



Planung des Forschungsdaten-Managements: Planung, Auswahl und Ingest

Gabriel Stöckle
ZAH Heidelberg
gst@ari.uni-heidelberg.de

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Planung

- Ziele des Projekts
- Beziehung zu vorhandene Daten
- Bewertung der eigenen Daten
- Kriterien für Datenformate

Auswahl

- Gründe für die Archivierung
- Auswahlkriterien
- Aufbewahrungsdauer

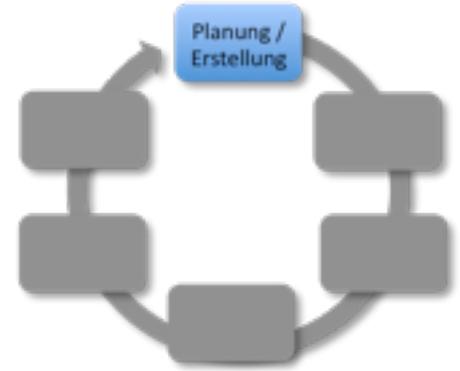
Ingest

- Verfahren der Datenübernahme
- Verantwortung



Planung und Erstellung

Unmittelbare **Wissenschaftliche Ziele** bestimmen, was erfasst und erstellt wird.



Nachnutzung und Nachvollziehbarkeit wird zunehmend thematisiert und stellt hohe/neue Anforderungen an Forscher.

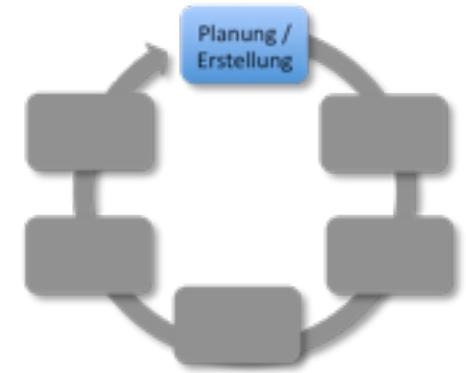
Planung des Daten-Managements so früh wie möglich im Projekt!



Planung und Erstellung - Vorhandene Daten

Frage:

Wie stelle ich meine Daten zur Verfügung und integriere diese in bestehende Datensätze?



zunächst:

Recherche vorhandener vergleichbarer Datensätze

dann:

Auswahl der eigenen Daten nach

- Wiederherstellbarkeit
- Erstellungs- und Erfassungsaufwand



Planung und Erstellung - Bewertung

Zur Archivierung, Weitergabe und Verarbeitung wissenschaftlicher Ergebnisse können unterschiedliche Daten genügen.

- ***Experimente, Beobachtungen:***

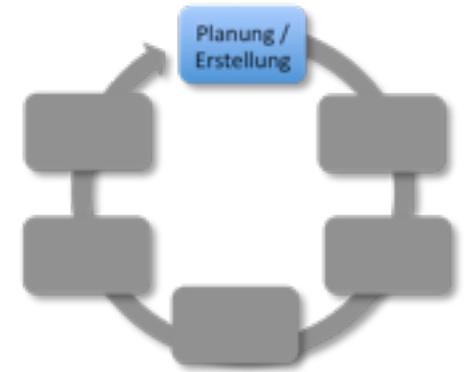
Neben aufbereiteten Daten werden Roh- und Kalibrationsdaten benötigt!

- ***Modelle oder Simulationen:***

Modellparameter, Hardware-Charakteristiken und Ergebnisse genügen, Rohdaten sind meist nicht archivwürdig.

- ***Abgeleitete Daten oder Referenzdaten:***

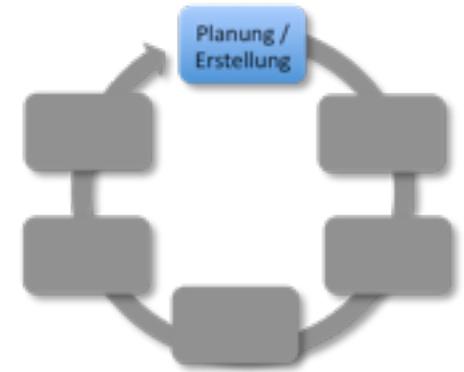
Können getrennt verarbeitet werden, sollten aber im Kontext der Originaldaten gesehen werden.



