

IPBES TCA Chapter 2. Data Management Report 2.13 Intercoder reliability within the vision database /IPBES transformative change assessment

Versioning: 4.0.0

DOI: <https://doi.org/10.5281/zenodo.14258309>

Project leaders

Timothée Fouqueray (timothee.fouqueray@uqo.ca)

Iago Otero (iago.otero@unil.ch)

Sebastian Villasante (sebastian.villasante@usc.es)

Researchers

Vera H. Hausner (vera.hausner@uit.no)

Oleksandr Karasov (oleksa.karasov@gmail.com)

Lelani Mannetti (lelani.mannetti@gmail.com)

Gervasio Piñeiro (pineiro@agro.uba.ar)

Lynne Shannon (lynne.shannon@uct.ac.za)

Catherine D. Ward (catherine.d.ward@gmail.com)

Sandra Waddock (sandra.waddock@bc.edu)

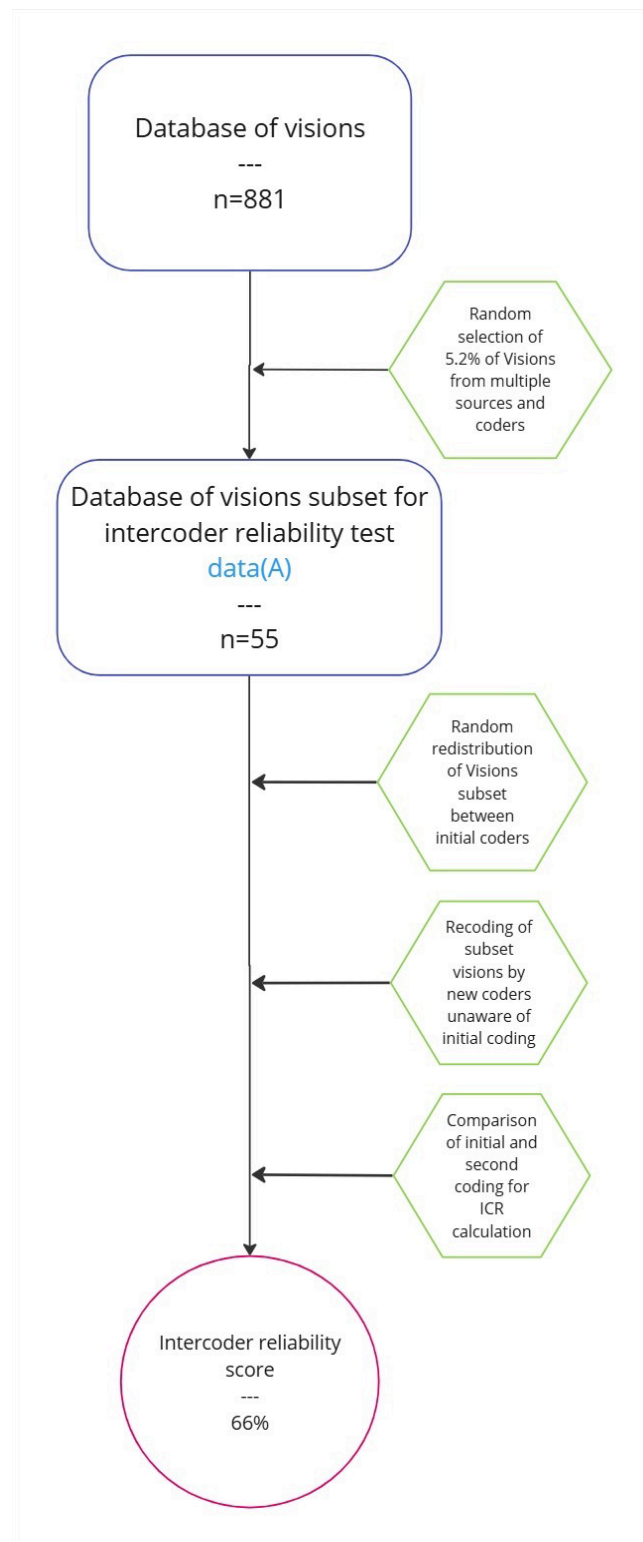
Data curator:

Laurence Perianin (laurence.perianin@umontpellier.fr)

Description: The review corresponds to the IPBES Transformative Change Assessment. The IPBES scoping document for the transformative change assessment describes that Chapter 2 should examine the implications of different visions for sectors, sub-systems (including economic/market, financial, political, legal/judicial, educational, Indigenous and local systems, and ecosystems) and the interactions between them, at different spatial scales.

In this context, Chapter 2 experts collected *Visions of a sustainable world for nature and people* from multiple sources. They coded each vision according to 13 categories of variables (eg., geographic scope of vision). To assess the consistency of vision coding, an intercoder reliability (ICR) evaluation was conducted. ICR is widely recognized as a best practice to evaluate the quality of coding applied to categorical data, ensuring that the coding frame and its application are credible, consistent, and transparent.

Process overview



Protocol

Description: The intercoder reliability test aims at assessing the consistency of coding across coders and categories, ensuring that all important vision types are included. This involves the random selection of 55 visions (5.2% of the total dataset) from multiple sources and coders to ensure that the dataset and all the coders involved have not been biased by categorical nor personal specificities. We obtained an overall intercoder reliability score of 66%, highlighting the robustness of the initial Visions coding with regards to the very large diversity of visions' sources (16 categories) and coders (15 experts).

