



**3<sup>rd</sup> International Conference  
on Public Policy (ICPP3)**

**June 28-30, 2017 – Singapore**

**Panel T14P03 Session 1**

*The Data/Sensor Revolution and Public Policy*

**Title of the paper**

*Harnessing the Deluge and Drought of Text Data for Policy*

*Analysis: An Ontological Approach*

**Authors**

*Chetan Singai, National Law School of India University, India,  
chetansingai@nls.ac.in*

*Thant Syn, Texas A&M International University, USA,  
thant3303@gmail.com*

*T R Kumara Swamy, Bangalore University, India,  
kumaraswamytr31@gmail.com*

*Ajay Chandra, National Institute of Advanced Studies, India,  
ajaychandra69@gmail.com*

**Date of presentation**

*June 30, 2017*

## **Abstract**

The advent of the internet and the electronification of documents has created a deluge and a drought of text data for policy analysis. The deluge has also made it difficult to separate the signal from the noise. Easy accessibility of almost any data that are available has created the deluge. Easy means of collecting a variety of data and making them available has added to the deluge. The deluge has also made it difficult to separate the signal from the noise, best exemplified by the ratio of relevant to irrelevant items in a Google search result. The overwhelming quantity of irrelevant data often creates a drought of relevant data – there is woeful data poverty amidst a mirage of plenty which may also be dysfunctional. The situation can be corrected only by studying the ‘big picture’ of a domain to determine: (a) its ‘bright’, ‘light’, and ‘blind/blank’ spots, and (b) the antecedents and consequences of the differences in emphases using an ontological framework.

## **Introduction**

Higher education in India (HEI), rooted deep into the colonial legacy (Altbach and Selvaratnam, 2002), has undergone substantial growth and has increased its institutional capacity in the last six-and-a-half decades. At present, the higher education sector constitutes of 30.15 million students enrolled in 760 universities and 38498 colleges (AISHE, 2016) in comparison to 369,000 students studying in 27 universities and 695 colleges in 1950 (Agarwal 2009). The country now claims to be the third largest system in the world after China and USA in terms of enrolment (Tilak, 2012) and largest in the world regarding a number of higher education institutions (UGC, 2016). In the last six-and-a-half decades the higher education landscape has expanded into some complex typologies of institutions. There exist universities (public state (federal), central, private, deemed, open and the institution of national importance) and colleges (University Constituent and Affiliated (government aided,

private aided and un-aided and autonomous). Such a complex structure of higher education and institutions necessitates a robust method to map and analyse the state of these institutions either in part or/and in whole.

The urban public university (here after University) was established in 1960's in a major Indian city as one of the largest state public universities in the country with 650 affiliated colleges and two constituent colleges. The magnitude of the affiliation system was one of the main reasons for deficiency in quality of teaching, research, extension and administration. To this end, the government of the day promulgated a gazette notification to 'trifurcate' the University in August 2015 (KSU (Amendment) Act, 2015). Such a move was necessary to improve administrative efficiency, academic quality and responsiveness to students and the overall functions of the institution.

According to the electoral district or assembly constituencies - a territorial subdivision for electing members to a legislative body - the jurisdiction of each of universities was determined. According to the gazette notification, the University has been allocated 232 colleges under its jurisdiction.

We present a method of harnessing the deluge and drought of text data for policy analysis using an ontology, with an example of envisioning an emerging urban public state university – University - created by the trifurcation of an existing university. The method is like that used to map the state of higher education system in a state in India and the universities in Chile (Coronado, La Paz, Ramaprasad, & Syn, 2015; Hasan, Ramaprasad, & Singai, 2014).

## **Ontological Framework**

An ontological framework for representing the mix of vision/mission of the units that would constitute the University shown in Figure 1. It is presented using six dimensions

represented by the six columns shown in the figure. They are, from left to right: (a) Aspiration – aspiration level of the institution, (b) Type – the type of institution, (c) Scope – geographical scope of the institution, (c) Function – functions of the institution, (d) Focus – foci of the institution, and (e) Outcome – desired types of advancement., there are eight possible aspirations. The six dimensions can be seen as six parts of an institution’s vision/mission. (In the rest of the paper we will capitalize the names of dimensions and the categories within. Thus, Sciences will refer to the category in the Focus dimension/column; sciences will refer to the collection of science disciplines.)

