EHB

EIDGENÖSSISCHE
HOCHSCHULE FÜR
BERUFSBILDUNG

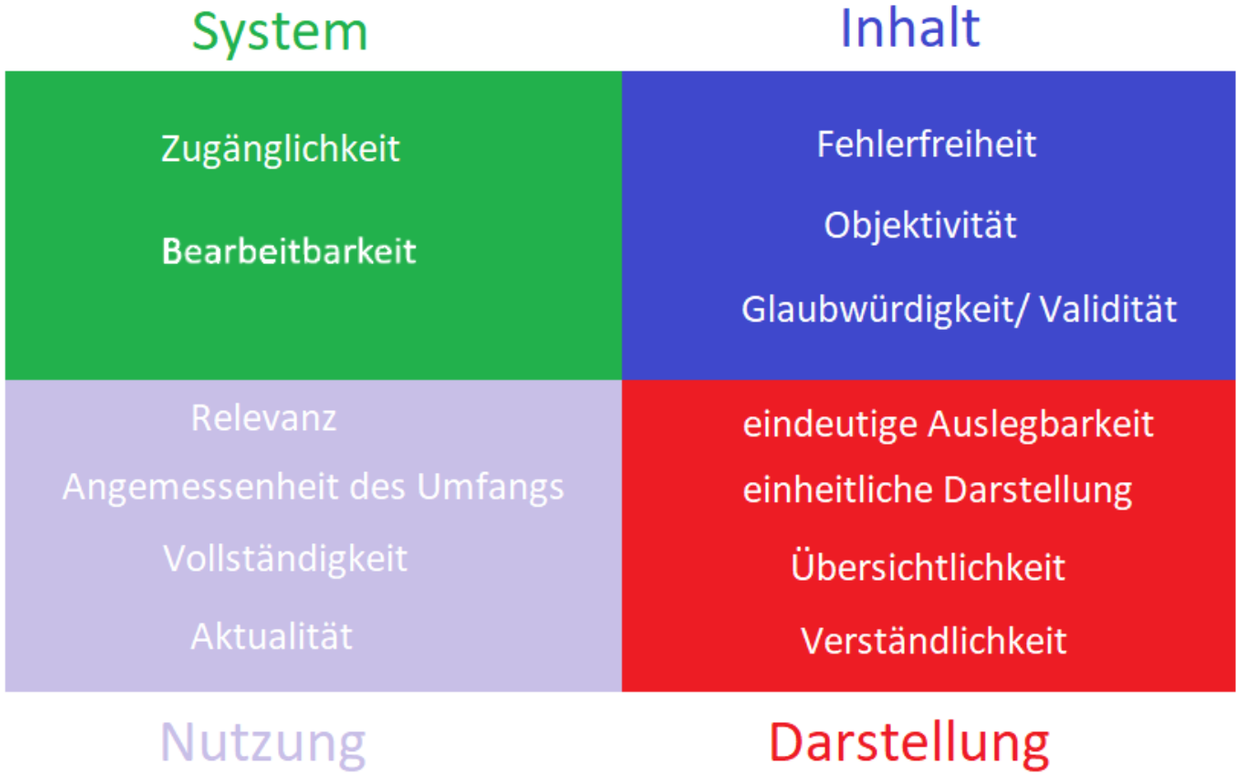
Schweizer Exzellenz in Berufsbildung

CRITICAL DATA LITERACY
BAUSTEIN 2: DATENQUELLEN & DATENQUALITÄT
KRITERIEN FÜR DIE QUALITÄT VON DATEN UND INFORMATIONEN

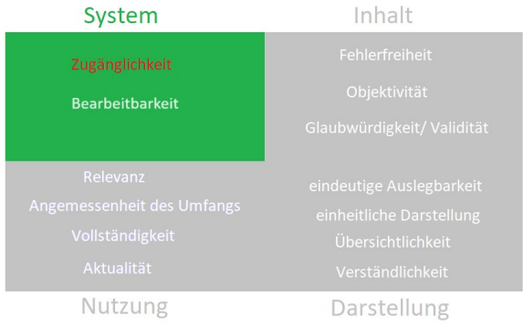
Mathias Missler

Deutsche Gesellschaft für Informations- und Datenqualität/ Massachusetts Institute of Technology

- Kriterien zur Datenqualität basieren auf Arbeiten der DGIQ und des MIT
- 20 verschiedene Kriterien der Datenqualität
- 13 sind von grösserer Relevanz



