REGIONALIZING THE GLOBAL INNOVATION INDEX:

A PROTOTYPE SUB-NATIONAL APPROACH FOR THE UNITED STATES

Courtney Bower
Fellow, Portulans Institute
PhD Student, Regional Science, Cornell University

The Global Innovation Index (GII) is a leading tool for understanding country-level determinants of innovation. However, while the national scale is useful for many researchers and decision-makers, a comprehensive index for use at the sub-national scale is needed to understand the regional differentials in innovation that exist within a country. To address the need for greater visibility into sub-national innovation differentials, this project explores the creation of a sub-national innovation index, drawing on the structure and methodology of the GII. Termed the Sub-National Innovation Index (SNII), this project aims to develop a regionally focused innovation index for use with the eight Bureau of Economic Analysis (BEA) regions of the United States.

Below is a first look at the project: its goals, roadmap, construction, and initial results. The initial results include a single BEA region in radar chart form, along with details for how these results were obtained. The chosen BEA region is the U.S. Northeast, which is composed of the following six states: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont. Indicator data comes primarily from 2021, though the entire year range of the sourced data is from 2017 to 2023. Importantly, the process and results of this project are tentative, and the final results for each of the eight BEA regions, along with any updates to the methodology or indicator construction, will be presented in a subsequent post.

Overview: Research Question and Goals

The major objective of this project, conducted as part of the Academic Network Fellow program, was to develop a representative subset of the GII that could be effectively translated to the sub-national level. This required considering the design of the GII, locating a suitable mechanism for condensing the GII to a representative sample of indicators, translating the indicators into sub-national form, and constructing a methodology to address issues around data availability and indicator proxies.

The guiding research question of this project was the following: Can major representative elements of the GII be sub-nationalized, retaining the general strengths of the GII while offering sub-national actors' insight into regional innovation? This question was motivated both by the absence of research around sub-national variants of the GII and the need to understand regional innovation differentials in a rapidly transforming world. The goal of the project was to create a proof-of-concept sub-national innovation index, based on the GII, tailored to the specifics of the eight US BEA regions.

Project Roadmap

Figure 1, below, details the seven steps of this project. Step one began with a review of GII structure, methodology, and published research literature. Initial questions included to what extent researchers had pursued either (*i*) constructing an index based on the GII for use elsewhere, or (*ii*) condensing the GII to a representative sample of the 80 indicators.

