DIO
Data Intelligence Offensive
Verein zur Förderung der Datenwirtschaft

Workshop
Deep Dive into Data

für Antragsteller*innen bei der IKT der Zukunft Ausschreibung 2020

Fachinput für Paneldiskussion Data Management Plan

Axel Quitt, DIO Data Steward / Consultant

Inhalt

- DMPOnline als Werkzeug für DMP-Erstellung
- Vertraulichkeit
- Best Practice
- DMP Grundstruktur
- Praxistipp

Verpflichtender Datenmanagement-Plan

„Antragsteller*innen bei diesem Schwerpunkt sind verpflichtet, einen Datenmanagementplan (DMP) als Annex zur Projektbeschreibung vorzulegen.

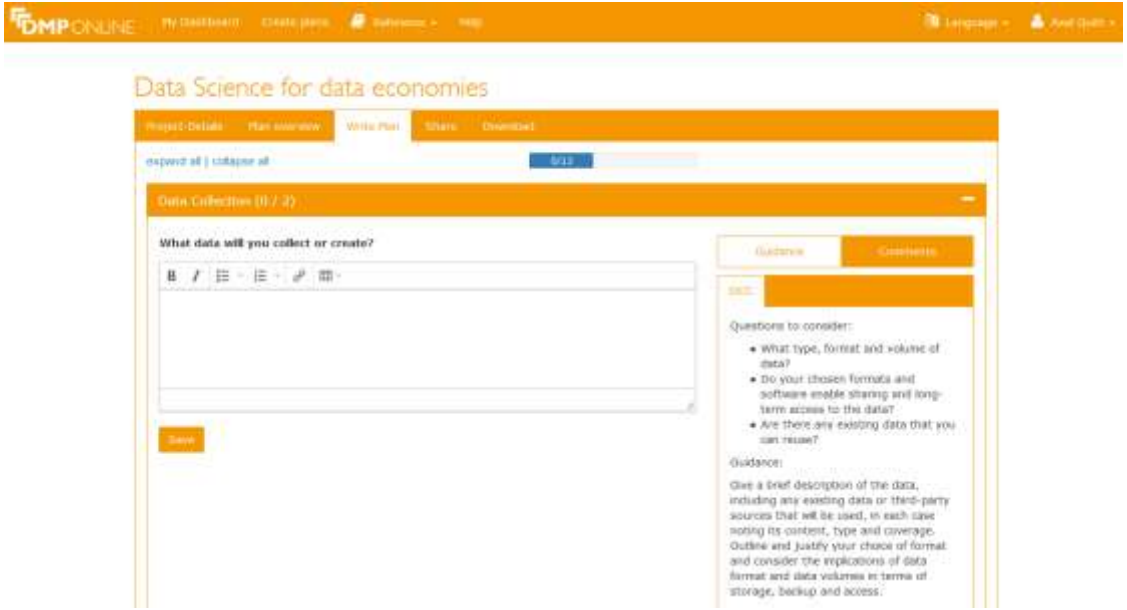
Ein Datenmanagement-Plan beschreibt,

- welche Daten im Projekt gesammelt, erarbeitet oder generiert werden,
- wie mit diesen Daten im Projekt umgegangen wird,
- welche Methoden und Standards dabei angewendet werden,
- wie die Daten langfristig gesichert und gepflegt werden, und
- ob es geplant ist, Datensätze Dritten zugänglich zu machen und ihnen die Nachnutzung der Daten zu ermöglichen (sog. „Open Access zu Forschungsdaten“ oder auch in „Datenkreisen“ – …). [1] “

Quelle: [1] Ausschreibung IKTdZ

Das Online-Tool DMP Online

- https://dmponline.dcc.ac.uk/plans
- Online Formular Service
- Data Management-Plan Struktur
- Ausfüllhilfe inkludiert
- Fragen und Richtlinien unterstützen bei DMP Erstellung
- Content ausschließlich Englisch
- Zusätzlich unterstützt es die Zusammenarbeit im Team