The aspiration of an institution defines the reputation it seeks. We have used a simple four-level categorization of aspiration – premier, leading, advanced, and basic. In each category, an institution may be the only one or one among many as connoted by the article preceding the adjective. The categories are subjective; there are no specific benchmarks to distinguish the three. In India, for example, the IITs and IIMs are considered to be premier institutes; the NITs are leading institutes, and there is a large number of other accredited degree-granting institutions without a distinctive reputation which may be advanced or basic. An institution which is premier globally will also be premier in the state and nationally. On the other hand, an institution which is premier nationally may be leading in the state or advanced globally. Thus, in specifying the aspiration it would be necessary to do so with reference to the scope. While the IITs and IIMs are premier institutes nationally they may be leading institutes regionally and globally.

	<u>Aspiration</u>	<u>Type</u>	<u>Scope</u>	<u>Function</u>	<u>Focus</u>	<u>Outcome</u>	
[A/An]	Premier	[+] Central/University	[college in] The World	[for] Education	[in] Sciences	[for] Scientific	[advancement]
	Leading	Constituent	India	Research	Social Sciences	Technical	
	Advanced	Affiliated	State	Service	Professions	Economic	
	Basic	Private-Aided	City		Humanities	Social	
		Private-Unaided	University		Arts	Cultural	
		Govt.-Aided			Vocations		
		Autonomous					

## **Figure 0: An Ontological Framework for Mapping the University Data**

The Type of institution defines its governance based on the Central and State government regulations. The types of institutions constituting the University have been broadly classified as the Post Graduate (PG) Center, the Constituent colleges, and the Affiliated colleges. The Affiliated colleges can be further classified as Autonomous, Government, Private, and B.Ed. colleges. The Private colleges may be Aided or Unaided. They are defined as follows:

PG Center – The location of post graduate programs directly administered by the University.

Constituent colleges – These are affiliated to and maintained by the University.

Affiliated colleges – These are institutions which can run degree programs but are not empowered to provide degree on their own. They necessarily must be attached to the University for the purpose of awarding degree. In addition to this, the college receives the UGC grants for salaries for permanent staff and infrastructural facilities.

Autonomous colleges – These are affiliated to the University, but do not follow its syllabus or curriculum (conducts own exams/evaluation system). They obtain with financial support from the UGC (infrastructure and salaries for permanent staff).

Government colleges -- Institutions, which are managed and financed by the State Government (Department of Collegiate Education) and University Grants Commission (20:80 scheme, respectively). Ex. Infrastructure funding from UGC and appointment of faculty from DCE, State Government.

Private-Aided colleges – Institutions, which are managed by an individual, Trust, Society or other private organization and which receive regular maintenance grants from State and Central Governments (the UGC).

Private-Unaided colleges – Institutions, which are managed by an individual, Trust, Society or other private organization, which either do not receive any grant or receive a onetime ad-hoc grant for a specific purpose like building construction, strengthening of library or laboratory, subsidy towards teacher salary, etc. They but not receiving a regular maintenance grant. They are self-financed institutions.

B.Ed. College – Colleges for the specific purpose of training teachers.

The scope of an institution defines its geographical reach -- how far its reputation extends, where its students are placed, where its students are recruited, where its services are provided, where its faculty is recruited, where their research is presented and published, etc. The scope of an institution may be broadly stated as The World, India, State, City, or the University. The aspiration of an institution is concerning its defined scope. An institution may aspire to be the ‘a premier one in India’, whereas another may aspire to be ‘a basic one in the City’. While geographical boundaries are getting blurred creating a global market for higher education, they have not disappeared and are unlikely to do so soon. They continue to play a role in the strategy of an institution.

