

HOCHWERTIGE DATENSÄTZE IN DEUTSCHLAND - ZUSAMMENFASSUNG

STUDIE IM AUFTRAG DES BUNDESMINISTERIUMS FÜR WIRTSCHAFT UND ENERGIE

Ausgangssituation und Ziel der Studie

Die novellierte EU-Richtlinie über offene Daten und die Weiterverwendung von Informationen des öffentlichen Sektors (PSI-RL) sieht Durchführungsrechtsakte zur Bestimmung sogenannter hochwertiger Datensätze (High Value Datasets – HVD) vor, die sechs Themenbereiche betreffen. Um Verhandlungen der EU-Mitgliedsstaaten und der EU-Kommission zur Bestimmung dieser Datensätze vorzubereiten und interessierte Kreise an der Diskussion zu beteiligen, beauftragte das Bundesministerium für Wirtschaft und Energie diese Studie. Sie identifiziert *potenziell* hochwertige Datensätze, deren technische, rechtliche und ökonomische Rahmenbedingungen sowie etwaige Anpassungsbedarfe in Folge einer Klassifizierung.

Auswahl der HVD

Ausgangspunkt für die Untersuchung waren bereits existierende Vorschläge für HVD, die in der PSI-RL bzw. in einer von der EU-Kommission beauftragten, zeitgleich angefertigten und zum jetzigen Zeitpunkt (Stand: Dezember 2020) nicht veröffentlichten Studie genannt wurden. Für eine erste Einschätzung des Weiterverwendungspotenzials unterschiedlicher Datenkategorien wurde eine Online-Umfrage durchgeführt. Anschließend wurden konkrete Datensätze aus Deutschland benannt, die den genannten Vorschlägen für potenzielle HVD bestmöglich entsprechen. Die Auswahl der Datensätze sowie damit verbundene Potenziale und Herausforderungen wurden durch einen Projektbeirat sowie Stakeholder-Workshops evaluiert. Dabei flossen die Sichtweisen von Datenbereitstellern und -nutzern ein. Auf dieser Basis wurden abschließend Handlungsempfehlungen erarbeitet.

Ergebnisse

Ob Datensätze *hochwertig* sind, entscheidet sich nach der Logik der PSI-RL primär danach, wie *hoch* deren *Mehrwert*potenzial im Falle einer offen lizenzierten und technisch guten Bereitstellung ist. Positive Effekte einer HVD-Klassifizierung solcher Datensätze können neue Dienstleistungen, Produkte oder Geschäftsmodelle, effizientere Produktionsabläufe oder mehr Wettbewerb sein. Aber auch sozioökonomische Potenziale, wie mehr Transparenz und Wissen für Bürger*innen oder die Schonung natürlicher Ressourcen, sind mögliche positive Folgen. Dementgegen kann eine HVD-Klassifizierung auch zu zusätzlichen Arbeitsaufwänden und Einnahmeverlusten für Datenbereitsteller führen und bestehende Finanzierungsmodelle beeinträchtigen. Neben der obligatorischen juristischen Prüfung, ob die PSI-RL überhaupt auf spezifische Datensätze Anwendung findet, wurden in dieser Studie daher auch die erwartbaren positiven Effekte mit den zu erwartenden Kosten einer HVD-Klassifizierung abgewogen.

Rechtlich zeigte sich dabei, dass die PSI-RL auf nahezu alle untersuchten Datensätze anwendbar ist. Dies gilt auch für Handelsregisterdaten, da die datenbereitstellenden Gerichte öffentliche Stellen im Sinne der PSI-RL sind. Allerdings enthält das untersuchte Handelsregisterpersonenbeziehbare Daten. Hier ist eine weitere Klärung durch die EU- und nationalen Datenschutzstellen erforderlich. Geeignete Anonymisierungsverfahren können diese Problematiken technisch adressieren. Diese sind jedoch auch mit zusätzlichem Umsetzungsaufwand verbunden und können den Nutzen der Daten verringern.

Klare und möglichst offene Lizenzbedingungen sind für die Weiterverwendbarkeit von Daten ein zentrales Erfolgskriterium. Die rechtliche Analyse des Status quo zeigt hierbei, dass einerseits zahlreiche der untersuchten Datensätze bereits jetzt frei unter Namensnennung verwendbar sind. Andererseits unterliegen einige Datensätze Lizenz- und Nutzungsbestimmungen, die eine Weiterverwendung erfahrungsgemäß eher erschweren. Dazu tragen u.a. fehlende Angaben zum Lizenztyp bzw. Urheberrecht sowie übermäßig enge oder unklare Vorgaben zur Kombination von Datensätzen mit abweichenden Lizenzen bei. Diesen Umständen könnte der Durchführungsrechtsakt durch den Verweis auf einheitliche Standards und Musterlizenzen begegnen.

