PET-unterstützte Freigabeverfahren für Offene Daten

Ulrich Greveler

Hochschule Rhein-Waal
Labor für IT-Sicherheit
mail@ulrich-greveler.de

Zusammenfassung

Der Vorgang, Daten der öffentlichen Verwaltung in proaktiver Weise oder auf Wunsch von Bürgerinnen und Bürgern, Organisationen oder Unternehmen öffentlich zur Verfügung zu stellen (Open Data), setzt einen geordneten Freigabeprozess voraus, der die Interessen aus Sicht des Datenschutzes, der Nutznießenden und der Öffentlichkeit ausreichend berücksichtigt. In diesem Beitrag wird eine datenschutzfördernde Technologie (PET) vorgeschlagen, um den Schutz der Privatsphäre bei der Publikation von Verwaltungsdaten zu stärken. Damit soll in erster Linie eine unbeabsichtigte bzw. fahrlässige Veröffentlichung von unzureichend anonymisierten Datensätzen im Hinblick auf eine semi-automatisierte Bereitstellung behördlich erhobener Datensätze verhindert werden. Die in diesem Beitrag vorgeschlagene Lösung OD-PET liest den Datensatz (der i. a. als CSV-Datei vorliegt) ein, und trifft eine algorithmische Entscheidung, ob eine Auffälligkeit vorliegt, die eine manuelle Prüfung des Datensatzes zur Auflage macht. Dabei werden eine Verifizierung des Tabellenschemas, ein Durchsuchen der Datensätze im Volltext und eine Validierung des Ausgabeformates jeweils automatisiert vorgenommen. OD-PET kann daher als technischer Mechanismus im Rahmen eines Upload-Prozesses eingesetzt werden, der eine letzte technische Prüfung vornimmt, bevor die Veröffentlichung tatsächlich ausgeführt wird.

1 Einführung

Mit der im Jahr 2017 erfolgten Verabschiedung eines Open-Data-Gesetzes des Bundes stellt Deutschland nun in umfangreicher Weise maschinenlesbare Verwaltungsdaten von Bundesministerien, Bundesämtern und öffentlichen Einrichtungen zur freien (wenn auch meist lizenzgebundenen) Verfügung.

Die Pflicht zur Freigabe oder auch nur die Möglichkeit, Daten auf Wunsch von Bürgerinnen und Bürgern, Organisationen oder Unternehmen öffentlich zur Verfügung zu stellen setzt einen geordneten Freigabeprozess voraus, der die Interessen aus Sicht des Datenschutzes (Feststellung, dass kein Personenbezug der Daten vorliegt oder dass eine Anonymisierung erfolgt ist), der Nutznießenden (Maschinenlesbarkeit, nützliche Metadaten, nicht einschränkende Lizenzbedingungen) und der Öffentlichkeit (Transparenz staatlichen Handelns) ausreichend berücksichtigt.

Offene Verwaltungsdaten werden als Bestandteil der Digitalisierungsstrategie der Bundesregierung und der Landesregierung wahrgenommen. Sie fördern die Entwicklung neuartiger Geschäftsmodelle, schaffen günstige Bedingungen für die wachsende Digitalwirtschaft (vgl. [DBP+16]) und stellen eine Grundlage von digitalen Ansätzen des zivilgesellschaftlichen und bürgerschaftlichen Engagements dar [Cin12].

P. Schartner · A. Baumann (Hrsg.) · D•A•CH Security 2017 · syssec (2017) 92-100.

Der Schutz der Privatsphäre ist ein vorrangiges Ziel bei der Speicherung, Verarbeitung und insbesondere bei der Veröffentlichung von Verwaltungsdaten und genießt daher eine hohe politische Aufmerksamkeit [Wew12].

In diesem Beitrag wird ein konstruktiver Ansatz über eine *Privacy-Enhancing Technology* (kurz: PET), d.h. über eine datenschutzfördernde Technologie vorgeschlagen, um den Schutz der Privatsphäre bei der Publikation von Verwaltungsdaten zu stärken.