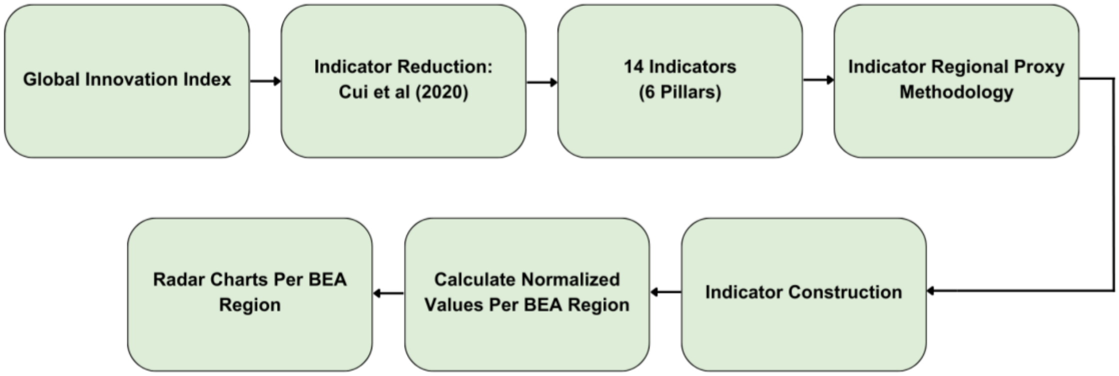

Figure 1: Sub-National Innovation Index (SNII) Project Roadmap

From the research in step one, it was determined that while variants of the GII had been produced, none were designed for the sub-national scale. Additionally, research showed that the GII could be condensed to a representative sample. Step two was the review and selection of a GII-reduction method, with Cui et al. (2020) being selected, resulting in step three's 14 indicators that were shown to be largely predictive of future GII rankings. Step four was the design of a method for reviewing each of the 14 indicators to determine to what extent the indicator would need to be modified for use at the U.S. sub-national level. In step five, each indicator was constructed using available U.S. datasets and modified, as needed, based on whether the indicator needed minor changes, moderate changes, or significant changes. As discussed later, some indicators, specifically composite indicators, needed to be reformulated as proxies for use at the sub-national scale. After applying the methodology and sourcing the data, each indicator was calculated, and scores were normalized on a 0 to 10 scale. The final step, number seven, is where the final scores were aggregated and visualized in a radar chart.

Indicator Choice and Proxy Methodology

One of the early concerns in formulating this project was that the GII datasets were often collected at the national scale, meaning that any sub-national scale index would be unable to use GII data. This meant that a sub-national index would, in large part, need to be constructed from the ground up for any given country. One can imagine that, with sufficient funding and access to accurate regional data, a country could formulate a sub-national innovation index to compare regional innovation differentials similar to what the GII does for countries. While these challenges aren't technically insurmountable, they are probably infeasible and impractical without sufficient data or the resources to produce the data.

Initial research suggested the possibility that the GII could be condensed to a representative sample of indicators, which would make the creation of a sub-national innovation index a far less daunting task than with the original 80 indicators. Several researchers, notably Cui et al. (2020) and Pence et al. (2019), pioneered methods that used artificial neural networks to estimate GII scores. A byproduct of this work was the identification of a subset of indicators that predict subsequent yearly GII ranks. Pence et al. (2019) identified 27 indicators, while Cui et al. (2020) identified 14 indicators. Figure 2 lists the 14 indicators identified by Cui et al. (2020) used in this project.

GII Pillars	Selected Indicators		
1. Institutions	Regulatory Quality (1.2.1)		
2. Human Capital and Research	Government Funding per Secondary Student (2.1.2)	Tertiary enrolment (2.2.1)	QS university ranking (2.3.4)
3. Infrastructure	Online E-Participation (3.1.4)	Environmental Performance (3.3.2)	ISO 14001 environmental certificates (3.3.3)
4. Market Sophistication	*none*		
5. Business Sophistication	Research talent in business enterprise (5.3.5)		
6. Knowledge and Technology Outputs	New businesses/th pop. 15–64 (6.2.2)	ISO 9001 quality certificates/bn PPP$ GDP (6.2.4)	High-tech exports, % total trade (6.3.3)
7. Creative Outputs	ICTs and organizational model creation (7.1.4)	Cultural and creative services exports, % total trade (7.2.1)	Entertainment and media market (7.2.3)

Figure 2: Selected SNII Indicators (14) and Their GII Pillars, or Dimensions (6)

Identifying a representative sample of GII indicators is the first challenge in building a sub-national index. Three distinct approaches offer varying insights and methodologies to address this challenge: Artificial Neural Networks, Principal Component Analysis, and Factor Analysis. Artificial Neural Networks leverage complex, non-linear relationships within data, prioritizing indicators based on their predictive strength for specific outcomes like GII scores, making it a highly data-driven predictive approach. Principal Component Analysis simplifies the dataset by transforming original variables into a smaller set of uncorrelated principal components, focusing on variance explanation within the data for simplification and interpretation rather than direct prediction. Factor Analysis, on the other hand, seeks to identify latent factors underlying observed variables, explaining correlations among them. Each method provides unique insights into the selection of indicators, from predictive accuracy and variance explanation to uncovering latent constructs.

The next challenge comes in translating the methodology of each selected indicator from the national scale to the sub-national scale. Not all indicators in the GII are created equal. Eleven of the 80 indicators in the GII are composite indexes composed of several data points, while five of the 80 are survey question responses from subject matter experts. Many, if not most, of the indicators reference datasets compiled by international institutions such as the World Bank, UNESCO, and the World Intellectual Property Forum (WIPO), among others. The methodologies used in the indicators align with the available data, and so have an international scale and scope not necessarily applicable at the sub-national level.