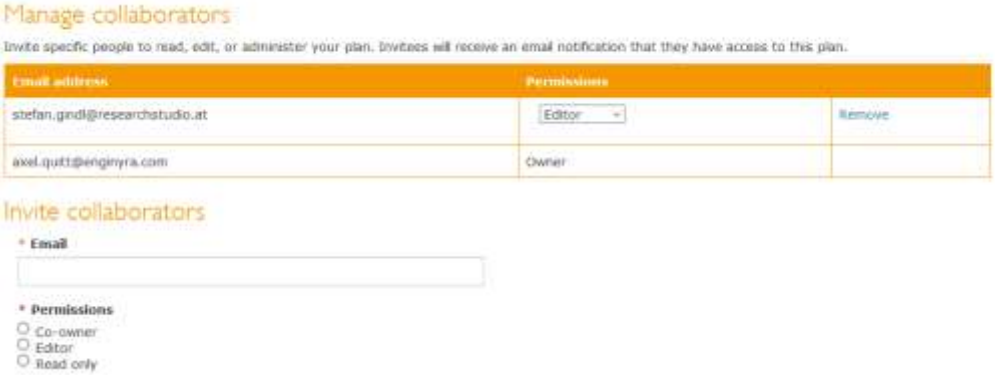


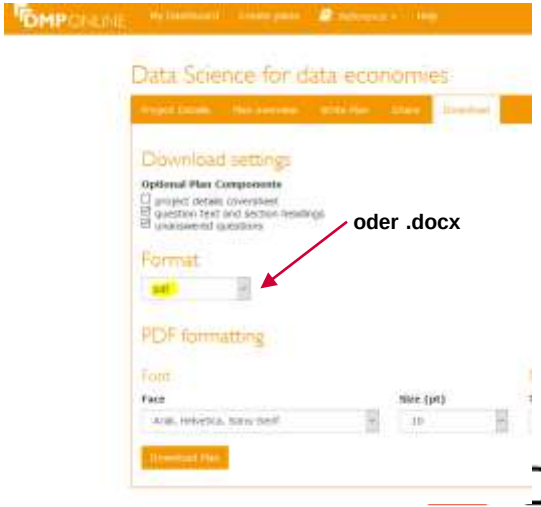
Was ist, wenn ich meinen DMP nicht der Cloud anvertrauen möchte?

Man kann seinen DMP direkt im DMP Online Tool ausfüllen, aber man muss nicht.

Das DMP Online Tool bietet die Möglichkeiten, dass ein Team gemeinsam am DMP arbeitet und es von dieser Plattform direkt publiziert wird.

Will man höchste Vertraulichkeit, kann man die Struktur auch als Word-Datei herunterladen (sogar mit Fragen und Richtlinien) und „Private“ auf eigener Infrastruktur arbeiten

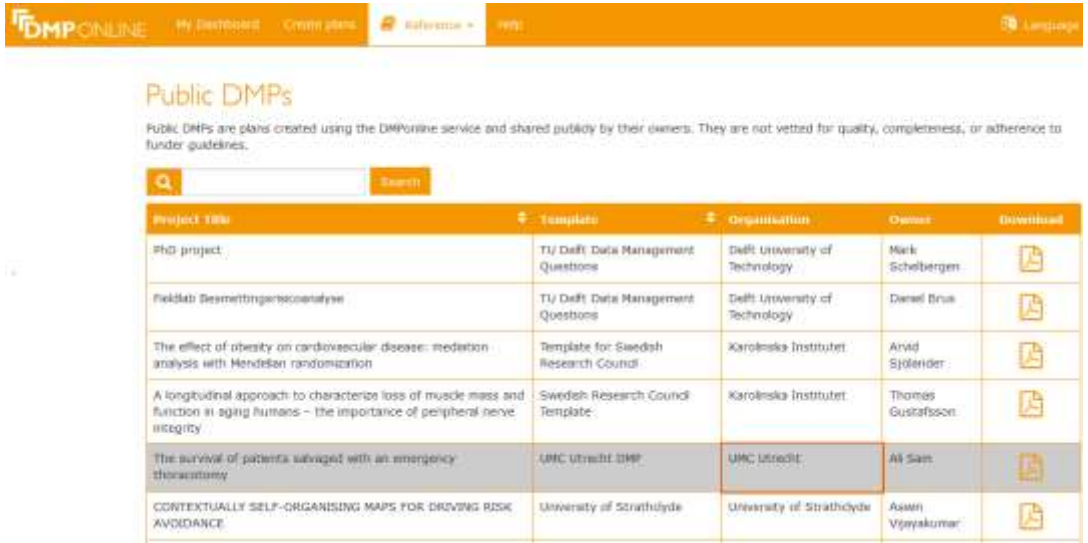




Bietet DMPOnline Best Practice Beispiele?