Die hierbei betrachteten Daten liegen in einer Tabellenstruktur vor und wurden aus Fachanwendungen oder Softwarelösungen der Verwaltung exportiert. Nicht betrachtet werden Daten, die als Bilddatei, als Tondokument oder audiovisuelle Darstellung (z.B. Videosequenz) vorliegen.

2 Personenbezug von Verwaltungsdaten

Open by default bezeichnet die Bereitstellung aller Daten, die die öffentliche Verwaltung erhebt, solange kein Ausnahmetatbestand vorliegt, in standardisierter und maschinenlesbarer Form ohne Einschränkung oder Monetarisierung der Nutzung oder Weitergabe.

Ausnahmetatbestände, die eine Bereitstellung ausschließen oder hinsichtlich der Nutzung einschränken sind gesetzlich geregelt oder entspringen politischen Entscheidungen der Leitungsebene. Außerhalb von Spezialgesetzen und Einzelregelungen lassen sich folgende Kategorien bilden, die sich in dieser oder ähnlicher Form bereits in Landesgesetzen zur Informationsfreiheit wiederfinden.

- Personenbezug von Daten (sofern keine Einwilligung besteht oder übergeordnete Rechtsgüter eine Einwilligung ersetzen)
- Betriebs- und Geschäftsgeheimnisse öffentlicher Unternehmen
- Urheberrechte
- Geheimschutz, Vertraulichkeitseinstufungen
- Schutz eines behördlichen Entscheidungsbildungsprozesses
- Vertragliche Regelungen zwischen der öffentlichen Hand und privaten Institutionen

Dabei stellt der Datenschutz als Freigabekriterium die wesentliche Hürde dar, denn eine Verletzung schutzwürdiger Interessen der betroffenen Personen wäre nach Veröffentlichung nicht mehr heilbar. Es muss jedoch auch konstatiert werden, dass weitere Motive von Beschäftigten der öffentlichen Verwaltung, die noch bis vor wenigen Jahren von einem Grundsatz des Amtsgeheimnisses ausging und Einblicke in Aufzeichnungen und Akten für interessierte Dritte nicht kannte, vorliegen können, die als „Datenschutzbedenken“ oder in ähnlicher Weise bezeichnet werden, obwohl kein Tatbestand vorliegt, der unter die Vorgaben des Bundesdatenschutzgesetzes fiele. Transparente Verwaltungsprozesse können von Beschäftigten als ungewohnt und unangenehm empfunden werden, was eine Suche nach Gründen, keinen weitgehenden Einblick zu gewähren, motiviert. Die vorgebliche datenschutzrechtliche Prüfung kann dann als willkommene Begründung herhalten, weil man sich der Mächtigkeit dieses Instrumentes wohl bewusst ist.

Eine Umsetzung der Freigabeanforderungen aus datenschutzrechtlicher Sicht stellt die Definition eines Freigabeprozesses dar, der eine Überprüfung der rechtlichen Vorgaben enthält. Diese Überprüfung kann manuell oder halb-automatisch über geeignete technische Werkzeuge erfol-

gen: Beispielsweise kann sich die überprüfende Person über eine Preview-Funktion des Datenportals technisch vergewissern, dass bestimmte Spalten aus den Datensätzen tatsächlich entfernt wurden, bevor die Veröffentlichung erfolgt.

Dem Autor bekanntgewordene Freigabeprozesse von offenen Daten der Kommunen sehen, sofern ein Personenbezug vorliegt, die einfache Entfernung der personenbezogenen Attribute vor, d.h. in der Tabellenstruktur werden entsprechende Spalten entfernt; zeilenbezogen wird bei statistischen Daten überprüft, ob eine Personenbeziehbarkeit besteht (bspw. bei der Angabe der Anzahl von Betroffenen, die dann einen Schwellwert überschreiten muss, damit die Zeile nicht entfernt wird). Komplexere Verfahren, die bei medizinischen Studien mit Patientendaten Anwendung finden und in diesem Kontext etabliert sind, wie Pseudonymisierungsverfahren, Verrauschen der Daten, Hinzufügen erfundener Daten etc. werden bei Kommunalverwaltungen – soweit der Autor Gespräche mit Verantwortlichen führen konnte – noch nicht eingesetzt. In diesem Beitrag werden allein mögliche Fehler der einfachen Anonymisierungsverfahren kommunaler und ministerialer Verwaltungen betrachtet.

