.1%	economic	4.1%	knowledge	3.4%
3	property	2.7%	energy	3.0%	social	3.1%	technology	3.3%
4	intellectual	2.7%	problem	2.4%	different	2.6%	policy	3.2%
5	right	2.6%	climate	2.0%	country	2.6%	concept	2.5%
6	community	2.4%	focus	2.0%	global	1.9%	evaluation	2.2%
7	registration	2.3%	perspective	1.8%	level	1.8%	develop	1.6%
8	international	2.2%	examine	1.6%	regional	1.5%	public	1.5%
9	procedure	2.0%	discuss	1.5%	governance	1.5%	university	1.3%
10	examination	1.8%	sustainable	1.4%	develop	1.1%	trend	1.3%
11	trademark	1.7%	infrastructure	1.3%	protection	1.0%	design	1.3%
12	trade	1.7%	explore	1.2%	world	0.9%	program	1.2%
13	industrial	1.5%	resource	1.0%	special	0.9%	transfer	1.2%
14	patent	1.4%	sector	0.9%	build	0.8%	institution	1.1%
15	report	1.3%	behavior	0.8%	effect	0.8%	base	1.1%
16	formal	1.2%	approach	0.8%	benefit	0.8%	sti	1.0%
17	file	1.1%	affect	0.8%	poverty	0.8%	private	0.9%
18	aim	1.0%	water	0.8%	attention	0.8%	institutional	0.9%
19	evaluation	1.0%	natural	0.8%	national	0.8%	measure	0.8%
20	search	0.9%	plan	0.8%	economy	0.8%	national	0.8%

Table 4: Estimated Latent Topics using LDA (3)

ID	Policy science Topic 9 (T_9)		Research design Topic 10 (T_10)		Quantitative methodology Topic 11 (T_11)		Qualitative methodology Topic 12 (T_12)	
1	policy	18.5%	research	9.3%	analysis	6.7%	student	6.8%
2	public	5.7%	lecture	4.3%	method	5.5%	class	3.1%
3	understand	5.5%	learn	2.8%	student	3.5%	develop	2.5%
4	health	5.1%	plan	2.3%	theory	2.6%	information	2.4%
5	make	3.1%	study	2.2%	data	2.3%	discussion	2.2%
6	political	1.8%	basic	2.2%	model	2.1%	seminar	1.7%
7	process	1.6%	field	2.0%	include	1.6%	presentation	1.5%
8	issue	1.6%	problem	1.9%	topic	1.6%	paper	1.4%
9	decision	1.3%	case	1.8%	apply	1.5%	need	1.3%
10	government	1.3%	economics	1.7%	provide	1.4%	study	1.3%
11	intervention	1.1%	base	1.6%	technique	1.3%	write	1.3%
12	politics	1.0%	conduct	1.4%	aim	1.3%	skill	1.2%
13	approach	0.9%	practice	1.1%	cover	1.2%	read	1.1%
14	impact	0.9%	theory	1.1%	exercise	1.2%	present	1.1%
15	evidence	0.8%	time	1.1%	various	1.2%	relevant	1.0%
16	evaluate	0.7%	ethics	0.9%	analyze	1.2%	group	1.0%
17	influence	0.7%	acquire	0.9%	quantitative	1.1%	major	1.0%
18	able	0.7%	range	0.9%	example	1.1%	final	0.9%
19	follow	0.7%	relate	0.9%	practical	1.0%	experience	0.9%
20	goal	0.6%	negotiation	0.8%	introduce	0.9%	review	0.8%

Table 5: Composition of topics by course (examples)

Title	T_1	T_2	T_3	T_4	T_5	T_6
Introduction to Quantitative Methods (GRIPS)	5.0%	4.0%	7.0%	5.0%	4.0%	4.0%
Public Economics (GRIPS)	6.6%	7.7%	6.6%	4.5%	9.9%	5.6%
Intellectual Property Rights and Regulation (METU (MS))	1.8%	1.8%	2.2%	1.8%	78.7%	1.8%
Advanced Scientometrics (Stellenbosch)	8.5%	4.3%	6.0%	4.3%	5.2%	9.4%

