

# Empfehlungen für die Datenbereitstellung im mFUND (Merkblatt für Fördernehmer)

## 1. Auswählen: Welche Daten kommen für eine Bereitstellung grundsätzlich in Frage?

Es geht bei der Bereitstellung grundsätzlich nur um solche Daten, die im Rahmen des durch den mFUND geförderten Projektes entstanden sind. Selbstverständlich steht es Ihnen frei, weitere Daten mit Bezug zum Projekt oder davon unabhängige Daten zusätzlich auf der mCLOUD zur Verfügung zu stellen.

Die bereitzustellenden Daten sollen die grundsätzlichen Open-Data-Prinzipien<sup>1</sup> erfüllen:

- Elektronisch gespeichert  
*Die Daten sollen bereits elektronisch gespeichert vorliegen. Zeichnungen, Notizen, jegliches „analoges“ Material ist nicht gemeint. Eine Digitalisierung ausschließlich zum Zwecke der Datenbereitstellung ist nicht erforderlich.*
- Maschinenlesbar  
*Es sollen nur solche Daten bereitgestellt werden, die maschinenlesbar, also automatisch interpretierbar und analysierbar sind. Nicht gemeint sind in diesem Sinne also Dokumente (reine Schriftstücke, E-Mails, etc.), Fotos oder Dateien, die in nicht weiter bearbeitbaren Formaten vorliegen (z.B. PDF-Dokumente).*
- In Sammlungen strukturiert  
*Wenig hilfreich für Dritte sind völlig unstrukturierte Einzeldaten (z.B. einzelne Messwerte, deren Zusammenhang nicht erkennbar ist). Daher sollten nur solche Daten weiter betrachtet werden, die bereits in einem Sinnzusammenhang strukturiert sind.*

Alle Daten, die die vorgenannten Bedingungen (1. Auswählen) erfüllen, sollten im weiteren Verlauf einer Prüfung (2. Prüfen) unterzogen werden. Folgende Daten sollten betrachtet werden:

- Rohdaten  
*Für den Zweck der Bereitstellung als Open Data sind Rohdaten besonders interessant, da diese noch keine Bearbeitung und Interpretation enthalten. Ein Beispiel hierfür sind Sensor-Messdaten. Grundsätzlich sind die Rohdaten zu veröffentlichen. Manchmal handelt es sich dabei jedoch um sehr große Datenmengen, die zudem stark fehlerbehaftet sein können. Daher ist alternativ die Veröffentlichung von aufbereiteten, plausibilisierten Daten zu prüfen.*
- Aufbereitete, plausibilisierte Daten  
*In vielen Fällen werden Rohdaten erst einer Qualitätsprüfung, Fehlerkorrektur und Strukturierung unterzogen, bevor sie weiterverarbeitet werden. Für die meisten Nachnutzer werden solche Daten interessanter sein, als die Rohdaten, da bei Rohdaten jeder Nutzer die Vorverarbeitungsschritte erneut ausführen müsste.*

---

<sup>1</sup> Einen guten Überblick über die Prinzipien von offenen Daten gibt die Konrad-Adenauer-Stiftung unter <http://www.kas.de/wf/de/71.15334/>. Der Fokus liegt in dieser Beschreibung auf „Verwaltungsdaten“, aber die Grundsätze lassen sich weitestgehend auch auf private Daten übertragen.

- Ergebnis- und Analysedaten  
*Bei der Verarbeitung von Daten entstehen oftmals verschiedene Zwischenstände, die beispielsweise nach einer Kalibrierung oder Validierung anfallen. Auch durch die Veredelung von Daten, bspw. durch Kombination mit anderen Daten, durch statistische Auswertung oder Interpretation können neue Datenprodukte entstehen. Auch für diese ist eine Veröffentlichung zu prüfen.*