The three broad functions of an institution of higher education are Education, Research, and Service. Service is also labeled Extension in India. The three are defined as follows:

Education – The propagation of knowledge through undergraduate, graduate, professional, and other programs.

Research – The discovery and dissemination of new knowledge through publications of papers, books, and monographs, presentations, conferences, etc.

Service – The application of knowledge to solving the problems of society through extension, social, clinical, consulting services and the like.

All higher education institutions emphasize the three functions but in different measure. These days research tends to get the greatest attention in the context of rankings and ratings; in some instances, especially in professional schools, teaching may get equal attention. While in the past the service role of an institution had been relegated to third place, the recognition of its role in economic and social development is increasing the attention it is given. An important factor determining the attention given these functions is their role in revenue generation, especially in these resource constrained times. In the sciences, for example, research can generate significant revenues through grants and contracts; in the professions such as business, teaching can generate revenues through the large demand for their programs; and in some professions, such as medicine, clinical service to patients can generate revenues. Thus, a complex set of considerations determine the emphasis placed on the three functions in a higher education institution.

An institution may focus on one or many disciplines. They are broadly classified as Sciences, Social Sciences, Professions, Humanities, Arts, and Vocations. They are defined as follows:

Sciences – 'Hard sciences' such as physical, natural, biological, and the mathematical sciences.

Social Sciences – 'Soft sciences' such as economics, psychology, sociology, political science, and linguistics.

Professions – Professions related to engineering, medicine, law, social work, nursing, applied health, and public health.

Humanities – Poetry, literature, languages, etc.

Arts – Painting, fine arts, sculpture, etc.

Vocations – Vocations related to engineering, medicine, law, social work, nursing, applied health, and public health.

Many comprehensive institutions focus on all of them; the specialized institutions, often called institutes, focus on one of them or a subcategory within. The four categories differ in many ways – their emphasis on research, teaching, and service; their sources of revenue; their expenditures; their research productivity benchmarks; their sources of reputation; and, so on. Thus, the ‘business model’, if one may use that term, can vary significantly across the four categories. There can also be similar significant variations within each category. A vision of a comprehensive institution must accommodate these differences – one vision will not fit all its foci. On the other hand, a specialized institute can have a singular vision corresponding to its focus.

An institution may seek one or more outcomes of the performance of its functions in the various focus area. They may be broadly classified as Scientific, Technical, Economic, Social, and Cultural development. They are defined as follows:

Scientific advancement – Development of the society's capacity to generate and apply scientific knowledge for its own advancement.

Technical advancement – Development of the society's capacity to generate and apply technical knowledge for its own advancement.

Economic advancement – Development of the society's capacity to generate and apply economic knowledge for its own advancement.

Social advancement – Development of the society's capacity to generate and apply social knowledge for its own advancement.

Cultural advancement – Development of the society's capacity to generate and apply cultural knowledge for its own advancement.

The framework can be used to describe the aspirations of the university's institutions using structured Natural-English sentences derived from it. Each descriptor is a concatenation of a word from each dimension (column), left to right, with the adjacent words/phrases. Thus, an aspiration of one of the institutions could be to become: A leading affiliated (autonomous) college in India for education in sciences for scientific development. Another, may aspire to be simply: A basic affiliated (private-unaided) college in the City for education in humanities for social development. The 12,600 potential descriptors represent all possible  $4*7*5*3*6*5$  combinations. Many of these combinations may be irrelevant to or non-existent in the University. They may be ignored. The relevant combinations can be used to systematically map the aspirations of the institutions as described next.

The ontological framework for mapping the University's institutions offers a number of distinctive advantages. It provides a way to describe an institution, and the system, using a set of structured natural language sentences, each of which is semantically valid – makes intuitive sense. Thus, while the encoding of an institution's vision statement using the statements concatenated from the ontological framework may not be identical to the original, one can compare the two semantically and establish their equivalence.

The large number of possible concatenations encapsulated in a very parsimonious framework provides the flexibility necessary to encode a wide range of institutions in the higher education system. Thus, it allows one to map the parts (institutions) and the whole (the university) systemically and systematically. Consequently, the system can be studied at different levels of granularity – one can study it at the institution level, an aggregation of institutions by different criteria, an aggregation of all institutions, etc.