EHB



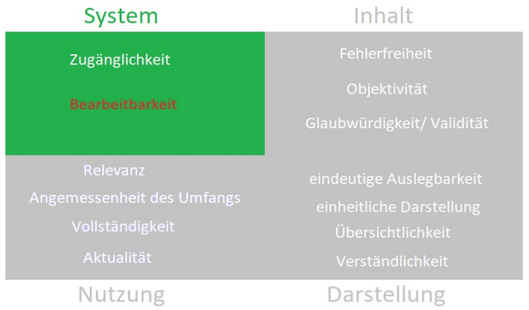
Dimension System, 1. Kriterium: Zugänglichkeit

- Zugänglichkeit besteht, wenn Daten für jeden einfach abzurufen sind

- Keine Portale mit Bezahlschranken oder Institutionen-Logins



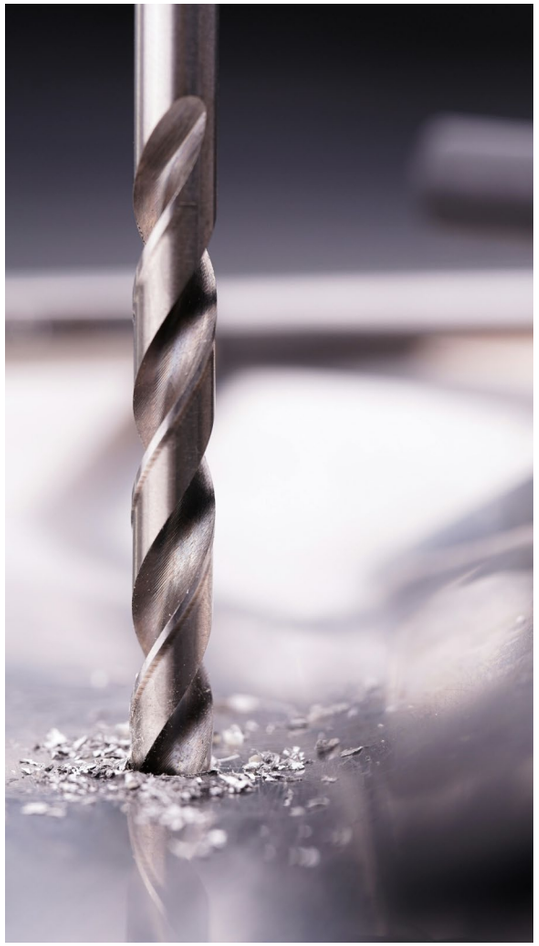
EHB

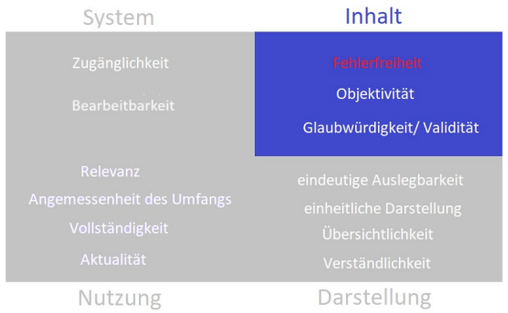


Dimension System, 2. Kriterium: Bearbeitbarkeit

- Bearbeitbarkeit besteht, wenn es die Möglichkeit der einfachen Änderung der Daten gibt und diese für unterschiedliche Zwecke genutzt werden können

- Datensätze, z.B., die nur in SPSS bearbeitet werden können und z.B. nicht in STATA, führen eventuell zu unzureichender Bearbeitbarkeit





Dimension Inhalt, 3. Kriterium: Fehlerfreiheit

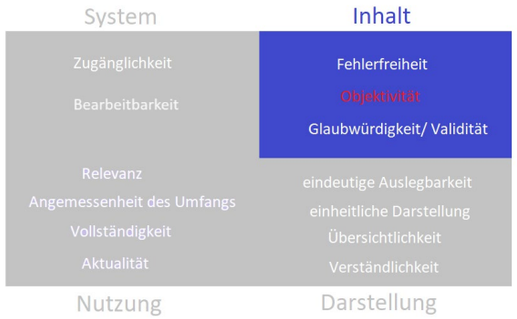
- Fehlerfreiheit ist gegeben bei Übereinstimmung mit der Realität

- Kriterium ist aus sozialwissenschaftlicher Sicht nicht so leicht zu überprüfen – die Metaphysik des Sozialen ist schwer quantifizierbar.

- Für Daten kann grundsätzlich gelten: Plausibilität von Variablen und Informationen.