3 Verwandte Ansätze

Die Stadt San Francisco veröffentlicht ihr „Open Data Release Toolkit“ [Fin16], das u. a. einen Entscheidungsprozess zur Freigabe offener Daten vorsieht. In der Abbildung 1 ist der Ablauf grafisch dargestellt. Zunächst wird geprüft, ob Tatbestände vorliegen, die unter eine Vorgabe zur Vertraulichkeit des Datensatzes fallen. Trifft dies zu, wird der Datensatz unter Verschluss gehalten. In den anderen Fällen wird in Abhängigkeit von der angenommenen Nützlichkeit der Daten anhand eines Matrix-Schemas entschieden, ob die *vermutete Erwartung* der betroffenen Personen hinsichtlich der potentiell schädlichen Wirkung einer Veröffentlichung der Daten in einem angemessenen Verhältnis zum geschätzten Aufwand steht, eine De-Anonymisierung der Daten vorzunehmen.

Dieser pragmatische und steuerbare Ansatz steht dem deutschen Verständnis von Datenschutz, das eine Verarbeitung oder Verbreitung personenbeziehbare Daten ohne Rechtsgrundlage grundsätzlich verbietet und explizite Erlaubnisvorbehalte vorsieht, entgegen. In Deutschland dürften Datensätze, die Angaben zu Personen enthalten, keineswegs auf der Ermessensgrundlage veröffentlich werden, dass eine Einschätzung der Behörde vorliegt, dass die betroffenen Personen allgemein nicht erwarten, dass diese Daten unzugänglich bleiben sollten. Vielmehr müssten *alle* Betroffenen explizit zustimmen, dass der Datensatz geöffnet wird, wenn kein anderer Erlaubnisvorbehalt vorliegt.

Manske und Knobloch [MaKn17] schlagen vor, ein Drei-Phasen-Modell zur Risikoabschätzung für Open Data anzuwenden. Für jede Phase werden Einzelmaßnahmen vorgesehen und voneinander abgegrenzt:

1. Vor der Veröffentlichung der Daten: Es ist eine Entscheidung zu treffen, ob die Datensätze „geöffnet werden dürfen beziehungsweise sollten“[1].
2. Bei der Veröffentlichung: Es sind Datenschutzmaßnahmen im Zuge der Veröffentlichung von Daten zu treffen.

[1] Ob sich „sollten“ hierbei auf eine Ermessensentscheidung der Behörde bezieht, bleibt unklar. Dies dürfte aber nach Verabschiedung des Open-Data-Gesetzes des Bundes und dem Erlass ergänzender Verordnungen ohnehin hinreichend bestimmt werden.

3. Nach der Veröffentlichung: Eine Steuerung der Nutzung bereits geöffneter Daten wird vorgesehen.

Sämtliche Maßnahmen sollen dabei auf „laufenden Geschäftsprozessen“ aufbauen, in denen Datenschutzprinzipien – wie Datenvermeidung und Datensparsamkeit, etwa bei der Beschaffung von IT-Systemen – ohnehin verankert sind [MaKn17].

