A lack of open data standards for large infrastructure projects hampers social-ecological research in the Brazilian Amazon

Comments to Authors

The manuscript addresses the lack of open data standards for large infrastructure projects in the Brazilian Amazon, which hinders both social and ecological research in the region. The subject is important, the manuscript is well written, and the figures are of good quality, but the authors should make corrections and clarify some issues.

Highlights

Line 24: "Infrastructure projects across the globe spur economic development but also lead to social-ecological degradation".

Sugestion: This highlight could be written as follows: Infrastructure projects across the globe spur economic development but also cause serious socio-ecological problems.

Comment: The language could be moderated, as not all infrastructure projects lead to socio-ecological degradation.

Line 26: "Researchers and practitioners need current and comprehensive data to better understand and mitigate social-ecological changes from infrastructure".

Sugestion: This highlight could be written as follows: Researchers and practitioners need current and comprehensive data to better understand, mitigate, and monitor the negative impacts of socio-ecological infrastructure.

Comment: Monitoring negative impacts is very important to avoid, for example, situations of socio-environmental risks.

Line 31: "To promote better research outcomes, governments and NGOs should ensure datasets on infrastructure are accessible, current, comprehensive, and include such vital information as the project's geographic extent, dates of construction and operation, project type, and essential technical data".

Sugestion: This highlight could be written as follows: To promote better research outcomes, governments and NGOs should ensure datasets on infrastructure are accessible, current, comprehensive, reliable and include such vital information as the project's geographic extent, dates of construction and operation, project type, and essential technical data.

Comment: Public domain data, in addition to being accessible, current and comprehensive, must be reliable so that errors are not made and compromise research.

Regarding the formatting of the manuscript:

Comment: Some of the structure of the manuscript does not comply with PeerJ standards.

- According to the standards, the text of the manuscript should be aligned to the left.

- The authors need to remove the numbering of the items and subitems. In any case, if the numbering were maintained, it would be incorrect in items 3 and 4.

Line 256: "3.3 Comparison of attributes in literature review versus survey"

Line 367: "3.2.2 Data accessibility and quality"

Line 307: "3.3 Critical attributes for infrastructure datasets"

Line 323: "3.4 Evaluation of public datasets"

Line 350: "4. Discussion"

Line 351: "4.2 Data quality and availability for the Amazon"

Line 381: "4.2 Gaps in data availability"

Line 398: "4.3 Accuracy of assessed data"

Line 419: "4.4 Critical data attributes for social-ecological research"

Line 456: "4.5 Caveats"

Line 468: "4.6 Conclusion and future directions"

Comment: The conclusion was presented as a topic within the discussion. The conclusion should be presented in a separate topic in the manuscript.

Abstract

Line 39 - 40: "...development can improve human health and living standards, it may also lead to environmental degradation and social change".

Comment: Please, provide examples of the environmental degradation and social change that infrastructure projects can bring about.

Line 48 - 49: "Data were often difficult to find..."

Comment: Authors should indicate where these data were researched, for example, on Brazilian government websites and others.

Line 56 - 57: "...planning, decision-making, and monitoring of existing and future infrastructure projects in Brazil and other developing countries".

Comment: Are the authors referring to infrastructure in operation and in planning? Make it clearer for better understanding.

Introduction

Comment: In the manuscript abstract, the authors refer to infrastructure planned or under construction. In the introduction, the authors mention infrastructure in operation, under construction or planned. Please, explain which phases of infrastructure were considered for the development of the research (planning, construction and operation).

Line 87 - 88: “These social-ecological impacts cannot be properly identified and quantified without information about the infrastructure projects themselves”.

Comment: Authors should briefly explain what these social-ecological impacts would be. These are important details to improve the description of the impacts of infrastructures.

Line 100: “Even when infrastructure data are available...”

Comment: In this paragraph, the authors could explain the meaning of “open data” and the meaning of “poor quality data”.

Line 122 - 123: “Unfortunately, most Brazilian government portals are not in compliance with international open government data criteria (Di Ciommo, 2015)”.

Comment: This information from Di Ciommo (2015) is already 10 years old. Is the information still appropriate for this situation and has nothing changed in Brazil? The authors could look for a more current source.

Line 128 - 130: “Using infrastructure development in the Brazilian Amazon as a case-study, we conducted a systematic review of how infrastructure data have been used in social-ecological research in academic publications”.

Comment: I suggest that the authors write a brief text contextualizing the Brazilian Amazon in terms of population size, number of states in the federation, biomes, number of protected areas, as one of the main focuses of the manuscript is socioecological research.

Materials & Methods

2.1 Systematic literature review

Line 146 - 147: “We only considered studies published after the passage of the Brazilian Federal Access to Information Law in 2011 (Lei. 12.527/2015)”.

Comment: I thank you for providing the raw data, however, when I accessed the articles searched at https://doi.org/10.5281/zenodo.10626908, I noticed that 4 articles were published in 2011. Therefore, I believe that the authors of these articles did not access the data available on the Transparency Portal, as the Law was only enacted in November 2011. I suggest reviewing these review articles to assess whether they are appropriate for be part of the analysis.

