



HOW TO DOCA - the Database of Variables for Content Analysis

A Practical Guide for Researchers, Students, and Lecturers

What is DOCA?

The open-access database of Variables for Content Analysis (DOCA) is a methodological infrastructure that compiles, systematizes, and critically evaluates content-analytical variables used across communication research. In doing so, it serves as the basis for how variables are developed and operationalized.

You can use DOCA to orient yourself across a wide range of content-analytical variables within thematic sections, to search specifically for variables you want to address in your research and see how they are operationalized, or to search for methods of content analysis for teaching purposes. The goal of DOCA is to provide a starting point for content-analytical questions, and to contribute towards a standardization of research practices and equal opportunities between all researchers. The platform provides entries for individual variables (e.g., actors, topics, issues, etc.), as well as for more complex theoretical constructs (e.g., americanization, stigmatization).

Content

How is DOCA structured?.....	1
How is a DOCA entry structured?	2
Use case 1: You are a student or researcher looking for suitable and tested variables for your content analysis	3
Use case 2: You have expertise in a specific field of research in which standardized content analysis is applied and would like to become an author	4
Use case 3: You teach the method of content analysis and want to use the database.	5
Glossary	6

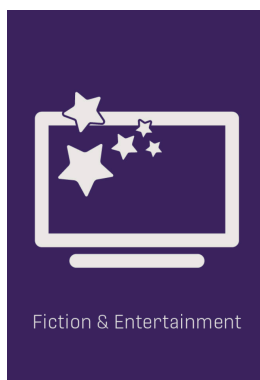
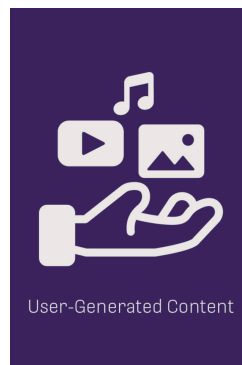




Structure of DOCA

The **variables** are systematized into six thematic sections of the database:

- Basic Variables
- Variables for research on news & journalism
- Variables for research on fiction & entertainment
- Variables for research on (professional) communicators & strategic communication
- Variables for research on user-generated media content
- Automated Content Analysis





Structure of a DOCA entry

Structure of a DOCA Entry

Each database entry follows (more or less) the following structure:

- A brief description of the variable or theoretical construct
- A brief description of the most common field of application / theoretical foundation
- (If applicable) combination with other methods of data collection
- Two sample operationalizations including information about the selected studies (research question(s); analyzed media type) and information about the variable or construct (level of analysis, reliability score, categories / values)
- (If available) the codebook / code

Construct

Scientific evidence/uncertainty (Science and Health Communication)

Sabrina H. Kessler

<https://orcid.org/0000-0003-1858-7041>

Author(s) & DOI

DOI: <https://doi.org/10.34778/2h>

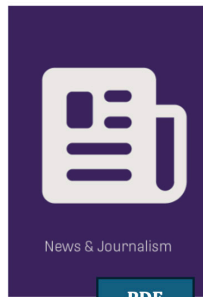
Keywords: scientific evidence, scientific uncertainty and certainty, frames of depicted evidence, science communication, health communication, formal-abstract frames, scientific findings, evidence sources

Abstract

The presented scientific evidence and uncertainty in science communication can be achieved by either different variables (e.g., Brechman, Lee, & Cappella, 2009, 2011; Guenther, Bischoff, Löwe, Marzinkowski, & Voigt, 2019; Kessler, 2016) or identifying frames (for thematic frames, see Ruhmann, Guenther, Kessler, & Milde, 2015; for formal-abstract frames, see Kessler, 2016).

Field of application/theoretical foundation:

Evidence and (un)certainly are integral components of scientific findings and science in general. Scientific evidence can be defined as a continuum, ranging from scientific uncertainty to certainty and from weak to strong evidence. Media content analyses are investigating the extent to which information is given in media articles that provide indications of the evidence or uncertainty of scientific findings. Content Analyses also measure how evident scientific findings are presented in the media.