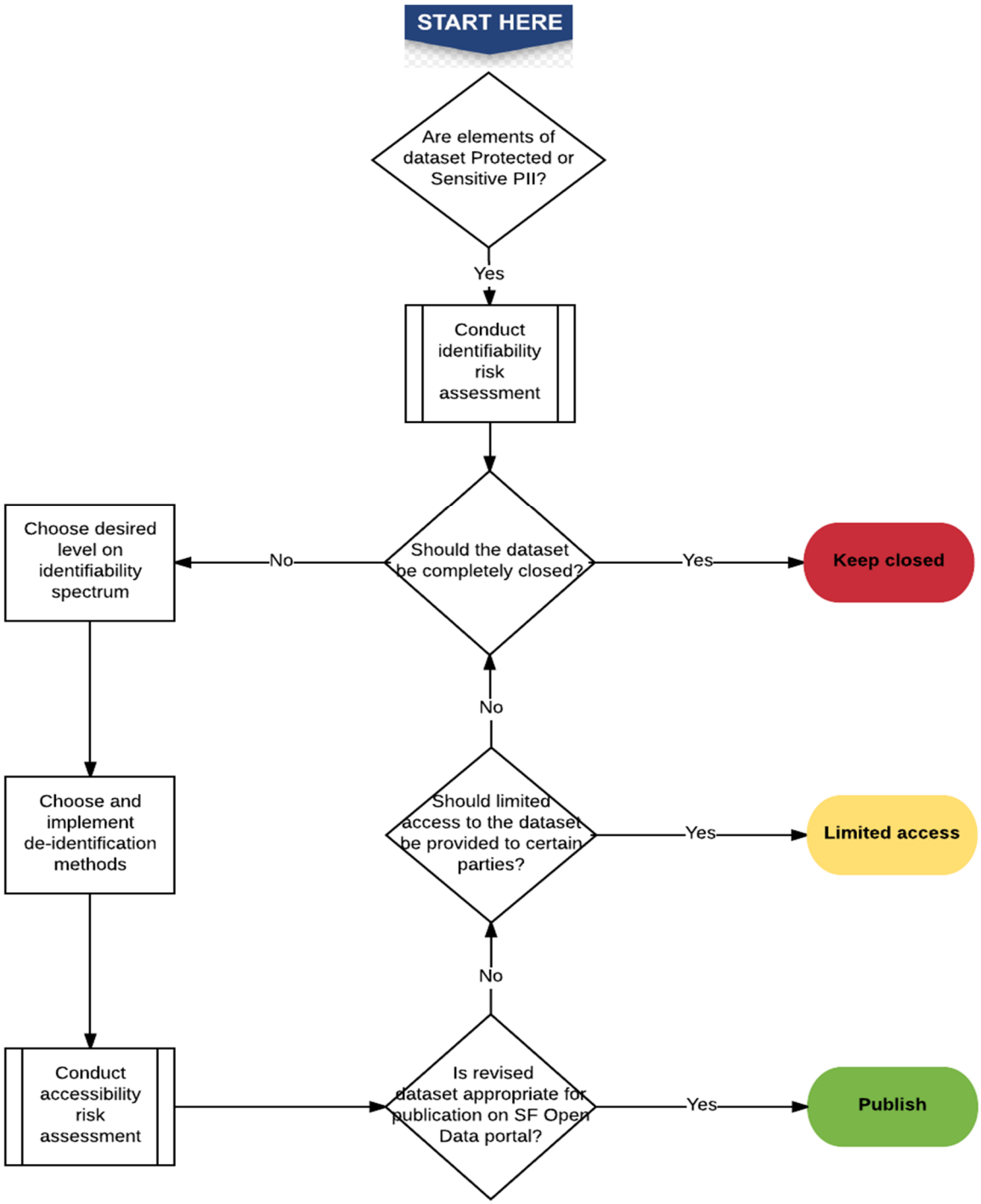

Abb. 1: Prozessbeschreibung des Open-Data-Freigabeprozesses von San Francisco, entn. [Fin16, p. 4]

Zur Unterstützung der ersten Phase wird ein Ampelsystem zur Kategorisierung von Datensätzen nach potenziellem Datenschutzrisiko vorgezeichnet. Dieses dient der Identifizierung von Datensätzen, die zunächst eine Prüfung durchlaufen müssen und ggf. erst nach Anwendung von Schutzmaßnahmen geöffnet werden können. Dies ist vergleichbar mit dem Prozessschritt des Open-Data-Freigabeprozesses von San Francisco, der eine Anonymisierung und eine Risikobewertung vorsieht, nachdem festgestellt wurde, dass keine Vorgabe zur Vertraulichkeit des Datensatzes anwendbar ist.