## **2. Prüfen: Worauf sollte vor einer Veröffentlichung von Daten unbedingt geachtet werden?**

Einige Eigenschaften von Daten können einer Veröffentlichung im Sinne von Open Data entgegenstehen. Um insbesondere den Datenschutz, aber auch andere rechtliche Aspekte zu wahren, sind die Daten vor einer möglichen Veröffentlichung mindestens auf folgende Aspekte hin zu prüfen (ggfs. wird diese Liste erweitert):

- Beinhalten die Daten personenbezogene oder personenbeziehbare Daten im Sinne des Bundesdatenschutzgesetzes (BDSG)?  
*Hierzu zählen alle Informationen, die einen Rückschluss auf natürliche Personen zulassen. Manchmal ist dieser Rückschluss erst durch die Kombination mit anderen Daten möglich. Daher ist zu prüfen, ob bei den zu veröffentlichenden Daten diese Gefahr besteht. Hier kann nur nach bestem Wissen und Gewissen entschieden werden, da niemand einen vollständigen Überblick über mögliche Kombinationsmöglichkeiten mit anderen Daten besitzt. Sollte aber ein Personenbezug in den Daten vorliegen oder erahnt werden, so ist zu prüfen, ob durch geeignete Maßnahmen der Personenbezug entfernt werden kann. Hierzu zählt die Anonymisierung z.B. durch Aggregation. Grundsätzlich kann auch die Einwilligung von Betroffenen eingeholt werden.*
- Würde eine Veröffentlichung der Daten das geistige Eigentum, Urheberrechte Dritter oder vertragliche Bedingungen verletzen?  
*Wenn durch eine Veröffentlichung von Daten Urheberrechte Dritter, Betriebs- und Geschäftsgeheimnisse, Marken- oder Patentrechte verletzt würden, ist zu prüfen, ob eine Einwilligung der Betroffenen eingeholt werden kann. Andernfalls können solche Daten nicht veröffentlicht werden. Gleiches gilt, wenn Daten aufgrund von Verträgen, die Sie eingegangen sind, nicht weitergegeben werden dürfen.*
- Handelt es sich um vertrauliche Daten?  
*Gibt es bei den Daten eine Geheimhaltungs- oder Vertraulichkeitspflicht oder wurden die Daten vertraulich erhoben oder übermittelt und besteht das Interesse Dritter an einer vertraulichen Behandlung fort, gilt das gleiche wie bei personenbezogenen Daten: Anonymisieren, Einwilligung einholen oder Veröffentlichung unterlassen.*
- Würde eine Veröffentlichung der Daten den Schutz öffentlicher Belange verletzen?  
*Es kann vorkommen, dass sich die Veröffentlichung von Daten negativ auf sog. öffentliche Belange auswirkt. Hierzu zählt beispielsweise die öffentliche Sicherheit, die dadurch gefährdet würde, wenn detaillierte Informationen über kritische Infrastrukturen veröffentlicht würden. Ein weiteres Beispiel wäre, wenn durch eine Datenveröffentlichung laufende Gerichtsverfahren beeinträchtigt würden. Sollte ein solcher Fall bei den zu veröffentlichenden Daten zu erwarten sein, ist entweder eine Reduktion des Informationsgehalts zu prüfen oder eine Veröffentlichung zu unterlassen.*

- Hat eine Veröffentlichung der Daten negative Auswirkungen auf das eigene Geschäftsmodell?

*In vielen mFUND-Projekten geht es um die Entwicklung datenbasierter Innovationen. Dabei können Daten entstehen, die fester Bestandteil einer im Rahmen des Projektes entwickelten Geschäftsidee sind. Wenn eine Veröffentlichung negative Auswirkungen auf die Tragfähigkeit dieser Idee hätte, sollte eine Reduktion des Informationsgehalts vor Veröffentlichung in Betracht gezogen werden oder auf eine Veröffentlichung verzichtet werden.*