Previous step – coding of the visions database

Each coder independently coded the selected visions using a pre-defined coding framework that captures key attributes such as vision description, geographical location and associated drivers. Coders adhere to quality criteria defined by previous research and relevant biodiversity reports. This process involved a detailed content analysis during which coders assessed the transformative potential of each vision by evaluating aspects such as systematic, plausible and motivational potential (see Data Management Report 2.0 Visions database analyses).

Step 1: Random, but representative subset from the database of visions

Method: To ensure coding consistency, an inter-coder reliability test was carried out (Huberman, Miles 1991). There are varied opinions on the subset size to be re-coded (Huberman, Miles 1991); however, considering the very high number of Visions included in the database, we subsampled 5.2% of our data (A) which we considered enough to highlight potential categorical or personal biases. These visions were randomly selected according to the following criteria:

- Representation of the various sources of visions (peer-reviewed literature, grey literature, manifestos, fiction, etc.).
- Inclusion of visions from the early and late phases of coding (to test the training effect).
- Representation of the visions coded by all the coders involved.
- Inclusion of certain non-positive visions in order to test the consistency of this filter in the database.

Location and format of the data:

3- IPBES_TCA_CH2_DMR 2.13_intercoder_reliability_VisionsSubset(A)

Table 1: Number of visions and their subsample used for ICR in different categories:

Category	Total number visions	ICR subset size	ICR subset percentage
Alternative Economic Visions	71	7	9,3
Artistic Expressions	22	3	11,5
Biomes	21	2	8,0
Declarations & Manifestos	22	4	14,8
Fictions	19	3	6,5
Grey Literature	32	5	11,9
Initiative & Social Movements	73	6	6,5
Ocean	304	9	2,6
Peer Reviewed Literature	76	5	3,9

Population Growth	6		
Rural	69	4	5,7
Social Media	6	1	16,7
Spiritual & Religious Visions	33	2	6,7
Technology	89	1	1,0
Urban	23	3	8,6
Water	5		
TOTAL	881	55	5,2

Following Miles and Huberman's (1991) method, ICR was calculated by comparing the proportion of categorical data on which coders reached an agreement. The overall ICR score for all coding categories was 66%, with higher consistency in some categories and lower agreement in others, such as the visualization process and transformation categories (Table 2).

Data curation measures were taken to standardize entries and resolve coding discrepancies, such as differing descriptions of geographic locations.

Step 2: Visions recoding and Data curation and standardization

Experts re-coded the visions that were randomly distributed to them, without knowing the initial codes. To correct coding inconsistencies, the experts curated the data, which was necessary to correct empty cells or “NA” in ‘No’ cells, and to normalize language discrepancies. For example, to describe the geographical scope of a vision, the first coder could use ‘UK’, while the recoder indicated ‘United Kingdom’. In addition, the coding grid for positive visions included more data units than negative visions (eg., “Main purpose of the vision” did not apply to visions of a unsustainable world for nature and people). Thus, if the two coders disagreed on the positivity of a given vision, its ICR was calculated only based on the data units associated with negative visions.

Step 3: Intercoder reliability calculation

The ICR scores were then averaged by data category (for example, for all data units grouped under the ‘IUCN vision biomes’ category), and an overall score of 66% was obtained by averaging all ICRs (see Table 2).

Table 2. Intercoder reliability (ICR) scores. The categories as in columns indicate mean ICR scores per data category. For percentage agreement comparison, there is no standard threshold agreed on for what indicates acceptable ICR, but authors report that a 0.61-0.80 score is substantial, while it is “nearly perfect” above 0.81 (O’Connor and Joffe, 2020). Data categories below 0.60 were removed from the analysis (ie., “Visioning process” and “Transformation” categories).

OVERALL	Description of Vision	Geographic Scope of Vision	Location of Vision	UN Region of Vision	IUCN Biomes of Vision	Other Content Areas of Vision	Positive Visions	Main Purpose of Vision	Other Purposes of Vision	Direct Drivers	Indirect Drivers	Visioning Processes	Transformation
0,66	0,61	0,63	0,64	0,69	0,74	0,74	0,84	0,63	0,74	0,69	0,64	0,45	0,42

Categories below the reliability threshold (> 60%) were excluded from further analysis.

References

- Huberman, M., Miles, M.B., 1991. Qualitative Data Analysis: A Collection of New Methods.
- O'Connor, C., Joffe, H., 2020. Intercode Reliability in Qualitative Research: Debates and Practical Guidelines. *Int. J. Qual. Methods* 19, 1609406919899220. <https://doi.org/10.1177/1609406919899220>

Definition of files

ID	Name	File type	Description
1	1-IPBES TCA Chapter 2 DMR 2.13 intercode_reliability	PDF	DMR
2	2-IPBES_TCA_CH2_DMR 2.13_intercode_reliability_process	JPG	Flowchart of the process
3	3- IPBES_TCA_CH2_DMR 2.13_intercode_reliability_VisionsSubset(A)	PDF	Initial sample of articles