**Measure 1: Task 4**

**Erstellung einer Übersicht über die Datenverarbeitungsstandards**

**Strukturierte Liste:**

1. **Bestehender technischer Empfehlungen**
2. **Von Richtlinien**
3. **Standards für die Datenverarbeitung in der historischen Forschung**

**Antrag:**

1. Assessing the status quo
2. Nachhaltiges Qualitätsmanagement
3. Verschiedene Datentypen und Phasen des Datenlebenszyklus
4. Aktuelle Richtlinien und Prinizipien
5. Historical source criticism
6. Best practices and obstacles to compliance with quality measures
7. Outcome: **Task 4**

**Richtlinien**

Data Governance Act: <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained>

Data Processing Principles: <https://www.clarin.eu/content/principles-data-processing>+

Datenverarbeitung zu wissenschaftlichen oder historischen Forschungszwecken und zu statistischen Zwecken: <https://dsgvo-gesetz.de/bdsg/27-bdsg/>

**Datenverarbeitung in der historischen Forschung:**

**Es geht hier eher um Verarbeitung von Textdaten.**

Hermeneutische Reichweite: Mit digitalen Methoden ein besseres Textverständnis bekommen.

Handwritten Text Recognition: Sprach- und Mustererkennung mit Hidden Markov Modelle und N-Gramms, die dann in Interaktion mit dem Benutzer über Trainingstranskriptionen die Vorhersagegenauigkeit für Manuskripte eines Schreibers z.T dramatisch verbessern (Wettlaufer 2016: 6.). Gute HTR-Genauigkeit braucht eine Kombination von Verfahren wie Layout-Analyse, Textzeilen-Extraktion, Vorverarbeitungsoperationen, Training, lexikalische- und Sprachmodellierung, HMMs, usw.

Beispiele: Venice Time Machine, Transcription and Recognition Platform

Soziale Netzwerkanalyse: Gephi, nodeXL, Javascript Frameworks: SWigmajs

Semantische Technologien: Forschungsbeispiele CompHistSem, HSC, LOEWE => Quantitative Analysen auf der Basis von lemmatisierten Texten ermöglichen z.B die Beobachtung von semantisch relevanten Gebrauchsverschiebungen; das geschieht in einer komfortablem Umgebung. Linked Data kann heute genutzt werden, um digitale Ressourcen miteinander zu verknüpfen und damit Zusammenhänge evident zu machen.

Datenbank: Relationale Datenbanksystemen sind Sammlungen von Informationen, bei der Daten in vordefinierten Beziehungen organisiert sind. Beispiel: SQL Queries, um verschiedene Datenpunkte zu kombinieren.

Ergebnisse Kick-Off: <https://docs.google.com/document/d/1ForfbU6UCz2A6XI9yGMn0_cYQFsIP0pHfzsyZz-E3rs/edit?tab=t.0>

Datenqualitätsmanagement während des Datenlebenszyklus:

1. Datenerhebung: Klarstellung der Objectives der Organisation und Richtlinien zu Erfassung und Erhebung
2. Data Storage and Maintenance: Richtlinien zu Backups und Speichern von Daten => Fokus auf Redundanz von Daten, da die Kopie ein Backup sein kann, falls die Daten gelöscht werden.
3. Data Usage: Staatliche Richtlinien wie beispielsweise Datenschutzrechte
4. Data Sharing and Distribution: Wieder DSVGO
5. Data Archiving: Wie soll die Daten archiviert werden und wie lange => Data Retention Policies
6. Data Disposal: Daten, die nicht gebraucht werden, müssen rechtgemäß gelöscht werden.

**Leitfaden zur Nutzung von qualitativ hochwertigen Daten:**

Priorisierende Datenformate

Fraunhofer: <https://cdn0.scrvt.com/fokus/551bf951bf1982f5/0c96fbf464ef/NQDM_Leitfaden_2019.pdf>