The framework is extensible in many ways. New dimensions can be added to the existing ones; new categories can be added to the taxonomies within the existing dimensions; and new sub- and super-categories can be added to the existing dimensions. Thus, the focus

on the sciences, for example, can be sub-categorized into science disciplines for a finer understanding of the institutions and the system. Extending the ontological framework will increase the possible combinations exponentially, and consequently increase its complexity. Hence, any extension must balance the benefits of ease of understanding and application of parsimony with the cost of increased complexity.

Many of these advantages are illustrated by the application in the subsequent sections. Using these properties of the ontological framework we will seek to portray the complexities of harnessing the deluge and drought of text data in formulating strategies for an emerging university.

## **Method**

There is little systematically organized data about the emerging university and its parent. We collected and collated a large volume of data about the constituent units of the emerging university by using (a) some of the authors' knowledge of the domain and sources of data, (b) online search capabilities, and (c) personal knowledge of a few experienced people. The data while not complete is perhaps the most comprehensive about the university. They were mapped onto an ontological framework, like the one used in earlier studies, to discover the 'bright', 'light', and 'blind/blank' spots in the emerging university. The ontological maps at different levels of granularity were used to recommend a strategic plan for the university. We describe the above steps in detail below.

### ***Collating the Data***

Two of the contributors to the study under the guidance of one of the primary authors made an exhaustive effort to collect all available secondary data about the institutions that will constitute the University. In the following we describe the process of data collection.

The first step was to collate the list of institutions and determine the type of each. It was a significant challenge since such a list is not available and the jurisdiction of the University was still being formalized at the time of the study. The list used in the study is based on the geographical location of the institution and the stated boundary of the University. The final list and the typing were validated through many iterations with other contributors knowledgeable about the history and structure of BU the parent of the University.

The second step was to collect all publicly available and accessible data about these institutions. The primary source of these data was the institutions website. Other sources were NAAC (National Assessment and Accreditation Council) reports, Internal Quality Assessment reports, Annual reports of the institutions and Newspaper articles.

The third step was to organize all the data electronically, by institution, at a single location using Zotero. This database was updated when new data became available. It was also the basis for coding the institutions described in the next section.

### ***Mapping the Data***

Two of the contributors under the supervision of one of the primary authors reviewed the data about each institution. Subsequently, based on their study of the institution's data, they mapped each institution by its Aspiration, Type, Scope, Function, Focus, and Outcome. Each institution was coded on a single or multiple elements of each dimension, as appropriate. Thus, an institution could be mapped for both Education and Research in Function. Each of the two contributor-mapper's coding was validated by the other, and subsequently by one of the primary authors.

All three are deeply familiar with its higher education institutions and socio-economic context. Their familiarity helped improve the validity of the data and the mapping. It helped determine gaps and errors in the data and correct them. It also helped interpret the data better in the context of the University.

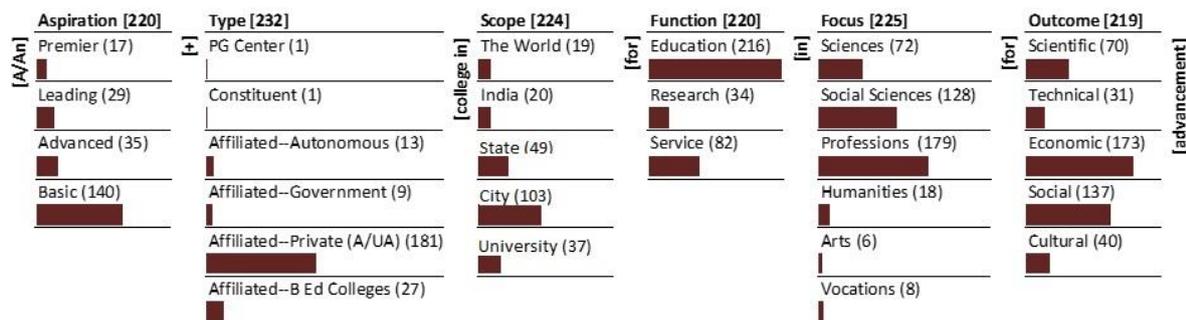
## Analyzing the Data

The results of the mapping are presented as seven visualizations using the framework – one comprehensive of the all the institutions, and one each for the six types of institutions (aided and unaided private affiliated colleges have been combined). In the following section, we discuss each map and its interpretation.

