



Foto: Thomas Josek

Datenmodellierung

Woche 3: Datenbanken

Institut für Digital Humanities, Historisch-Kulturwissenschaftliche Informationsverarbeitung | Prof. Dr. Øyvind Eide

Zu den Hausaufgaben

Eine Hälfte von Ihnen (Nachname A-L) hat S. 3 bis 16 von:

Julia Flanders und Fotis Jannidis. 2018. "Data modeling in a digital humanities context. An introduction." S. 3–25 in: Julia Flanders und Fotis Jannidis (Hg.): The Shape of Data in Digital Humanities: Modeling Texts and Text-based Resources, London und New York.

gelesen.

Die andere Hälfte von Ihnen (Nachname M-Z) hat S. 26 bis 36 von:

Fotis Jannidis und Julia Flanders. 2018. "A gentle introduction to data modeling." S. 26–95 in: Julia Flanders und Fotis Jannidis (Hg.): The Shape of Data in Digital Humanities: Modeling Texts and Text-based Resources, London und New York.

gelesen.

- Diskutieren Sie jetzt in den jeweiligen Gruppen die im folgenden gestellten Fragen. Erstellen Sie ein Handout dazu (digital oder handschriftlich), welches später auf ILIAS hochgeladen wird.
- In ca. 20 Minuten berichtet jede Gruppe im Plenum der anderen Gruppe.

Fragen für “Data modeling in a digital humanities context. An introduction” (S. 3–16) / Nachnamen A-L

1. Was meinen die AutorInnen mit der Aussage “Debates about method are ultimately debates about our models”?
2. Inwiefern ist der kritische Apparat einer Edition ein Modell?
3. Was sind die Unterschiede zwischen “informal”, “formalisable” und “formal” models?
4. Welche Herausforderungen sehen die AutorInnen in der geisteswissenschaftlichen Datenmodellierung?
5. Wieso ist es hilfreich, ein Modell von den Daten zu haben? Wieso kann man von “intelligence in the data” sprechen? Welche Vorteile hat die Formalisierung der Daten?
6. Welche Kriterien folgt man, wenn man ein Datenmodell wählt?
7. Was sind die Risiken, wenn man nicht formale Modelle benutzt, wie z.B. wenn man einen Text in Microsoft Word schreibt?
8. Welche Beziehung existiert zwischen Datenmodellen und Tools?
9. Was muss man beachten, wenn man ein Tool für die Datenmodellierung benutzt?
10. Sind Modelle immer explizit sichtbar?
11. Wieso sollte man Daten wenn möglich Tool-unabhängig modellieren?
12. Wieso sollte man aber im Prozess der Modellierung die Tools nicht komplett ausschließen?

Fragen für “A gentle introduction to data modeling” (S. 26–36) / Nachnamen M-Z

1. Kann ein Computer den Text eines Briefes ohne Weiteres als solchen erkennen? Oder braucht er da “Unterstützung”?
2. Wie können wir den Prozess der Erkennung unterstützen?
3. Die AutorInnen nennen drei Merkmale von Modellen, welche?
4. Was braucht ein Modell, um prozessiert werden zu können?
5. Was ist die Beziehung zwischen einem nicht formalen und einem formalen Modell?
6. Was sind die Vorteile eines formalen Modells?
7. Was ist der Unterschied zwischen data und process modelling?
8. Was sind Metadaten?
9. Was ist der Unterschied zwischen inline und stand-off Annotationen?
10. Die AutorInnen sprechen von modellierter Instanz, Datenmodell und Metamodell: Was bedeuten diese Begriffe?
11. Welche Rolle spielen data type und data structure in der Datenmodellierung?
12. Was ist der Unterschied zwischen discrete und continuous data? Wieso ist das wichtig?