Title	T_7	T_8	T_9	T_10	T_11	T_12
Introduction to Quantitative Methods (GRIPS)	4.0%	4.0%	6.0%	7.9%	44.8%	4.0%
Public Economics (GRIPS)	6.6%	8.8%	7.7%	15.2%	12.0%	8.8%
Intellectual Property Rights and Regulation (METU (MS))	1.8%	2.6%	1.8%	1.8%	1.8%	2.2%
Advanced Scientometrics (Stellenbosch)	14.4%	11.1%	7.7%	10.2%	9.4%	9.4%

These topics can be categorised into broadly-defined groups. Fig1 shows the clustering of topics using the Ward method. This is calculated based on variance of each topic which is defined as normalised sum of variances of top words' attached to each topic. One distinct group is *Quantitative research methodologies* (T_11), which shown to be distant from the other topics. By contrast, *Policy science* (T_9), *Science and society* (T_1), *Innovation and policy* (T_8), *Law and ethics* (T_2), and *Qualitative research methodologies* (T_12) are categorised as neighbouring topics; their proximity can be partly explained by the fact that they share common methodologies. *Development and STI* (T_7) is similar to the above topics. *Intellectual assets* (T_5) and *Technology adoption to society* (T_3) may be categorised as a similar group. Together with *Energy and Environment* (T_6), *Research design* (T_10), and

Technology and management (T_4), T_5 and T_3 constitute a broadly defined similar topic group.

Fig 1: Clustering topics using the Ward method

Appendix 2 shows the composition of topics by education programs, calculated as a simple average of courses by programs. These are also visualized in radar charts and shown on our website (https://scirex.grips.ac.jp/sti-syllabi-radar). They are valuable because they show each programme's coverage of educational content.

Figure 2 is a map of education programs and Figure 3 is a cluster dendrogram of programs by type. METU (MS and PhD), HSE and POLY may be clustered as one group with more management-intensive course components, including *Technology adoption* (T_3), *Technology and management* (T_4), and *Innovation and policy* (T_8). Michigan and CMU may also be defined as one group with a strong focus on *Energy and environment* (T_6) and *Policy Science* (T_9). UNU-MERIT also constitutes a single group, due to its intensive coverage of *Development and STI* (T_7).

Figure 2: Mapping STI education programmes

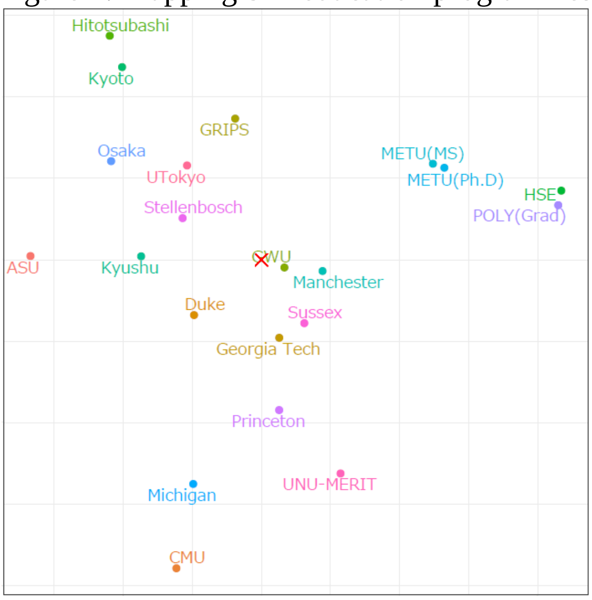
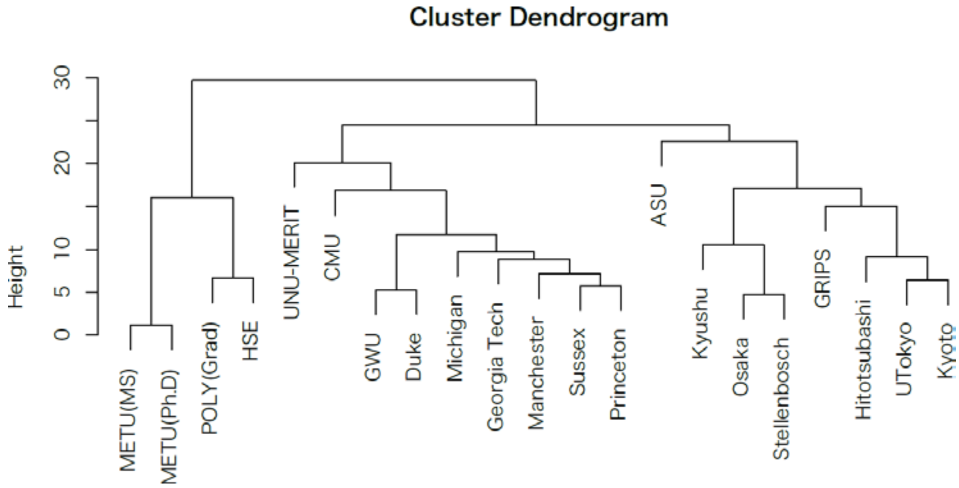