Line 152 – 154: "We specifically looked for studies focusing on environmental impacts, management, or conservation in relation to current or planned infrastructure projects in the Brazilian Amazon”.

Comment: As previously mentioned, it is important to clarify whether these infrastructure projects are in planning, construction and/or operation. Each phase of an infrastructure project will result in different data, and impacts.

2.2 Key informant survey

Line 169: “...we collected demographic information and asked participants questions about...”

Comment: In addition to demographic information, the authors should have identified in which Amazonian state the research was conducted. This would have provided information about the state with the largest set of infrastructure data and the state with the least infrastructure data. Thus, this information could have shown the gaps found in the provision of data.

Line 183 - 186: "To examine differences in data quality between data retrieved from government versus non-government sources, we only considered answers from participants who reported retrieving data exclusively from a government repository or exclusively from a non-government repository".

Comment: For a better understanding, I suggest informing which were the government repositories and whether the research was carried out virtually on the websites of municipal, state and/or federal institutions.

Results

Line 230 – 234: "Sixty-two studies fit our criteria for the systematic literature review of articles that have investigated social-ecological impacts of infrastructure in the Brazilian Amazon. Together, the articles used infrastructure data 94 times, requiring 236 attributes about those infrastructure projects (Figure A-1B). Hydropower projects were the most common infrastructure category investigated (43 datasets), followed by roads and highways (18 datasets)".

Comment: 62 studies were reported, but when adding hydroelectric projects (43) and roads and highways (18) the total is 61 studies. Check and correct.

Critical attributes for infrastructure datasets

Line 308 – 310: "Based on the feedback from the literature review and surveys, we propose a list of critical attributes that should be included in all infrastructure datasets. At a minimum, all infrastructure datasets should include..."

Comment: In Brazil, according to the National Environmental Council (CONAMA), all environmentally modifying activities (e.g., hydroelectric plants and highways) will need to meet a series of responsibilities, basic criteria and general guidelines for Environmental Impact Assessment. And it is the responsibility of the state agency (e.g., Environmental Secretariats) or federal agency, such as IBAMA (Brazilian Institute of Environment and Renewable Natural Resources) to license the activities. Therefore, these attributes proposed by the authors in the manuscript are already required by Brazilian environmental policy.

329 - 330: "Neither dataset contained all the critical attributes we proposed (Table 1)".

Comment: As mentioned previously, in Brazil, the responsibility for environmental licensing lies with the state or federal environmental agency (IBAMA). Therefore, when dealing with state information (in the case of state licensing), the data should be available at the State or Municipal Environmental Secretariats, depend on the type of Project.

Discussion

Comment: Regarding the systematic literature review and information collection, together with researchers and institutions, I suggest that the authors write a brief text explaining which states in the Brazilian Amazon presented the largest set of infrastructure data. This is important to show

which states represent greater data accessibility, good governance and good proximity to researchers. All states in the Amazon have universities and research institutes. It would be important to know the focus of the systematic review researchers, such as which infrastructure project was studied, in which phase of the project the research was developed and the state or municipality where this project is located in the Brazilian Amazon. For instance, make a Table, resuming the informations.

457 - 461 “While considerable effort was made to obtain a representative sample of stakeholders interested in all infrastructure types in the Amazon region, the survey population was biased toward a hydropower focus. However, we note that the literature review results also were also skewed towards dams. Thus, this may simply reflect the general focus of the scientific community on dams in the Amazon region in the context of the hydropower boom...”

Comment: I suggest that the authors provide more information on this topic, as the Amazon has great energy potential and there is very strong pressure from the country on the region, and this could result in negative impacts on the population and the environment.

Conclusion

Comment:

In their conclusions or discussions, I suggest that the authors consider the following points:

In general, the authors considered the federal government sphere in their analyses for the acquisition of data on large infrastructure projects. The municipal and state spheres are important, such as these spheres are responsible for the technical and administrative process of environmental impact assessment. Data on infrastructure projects are generally held by environmental agencies in the Brazilian states. For example, there are river basins that belong only to strictly state areas. Therefore, normally, state laws restricted to that region are applied.

The states of the federation have major responsibilities, since they are responsible for the management, monitoring, inspection and even suspension of licenses for infrastructure projects. Suspension occurs due to failure to comply with some requirement set forth in the licensing phase (preliminary license, construction license and operating license).

It is important to discuss the responsibilities of municipal and state institutions regarding the availability of infrastructure data, since the lack of communication between the different government spheres results in incomplete and insecure data disclosure.

The availability of reliable data depends on investment in technology, improvements in the information system and technical training, especially in the municipalities that receive the infrastructure.

Furthermore, it is important to establish and/or strengthen partnerships between governments and research and teaching institutions, as the Brazilian Amazon has expertise in several areas of knowledge with trained professionals, equipment and laboratories that contribute to the preparation of Environmental Impact Studies for large infrastructure projects that generate socio-ecological and other data.