Planung und Erstellung - Formatkriterien

Die Wahl eines „**guten**“ Datenformats entscheidet oft über Nachnutzbarkeit und Interoperabilität

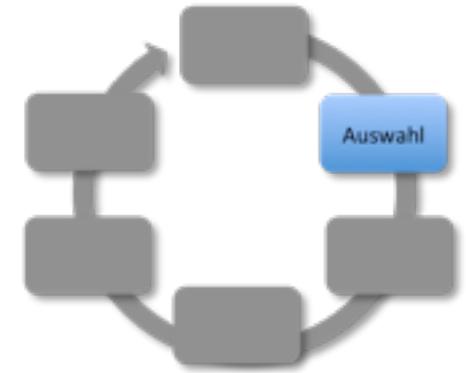
- **Einfach**: leichtere Verarbeitung
- **Flexibel**: Das Format ist für möglichst viele Daten anwendbar
- **Verbreitung** in der Community, Software und Tools existieren
- **Formale Spezifikationen** existieren und sind frei verfügbar
- Kann **unabhängig** von Hard- und Software verwendet werden
- **Datenintegrität, Provenienz, Robustheit**





Auswahl und Aufbewahrungsdauer

Aus welchem Grund archiviere ich Daten?



- **Arbeitskopie**

Ziel: Sicherung der Arbeit während des Projekts

- **Nachweis, Teil einer Publikation**

Ziel: Forschungsergebnisse überprüfen/darstellen

- **Nachnutzung**

Ziel: Community building, Zitierung

- **Dokumentation**

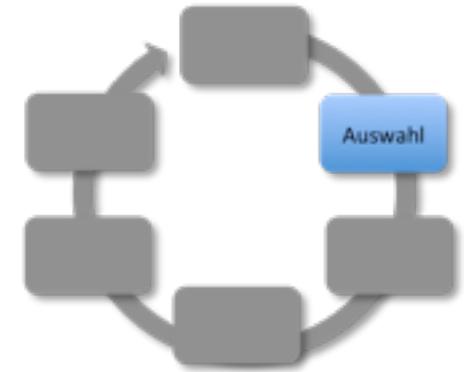
- **gesetzliche Auflagen**

inhaltliche Archivwürdigkeit

- Erzeuger legt Qualitätskriterien fest
- Dialog mit Datenzentrum
- ggf. Rückkopplung Nutzer
- QK sollten *transparent* sein

Beispiele:

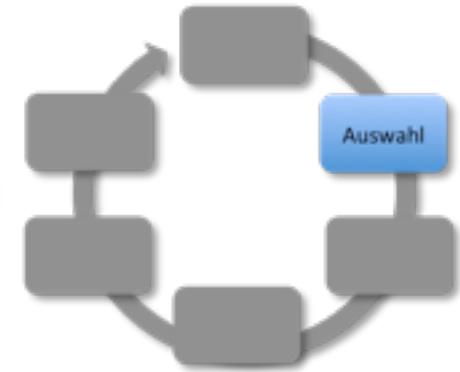
- Methode, verwendete Normen und rechtliche Grundlagen und
- Rohdaten, Provenienzdaten und Fachgutachten werden mit den Daten geliefert.



technische Archivfähigkeit

Daten befinden sich in einem einem vom Archiv vorgegebenen Format.

Metadaten sind vorhanden und verarbeitbar.



wichtige Fragen zur Datenauswahl

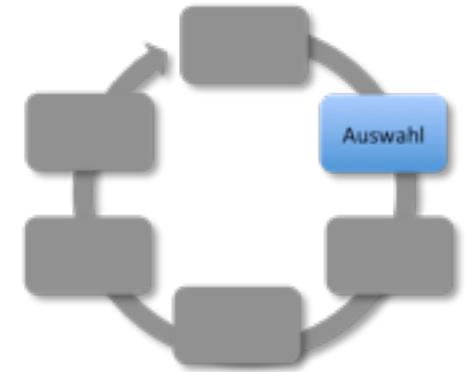
- Wann erfolgt die Auswahl?
- Wer ist verantwortlich?
- Welche Auswahlkriterien?
- Wie ist das technische Verfahren?