Gemeinsame Zusammenfassung

Bitte fügen Sie Ihre Ergebnisse/Handouts in ILIAS ein:

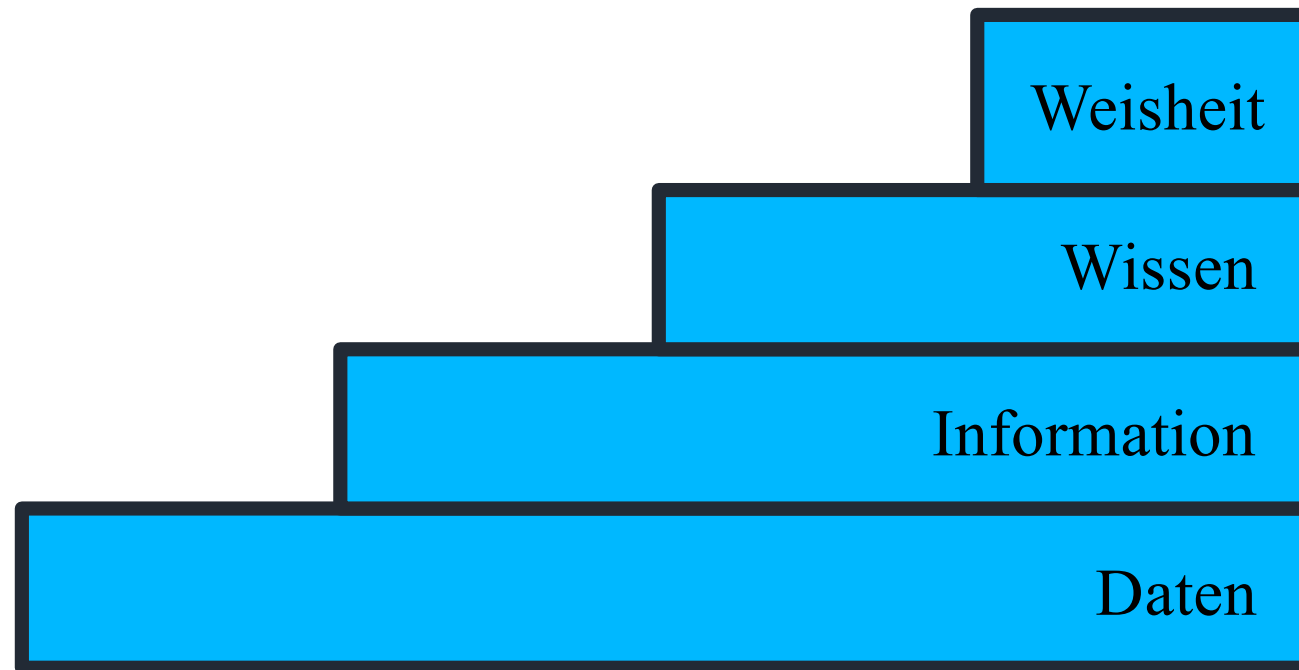
https://www.ilias.uni-koeln.de/ilias/goto_uk_xpdl_6306874.html

Danke dafür!

Daten

Fragen für Überlegung

- Was sind Daten?
- Was sind Daten in den Geisteswissenschaften?
- Wie verhält sich Daten zu Quellen?



Daten – Weiterführende Literatur (optional)

- Christof Schöch. 2013. “Big? Smart? Clean? Messy? Data in the Humanities”. Journal of Digital Humanities.
<https://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>
- Johanna Drucker. 2011. "Humanities Approaches to Graphical Display." Digital Humanities Quarterly 5, 1.
<http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>
 - “[...] the concept of data as a given has to be rethought through a humanistic lens and characterized as capta, taken and constructed.”

Zusammenfassung Datenmodellierung

- Modell als Repräsentation von einer oder mehreren Instanzen und deren Verbindungen. Es wird von jemandem für einen bestimmten Zweck gemacht und benutzt. Es folgt bestimmten Regeln, die von einer Gruppe geteilt werden → ein Modell ist also in einem semiotischen System integriert
- Modelle und Modellierung sind explizite erklärende, explorative und empirische Strategien der Untersuchung
- Modellierung ist ein iterativer Prozess der Integration von Informationen und deren Exploration mit wiederholten Schleifen von Tests, Feedback und Anpassung
- Wir können Daten modellieren (= Datenmodellierung) oder Prozesse