Für die ökonomische Bewertung zukünftiger Potenziale sind die Kosten sowie lizenzrechtlichen Bedingungen der aktuellen Datenbereitstellung wesentlicher Ausgangspunkt. Hier zeigt sich durch die Studie, dass die untersuchten, potenziellen HVD in zwei Gruppen fallen:

1. Für Datensätze, die bereits zum jetzigen Zeitpunkt kostenfrei, technisch gut aufbereitet und ohne Weiterverwendungsbeschränkungen verfügbar sind, entstehen durch eine HVD-Klassifizierung keine oder nur geringe zusätzliche Potenziale. Da die wesentlichen Bedingungen für eine einfache Weiternutzung bereits erfüllt sind, werden deren Potenziale, bspw. in Geschäftsmodellen, bereits genutzt. Für solche Daten sind jedoch häufig auch die Herausforderungen einer HVD-Klassifikation gering.
2. Für Datensätze, die zum jetzigen Zeitpunkt nur gegen Bezahlung, technisch nicht gut aufbereitet oder mit einschränkenden Weiterverwendungsbedingungen verfügbar sind, bietet eine HVD-Klassifizierung zum Teil mittlere bis hohe Potenziale. Profitieren könnten insbesondere neue, datengetriebene Anwendungen für Geschäftsmodelle, die auf Bodenschätzungsdaten, Geländemodellen oder Handelsregisterdaten aufbauen. Je nach den erforderlichen Anpassungen können hier, bspw. durch Gebühren- und Einnahmeausfälle, teilweise aber auch hohe Folgekosten für einzelne Datenbereitsteller entstehen. Insbesondere eine HVD-Klassifizierung von Katasterdaten wäre in Deutschland vor diesem Hintergrund herausfordernd. Anfallen können dabei zunächst erhöhte Investitionskosten, vor allem in Form monetärer Kosten auf Seiten der Datenbereitsteller. Darüber hinaus müssten im Falle der Katasterdaten landesrechtliche Regelungen harmonisiert werden. Neben der Anpassung von etwaigem Kostenrecht und damit einhergehenden Einnahmeverlusten müssten einheitliche technische Standards erfüllt werden. Die damit verbundenen Kosten übersteigen möglicherweise die erwartbaren positiven Effekte.

Technisch gesehen erfüllt bereits heute ein großer Teil der untersuchten Datensätze die von der PSI-RL geforderte Bereitstellung von HVD in maschinenlesbaren Formaten und per Programmierschnittstelle (API). Dies gilt insbesondere für meteorologische Daten und Statistikdaten. Die Analyse zeigt aber auch, dass immer noch viele der untersuchten Datensätze diese Anforderungen nicht erfüllen. Dabei würde gerade die Erfüllung dieser Bedingungen dafür sorgen, dass entsprechende Daten für Datennutzer besser erreichbar und weiterverwendbar wären. Um die mit einer HVD-konformen Bereitstellung verbundenen, weitergehenden Mehrwertpotenziale optimal zu realisieren, müssen allerdings in der Regel auch weitergehende Anforderungen erfüllt sein. Dazu gehören u.a. eine gute Dokumentation der Daten und Schnittstellen, deren klare programmatische Handhabbarkeit sowie die Performanz und Zuverlässigkeit der API. Aus technischer Perspektive sind die damit verbundenen qualitativen Verbesserungen der Datenbereitstellung grundsätzlich anzustreben.

Ökonomisch sind die dadurch erwartbaren Zusatzeffekte allerdings mit gegebenenfalls erheblichen Zusatzinvestitionen abzuwägen. Besonders kritisch ist, dass oftmals große Mehrwertpotenziale bereits durch eine lizenzrechtlich offene, kostenfreie Datenbereitstellung (per Download) erreicht werden können. Folglich wurde in Workshops von einigen Stakeholdern hinterfragt, ob eine technisch aufwändigere, kostenintensivere Bereitstellung per API in Relation zur Bereitstellung per Download unter Effizienzgesichtspunkten vertretbar ist. Allein aus technischen Anforderungen, wie einer umfangreichen Dokumentation, Metadatenpflege und einer skalierbaren Datenbankstruktur, lässt sich allerdings aus Sicht dieser Studie nicht die Einschätzung ableiten, dass ein API-Betrieb zu aufwändig und kostenintensiv wäre. Denn diese Anforderungen sollte bereits eine qualitativ gute Datenbereitstellung per Download grundsätzlich erfüllen - wobei die detaillierte Datensatzanalyse in dieser Studie auf erhebliche Umsetzungsdefizite hinweist¹. Sofern diese Bedingungen bereits von Datenbereitstellern erfüllt sind, reduzieren sich auch die notwendigen, zusätzlichen Investitionen für eine Bereitstellung per API. Möglicherweise tragfähig erscheint das Argument eines