EHB

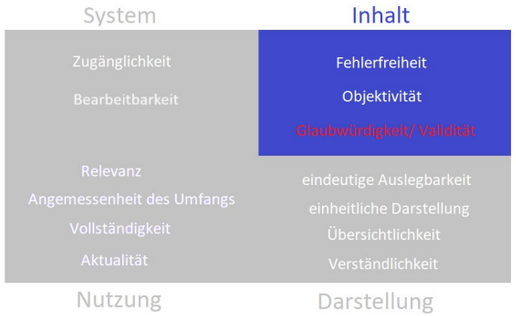


Dimension Inhalt, 4. Kriterium: Objektivität

- Objektivität besteht bei sachlicher und wertfreier Information.
- Bei z.B. Modellen mit vielen subjektiven Wertungen ist das Kriterium der Objektivität vermutlich nicht gegeben.
- Zu beachten ist allerdings: Vollkommene Objektivität ist für Menschen unerreichbar.
- Deswegen muss versucht werden einen möglichst hohen Grad an Objektivität anzustreben.



EHB

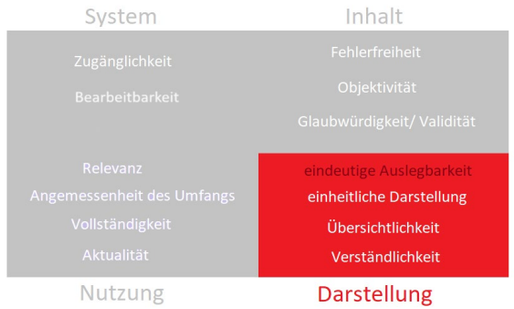


Dimension Inhalt, 5. Kriterium: Glaubwürdigkeit/ Validität

- Glaubwürdigkeit besteht, wenn:
  - Zertifikate vorhanden sind
  - Die Datengewinnung mit hohem Aufwand und Sorgfalt betrieben wurde
- Der Datensatz des Schweizer Haushaltspanels (SHP) oder die Allgemeine Bevölkerungsumfrage (ALLBUS) der GESIS entsprechen diesen Kriterien



EHB



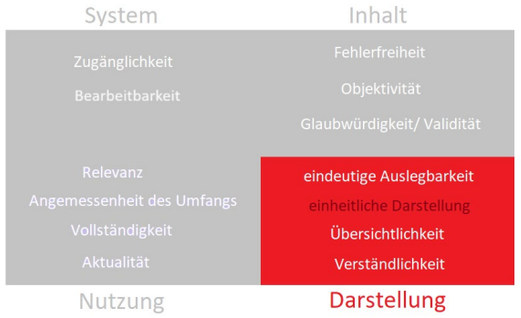
Dimension Darstellung, 6. Kriterium: eindeutige Auslegbarkeit

- Interpretierbarkeit ist gegeben, wenn Daten eindeutig, gleich und fachlich korrekt dargelegt sind

- Kriterium Darstellung ist teilweise subjektiv, da Menschen naturgemäss unterschiedliche Perspektiven haben



EHB



Dimension Darstellung, 7. Kriterium: einheitliche Darstellung

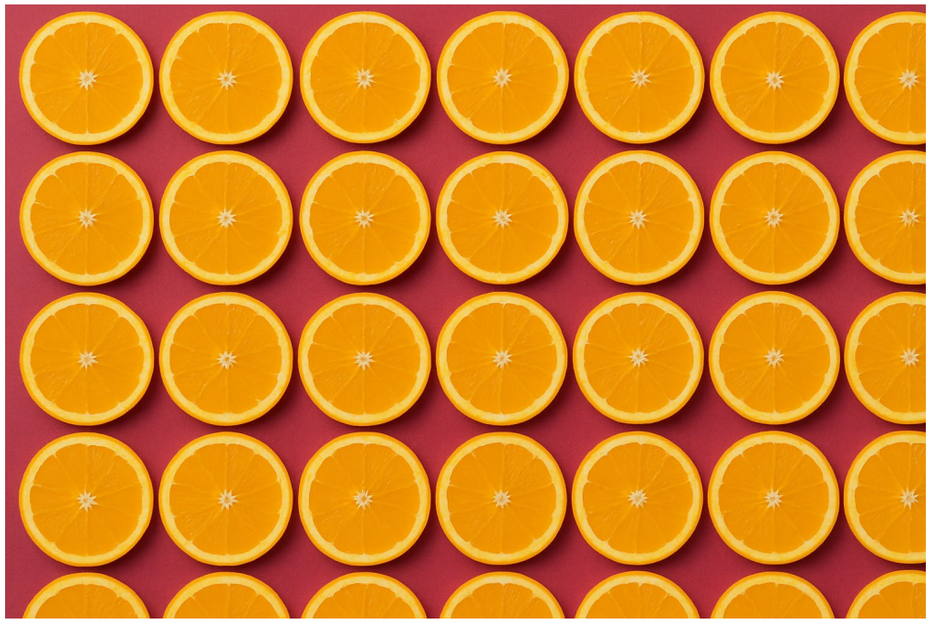
- Dieses Kriterium ist erfüllt, wenn Daten folgende Eigenschaften besitzen:
  - Dasselbe Format
  - Konsistente, fortlaufende Darstellung
  - Kompatibilität zu vorangegangenen Daten