Die Steuerung der Nutzung bereits geöffneter Daten stellt dabei ein Konzept dar, das in vergleichbaren Ansätzen meist nicht enthalten ist. Es sieht vor, dass in regelmäßigen Abständen die öffentlich bereitgestellten Datensätze auf „Risiken der De-Anonymisierung geprüft" werden, was auch durch externe Experten erfolgen kann. Fälle, in denen Risiken zum Tragen kommen, werden gemäß des Prozessvorschlags gesammelt und ausgewertet, um zukünftige Fehler zu vermeiden und ein (im Vorschlag so benanntes) Frühwarnsystem zu entwickeln. Die Verbreitung und Nutzung deanonymisierter Datensätze soll zudem sanktioniert werden.

Green et al. [GCE+17] verweisen auf das in der IT-Sicherheit etablierte Konzept des Penetrationstests und schlagen vor, dieses in ähnlicher Form auf offene Daten anzuwenden: Ein „re-identification testing" soll dabei vor der endgültigen Freigabe stattfinden. Vorgenommen werden soll dieser aufwändige Test von internen Data Scientists oder vertrauenswürdigen externen Mitgliedern der lokalen Open-Data-Community. Ein vorgeschlagenes Gerüst für einen solchen Testvorgang ist das „motivated intruder"-Modell, das (hier zitiert nach dem *UK Information Commissioner's Office*) eine Person abstrakt beschreibt, "who starts without any prior knowledge but who wishes to identify the individual from whose personal data the anonymised data has been derived" [Ico12]. Die hierbei entstehenden Aufwände sind für Kommunen, die tausende von Datensätzen publizieren, auf den ersten Blick kaum tragbar; in der Praxis beobachten wir jedoch, dass die publizierten Datensätze einander sehr ähnlich sind, so dass viele Re-Identifizierungs-Testergebnisse Datensatz-übergreifend und überregional übertragen werden können.

4 PET-unterstützte Freigabeverfahren

Wir schlagen in diesem Beitrag vor, Freigabeverfahren für Offene Daten um datenschutzfördernde Technologien zu ergänzen. Dabei sollen weder datenschutzrechtliche Prüfungen noch fachliche Bewertungen der bereitstellenden Behörde durch technische Verfahren ersetzt werden. Vielmehr soll die zusätzliche Anwendung PET-unterstützter Verfahren eine unbeabsichtigte bzw. fahrlässige Veröffentlichung von unzureichend anonymisierten Datensätzen im Hinblick auf eine zukünftige massenhafte und semi-automatisierte Bereitstellung behördlich erhobener Datensätze verhindern helfen.

Unser Vorschlag realisiert dabei in technischer Hinsicht teilweise den im Abschnitt zuvor benannten Vorschlag von Green et al. [GCE+17], Re-Identifizierungs-Tests auf zu publizierenden Daten vorzunehmen. Solche Tests setzen allerdings hohe personelle Aufwände voraus. Da eine Identifizierung jedoch in einigen Fällen mit technischen Hilfsmitteln – teilweise automatisiert – vorgenommen wird, können einzelne Tests, wenn sie erst einmal vorbereitet und in das System, das den Freigabeprozess steuert, integriert sind, auch ohne wiederkehrende manuelle Aufwände vorgenommen werden.

4.1 Algorithmische Lösung

Die in diesem Beitrag vorgeschlagene Lösung (im Folgenden als OD-PET bezeichnet) verarbeitet Datensätze vor der Veröffentlichung auf einem Open-Data-Portal. Der Datensatz (der i.a. als CSV-Datei vorliegt) wird eingelesen und es wird eine algorithmische Entscheidung getroffen, ob es eine Auffälligkeit gibt, die eine (erneute) manuelle Prüfung des Datensatzes zur Auflage macht. OD-PET kann daher als Mechanismus im Rahmen eines Upload-Prozesses eingesetzt werden, der eine letzte technische Prüfung vornimmt, bevor die Veröffentlichung tatsächlich ausgeführt wird.

OD-PET ist dabei in folgende Einzelfunktionen zerlegbar, die sukzessive auf den Datensatz angewandt werden.