¹ Beispielsweise werden einige der untersuchten Daten lediglich in unstrukturierten und nicht maschinenlesbaren Formaten bereitgestellt. Andere Datensätze liegen zwar im CSV-Format vor, weisen aber aufgrund diverser Formatierungen erhebliche Qualitätsmängel auf, die die Weiterverwendbarkeit der Daten erheblich einschränken (siehe Tabellen „Technische Rahmenbedingungen“ im Anhang).

zu kostenintensiven, weil für Nutzer kostenfreien API-Zugangs nach HVD-Kriterien nur, wenn etwa der laufende Betrieb der erforderlichen Infrastruktur selbst sehr kostenintensiv ist. Das ist vor allem in Anwendungsszenarien der Fall, bei denen *sehr* große Datenmengen übertragen werden müssen, wie bspw. bei einigen Geo- und Satellitendaten. Wenn solche Datensätze bereits offen verfügbar sind und keine besonderen weiteren Mehrwertpotenziale erkennbar sind, ist daher in diesen Fällen eine HVD-Klassifizierung besonders kritisch zu bewerten. Dabei wird allerdings auch die Möglichkeit zu berücksichtigen sein, dass gerade eine Bereitstellung per API einen gezielteren Datenzugriff ermöglicht, in dessen Folge weniger Daten übertragen werden müssen. Damit würde die Datennutzung auf Nutzerseite erheblich erleichtert und folglich Mehrwertpotenziale eröffnet.

Die konkrete Definition, welche Datensätze HVD sind und insbesondere welche technischen Modalitäten bei der Bereitstellung zu gewährleisten sind, sollten daher im Durchführungsrechtsakt klar bestimmt werden. Dabei sind die Grenzen der PSI-RL zu beachten, sodass im Durchführungsrechtsakt etwa keine Pflicht zur Archivierung alter Datensätze oder inhaltliche Aufbewahrungspflichten neuer, d.h. nicht bereits bestehender Datensätze, etabliert werden dürfen. Soweit dies für die effektive Nutzung bestimmter Datensätze erforderlich ist, muss dies über sektorale Regulierung auf EU- oder nationaler Ebene adressiert werden.

Im Verhältnis zu sektoralen Regelungen auf EU-Ebene, insbesondere der INSPIRE- und IVS-Richtlinien, ist letztlich festzuhalten, dass die PSI-RL einen Mindeststandard für Weiterverwendungsbedingungen darstellt. Existierende sektorale Regelungen müssen bei der Konkretisierung der PSI-RL in Form von Durchführungsrechtsakten berücksichtigt werden. Im Durchführungsrechtsakt sind daher die technischen und organisatorischen Vorgaben der bestehenden sektoralen Regelungen möglichst so abzubilden und zu berücksichtigen, dass Doppelstrukturen und -arbeiten für die Datenbereitsteller bei nur geringem Zusatznutzen vermieden werden. Dies gilt auch für IVS-Datensätze aus dem Bereich Mobilität, da auch bei einer Qualifikation als HVD das abgewogene sektorale Regime in entscheidender Hinsicht nicht modifiziert wird: Die Pflicht, (bestimmter) privater Unternehmen, ihrerseits Daten bereitzustellen bleibt auch nach einer Qualifikation als HVD bestehen..

Empfohlene Datensätze

Die abschließende Tabelle listet alle in der Studie exemplarisch untersuchten Datensätze auf. Dabei wurde für alle Datensätze mindestens ein niedriges positives Mehrwertpotenzial identifiziert. Sofern diese Datensätze auch im Anwendungsbereich der PSI-RL liegen, kann deren HVD-Klassifizierung daher grundsätzlich empfohlen werden. Je nach Datensatz könnten sich aus einer solchen Klassifizierung allerdings unterschiedliche Handlungsbedarfe ergeben. In der zweiten Spalte sind Datensätze erfasst, deren HVD-Klassifizierung keine Handlungsbedarfe nach sich ziehen würde. Diese Datensätze sind zweifelsfrei von der PSI-RL erfasst, ihre Klassifizierung als HVD ist aus ökonomischer Sicht sinnvoll, da sie über entsprechendes Potenzial verfügen und keine wesentlichen Kosten zu erwarten sind, und die Daten werden bereits maschinenlesbar und per API bereitgestellt. In der dritten Spalte sind Datensätze aufgelistet, die Potenzial haben, deren Klassifizierung als HVD aber zu weiteren Handlungsbedarfen führen würde. Dies bedeutet aber ausdrücklich nicht, dass diese Datensätze per se als HVD ungeeignet wären. Wichtig ist in diesen Fällen vielmehr, die Wechselwirkungen technischer, rechtlicher und ökonomischer Interventionen mit Hilfe wirtschafts- und rechtspolitischer Instrumente gezielt zu optimieren. In der letzten Spalte werden Datensätze aufgeführt, bei denen die Analyse ergeben hat, dass sie nicht im Anwendungsbereich der PSI-RL liegen.