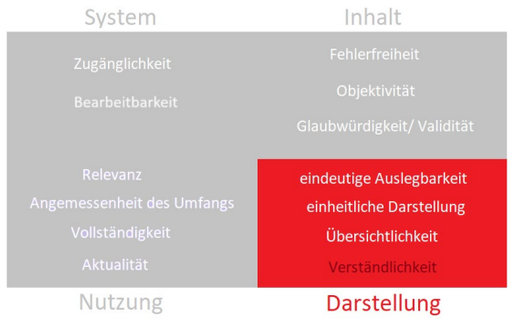


Dimension Darstellung, 8. Kriterium: Übersichtlichkeit

- Daten sind übersichtlich, wenn ein passendes und leicht fassbares Format gewählt wurde
- Excel Tabellen mit über 10.000 Zeilen widersprechen diesem Kriterium



EHB



Dimension Darstellung, 9. Kriterium: Verständlichkeit

- Das Kriterium der Verständlichkeit ist erfüllt, wenn Daten
  - unmittelbar verstanden und eingesetzt werden können und
  - unnötige Fachausdrücke vermieden werden



EHB



Dimension Nutzung, 10. Kriterium: Relevanz

- Daten sollen notwendige Informationen liefern
- Zusätzlich sollen sie einen starken Bezug zum ausgewählten Thema haben



EHB



Dimension Nutzung, 11. Kriterium: Angemessenheit des Umfangs

- Daten sind von angemessenem Umfang, wenn die Menge der verfügbaren Informationen den gestellten Anforderungen genügt

- Diese können je nach Thema der wissenschaftlichen Arbeit oder des Arbeitsprozesses variieren



EHB



Dimension Nutzung, 12. Kriterium: Vollständigkeit

- Dieses Kriterium ist erfüllt, wenn Daten zu den festgelegten Zeitpunkten und in den jeweiligen Prozessschritten vorhanden sind

- Werden wichtige Daten ausgespart, wie z.B. die Variablen „Alter“ oder „Geschlecht“ bei einer Umfrage, steht dies in Konflikt mit der Vollständigkeit



EHB



Dimension Nutzung, 13. Kriterium: Aktualität

- Daten können als aktuell bezeichnet werden, wenn sie die tatsächlichen Eigenschaften eines Objekts oder Themas zeitnah abbilden

- Eine Datenquelle aus dem 19. Jahrhundert dürfte also bei der Beschreibung von Phänomenen des 21. Jahrhunderts zu Problemen führen



EHB

Kontextabhängigkeit

- Kein starres Bewertungsmuster
- Anwendungskontext abhängig von Fragestellung
- Berufsbildung verlangt andere Voraussetzungen als eine wissenschaftliche (Seminar)Arbeit
- Bei Zeitknappheit lohnt es sich, die Aufmerksamkeit auf die vier am häufigsten verwendeten Kriterien der Vollständigkeit, Genauigkeit, Glaubwürdigkeit/ Validität und Verfügbarkeit zu richten



EHB

Literatur

- Rohwedder, Jan Philipp; Piro, Andrea; Schmid, Joachim (2007): DGIQ Projekt – IQ-Definition. Definition des Begriffsfeldes Informationsqualität bzw. Datenqualität. URL: https://silo.tips/download/dgiq-projekt-iq-definition. Zuletzt aktualisiert: 22.11.2007. Heruntergeladen am 20.02.2023.

- Wang, Richard Y.; Guarascio, Lisa M. (1991): Dimensions of Data Quality. Toward Quality Data by Design. Massachusetts: Massachusetts Institute of Technology.

EHB