Verfügbar ist ein Verzeichnung von vielen bereits publizierten Data Management Plänen

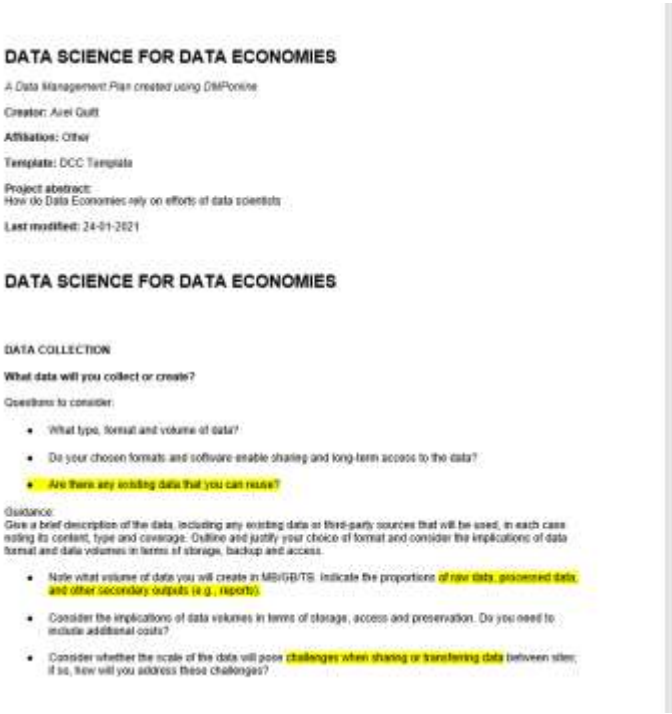
Eine Bewertung über deren Qualität oder Vollständigkeit wird NICHT angeführt



Grundstruktur eines Data Management Plans in DMPOnline

Praxistipp

- Kopiere Fragen ins Formular und Download
- Setze Schwer-punkte (z.B. hier in Gelb)



DATA SCIENCE FOR DATA ECONOMIES

A Data Management Plan created using DMPonline

Creator: Axel Guitt

Affiliation: Other

Template: DCC Template

Project abstract:
How do Data Economies rely on efforts of data scientists

Last modified: 24-01-2021

DATA SCIENCE FOR DATA ECONOMIES

DATA COLLECTION

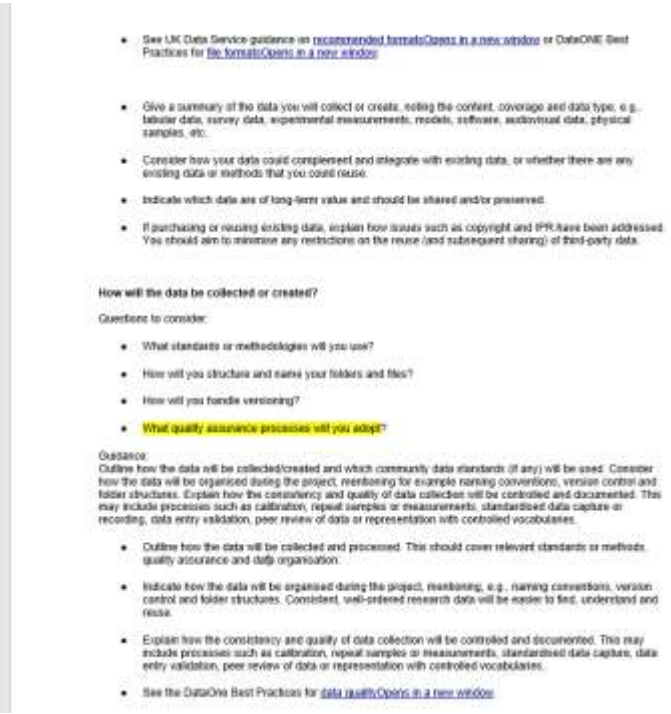
What data will you collect or create?