PDF

Kessler Scientific evidence/uncertainty

Appendix 1 Codebuch Kessler

2016

Appendix 2 Günther et al. 2019

Codebook & Coding

Variables: "For each represented research result, a variable collected the main (hypo-)thesis of the research study, the direction of the result (for or against the thesis), as well as the relevant criteria to assess evidence. [...] For each result, it was also relevant to collect to which extent scientific certainty or scientific uncertainty was discussed. In the current study, an explicit statement referring to (un)certainly was differentiated from an implicit statement (subjunctive, speculative language as an indicator of uncertainty versus imperative as an indicator for certainty). This was supplemented by collecting the justifications for (un)certainly that were given for the scientific results." (p. 10)

Level of analysis: news article

Variables and values:

Values

Variables

- reported relevant criteria to assess scientific evidence: theoretical assumptions/(hypo-)theses; pilot study/a study never done before; research design: experiment, case study, etc.; research and measurement instruments; quality criteria, such as reliability; quality criteria, such as validity; references to significance (statistic values); objectivity; information about sample (size); time of study; explicit depiction of the research setting; number of studies done; information about how results were obtained; limitations, such as knowledge gaps; comparisons to other studies; funding source(s); reference to the investigating researcher(s); reference to the publication/ journal/ conference; future scenarios, specific applications
- reported explicit justifications for scientific (un)certainly: preliminary data, knowledge gap(s); (poor) methodological quality; contrasting findings of research; contrasting interpretation of same dataset; conflicting viewpoints of researchers; doubt whether data can be applied to humans; effect on humans not clear; effect on nature not clear; lack of technical/scientific opportunities; justifications for certainty; certain single result(s); sufficient data; (strong) methodological quality; results pointing in the same direction; successfully replicated findings; application for humans clear; effect on humans clear; effect on nature clear; highly experienced researcher(s)
- implicit statement referring to (un)certainly: no implicit representation vs. implicit representation

Reliability: "Four experienced coders coded the articles of the sample after several intensive training sessions. Intercooder reliability was calculated according to Holsti for 26 articles (20 percent of the sample) and the following satisfactory results were obtained: formal variables: 0.97; criteria relevant to assess evidence: 0.92; uncertainty (explicit and implicit): 0.95; certainty (explicit and implicit): 0.92." (p. 10)

Codebook: in the appendix (in German)

Reliability



You are a **researcher** looking for suitable and tested variables for your content analysis

DOCA can be used as a starting point to explore how similar constructs have been operationalized across different studies. Researchers can compare existing approaches and derive new or refined operationalizations tailored to their specific research context.

Example: You want to analyze political topics across fictional content (DOCA Entry).

- You can either use the search function, or, if possible, directly look into one of the thematic sections of the database.
- You'll find the search function in the menu bar, or here: hope.uzh.ch/doca/search
- After searching for "political topics", you'll find that the first entry may represent what you are looking for. By clicking on it, you'll see:
 - Abstract with an overview and a field of application, references and example studies
 - More information is provided in (at least) one of the example studies. Here you will find information about the variable, definitions, level of analysis, scale and reliability, if available.
- The construct in this example, political topics, spans across three thematic variables of politics: polity, policy, and politics. Thus, the entry provides the values for all three variables in this context, as well as respective reliability measures for each variable.
- You can choose if the variable / codebook suits your purpose. Even if the variable does not correspond exactly to the one you are looking for, the example codings will certainly provide valuable guidance, and you can often operationalize your own variables by drawing on an already operationalized variable and adapting it to your specific context. Please note that DOCA only aims to standardize variables and thus enable comparison. The specific characteristics of the respective research questions and research designs should be taken into account in each case.
- Please cite the DOCA entry where appropriate
- On the bottom of the page, you'll find the references and most read articles by the same author



You have expertise in a specific field of research in which standardized content analysis is applied and would like to become an **author**

DOCA is an open access database which compiles, systematizes and evaluates relevant content-analytical variables and publications of communication and political science research areas and topics. At regular intervals, there is a thematic call, where authors are encouraged to submit a contribution in the form of a new variable or adding operationalizations to an existing variable. If you want to suggest a further variable, or are not sure if your entry fits the scope or call of DOCA, you can reach out to us.