## Results

### *The University Colleges' Aspirations*

There are 232 institutions that constitute the University. It includes the PG Center (1), Constituent College (1), Affiliated-Autonomous Colleges (13), Affiliated-Government Colleges (9), Affiliated-Private-Aided/Unaided Colleges (181), and Affiliated-B.Ed. Colleges (27). The overall aspiration map of these institutions is shown in Figure 7. The number of colleges mapped to an element is shown in parentheses adjacent to the element. The bar below each element is a visual representation of the relative emphasis on the element relative to the overall maximum.



**Figure 7: The University Colleges' (232) Aspirations**

The following may be observed about the institutions from the map:

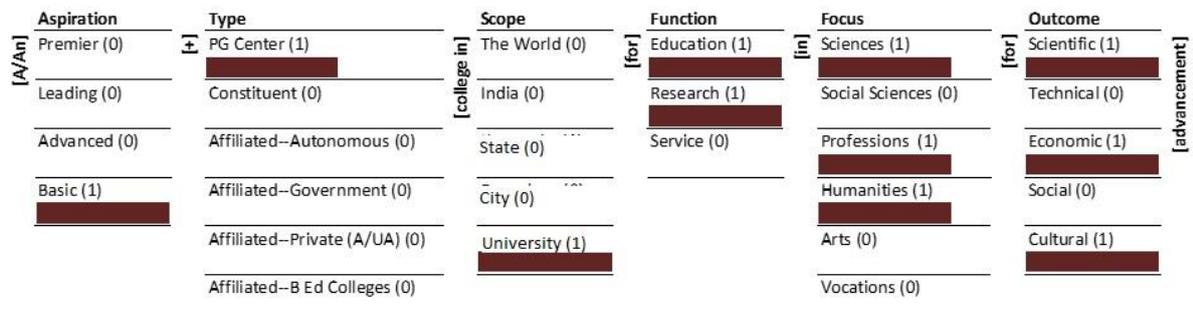
- They dominantly have basic aspirations with a few having advanced, leading, premier aspirations;
- They are dominated by affiliated, private (aided and unaided) colleges;

- They are dominantly focused locally (the City, the State, the University); some have national and global aspirations;
- They are dominantly focused in education, then service, and last research;
- They are dominantly focused on the professions, social sciences, and sciences; not on humanities, vocations, and arts; and
- They are dominantly focused on economic, social, and scientific development; not on technical and cultural development.

### ***The University PG Center’s Aspiration***

The University PG Center’s aspirations are shown in Figure 8. They may be characterized as follows:

- Basic;
- Focused locally;
- Focused on education and research, not service;
- Focused on the professions, humanities, and sciences; not on social sciences, arts, and vocations; and
- Focused on scientific, economic, and cultural development; not on technical and social development.



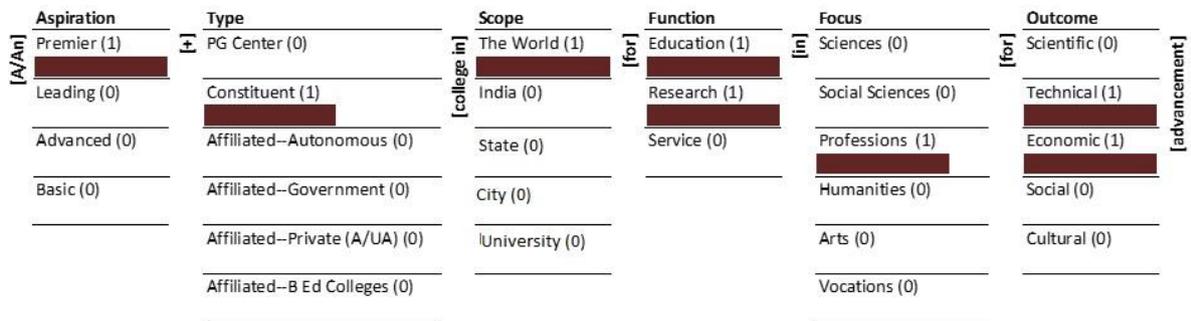
**Figure 8: The University PG Center’s (1) Aspirations**