Given the inherent structure and limitations of the GII data and design, this project needed a methodology for translating the selected GII indicators into a suitable sub-national variant. Naturally, this process is difficult, as some indicators can be reasonably approximated, while others, such as composite indexes and surveys, might have little to no sub-national analog. Figure 3 below shows an early-stage workflow created and used to assess to what extent a GII indicator would need to be transformed in order to function in a U.S. sub-national context.

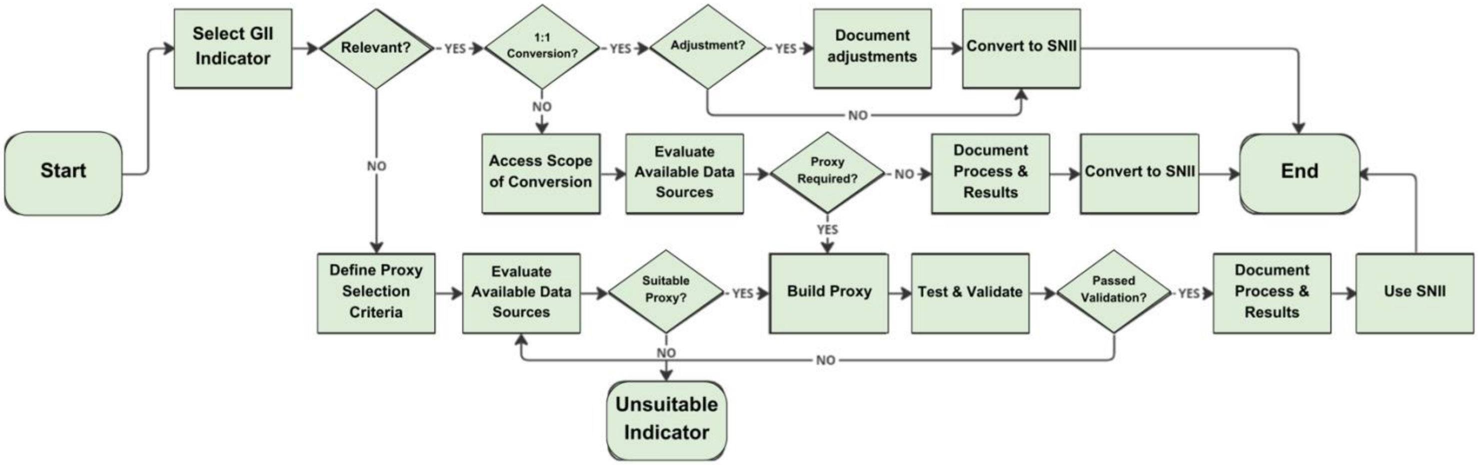

Figure 3: SNII Proxy Workflow

The workflow outlined in Figure 3 is rudimentary but does start to engage with the important components of creating a proxy indicator. Major components include the following:

1. Relevance: Evaluate the relevance of each GII indicator at the sub-national level. If an indicator is irrelevant or unmeasurable, proceed to select a proxy.
2. Proxy Selection Criteria: Define clear criteria for proxy selection that align with the conceptual meaning of the original indicator. Criteria may include data availability, regional applicability, and statistical correlation with the original indicator.
3. Data Source Evaluation: Identify potential data sources for proxy indicators, giving preference to those with comprehensive coverage, high reliability, and regular updates.
4. Iterative Testing: Evaluate the proxies within the index construction process, comparing results with national GII outcomes to ensure consistency and validity.
5. Statistical Validation: Conduct statistical tests, such as correlation analysis, to validate that the proxy maintains a strong and significant relationship with the concept it aims to represent.

Using the workflow in Figure 3 as a guide, four of the 14 indicators required no or minor adjustments, given the data in the GII was available at the sub-national level in the U.S. Six of the 14 indicators would require a full proxy, as the indicators were either a composite index, a survey, or had data sources only publicly available at the national scale. The final four indicators required a partial conversion, or moderate adjustment, with one prominent example being the use of U.S.-specific industry codes (i.e., the North American Industry Classification System, or NAICS) instead of an international variant. Figure 4 below lists each indicator, the level of the proxy, the data sources, and notes regarding the conversion from the GII to the U.S. sub-national level.