Zusammenfassung Datenmodellierung

- Datenmodellierung basiert auf einer langen Geschichte von Beobachtung und Repräsentation von Daten in den Geisteswissenschaften (obwohl Geisteswissenschaftler:innen Daten nicht immer als solche erkannt haben)
- Auch Modelle wurden schon in den Geisteswissenschaften benutzt (z.B. kritische Apparate), aber sie waren immer informelle (oder formalisierbare) Modelle, d.h., sie haben auf impliziten Regeln basiert
- Datenmodelle interagieren mit Tools, können in Tools integriert oder durch Tools verarbeitet und visualisiert werden. Die Modelle selbst können aber toolabhängig oder -unabhängig sein
- Datenmodelle sind eine Art, “intelligence in the data” zu etablieren: Wir können unser Wissen in die Modelle einfließen lassen, damit wir dann Daten und Wissen über die Daten prozessieren und verarbeiten können

(Relationale) Datenbanken

Themenüberblick Sitzung 3, 4 und 5

- Zweck von Datenbanken
- Begrifflichkeiten
- Verschiedene Datenbankmodelle, Pro und Contra
- Relationale Datenbanken
- Entity-Relationship-Modell, Relationale Algebra
Entwicklung einer relationalen Datenbank
- Anforderungserhebung, Konzeptuelles
Datenbankschema, Entity-Relationship-Diagramm,
Logisches Datenbankschema, Schlüssel, Kardinalität,
Normalisierung (Normalformen) Implementierung, SQL,
Erstellung und Abfragen
- Übungsaufgabe (Studienleistung)

Datenbanken – Literatur (optional)

Klinke, Harald. 2017. “Datenbanken.” S. 109–127 in: Fotis Jannidis, Hubertus Kohle, Malte Rehbein (Hg): *Digital Humanities: Eine Einführung*.

<https://link.springer.com/book/10.1007/978-3-476-05446-3>

Wozu Datenbanken?

- Datenbanken dienen der effizienten elektronischen Datenverarbeitung:
 - Speichern von großen Mengen von Daten: sicher, dauerhaft, widerspruchsfrei, effizient, konsistent
 - Zugriff auf Daten: Eingabe, Update, Abfragen, Auswertung, Zugriffsrechte

Wo finden wir Datenbanken?

- KLIPS 2
- Sparkasse
- Deutsche Bahn
- lehre.idh
- REWE
- OpenStreetMap
- ...