- Verifizierung des Tabellenschemas nach Bearbeitung. Bei (relationalen) Datenbanken legt ein Schema die Tabellenattribute und die Integritätsbedingungen fest. Dies umfasst die Festlegung von Wertebereichen einzelner Attribute sowie die Fremdschlüsselbeziehungen. Dabei werden die drei folgenden Einzelschritte ausgeführt.
  - **schema_attr:**
    OD-PET prüft: Sind Attribute enthalten, die regelmäßig Personenbezug aufweisen? (Hier wird eine schwarze Liste von Attributbezeichnern angewendet, z.B. „Name“, „Geburtsdatum“, Ausweisnummer, etc.)
  - **schema_foreign:**
    OD-PET prüft: Sind Fremdschlüssel enthalten, die einen Personenbezug nahe legen? (Fremdschüsselbeziehungen zu Tabellen, die ihrerseits Personendaten enthalten, z. B. Vorgangsnummer oder Aktenzeichen, legen den Verdacht nahe, dass die Spalte mit den Schlüsseln versehentlich nicht entfernt wurde.)
  - **schema_clone**:
    OD-PET prüft: Gibt es bereits Datensätze auf der eigenen oder weiteren vertrauenswürdigen OD-Plattformen, die dieselben Attribute-Bezeichner aufweisen? (Der Ansatz bezieht sich auf geprüfte Präzedenzfälle, die bereits publiziert wurden. Der neue veröffentlichungsfreie Datensatz soll einem dieser Präzedenzfälle entsprechen. Ist dies nicht der Fall, liegt ein neuer Präzedenzfall vor, der eine aufwändige manuelle Überprüfung rechtfertigt.) Bsp.: Für den Datensatz aus Abbildung 2 würde überprüft, ob es bereits einen publizierten Datensatz gibt, der die gleiche Attributmenge *{Geburtsort, Geburtsland, Anzahl}* aufweist.
- **scan_attr:**
  Durchsuchen der Daten in der veröffentlichungsreifen, maschinenlesbaren Form nach typischen Attributwerten, die einen Personenbezug nahe legen[2], z.B. gemäß einer Liste von Vornamen und Nachnamen; angereicherte Datumsangaben, die Geburtsdaten darstellen könnten (z.B. können Substrings wie „11.09.1970 in“ über reguläre Ausdrücke aufgedeckt werden), Mailadressen, Telefonnummern, amtliche Kennzeichen, Personalnummern und Kundennummern. Diese Suche wird ohne Berücksichtigung der Attributbezeichner vorgenommen.

[2] Vgl. dokumentierten Fall eines Hackathons im Jahre 2015, bei dem bei der Freigabe von Notruf-Datensätzen versehentlich auch personenbezogene Daten der anrufenden Personen, bspw. die Rufnummer, enthalten waren. https://responsibledata.io/reflection-stories/open-data-hackathon/

- **validate_format:**
  Validierung des Ausgabeformates. OD-PET prüft: Liegt ein erlaubtes Format vor und entspricht das Format dem, das durch die Dateiendung angegeben wird? (Dies verhindert, dass falsche Dateien aus dem Workflow publiziert werden, beispielsweise dass versehentlich eine unbearbeitete Excel-Datei anstatt der daraus generierten und bearbeiteten CSV-Datei hochgeladen wird.)

_id	Geburtsort	Geburtsland	Anzahl
47	Herne	005 Nordrhein-Westfalen	138
48	Breslau	152 Polen	135
49	Kattowitz	152 Polen	131
50	Xanten	005 Nordrhein-Westfalen	130
51	Bartin	009 Bayern	126
52	Neuss	005 Nordrhein-Westfalen	125
53	Kempen	005 Nordrhein-Westfalen	124
54	Aachen	005 Nordrhein-Westfalen	121
55	Leipzig	014 Sachsen	119
56	Damaskus	475 Syrien	117
57	Gleiwitz	152 Polen	116
58	Eregli	163 Türkei	111
59	Danzig	152 Polen	110
60	Bagdad	438 Irak	108
61	Kleve	005 Nordrhein-Westfalen	108
62	Mardin	163 Türkei	108

Abb. 2: Beispiel eines offenen Datensatzes der Stadt Moers: Geburtsorte der Bevölkerung