Selected Indicators	Sub-National Analogue	US-Based Data Sources	Notes
Regulatory Quality (1.2.1)	No (full proxy)	ASU DBNA Study	Major cities used over state-level. Simple averages used per city, per category
Government Funding per Secondary Student (2.1.2)	Yes (minor adjustments)	NCES & FRED	Chained GDP; primary and secondar education included
Tertiary enrolment (2.2.1)	Yes (minor adjustments)	NCES & Census	School population assumed, ages 18 to 24
QS university ranking (2.3.4)	Yes (minor adjustments)	QS University	QS World Rankings
Online E-Participation (3.1.4)	No (full proxy)	Digital States Survey 2022	Govtech.com survey assessment per state
Environmental Performance (3.3.2)	No (full proxy)	CDC environmental data	Full composite proxy, data only captures about 30% of the proxied EPI
ISO 14001 environmental certificates (3.3.3)	No (full proxy)	Factiva	Factiva database search of press releases; population data
Research talent in business enterprise (5.3.5)	Somewhat (partial proxy)	BLS data	Some employment numbers redacted; sector (2-digit) NAICS code; total employment, not FTE only
New businesses/th pop. 15–64 (6.2.2)	Yes (minor adjustments)	Census	Annual business applications by county
ISO 9001 quality certificates/bn PPP$ GDP (6.2.4)	No (full proxy)	Factiva	Factiva database search of press releases; population data
High-tech exports, % total trade (6.3.3)	Somewhat (partial proxy)	Census US Trade Online	Commodities, excludes services
ICTs and organizational model creation (7.1.4)	Somewhat (partial proxy)	BLS data & USNews state rankings	GII discontinued in 2022
Cultural and creative services exports, % total trade (7.2.1)	Somewhat (partial proxy)	Census US Trade Online	Combined US Census and USA Trade online; 8% of Census data withheld
Entertainment and media market (7.2.3)	No (full proxy)	BLS revenue data	No advetising, only consumer spending on entertainment

Figure 4: SNII Indicator Conversion Results

There is one important point to note about the conversion process and resulting U.S. sub-national indicators. Indicators ISO 14001 environmental certificates (3.3.3) and ISO 9001 quality certificates/bn PPP$ GDP (6.2.4) both use ISO certification counts, which are made public at the national level, but are proprietary – or not available – at the sub-national level. Certification counts per U.S. state would make both indicators relatively straightforward to convert to the sub-national level; however, without this data, a proxy needs to be created that approximates what the ISO certificate measures. The ISO 14001 environmental certificates indicator (3.3.3) and the ISO 9001 quality certificates/bn PPP$ GDP indicator (6.2.4) were proxied using press releases for both certification types per region.

Initial Results

Figures 5 through 8 show the tentative results for the first test region of this project: the BEA Northeast region, which contains the six states mentioned above. Results were first calculated per indicator (Figure 5) and visualized in radar chart form (Figure 6). Appendix A provides additional details about how each indicator was composed and calculated.

SNII - United States, BEA Region "Northeast"	
Regulatory Quality (1.2.1)	6.3
Government Funding per Secondary Student (2.1.2)	3.1
Tertiary enrolment (2.2.1)	7.1
QS university ranking (2.3.4)	9.5
Online E-Participation (3.1.4)	8.7
Environmental Performance (3.3.2)	8.2
ISO 14001 environmental certificates (3.3.3)	1.7
Research talent in business enterprise (5.3.5)	10
New businesses/th pop. 15–64 (6.2.2)	1.7
ISO 9001 quality certificates/bn PPP$ GDP (6.2.4)	3.3
High-tech exports, % total trade (6.3.3)	2.4
ICTs and organizational model creation (7.1.4)	5.2
Cultural and creative services exports, % total trade (7.2.1)	1.2
Entertainment and media market (7.2.3)	2