Sollten Belange gegen eine direkte Veröffentlichung sprechen, sind die Daten sofern möglich soweit zu bearbeiten, dass einer Veröffentlichung keine Belange mehr entgegenstehen. Der Umfang nicht veröffentlichter Daten ist auf ein Minimum zu reduzieren. **Final nicht zur Veröffentlichung geeignete Daten sind unter Angabe von Gründen nachvollziehbar mit den Zwischen- und Abschlussberichten des Projektes zu erläutern.**

### **3. Veröffentlichung vorbereiten: Wie können die Daten veröffentlicht werden?**

Für die Veröffentlichung von Daten als Open Data sind sowohl die Daten selbst, als auch eine Beschreibung der Daten (Metadaten) bereitzustellen. Daher sollte nach der Prüfung (Schritt 2) die Veröffentlichung der Daten vorbereitet werden. Ziel seitens des mFUND ist es, möglichst viele Daten aus den Projekten über das Open-Data-Portal des BMVI ([www.mCLOUD.de](http://www.mCLOUD.de)) verfügbar zu machen. Hierfür muss zunächst für jeden zu veröffentlichenden Datensatz eine zentrale Frage beantwortet werden können:

- Gibt es im Projekt (bzw. im Unternehmen oder in der Institution) Möglichkeiten, die Daten an öffentlich zugänglicher Stelle längerfristig abzulegen?

*Kann diese Frage mit JA beantwortet und z.B. ein eigener Webserver oder eine öffentliche Cloud-Infrastruktur genutzt werden, kann direkt Schritt 4 (Metadaten erfassen) fortgeführt werden. Wenn Sie die Daten nicht selbst bereitstellen können, wenden Sie sich bitte an [mcloud@bmvi.bund.de](mailto:mcloud@bmvi.bund.de). Den weiteren Bereitstellungsprozess können wir dann im Dialog erörtern und ggfs. Alternativen aufzeigen.*

Zur Unterstützung des Bereitstellungsprozesses soll ein technisches Verfahren erarbeitet werden. Vorläufig gibt es jedoch nur die beiden Alternativen der eigenen Datenbereitstellung oder der Ablage in einem Cloud-Speicher, der nach Absprache über das BMVI bereitgestellt wird.

### **4. Metadaten erfassen: Welche Datenbeschreibungen werden für eine Veröffentlichung in der mCLOUD benötigt?**

Daten werden erst dann auffindbar und für Dritte nutzbar, wenn sie gut beschrieben sind. Daher bedarf es für alle Daten, die einer Veröffentlichung zugeführt werden können, einer Beschreibung mit Metadaten. Zur Unterstützung der Metadatenerfassung wird derzeit ein Online-Werkzeug erstellt. Solange dieses noch nicht nutzbar ist, senden Sie die Informationen bitte per E-Mail an [mcloud@bmvi.bund.de](mailto:mcloud@bmvi.bund.de). Folgende Details sollen erfasst werden (\* = Pflichtangaben):

- Titel\*  
*Name des Datensatzes.*
- Kurzbeschreibung\*  
*Prägnante Aussage über den Inhalt des Datensatzes. Wenn der Inhalt nicht selbsterklärend ist, sollte hier eine ausführlichere Beschreibung des Datensatzes, inkl. Hinweisen, die für die Nachnutzung erforderlich sind (z.B. Erläuterung kryptische Spaltennamen in Tabellen) angebracht werden.*
- Datenbereitsteller (Herausgeber)\*  
*Institution, Firma, Behörde, etc., die die Daten erstellt und über die Veröffentlichung entschieden hat.*
- Lizenz\*  
*Für eine Veröffentlichung und Nutzung der Daten kommen folgende Lizenzen in Betracht:*
  - [Datenlizenz Deutschland – Namensnennung – Version 2.0](#) (verpflichtet den Datennutzer, den jeweiligen Datenbereitsteller zu nennen)
  - [Datenlizenz Deutschland – Zero – Version 2.0](#) (ermöglicht einschränkungslose Weiterverwendung)
  - [Creative Commons – Namensnennung – Version 4.0](#) (verpflichtet den Datennutzer, den jeweiligen Datenbereitsteller zu nennen)
  - [Creative Commons – Namensnennung-Share Alike – Version 4.0](#) (verpflichtet den Datennutzer, den jeweiligen Datenbereitsteller zu nennen und bei einer Weiterveröffentlichung die gleichen Bedingungen einzuhalten)
  - [Creative Commons – Public Domain Dedication – Version 1.0](#) (Daten sind in die Gemeinfreiheit entlassen)
  - Weitere empfohlene Open-Data-Lizenzen (siehe <http://opendefinition.org/licenses/>)
  - Wenn eine der vorgenannten offenen Lizenzen nicht infrage kommt, nennen Sie hier die Nutzungsbedingungen, unter denen eine Weiterverwendung der Daten möglich ist.