### ***The University Constituent College’s Aspiration***

The University Constituent College’s aspirations are shown in Figure 9. They may be characterized as follows:

- Premier aspiration;

- Focused globally;
- Focused on education and research;
- Focused on the professions; and
- Focused on economic and technical development.



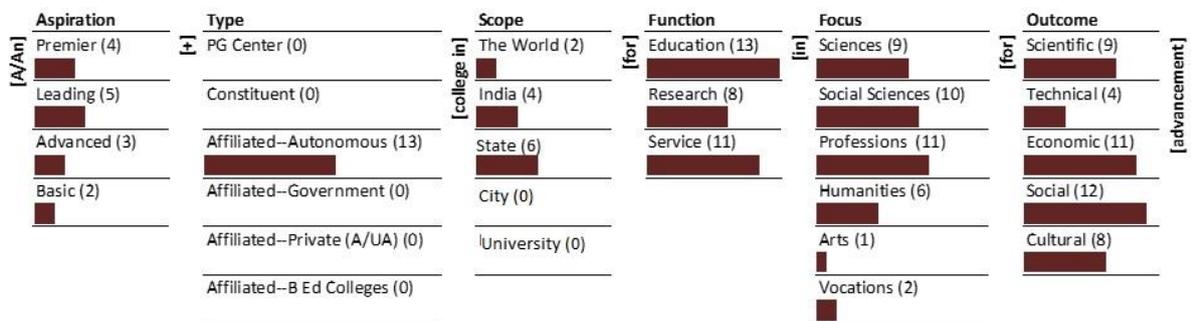
**Figure 9: The University Constituent College's (1) Aspirations**

### *The University Affiliated-Autonomous Colleges' Aspirations*

The University Affiliated-Autonomous Colleges' aspirations are shown in Figure 10.

They may be characterized as follows:

- Aspirations distributed across the spectrum – from basic to premier;
- Outward looking – focused on City, State, India, and the World;
- Dominantly focused on education, then service, and last (but significantly on) research;
- Dominantly focused on the professions, social sciences, and sciences; some focus on the humanities, vocations, and arts; and
- Dominantly focused on economic, social, scientific, and cultural development; significant focus on technical development.

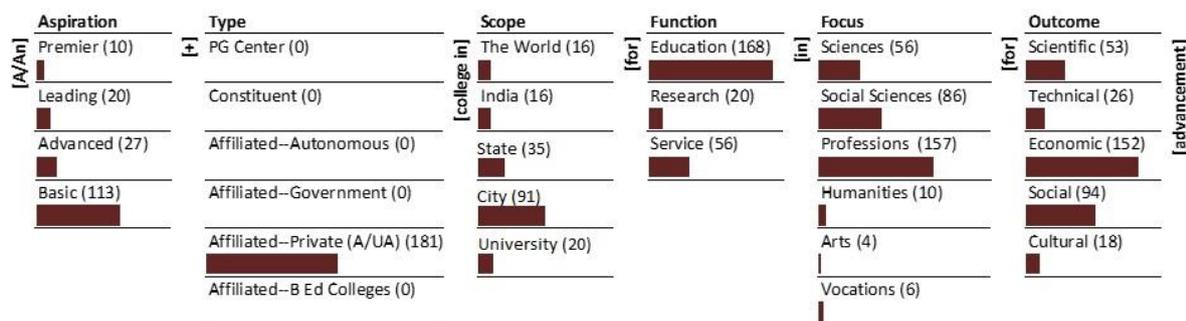


**Figure 10: The University Affiliated-Autonomous Colleges' (13) Aspirations**

## ***The University Affiliated-Private (Aided and Unaided) Colleges' Aspirations***

The University Affiliated-Private (Aided and Unaided) Colleges' aspirations are shown in Figure 11. They may be characterized as follows:

- Dominantly basic aspirations with a few having advanced, leading, premier aspirations;
- Dominantly focused locally (the City, the State, the University); some with national and global aspirations;
- Dominantly focused in education, then service, and last research;
- Dominantly focused on the professions, social sciences, and sciences; not on humanities, vocations, and arts; and
- Dominantly focused on economic, social, and scientific development; less on technical and cultural development.

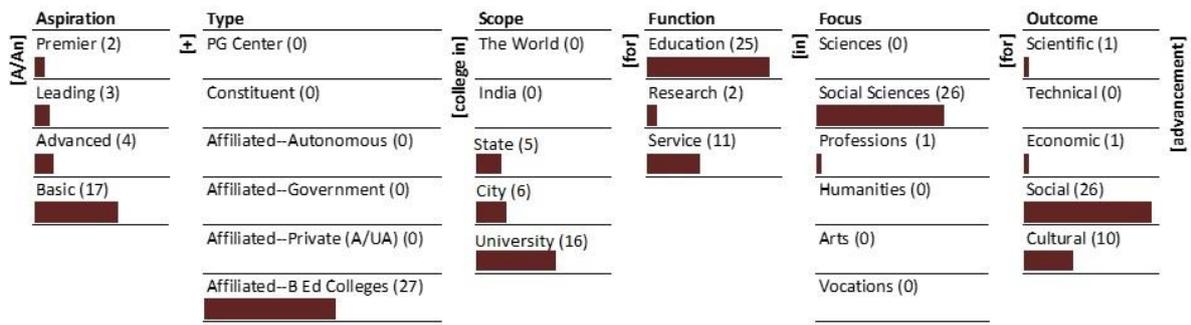


**Figure 11: The University Affiliated-Private (Aided and Unaided) Colleges' (181) Aspirations**

## ***The University Affiliate-B.Ed. Colleges' Aspirations***

The University Affiliated-B.Ed. Colleges' aspirations are shown in Figure 12. They may be characterized as follows:

- Dominantly basic aspirations with a few having advanced, leading, premier aspirations;
- Focused locally (the State, the City, the University);
- Dominantly focused in education, then service, and last research;
- Focused on the social sciences, some on professions; and
- Focused on social and cultural development; a little on economic and scientific development.



**Figure 11: The University Affiliated-B.Ed. Colleges' (28) Aspirations**

## Summary

The aspirations of the University's institutions vary significantly by the type of institution. They vary in their scope, emphasis on functions, focus and the desired outcomes. The vision for the University has to draw upon and extend these aspirations. In the following we present a vision for the University and a corresponding set of strategies and actions plans.

## Discussion

The vision of the University mapped across Aspiration, Type, Scope, Function, Focus and Outcomes are result of large amount of text data gathered from primary and secondary sources. The textual data has been systematically mapped onto the dimensions to construct a systemic perspective. Such an exercise provides a robust basis for an evidence-based analysis. The ontology framework of mapping the University data (figure 1) provides a critical lens to disaggregate the data across types of colleges under the jurisdiction of the University. Such an analysis is relevant to determine dormant patterns in examining the role and relevance of all the colleges in the University.

There are 232 institutions that constitute the University. It includes the PG Center (1), Constituent College (1), Affiliated-Autonomous Colleges (13), Affiliated-Government Colleges (9), Affiliated-Private-Aided/Unaided Colleges (181), and Affiliated-B.Ed. Colleges

(27). The affiliated-private-aided/unaided colleges comprise of the dominate profile of the University's affiliation system.

The University's key areas of function are primarily education followed by research and service. Education is one of the critical function in an affiliation system reminisced since the colonial university system.