Questions to consider:

- What type, format and volume of data?
- Do your chosen formats and software enable sharing and long-term access to the data?
- Are there any existing data that you can reuse?

Guidance:
Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.

- Note what volume of data you will create in MB/GB/TB. Indicate the proportions of raw data, processed data and other secondary outputs (e.g., reports).
- Consider the implications of data volumes in terms of storage, access and preservation. Do you need to include additional costs?
- Consider whether the scale of the data will pose challenges when sharing or transferring data between sites; if so, how will you address these challenges?



- See UK Data Service guidance on recommended formatsOpens in a new window or DataONE Best Practices for file formatsOpens in a new window

- Give a summary of the data you will collect or create, noting the content, coverage and data type, e.g., tabular data, survey data, experimental measurements, models, software, audiovisual data, physical samples, etc.
- Consider how your data could complement and integrate with existing data, or whether there are any existing data or methods that you could reuse.
- Indicate which data are of long-term value and should be shared and/or preserved.
- If purchasing or reusing existing data, explain how issues such as copyright and IPR have been addressed. You should aim to minimise any restrictions on the reuse (and subsequent sharing) of third-party data.

How will the data be collected or created?

Questions to consider:

- What standards or methodologies will you use?
- How will you structure and name your folders and files?
- How will you handle versioning?
- What quality assurance processes will you adopt?

Guidance:
Outline how the data will be collected/created and which community data standards (if any) will be used. Consider how the data will be organised during the project, mentioning for example naming conventions, version control and folder structures. Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies.

- Outline how the data will be collected and processed. This should cover relevant standards or methods, quality assurance and data organisation.
- Indicate how the data will be organised during the project, mentioning, e.g., naming conventions, version control and folder structures. Consistent, well-ordered research data will be easier to find, understand and reuse.
- Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture, data entry validation, peer review of data or representation with controlled vocabularies.
- See the DataOne Best Practices for data qualityOpens in a new window

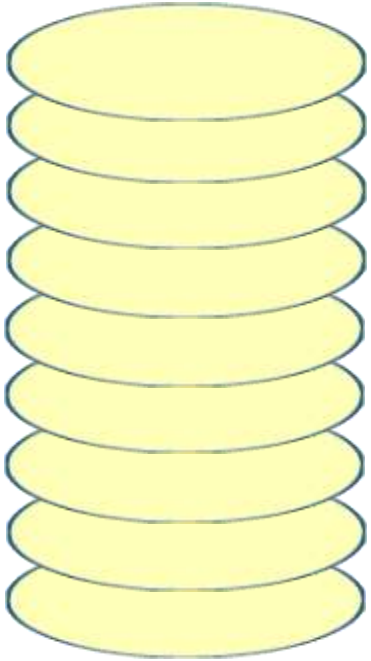
DIO
Data Intelligence Offensive
Verein zur Förderung der Datenwirtschaft

Weitere Ressourcen

Moderationsinput: Interessante Fragen die für Data Management Plan interessant sein können

- Welche Daten können wiederverwendet werden → im Sinne bereits existierender Daten, die Eingang in die Forschungsarbeit finden und im 2. Sinne – welche werden nach der Forschungsarbeit existieren und wiederverwendet werden können.
- Wie findet man diese, und unter welchen rechtlichen Rahmenbedingungen können sie genutzt werden → Link zu Moderation im Rechtsteil
- Herausforderungen beim Datenaustausch (Menge, Formate, Tools, Vorarbeiten) und Möglichkeiten sie zu überwinden → möglichst frühzeitig überlegen (Praxistipp)
- Wie verhandelt man, um Übereinstimmung/Einvernehmen zur Datenverwendung zu erzielen → Tipps und Erfahrungen von den Teilnehmern
- Betriebswirtschaftliche Betrachtungen, v.a. Kosten für Datenakquisition, Vorverarbeitung etc. und Stakeholderbetrachtungen gem. Datenmanagement-Layern.