Auswahl - Aufbewahrungsdauer

Die Aufbewahrungsdauer ist abhängig von den **Archivierungsgründen**.

Aber: diese können sich verändern!

- Nicht nur fachliche Kriterien beachten!
- Weitere Nachnutzer in Betracht ziehen!



Kosten/Nutzen

evtl. Nutzerstatistik führen und auswerten!

- Wer nutzt bzw. hat Interesse an den Daten?
- Wie oft werden Daten zitiert?

Gesetzlich vorgeschriebene **Löschungen** sind in jedem Fall durchzuführen!

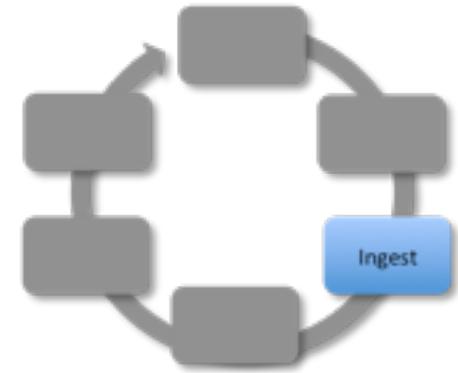


Ingest

Ingest: Einlagern von Daten ins Archiv

ein typischer Ingest-Workflow:

- **Transport** (Netzwerk, Sneaker-Net)
- **Vorbereitung**
- Vergabe eines archivweiten Identifiers
- Metadaten: Erzeugung, Extraktion oder Übernahme
- Test auf Schadsoftware
- ggf. Formatkonvertierung
- Prüfungen auf Vollständigkeit und Richtigkeit, etc.
- **Einfüllen**: Prüfsummen erzeugen, um ungewollte Veränderungen überprüfen zu können



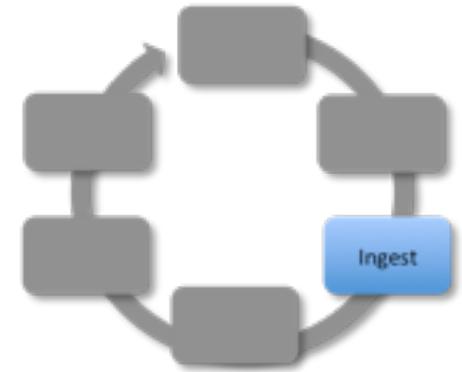
Jeder Schritt des Ingest sollte *dokumentiert* werden !

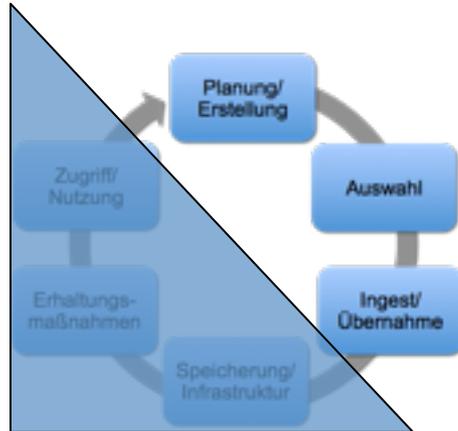
Ingest - Verantwortungsübernahme

Die rechtlichen Aspekte sollten in Form einer **Übernahmevereinbarung** zwischen Produzent und Datenzentrum geklärt sein!

Diese sollte enthalten:

- Liste der zu archivierenden Datensätze
- Liste der Datenobjekte
- Rahmenbedingungen (Aufbewahrungsdauer, Sperrfristen)
- Kostenschätzung
- Regelungen zu Urheberrecht und Haftung
- Zeitplan





Planung und Erstellung

- Frühe Planung des Datenmanagements zahlt sich aus
- Verwendung „guter“ Formate

Datenauswahl:

- Qualitätskriterien und Aufbewahrungsdauer klären

Ingest:

- Überenahmevereinbarung erstellen
- Dokumentation des Verfahrens



...

...