
Plan Overview

A Data Management Plan created using DMPonline

Title: Public Affairs Index (PAI) 2019

Creator: Meena Nair

Affiliation: IDRC

Template: DCC Template

Project abstract:

Measuring the quality of governance poses a challenge. To address this, Public Affairs Centre (PAC) pioneered and developed an evidence-based, data-driven framework to rank states on governance. This framework, the Public Affairs Index (PAI), besides adding to the discourse on governance, enables measurement of the quality of governance. PAI as a composite index captures the complexities of governance within and across the states of India. It is a statistical index that ranks the socially, culturally, politically and economically diverse states on standardised metrics based on a scale of 0 to 1; 0 being the worst performing state and 1 being the best performing state. Using an indigenous methodology and based on valid and reliable data, PAI adopts a rigorous process to calculate the scores of the states and derive the inter-se rankings. The quantum of data sets used to calculate the index, all drawn from the public domain and largely government data, are humongous, with each data set reflecting diverse aspects of governance. PAC has been annually releasing the PAI since 2016. PAI 2019 aimed to rebuild the methodology in a more scientific manner. One of the ways to achieve this has been to use statistical techniques such as Principal Component Analysis (PCA) and Factor Analysis (FA) to remove autocorrelations and apply a rigorous method of applying weights. The PAI 2019 has incorporated the Sustainable Development Goals (SDGs) in its framework and has also referred to the World Governance Index. The model of PAI 2019 has strived to make use of data analytics to its optimal and attempted to provide various correlations which would provide interesting insights to interested stakeholders.

ID: 49214

Last modified: 20-12-2019

Grant number / URL: -

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Public Affairs Index (PAI) 2019

Data Collection

What data will you collect or create?

Public Affairs Index (PAI) is a data driven framework to rank all the states of India on various parameters of Governance. The data collected for PAI covered various aspects of Governance including Health, Education, Infrastructure, Women, Children, Crime, Law and Order, Fiscal Management, Economic Freedom, Social Protection and Environment.

How will the data be collected or created?

The data for PAI is collected from secondary sources. All the data is collected from the Union/Central Government websites. PAI did not focus on the data that was published by the individual State Governments to avoid bias.

The secondary data was collected from the public domain. The data ranged from diverse sectors including Education, Health, Nutrition, Women, Children, Fiscal Responsibility etc. The raw data was standardised to ensure uniformity in comparing the performance of the states. Thus, the data set can be characterised by ratios, percentages and rate. After standardisation, the data were converted into the scale of 0 to 1. In addition, the data sets were converted into a normal distribution, i.e. in the scale of -1 and 1. Finally, weightages were applied on the indices to arrive at the scores and rankings.

Documentation and Metadata

What documentation and metadata will accompany the data?

The PAI data was documented in a well-drafted matrix. The time-period, data source, methodology and notes of each of the data points (each parameter) was documented in the matrix.

Metadata was not part of PAI.

Ethics and Legal Compliance

How will you manage any ethical issues?

The data for PAI was collected only from the Union Government from websites that published the data publicly. PAI did not include any personal information of the respondents nor did it include any sensitive information. Thus, PAC is confident that there were no ethical issues.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

All data that were collected from secondary sources, was given due credit. In the PAI report, PAC has kept a separate section on 'Annexures' where all the data sources are provided to tackle the issue of copyright.

Storage and Backup

How will the data be stored and backed up during the research?

The data analysis was conducted with the help of the R software. The codes are documented and have been saved in the cloud. The data has been recorded in a very systematic manner. All the excel files and the R files have been documented separately. In

addition, the raw data and the processed data has been recorded separately so that one can revisit the same seamlessly.

How will you manage access and security?

The data and the R codes are accessible to the team members. The codes are not yet made public for data security. Any individual will first have to share their details to the PAI team that includes name, organisation, purpose of downloading the data etc. Only on receipt of the above information, PAC would make the data accessible to the individuals.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

The data collected is also cumulatively used to find trends in the movement of the selected indicators. And therefore the original data collected have been saved in the computers of the PAI team as well as preserved by the Senior Systems Administrator at PAC.

The entire data set that has been collated for the PAI has long term value and should be retained. These data gives an opportunity to a diverse set of researchers and subject matter experts to explore and conduct various spatial and temporal analysis.

What is the long-term preservation plan for the dataset?

The database is referred to every year to add more recent data and is therefore mostly 'live'. The team needs to work out a long-terms preservation plan for the dataset.

Data Sharing

How will you share the data?

The PAI data can be accessed at the website which is available in the public domain (pai.pacindia.org). This website enables the stakeholders to download the data as per their areas of interest and conduct their independent research.

Are any restrictions on data sharing required?

PAC believes in transparency and hence the data was shared without any hesitation. However, we ask the following information of the individuals - name, organisation, purpose of downloading the data, while sharing the data with anyone.

Responsibilities and Resources

Who will be responsible for data management?

The Team Lead for PAI will be responsible for data management. The team lead is responsible to make sure that all the data and their respective codes are documented properly and accessible to others.

What resources will you require to deliver your plan?

1. A well-functioning PAI website
2. An internal backup system or a cloud to document all the raw data and R files