Figure 3: Clustering STI education programmes by type



Conclusion
As the data are far from complete, it is still quite difficult to draw definitive conclusions. However, this exercise may increase understanding of the 'core' of STI studies. Martin (2012) has described *Economics*, *Sociology*, *Management*, and *Organization studies* as core disciplines that made a significant contribution to building science policy and innovation studies (SPIS) in what he refers to as the 'pre-history period'. As the fields matured, several clusters emerged, relating to (i) the economics of innovation, technology, and growth; (ii) management of industrial innovation and the resource-based view of the firm; (iii) organisations and innovation; (iv) systems of innovation; (v) sociological and other contributions to SPIS; and (vi) measuring technology and innovation. These observations partially coincide with our results. However, we have also found other areas, such as *Law* and *Policy Science* that may play a more significant role in constituting science of science, technology, and innovation policy.

This research-in-progress paper reports on our efforts to develop curricula designed to enhance the human resources necessary for future evidence-based policymaking. To achieve this purpose, we have conducted both qualitative research (which has revealed shared terminology and significant questions for both policymakers and researchers) and quantitative research (which has defined the architecture of different STI education programmes). This report focuses on the quantitative topic-model method. We recognise that it is necessary to examine the relationship between educational programmes and their outcomes, including the social impact of graduates. Additional time is needed for observation. Meanwhile, it is important to develop consistent quantitative analyses to define the specialities of each STI programme. For a robust analysis, it will be important to include more time-series data and detailed information on the syllabi used by other relevant educational programmes.

References
Alexander, J., Hart, D.M., & Hill, C.T. (2016). *Enhancing the Usefulness of Science of Science and Innovation Policy (SciSIP): Research and Agenda-Setting Workshop.* Retrieved

from http://davidhart.gmu.edu/wp-content/uploads/2011/05/GMU-SRI-SciSIP-workshop-report-FINAL.pdf.

Blei, D.M., Ng, A.Y., & Jordan, M.I. (2013). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993–1022.

Ena, O., Mikova, N., Saritas, O., and Sokolova, A. (2016). A methodology for technology trend monitoring: the case of semantic technologies, *Scientometrics*, September 2016, Volume 108, Issue 3, pp 1013–1041, https://link.springer.com/article/10.1007/s11192-016-2024-0

Griffiths, T. (2014). *Gibbs sampling in the generative model of Latent Dirichlet allocation*, Technical Report. Doi: 10.1.1.138.3760.

Griffiths, T. and Steyvers, M. (2004). 'Finding scientific topics', *Proceedings of the National Academy of Sciences.*

Higher School of Economics (2018). Mapping the Radical Innovations in Food Industry: A Text Mining Study. *Higher School of Economics Research Paper* No. WP BRP 80/STI/2018, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3143721

Husbands Fealing, K., Lane, J.I., Marburger, J.H., III, & Shipp, S.S. (2011). *The Science of Science Policy: A Handbook.* Stanford, CA: Stanford Business Books.