Das BMVI empfiehlt die Verwendung der Datenlizenz Deutschland – Namensnennung oder Zero – Version 2.0.

- Schlagworte\*  
*Schlagworte, die den Datensatz prägnant beschreiben. Idealerweise zusätzlich passende mCLOUD-Kategorien (z.B. Klima und Wetter, Straßen, etc.) nennen. Nennen Sie als Schlagworte auf jeden Fall auch Ihr Projekt (Kürzel) und Ihr Förderkennzeichen. Sofern nicht bekannt, finden Sie dieses auf Ihrem Förderbescheid beginnend mit 19FXXXXX. Dies erleichtert später die Zuordnung von Daten zu konkreten mFUND-(Teil-)Projekten.*
- Link\*  
*Falls die Daten bereits offen bereitgestellt werden, sind die Links zu den Daten oder*

*Datendiensten anzugeben (Mehrfachnennungen möglich). Sollten die Daten noch nicht veröffentlicht sein (siehe 3: Veröffentlichung vorbereiten), kann diese Information frei gelassen werden.*

- Kontakt  
*Kontaktdaten des Datenbereitstellers (bitte ohne personenbezogene Informationen; am besten einfach eine Funktions-E-Mail-Adresse).*
- Quellenvermerk  
*Sollten bestimmte Anforderungen an die Nennung der Quelle bestehen (z.B. Quelle nicht nennen oder Quelle in besonderer Art und Weise nennen), ist dies entsprechend anzugeben.*
- Distribution  
*Handelt es sich um einen einzelnen Datensatz oder um eine Distribution (z.B. eine Datenreihe)?*
- Grad der Zugänglichkeit  
*Ist die Weiterverwendung offen für Jedermann oder gibt es Einschränkungen?*
- Letzte Aktualisierung  
*Datum, an dem die Datenstruktur zuletzt aktualisiert wurde.*
- Aktualisierungsfrequenz  
*Wie häufig wird die Datenstruktur aktualisiert?*
- Räumliche Abdeckung  
*geographische Region, die durch die Datenstruktur abgedeckt wird (z.B. Bund, Bundesland, Kreis, Kommune oder konkrete Koordinaten).*
- Abgedeckter Zeitraum  
*Zeitraum, der durch die Datenstruktur abgedeckt wird (idealerweise: Startzeitpunkt und Endzeitpunkt).*
- Datenformat oder Datenschnittstelle  
*Format, in dem die Daten gespeichert bzw. angeboten werden (z.B. CSV, Shapefile, GeoJSON für das Dateiformat oder WMS, WFS, AtomFeed, REST etc. für eine vorhandene Datenschnittstelle). Wenn die Daten in verschiedenen Formaten bereitgestellt werden können, sind hier Mehrfachnennungen möglich.*

## **5. Berichten: Was soll dokumentiert werden?**

Bitte berichten Sie in Ihren Zwischen- und Abschlussberichten kurz und knapp über den Fortschritt der Datenbereitstellung in Ihrem Projekt. Hierbei sollte insbesondere auch beschrieben werden, welche Daten nicht zur Veröffentlichung geeignet sind und worin die Gründe hierfür liegen.