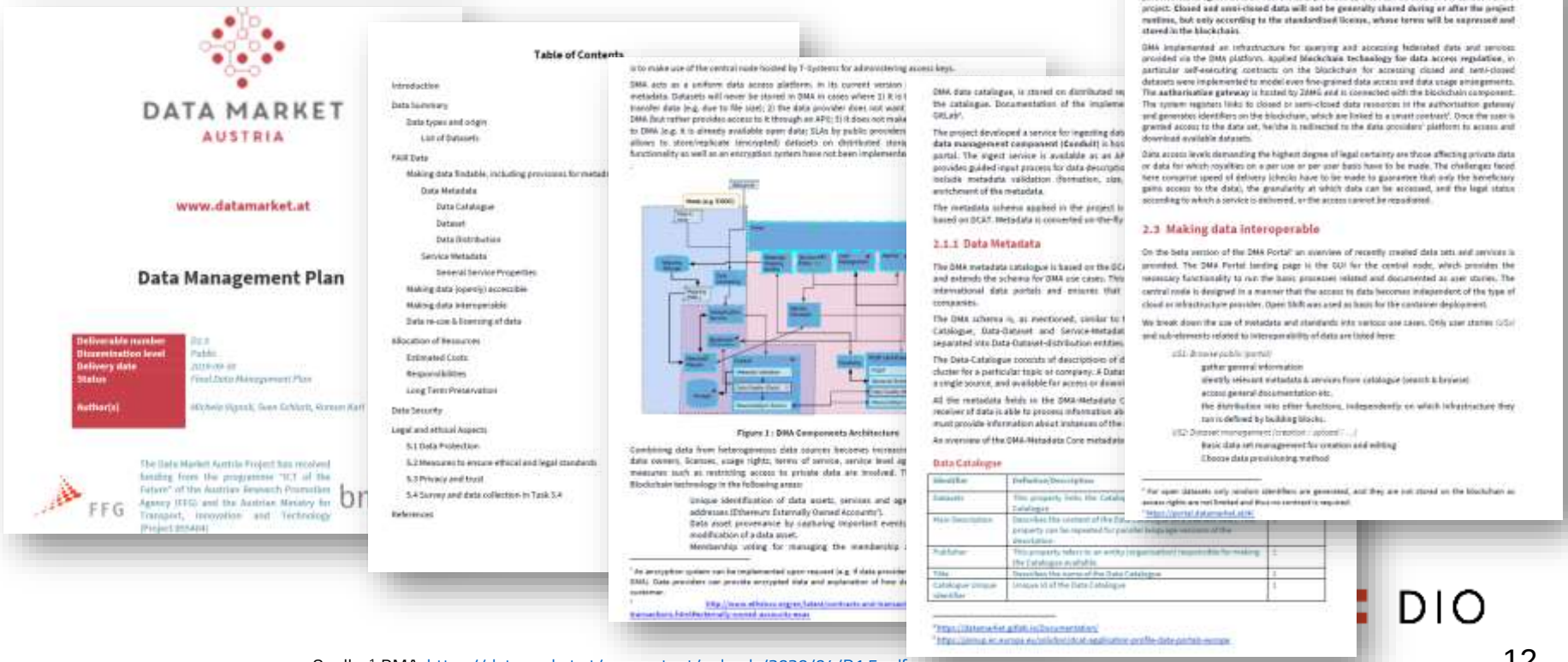
Ebenen des Datenmanagements



- Use Case Vision und Anforderungsdefinition
- Datenwert- / -preisbemessungsmethoden
- Stakeholder-Analyse national / international
- Umfeld-Analyse IST-Stand national / international – bereits bestehende DK
- Rechtliche Rahmenbedingungen für Use Case Umsetzung und Datenverwendung
- Datenmanagement – Datenaufbereitung, -anreicherung, -qualität
- Servicearchitektur / Technische Schnittstellen / Funktionalität
- Dokumentation Datenkreis – Managementebenen
- Infrastrukturarchitektur / Betriebsprozesse

DIO

Beispiel: Datamanagement-Plan des Data Market Austria (DMA) Forschungsprojekts [1]



Quelle: [1] DMA: https://datamarket.at/wp-content/uploads/2020/04/D1.5.pdf