Herausforderungen

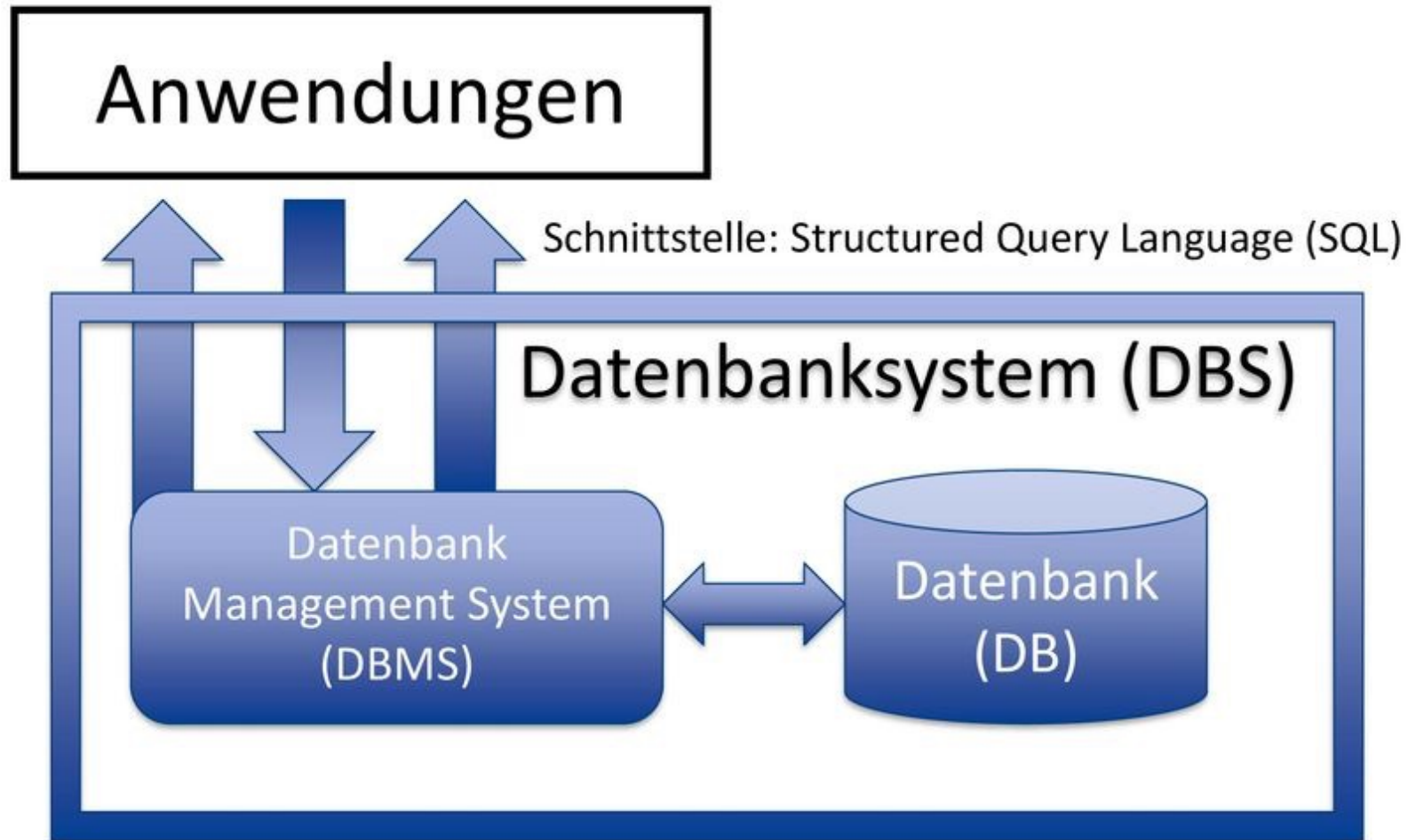
mit Beispielen für Lösungsstrategien

- Struktur
 - Modellierung
- Effizienz
 - Indexierung
- Konsistenz
 - keine Redundanz
- Integrität
 - Transaktionen
- Zugänglichkeit
 - Flexibilität

Datenbanken – Begriffe

- Datenbanksystem = Datenbank + Datenbankmanagementsystem
 - Datenbanksystem (DBS)
 - Datenbank (DB): die konkrete Datenbasis, die zu verwaltenden Daten
 - Datenbankmanagementsystem (DBMS): Software zur Verwaltung, bestimmt das Datenbankmodell sowie die Funktionalitäten und Performanz des Datenbanksystems
- Datenbankmodell → z.B. relationales Datenbankmodell
- Datenbanksprache → z.B. SQL

Fachbegriffe: **(DBS = DB + DBMS)**



Relationales Datenbankmodell

- Am weitesten verbreitetes Datenbankmodell
- Vorteile:
 - Flexibel
 - Schnelle Verarbeitung
 - Reduzierung von Redundanz
 - Verständlich
- Grundlage:
 - Tabellen = Relationen
 - Entity-Relationship-Modell (Chen 1976)
 - Relationale Algebra (Codd 1970)

Entity-Relationship-Modell (Chen 1976)

- Beschreibt für einen relevanten Ausschnitt der Realität die Struktur der Daten und ihrer Beziehungen zueinander
- **Entität**: individuelles, eindeutig identifizierbares Objekt oder Sachverhalt, z.B. die Person Øyvind Eide
- **Entitätstyp** (Klasse): Entitäten können zu Entitätstypen zusammengefasst werden, z.B. Personen
- **Attribut** (Eigenschaften): Entitäten haben Attribute, z.B. haben Personen das Attribut Vorname
- **Relation** (Tabelle): Relationen geben die Beziehungen zwischen Entitäten wieder