Most of the colleges constituting the University have basic aspirations, with a few having advanced, leading and very few of them aspiring to be a premier institution. The constituent, the autonomous colleges and very few of the private affiliated colleges seek to a premier institute with global outlook. According to the data at the disaggregate level, it is evident that the colleges seeking global aspiration emphasize on research as the core function followed by education and service. Such types, levels, scope and functions of the college are potential research centers for the University and its advancement.

Similarly, the PG center, government and private unaided affiliated colleges aspire to be basic institution in the University's jurisdiction, city or the State with education as its core function. Most of these colleges provide education in social sciences, basic sciences, arts and humanities leading to social, scientific, cultural advancements.

Thus, the data for the inception of a vision report for the University contain all the elements in the ontological framework. However, the data does not encapsulate the evidence systematically and systemically for the University. The logic of the ontology is one of the main outcomes of the mapping and analysis. There is deluge of data but not coherent and organized, which essentially means data drought. The ontological approach has been instrumental in sifting irrelevant data, by illustrating the 'big picture' of a domain determining: its 'bright', 'light', and 'blind/blank' spots.

## Conclusion

The overwhelming quantity of irrelevant data often creates a drought of relevant data. To obtain a valid ‘big picture’ of a domain, in this case an emerging university, one must systematically separate the signal from the noise, the wheat from the chaff. An ontological framework can be used as a match-filter to do so. A logically constructed ontological framework that is grounded in the domain can be used as a lens to manage the deluge of data. It can also be used to discover areas in which there is potentially a drought of data – the ‘blind’ and the ‘blank’ spots.

At present the mapping of the voluminous (primarily) text data onto the ontological framework is manual. That is perhaps the most important bottleneck in enhancing the volume of data that can be processed and the speed of processing. Fully or partially automating the process can help process more data quicker. The authors are working on tools like convolutional neural networks to aid the mapping.

Another major constraint is the availability, accessibility, and the quality of the data. Some data may be available but not accessible; other data may be available and accessible, but of poor quality. There are also data that are simply not available. The iterative process of mapping and applying the results to planning can create a feedback loop which over many cycles can improve the availability, accessibility, and the quality of the data.

In this research, we have worked with the best and the broadest data publicly available. It is incomplete and of uneven quality. The portraits painted are thus limited by the comprehensiveness and the quality of the data. Despite the limitation, they are a very significant advancement on the portraits in the minds of the stakeholders and decision makers. It is likely analogous to a slightly blurred, out of focus, black and white picture. In subsequent iterations, we should be able to obtain a sharp, focused, color picture of the domain.

## References

- Altbach, P. and V. Selvaratnam (Eds.) (2002) *From Dependence to Autonomy: The Development of Asian Universities*. Kluwer Academic Publishers: Netherlands.
- Agarwal, P. (2009). *Indian higher education: Envisioning the future*. Sage Publication: India.
- AISHE. (2016). *All India Survey on Higher Education (AISHE)*, Ministry of Human Resource of Development (MHRD). New Delhi: Government of India.
- Coronado, F., La Paz, A., Ramaprasad, A., & Syn, T. (2015). *Navigating the Complexity and Uncertainty of Higher Education Systems: Ontology Mapping of Chile's Universities Proceedings of HERDSA 2015*. Melbourne, Australia.
- Hasan, T., Ramaprasad, A., & Singai, C. (2014). *Rethinking Higher Education Research: Ontology Mapping of Higher Education Systems Proceedings of HERDSA 2014*. Hong Kong.
- KSU. (2015). *Karnataka State Universities (Amendment) Act (2015)*. Department of Parliamentary Affairs, Government of Karnataka Press: Bengaluru (10 August 2015).
- Ramaprasad, A., & Syn, T. (2015). Ontological Meta-Analysis and Synthesis. *Communications of the Association for Information Systems*, 37, 138-153.
- Tilak, J. B. (2012). Higher education policy in India in transition. *Economic and Political Weekly*, 47(13), 36-40.
- UGC. (2016). *University Grants Commission Annual Report 2015-16*. VIBA Press Pvt. Ltd., New Delhi.