Figure 5: Indicator Calculations for the BEA Northeast Region

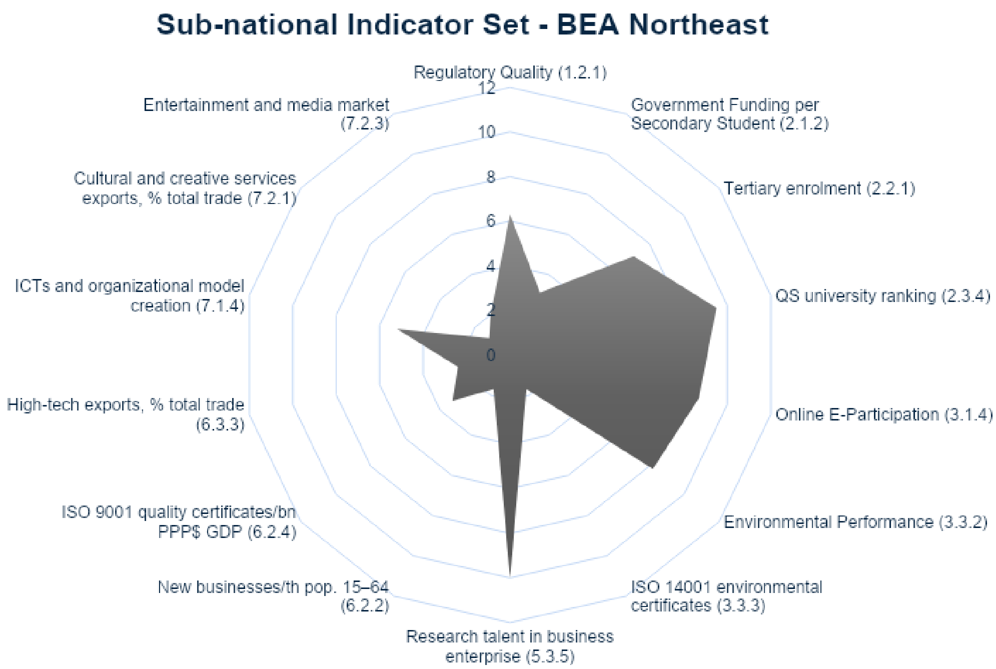

Figure 6: BEA Northeast Region Indicator Radar Chart

Next, each indicator was aggregated into its respective GII dimension (Figure 7) and again visualized (Figure 8). Careful readers may note that one of the seven GII dimensions, the Market Sophistication dimension, is missing. The absence of the Market Sophistication dimension is correct, as no indicators in this dimension were present in the GII-reduction method adapted from Cui et al. (2020).

SNII - United States, BEA Region "Northeast"	
GII Dimension	*Normalized Scores*
Institutions	6.3
Human Capital and Research	5.1
Infrastructure	6.2
Business Sophistication	10
Knowledge and Technology Outputs	2.466666667
Creative Outputs	2.8