JST-CRDS, (2010). Towards Realization of Evidence-based Policy Formation: Development of Science of Science, Technology and Innovation Policy, CRDS-FY2010-SP-13. (In Japanese). Retrieved from http://www.jst.go.jp/crds/pdf/2010/SP/CRDS-FY2010-SP-13.pdf.

Marburger, J.H., III. (2005). Wanted: Better benchmarks. *Science* 308(5725), 1087.

Martin, B.R. (2010). Inside the public scientific system: Changing modes of knowledge production. In R. Smits, S. Kuhlmann, & S. Philip (Eds.), *The Theory and Practice of Innovation Policy: An International Research Handbook*, (pp. 25–50). Cheltenham, UK: Edward Elgar.

Martin, B.R. (2012). The evolution of science policy and innovation studies. *Research Policy*, 41(7), 1219–1239.

Martin, B.R., Nightingale, P., & Yegros-Yegros, A. (2012). Science and technology studies: Exploring the knowledge base. *Research Policy*, 41(7), 1182–1204.

Martin, B.R. (2016). Twenty challenges for innovation studies. *Science and Public Policy* 43(3), 432–450. doi: 10.1093/scipol/scv077.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality, in Proceedings of NIPS

Okamura, A., Hayashi, S., Koshiba, H, and Nishimura, Y. (2017). Identifying core questions for STI studies: An exploratory international comparison of syllabi. *Science, Technology and Innovation Indicators 2017*. Paris: ESIEE.

https://sti2017.pars/wp-content/uploads/2017/11/sti-2017-book-of-abstracts.pdf

Organisation for Economic Co-operation and Development, (2016). *Blue Sky Forum: Agenda*. Retrieved March 22 2017 from: http://www.oecd.org/sti/blue-sky-2016-agenda.htm

Phan, Xuan-Hieu and Nguyen, Cam-Tu (2007). GibbsLDA++: A C/C++ implementation of latent Dirichlet allocation (LDA). Retrieved from http://gibbslda.sourceforge.net/

National Science and Technology Council & Office of Science and Technology Policy. (2008). *The Science of Science Policy: A Federal Research Roadmap*. Retrieved from http://cssip.org/docs/meeting/science_of_science_policy_roadmap_2008.pdf

SciREX Symposium (2015). *Toward Co-evolution of Science for Policy and Policy Making Process in Science, Technology and Innovation Policy - Challenges and Future Perspectives.* Retrieved from http://www.scirex2015.tokyo/programEG.html

Sen, A. (2015). *SoSP: Are We There Yet?* Presentation at the 2015 AAAS Forum. Retrieved from http://prezi.com/mmvsh7kff5ao/?utm_campaign=share&utm_medium=copy

Teich, A.H. & Feller, I. (2009). *Toward a Community of Practice: Report on the AAAS-NSF SciSIP Grantees Workshop.* Washington, DC: American Association for the Advancement of Science. Retrieved from http://www.aaas.org/sites/default/files/scisip_report_2009.pdf.

Appendix 1: List of STI education programs covered in the analysis

Japan:

GRIPS Innovation, Science and Technology Policy Programme (GIST), National Graduate Institute for Policy Studies (GRIPS) * Master's and doctoral level degree programmes. Retrieved from http://gist.grips.ac.jp/en/education/.

Science, Technology and Innovation Governance Programme (STIG), The University of Tokyo *Certificate programme. Retrieved from http://stig.pp.u-tokyo.ac.jp/en/program_curriculum.html.

Innovation Management Policy Programme (IMPP), Hitotsubashi University *Certificate programme. Retrieved from http://impp.iir.hit-u.ac.jp/syllabus/.

Programme for Education and Research on Science and Technology in Public Sphere (STiPS), Osaka University and Kyoto University * Joint certificate programme. Retrieved from http://stips.jp/english/.