Entity-Relationship-Diagramm

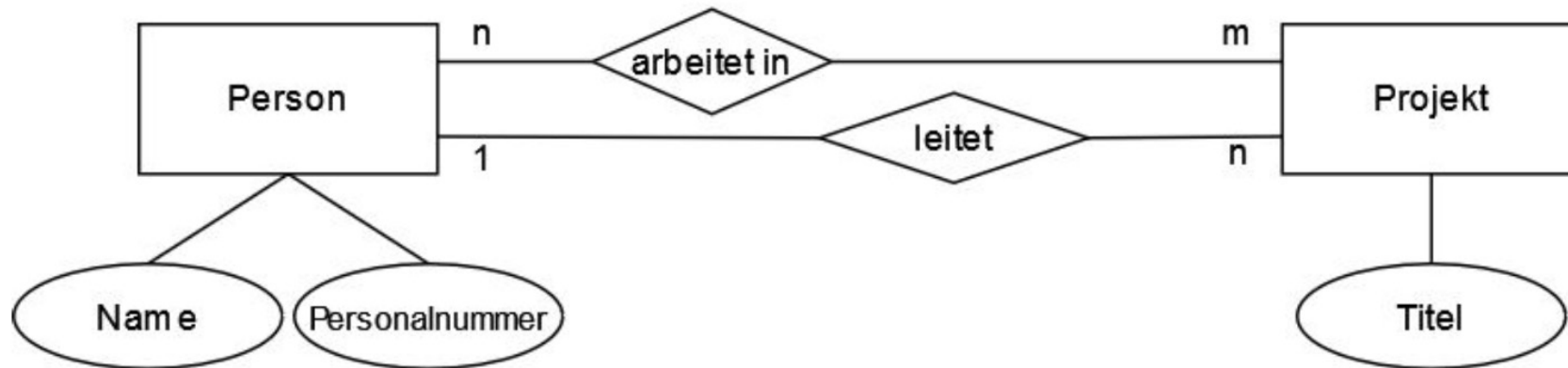
- Grafische Darstellung des Entity-Relationship-Modells
- Es existieren verschiedene Variationen

Kardinalität:

- 1 : 1 – one to one
 - z.B.: 1 Studierende*r hat 1 Matrikelnummer
- 1 : n – one to many
 - z.B. 1 Ort ist Geburtsort von n Personen
- n : m – many to many
 - z.B.: n Personen arbeiten in m Unternehmen
- Es kann auch Minimum und Maximum angegeben werden
 - z.B.: Jede Person wurde an genau 1 (min. 1 und max. 1) Ort geboren.
 - In einer Veranstaltung können 5 bis 20 Studierende teilnehmen.

Entity-Relationship-Diagramm: Beispiel

Beispiel: In einem Unternehmen sind Mitarbeitende an verschiedenen Projekten beteiligt. In dem Unternehmen ist festgelegt, dass mehrere Personen jeweils in mehreren Projekten arbeiten können, aber jedes Projekt von genau einer Person geleitet wird. Personen und Projekte haben jeweils Attribute.



Relationales Datenbankmodell: Beispiel

Beispiel: In einem Unternehmen arbeiten Mitarbeitende in verschiedenen Projekten.

Vorname	Nachname	Telefonnummer	Projekt	Projektstart
Brigitte	Beispiel	0221 12345	Superprojekt	01.07.2020
Brigitte	Beispiel	0221 12345	Duperprojekt	01.07.2020
Jane	Smith	0221 52478	Projekt 3000	15.08.2023
Joe	Smith	0221 21738	Superprojekt	01.07.2020

Relationales Datenbankmodell: Beispiel

Mitarbeiter*in:

Vorname	Nachname	Telefonnummer	Projekt
Brigitte	Beispiel	0221 12345	Superprojekt
Brigitte	Beispiel	0221 12345	Duperprojekt
Jane	Smith	0221 52478	Projekt 3000
Joe	Smith	0221 21738	Superprojekt

Projekt:

Titel	Projektstart
Superprojekt	01.07.2020
Duperprojekt	01.07.2020
Projekt 3000	15.08.2023