Figure 7: Aggregate Indicator Calculations for the BEA Northeast Region

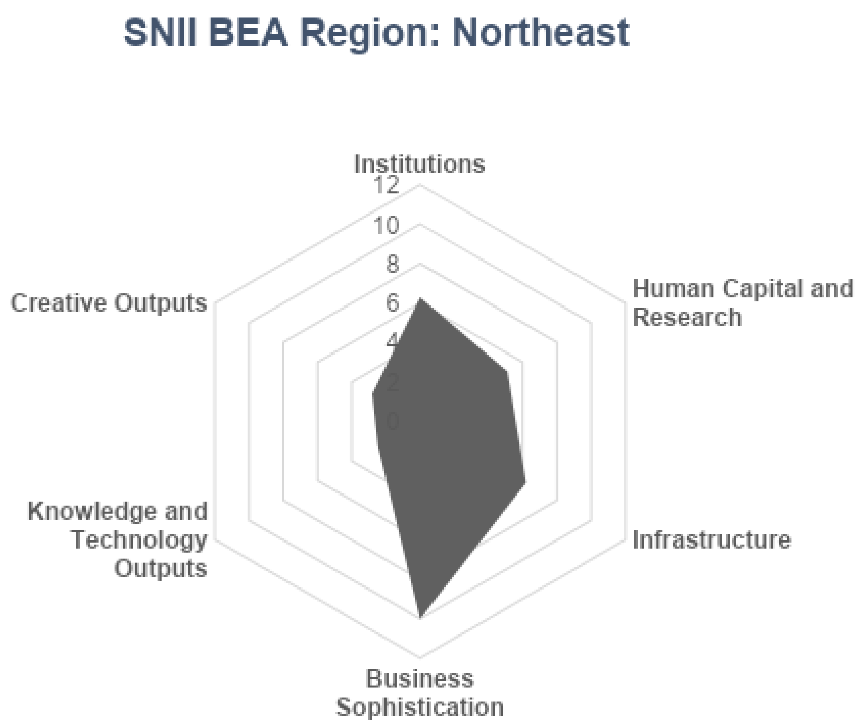

Figure 8: BEA Northeast Region Aggregate Indicator Radar Chart

Lessons Learned So Far & Project Next Steps

Work on this project has highlighted several crucial lessons in developing a sub-national innovation index. The first lesson is the inherent complexity involved in constructing the GII. Composed of 80 indicators of different types and drawing on a range of data sources, each with its specific methodology, the GII brings together many different assumed determinants of innovation. Reducing the total number of indicators makes the task of creating a sub-national index possible for a low-budget project but comes at the cost of removing part of the richness that makes the GII a useful tool. The second lesson is that much of the GII data is driven by international efforts to collect and make the data available for public use, which carries with it a preference for country-level analysis. Were national governments, or other institutions, interested in understanding regional innovation differentials within their countries, concerted effort would likely be needed to gather, and consolidate, the data required to build a sub-national, country-specific index. The third lesson concerns the use of proxies and a proxy workflow, which need to be carefully considered and designed to ensure indicators in the GII are adequately represented at the sub-national level.

Finally, as this is an initial stage of the project, it is prudent to withhold interpretation of the results of the Northeast BEA region. Based on the lessons learned, it is probably best to consider this project as either inspired by or informed by the GII, versus a condensed and adapted version of the GII for sub-national use. The above results can be expected to change following the inclusion of missing data, critique from researchers, and adjustments to the methodology. Once the methodology is finalized, missing data is included or proxied, and the results are tabulated for the eight BEA regions, this project could offer important insights to decision-makers interested in the sub-national innovation landscape in the U.S.

While building any index is a practice of making informed, defensible decisions, the absence of robust data and the nature of composite indicators made it expedient to have data availability drive what was included in this project. Given sufficient technical and financial resources, a sub-national innovation index could be built that is more methodology-driven versus data availability-driven (see Appendix B for a hypothetical budget).

References

Cui, R., Sun, J., Li, Y., Yang, K., & Wu, X. (2020). Data-driven approach with artificial neural network for Global Innovation Index re-evaluation. 2020 6th International Conference on Big Data and Information Analytics (BigDIA), Big Data and Information Analytics (BigDIA), 2020 6th International Conference on, BIGDIA, 88–93. https://ieeexplore.ieee.org/document/9384552.

WIPO (2022). Global Innovation Index 2022: What is the future of innovation-driven growth? *Geneva: World Intellectual Property Organization*

WIPO (2023). Global Innovation Index 2023: Innovation in the Face of Uncertainty. *Geneva: World Intellectual Property Organization*

Pençe, İ., Kalkan, A., & Çeşmeli, M. Ş. (2019). Estimation of the Country Ranking Scores on the Global Innovation Index 2016 Using the Artificial Neural Network Method. International Journal of Innovation and Technology Management, 16(04), 1940007. https://doi.org/10.1142/S0219877019400078.

Appendix A

SNII Indicator Calculation Method

Table 1 provides detail regarding the selected GII indicator, and the steps used to convert the indicator to the SNII.