Center for Science, Technology and Innovation Policy Studies (CSTIPS), Kyushu University * Certificate programme. Retrieved from http://www.sti.kyushu-u.ac.jp/english/index.php

UK:

MSc Innovation Management and Entrepreneurship, Manchester Institute of Innovation Research, The University of Manchester Alliance Manchester Business School (Manchester). Retrieved from https://www.mbs.ac.uk/media/ambs/content-assets/documents/masters/2016/ime-structure-2016.pdf.

Science and Technology Policy MSc, Science Policy Research Unit (SPRU), School of Business, Management and Economics, University of Sussex. Retrieved from http://www.sussex.ac.uk/bmec/internal/departments/spru/pgcourses/2015/N1501T/57010

USA:

Master of Science and Technology Policy, Consortium for Science, Policy & Outcomes (CSPO), Arizona State University (ASU). Retrieved from https://sfis.asu.edu/student-life/current-upcoming-courses.

Energy Science, Technology and Policy (EST&P), Institute of Technology, Department of Engineering and Public Policy (PPP), Carnegie Mellon University. Retrieved from https://enr-apps.as.cmu.edu/open/SOC/SOCServlet/search

International Science and Technology Policy Master of Arts, Elliott School of International Affairs, Center for International Science and Technology Policy (CISTP), George Washington University. Retrieved from https://elliott.gwu.edu/graduate-course-descriptions/international-science-technology

Master of Science in Public Policy, School of Public Policy, Georgia Institute of Technology. Retrieved from http://www.spp.gatech.edu/courses

MIT Graduate Certificate Program in Science, Technology and Policy, School of Engineering, Engineering Systems Division, Technology and Policy Program (TPP), Massachusetts Institute of Technology (MIT). Retrieved d from: http://tpp.mit.edu/index.php/home/masters-program/tpp-sm-curriculum/

Master of Arts in Bioethics & Science Policy, Duke Initiative for Science & Society, Duke University. Retrieved from https://scienceandsociety.duke.edu/learn/ma/curriculum/

Science, Technology, and Public Policy Program: Graduate Certificate Program, Gerald R. Ford School of Public Policy, University of Michigan. Retrieved from http://www.stpp.fordschool.umich.edu/graduate-certificate/coursework/

The Program in Science, Technology, and Environmental Policy: Ph.D. Program, Woodrow Wilson School of Public and International Affairs, Princeton University. Retrieved from http://www.princeton.edu/step/degrees/phd/

The Netherlands:

Master of Science in Public Policy and Human Development, United Nations University (UNU-MERIT), Maastricht University. Retrieved from https://www.maastrichtuniversity.nl/education/master-public-policy-and-human-development/courses-curriculum

South Africa:

Mphil/PhD in Science and Technology Studies, Centre for Research on Evaluation, Science and Technology, Stellenbosch University. http://www0.sun.ac.za/crest/ * Syllabi data was collected from personal email exchange with the faculties.

Canada:

MSc/PhD Département de mathématiques et de génie industriel, Polytechnique Montréal. http://www.polymtl.ca/magi/ * Syllabi data was collected from personal email exchange with the faculties.

Turkey:

MSc/PhD Science and Technology Policy Studies, Middle East Technical University (METU-TEKPOL). http://stps.metu.edu.tr/courses * Syllabi data was collected from personal email exchange with the faculties.

Belgium:

Bachelor Course, Ghent University. https://www.ugent.be/en * Syllabi data was collected from personal email exchange with the faculties.

Italy:

PhD in Economics and Management of Innovation and Technology, Bocconi University. https://www.unibocconi.eu/wps/wcm/connect/bocconi/sitopubblico_en/navigation+tree/home/programs/master+of+science/economics+and+management+of+innovation+and+technology/
* Syllabi data was collected from personal email exchange with the faculties.

Russia:

Master' Programme in Governance of Science, Technology, and Innovation, Institute for Statistical Studies and Economics of Knowledge, National Research University Higher School of Economics (HSE). http://www.polymtl.ca/magi/ * Syllabi data was collected from personal email exchange with the faculties.