Example: You perform a standardized content analysis with codebook to operationalize hostile comments on social media

- You'll find the "Call & Guidelines" section in the menu bar, or here: hope.uzh.ch/doca/call
- First, you see the most recent Thematic Call, and the description of the theme. Below, you'll find the submission guidelines.
- As an author, you decide which example operationalization of your contribution is chosen. To ensure consistency with other contributions, you are asked to consider the following criteria:
 - Topicality
 - Relevance (possible indicators: citation frequency, expert opinion, etc.)
 - Accessibility of variables
 - Quality (reporting of reliability coefficients)
- Each contribution follows (more or less) a specific structure:
 - A brief description of the variable or construct
 - A brief description of the most common field of application and/or theoretical foundation
 - (If applicable) References and/or combination with other methods of data collection
 - A sample operationalization. This includes:
 - Information about the selected study (research questions, analyzed medium)
 - Information about the variable or construct (level of analysis, reliability, categories)
 - (If available) Links / referral to Codebook(s), Protocol(s), Dictionaries, ...
- DOCA is structured into different sections (See: How is DOCA structured?). You are asked to specify to which section your contribution belongs, and to specify whether you submit an entry for a new variable or an existing variable.
- A contribution should be between 1'000 and 1'500 words, and preferably in English.



You **teach** the method of content analysis and want to use the database

DOCA is not only suited for researchers, it also serves as a helpful starting point to teach content-analytical basics and procedures in communication and political science. You can use DOCA to teach students how to look for relevant variables, and how to code specific variables of interest.

Example: You want to perform a standardized content analysis with codebook to operationalize hostile comments on social media

- You can start with defining a research question, which can be answered with content analysis. For example, “how rational is online discourse in the comments of climate change posts?”, or “how are political actors framed in online news during election campaigns?”
- Explore DOCA to identify relevant variables or operationalizations of your constructs of interest. You can browse the sections, or use the search function, searching for specific variables or constructs.
- When you identified the relevant variables on DOCA, you can review the definitions, example operationalizations and, if available, codebooks provided in each entry. Based on this information, you can decide how you want to measure each variable of interest.
- Before you start coding, you should develop a codebook for your purposes. You can use one of the existing ones, combine multiple codebooks if applicable, or use an existing codebook as a starting point for your content analysis.
- After that, you can select the corpus you want to analyze and train/teach the students for the subsequent analysis. Preferably, you also compute an intercoder reliability, both for the purpose of consistent coding and to teach the students the importance of reliability.
- Lastly, you can pursue interpreting the results of your content analysis. You can also reflect on how the operationalizations drawn from DOCA helped you answer the research questions, or if you identified shortcomings of the entry for your specific purpose.



Glossary

Construct: Latent theoretical concepts that researchers aim to measure empirically.

Codebook: A document containing definitions of constructs, variables, values and detailed coding instructions.

Reliability: A measure of the consistency and reproducibility of coding results. In manual content analysis, reliability refers to the extent to which different coders achieve similar results under comparable conditions (intercoder reliability). In automated content analysis, it refers to the stability and consistency of computational coding procedures across repeated applications.

Value: Distinguishable and exhaustive categories or expressions a variable can take; values should always be mutually exclusive.

Variable: A variable is a characteristic of the object of analysis that is to be examined in the course of the analysis. There are formal variables, such as date or genre, as well as content-related variables, such as the actors who speak or the positions expressed. An operationalized category used to classify and measure aspects of a construct.

Further information:

Oehmer-Pedrazzi, F., Kessler, S. H., Humprecht, E., Sommer, K., & Castro, L. (2023). *Standardisierte Inhaltsanalyse in der Kommunikationswissenschaft - Standardized Content Analysis in Communication Research: Ein Handbuch - A Handbook* (p. 478). Springer Nature.



„Ziel der Datenbank ist es, Wissen über die Erhebungsinstrumente und Anwendungsfelder der standardisierten Inhaltsanalyse und damit über eine der zentralen Methoden kommunikationswissenschaftlicher Forschung und Lehre für Wissenschaftler:innen und Studierende zu sammeln, zu systematisieren, zu reflektieren und im Sinne von Open Method und Open Access frei zugänglich zu machen.“

Editorial Team

Franziska Oehmer-Pedrazzi (principal editor), MILEVA INSTITUT für Digitales und Gesellschaft Schweiz | University of Applied Sciences of the Grisons FHGR

Sabrina H. Kessler (database manager), University of Zurich

Edda Humprecht, University of Jena

Laia Castro Herrero, University of Barcelona

Nicole Bizzotto, University of Zurich

Philippe Sloksnath, University of Zurich