Selected Indicators	**Steps**
Regulatory Quality (1.2.1)	1. Calculated values per indicator for each state. 2. Simple average of single indicator for all six states. 3. Simple average of all indicators for all six states.
Government Funding per Secondary Student (2.1.2)	1. GRP per capita calculated. 2. Calculated values for indicator per each state. 3. GRP per capita divided by calculated values.
Tertiary enrolment (2.2.1)	1. Calculated values per indicator for each state. 2. Simple average of single indicator for all six states. 3. Average fall enrollment divided by average total population aged 18 to 24.
QS university ranking (2.3.4)	1. Calculated values for top three universities in region. 2. Averaged scores.
Online E-Participation (3.1.4)	1. Grades for each state calculated. 2. Grades converted to numerical scale. 3. Simple average of all six states.
Environmental Performance (3.3.2)	1. Select indicators from CDC National Environmental Public Health Tracking Network. 2. Calculate nine indicators per state. 3. Normalize indicators per state. 4. Simple average of indicators for each state. 5. Simple average of total from each state.
ISO 14001 environmental certificates (3.3.3)	1. Conduct lookup for all ISO 14001 press releases in the BEA region using Factiva database. 2. Conduct lookup for all ISO 140001 press releases in the USA using Factiva database. 3. Divide region by USA counts. 4. Calculate GRP for the region. 5. Divide the region by USA counts by the GRP.
Research talent in business enterprise (5.3.5)	1. Identify all NAICS codes representing researchers in the business sector. 2. Calculate, using BLS data, researcher employment by state for all NAICS codes. 3. Sum research employment. 4 Identify, using BLS data, total employment by state for all NAICS codes. 5. Sum total employment. 6. Divide researcher employment by total employment for BEA region. 7. Calculate, using BLS data, research employment for the US. 8. Calculate, using BLS data, total employment for the US. 9. National average percent of researchers in business set to 100. 10. BEA NE is above the national average on a 0 to 100 scale.
New businesses/th pop. 15–64 (6.2.2)	1. Calculate number of businesses per state. 2. Calculate population counts for ages 15 - 64 by state. 3. For each state, divide business counts by population counts and multiple by 1000. 4. Calculate simple average for all six states.
ISO 9001 quality certificates/bn PPP$ GDP (6.2.4)	1. Conduct lookup for all ISO 14001 press releases in the BEA region using Factiva database. 2. Conduct lookup for

	all ISO 140001 press releases in the USA using Factiva database. 3. Divide region by USA counts. 4. Calculate GRP for the region. 5. Divide the region by USA counts by the GRP.
High-tech exports, % total trade (6.3.3)	1. Identify all NAICS codes representing high-tech. 2. Calculate export values for the entire year per state. 3. Add the states. 4. Calculate the entire US. 5. Divided the total state value by the US national value.
ICTs and organizational model creation (7.1.4)	1. Identify five questions from the Business Response Survey as proxy. 2. Calculate values per question per state. 3. Calculate the simple average of each question for all six states. 4. Sum all simple averages. 5. Determine values of broadband and internet access ranks per state. 6. Normalize each value on a scale of 0 to 100. 7. Calculate simple average normalized values. 8. Sum all questions and ranks and average final results by four.
Cultural and creative services exports, % total trade (7.2.1)	1. Identify all NAICS codes for cultural and creative exports using two systems: US Trade and US Census. 2. Calculate total values for a single year per state for each system. 3. Calculate national total values for both systems. 4. Divide the combined state value by the total US value.
Entertainment and media market (7.2.3)	1. Calculate estimate of size of US consumer media market using BLS data spending patterns. 2. Determine US % population in 15-69 age range. 3. Multiply spending patterns and age range. 4. Calculate estimate of size of BEA region consumer media market using BLS data spending patterns. 5. Determine BEA region % population in 15-69 age range. 6. Divide BEA region value by US value.

Table 1: GII to SNII Indicator Steps

Appendix B

SNII Budget Estimation

This section outlines a hypothetical budget for producing the SNII for the eight BEA regions of the United States. This budget assumes the project will be in partnership with a university and research institute that can offer cost offsets through in-kind contributions. The estimated direct cost budget for the project, including personnel, ranges from $150,000 to $325,000. The total in-kind contributions, primarily covering expertise, software licenses, and infrastructure, are estimated to range from $45,000 to $100,000. The total estimated project cost including direct costs and in-kind contributions is $195,000 to $425,000.