Themenbereich	Klassifizierung als HVD ohne Interventionen möglich	Klassifizierung als HVD erfordert Intervention in technischer (T), rechtlicher (R) oder ökonomischer (Ö) Hinsicht ²	Anwendungsbereich PSI-RL nicht eröffnet
Georaum	Geo_05 Flüsse; Geo_06 Nationale und lokale Daten;	Geo_01 Katasterdaten (Ö); Geo_02_alt Postleitregionen (T); Geo_03 Bodenschätzungsdaten (T,Ö); Geo_04 Bodenrichtwerte (T); Geo_07 Ortho-Bilder (Ö); Geo_08 Digitale Oberflächenmodelle (R,Ö); Geo_09 Digitale Geländemodelle (Ö)	Geo_02 Postleitdaten (R)
Erdbeobachtung und Umwelt	Erd_02 Luftdaten; Erd_03 Lärmkartierung; Erd_09 CORINE Landcover Deutschland; Erd_14 Erdbeben; Erd_18 Bodenzustand; Erd_19 Satellitenbilder	Erd_01 Luftemissionen (T); Erd_04 Gewässerqualität (T); Erd_05 Grundwasserbeschaffenheit (T); Erd_06 Öffentliche Wasserversorgung (T); Erd_07 Energieverbrauch (T); Erd_08 Abfallbilanz (T); Erd_10 Bodenflächen (T); Erd_11 Waldzustand (T); Erd_12 Holzeinschlag (T); Erd_13 Hochwasser (T,R); Erd_15 Naturschutzgebiete (T); Erd_16 Artenvielfalt (T,R); Erd_17 Fangquoten, Importe & Exporte (T)	
Meteorologie	Met_01 Numerische Wettervorhersage für Deutschland und Europa; Met_02 Wetterwarnungen und Vorabinformationen auf Gemeindeebene; Met_04 Raster der vierteljährigen Mittel der Lufttemperatur für Deutschland; Met_05 Regionalisierte Klimaprojektionen; Met_06 2m Temperatur an RBSN Stationen; Met_07 Radarkomposit RV (RADVOR-OP)	-	-
Statistik	Sta_01 Bevölkerungsstatistik; Sta_04 Bruttoinlandsprodukt; Sta_06 Gesundheitsstatistik; Sta_07 Schulstatistik; Sta_08 Einkommensstatistik	Sta_02 Haushaltsdaten (T); Sta_03 Arbeitslosenquote (T,R)	Sta_05 Ifo-Geschäftsklima (R)
Unternehmen und Eigentümerschaft	-	Unt_01 Handelsregister kostenfreier Zugang (T,R); Unt_02 Handelsregister kostenpflichtiger Zugang (T,R,Ö); Unt_04 Namen der Aktionäre (T,R)	Unt_03 Transparenzregister (R)
Mobilität	Mob_09 Anlagen und Einrichtungen für Schiffsverkehr	Mob_01 Verkehrszeichen (R,Ö); Mob_02 Straßennetz (R,Ö); Mob_03 Ladesäulenkarte (T,R); Mob_04 Radverkehrsanlagen (R,Ö); Mob_05 ÖPNV-Daten Bus und Bahn (R,Ö); Mob_06 Fahrplandaten (R); Mob_07 Stationsdaten (T,R); Mob_08 Beförderungsmenge und -Leistung nach Verkehrsträgern (T); Mob_10 Seegangsmessungen in Nord- und Ostsee (T); Mob_11 Breitbandversorgung (T)	-

² T = Daten nicht in maschinenlesbarem Format und / oder per API bereitgestellt

R = Anwendungsbereich bzw. Ausnahme von PSI-RL rechtlich unsicher bzw. nicht unerhebliche rechtl. Folgeanpassungen notwendig

Ö = Klassifizierung als HVD verursacht wesentliche Kosten.