Relationales Datenbankmodell: Beispiel

Mitarbeiter*in:

PersonalID	Vorname	Nachname	Telefonnummer
1	Brigitte	Beispiel	0221 12345
2	Jane	Smith	0221 52478
3	Joe	Smith	0221 21738

Projektmitarbeit:

PersonalID	Projekt
1	Superprojekt
1	Duperprojekt
2	Projekt 3000
3	Superprojekt

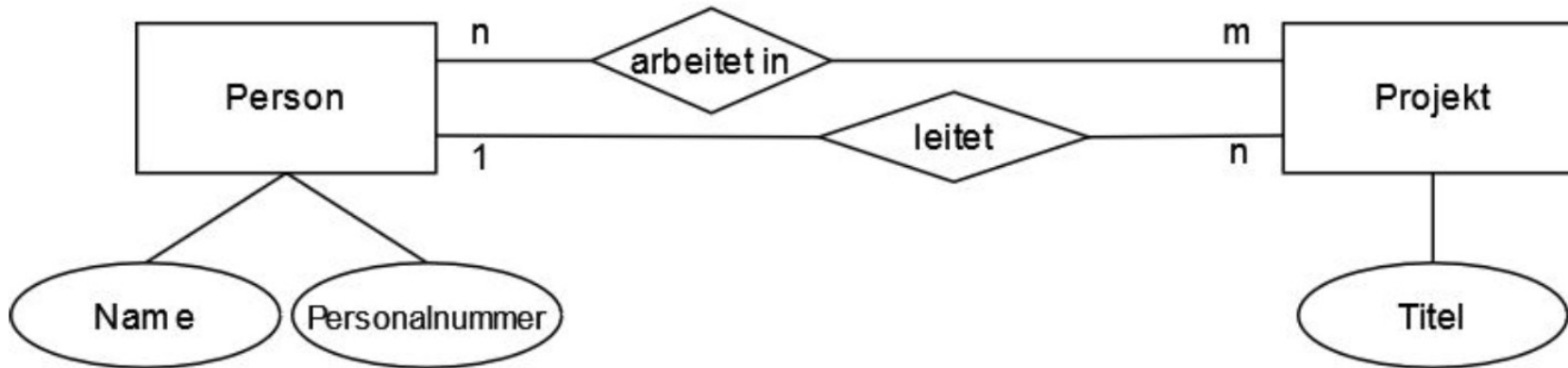
Projekt:

Titel	Projektstart
Superprojekt	01.07.2020
Duperprojekt	01.07.2020
Projekt 3000	15.08.2023

Übung: Anforderungserhebung und Konzeption einer relationalen Datenbank

- Wählen Sie ein Beispiel aus (weitere Details können Sie sich ggf. ausdenken):
 - Datenbank über Personen und ihre Haustiere
 - Datenbank von Bestellungen in einem Onlineshop
 - Datenbank über Filme und ihre Schauspieler*innen und Regisseur*innen
 - Datenbank der Gebäude der Universität zu Köln
- Überlegen Sie sich:
 1. Welche Entitäten sind dabei relevant?
 2. Welche Beziehungen bestehen zwischen den Entitäten?
 3. Welche Attribute haben die Entitäten?
 4. Welche Kardinalitäten haben die Beziehungen?
- Machen Sie sich Notizen und/oder erstellen Sie Diagramme

Übung: Anforderungserhebung und Konzeption einer relationalen Datenbank



- Überlegen Sie sich:
 1. Welche Entitäten sind dabei relevant?
 2. Welche Beziehungen bestehen zwischen den Entitäten?
 3. Welche Attribute haben die Entitäten?
 4. Welche Kardinalitäten haben die Beziehungen?
- Machen Sie sich Notizen und/oder erstellen Sie Diagramme

Übung: Anforderungserhebung und Konzeption einer relationalen Datenbank

Bringen Sie Ihre Notizen/Diagramme von dieser Übung gerne für die nächste Sitzung wieder mit, denn dann können Sie bei den nächsten Übungen weiter damit arbeiten.