Appendix 2

Topic components by programme (1)

Name	Science and society T_1	Law and ethics T_2	Technology adoption T_3	Technology and management T_4	Intellectual assets T_5	Energy and environment T_6
POLY (Grad)	6.7%	6.5%	11.6%	18.7%	7.8%	6.3%
Manchester	8.3%	6.1%	7.6%	11.9%	5.3%	7.0%
Sussex	8.0%	6.9%	6.9%	9.0%	6.2%	10.3%
GRIPS	8.6%	5.8%	7.2%	5.6%	7.0%	5.9%
Hitotsubashi	10.1%	5.2%	5.5%	5.9%	5.5%	6.1%
Kyoto	9.0%	7.4%	6.3%	8.4%	6.5%	6.2%
Kyushu	14.5%	5.8%	4.6%	5.6%	5.3%	12.0%
Osaka	15.5%	9.1%	6.1%	6.9%	5.2%	8.4%
U Tokyo	10.2%	7.9%	6.1%	8.6%	5.8%	6.7%
UNU-MERIT	5.6%	6.7%	6.8%	6.0%	6.2%	6.1%
HSE	6.1%	5.2%	7.9%	17.9%	6.4%	5.9%
METU (MS)	9.1%	6.8%	13.4%	8.5%	10.1%	5.9%
METU (Ph.D)	8.3%	6.7%	13.1%	8.6%	10.1%	5.7%
ASU	11.7%	6.5%	7.1%	5.2%	4.1%	4.8%
CMU	6.3%	8.8%	6.9%	7.7%	5.4%	19.8%
Duke	8.9%	13.8%	7.1%	6.3%	7.0%	7.2%
Georgia Tech	6.4%	6.5%	7.6%	9.1%	6.2%	7.7%
GWU	8.1%	12.0%	8.0%	7.7%	6.9%	6.7%
Michigan	6.9%	10.6%	7.2%	7.3%	5.7%	13.0%
Princeton	7.4%	8.6%	8.4%	7.2%	7.3%	10.2%
Stellenbosch	14.0%	8.2%	6.7%	6.3%	6.3%	7.9%

Topic components by programme (2)

Name	Development and STI T_7	Innovation and policy T_8	Policy science T_9	Research design T_10	Quantitative methodology T_11	Qualitative methodology T_12
POLY (Grad)	7.2%	9.5%	5.9%	6.3%	7.7%	5.9%
Manchester	9.0%	8.3%	8.9%	7.4%	10.0%	10.2%
Sussex	8.5%	8.9%	8.3%	7.1%	12.2%	7.6%
GRIPS	6.9%	7.4%	7.5%	11.6%	19.6%	6.8%
Hitotsubashi	6.3%	10.0%	5.9%	16.3%	11.0%	12.1%
Kyoto	6.2%	6.4%	9.0%	18.3%	7.2%	9.1%
Kyushu	5.1%	10.2%	13.7%	9.3%	9.1%	4.9%
Osaka	6.8%	6.9%	8.5%	12.6%	6.8%	7.1%
U Tokyo	7.0%	7.7%	10.8%	12.9%	8.6%	7.6%
UNU-MERIT	21.9%	7.7%	8.9%	6.0%	10.3%	7.7%
HSE	8.9%	13.9%	6.6%	6.2%	9.4%	5.8%
METU (MS)	6.5%	13.3%	6.2%	6.6%	7.4%	6.3%
METU (Ph.D)	6.5%	13.6%	6.5%	6.4%	8.0%	6.5%
ASU	7.8%	7.3%	8.2%	8.0%	6.2%	23.1%
CMU	7.1%	5.5%	6.7%	6.6%	7.4%	11.7%
Duke	7.6%	6.9%	9.5%	8.6%	8.2%	8.8%
Georgia Tech	7.3%	7.0%	15.6%	7.3%	11.4%	7.7%
GWU	7.6%	8.0%	8.0%	8.0%	11.9%	7.0%
Michigan	7.4%	6.1%	12.8%	6.7%	6.8%	9.4%
Princeton	10.1%	7.3%	10.7%	6.5%	8.8%	7.5%
Stellenbosch	9.6%	8.8%	8.8%	10.3%	6.4%	6.8%