- **person_min:**
  Mindestanzahl von Personen bei aggregierten Daten. Oft wird eine datenschutzrechtliche Anforderung durchgesetzt, dass eine Mindestanzahl von *n* Personen vorliegen muss, damit das Werte-Tupel publiziert werden darf. (Bsp.: 15 Personen aus einem bestimmten Geburtsort leben in einer Stadt; vgl. Abbildung 2 mit realen Daten aus Moers, bei denen eine Mindestanzahl von zehn Personen vorgegeben wurde, um die Personenbeziehbarkeit auszuschließen.) Da das Attribut, das die Anzahl der Personen angibt, meist einen eindeutigen Bezeichner trägt (z.B. Anzahl_Personen), kann OD-PET verifizieren, ob die Mindestanzahl in allen Tupeln erreicht wird.
- **check_sort:**
  Überprüfen, ob freizugebenden Datensätze gemäß eines veröffentlichten Attributs sortiert sind. Hier soll das Risiko gemindert werden, dass scheinbar unsortierte Datensätze tatsächlich nach einem Attribut sortiert wurden, das nach der Sortierung entfernt wurde (z.B. Nachname). Die noch vorhandene Sortierung ließe jedoch u.U. Rückschlüsse auf Personen zu, wenn die Position des Datensatzes in einer Tabelle auf das entfernte Attribut verweist (z.B. Verweis auf die Anfangsbuchstaben des Nachnamens aufgrund der Sortierung).

Der hier vorgestellte, letztlich heuristische Ansatz, der in der OD-PET-Lösung implementiert ist, führt unweigerlich zu *false positives*, die dann manuell überprüft werden müssen. Da Parameter jedoch anpassbar sind (Blacklists sind editierbar) kann der Anteil dieser falschen Zurückweisungen im Praxiseinsatz deutlich reduziert werden.

4.2 Integration in Portalsoftware

Die in der Praxis von vielen Kommunen und öffentlichen Einrichtungen eingesetzten Open-Source-Portallösungen wie CKAN (Comprehensive Knowledge Archive Network)[3], Dataverse[4] oder DSpace[5] stellen Web-basierte Katalog-Lösungen dar, die es den Beschäftigten der öffentlichen Verwaltungen und vergleichbaren Einrichtungen erlauben, ähnlich wie Redakteure eines Web-Content-Management-Systems Datensätze in strukturierter Form bereitzustellen, zu beschreiben und eine Suche sowie die Vorschaufunktion auf den Datensätzen zu aktivieren.

Im Workflow ist dabei jeweils ein Schritt vorgesehen, in dem der zu publizierende Datensatz als Datei ausgewählt und in den Katalog übertragen wird; zudem wird die Sichtbarkeit des Datensatzes konfiguriert (z.B. *public* oder *private*). Eine Integration könnte daher dergestalt erfolgen, dass die Sichtbarkeit per default auf *private* eingestellt wird und die Übertragung des Datensatzes eine OD-PET-Überprüfung auslöst. Eine spätere Änderung der Sichtbarkeit setzt dann voraus, dass die Überprüfung zuvor positiv erfolgt ist. Diese Anpassung der Open-Source-Portallösung ist zwar mit geringem bis mittleren Aufwand möglich, sie würde jedoch dazu führen, dass ein Update der Portal-Software nicht mehr ohne erneute Anpassung möglich wäre.

(Nichtrepräsentative) Gespräche mit kommunalen Beschäftigen zeigten, dass die in der Praxis zu publizierende Datensätze verwaltungsintern gesammelt und dann von einer sehr kleinen Personengruppe im Portal verwaltet werden. Hier bietet sich daher an, dass OD-PET auf dem Verzeichnis operiert, in dem die Datensätze abgelegt werden, und anschließend die Datensätze in ein anderes Verzeichnis verschiebt, das für den Web-Upload verwendet wird bzw. eine E-Mail an die Redakteure generiert, wenn die Überprüfung negativ verlaufen ist. Auf diese Weise ist keine Sourcecode-Anpassung der Portallösung erforderlich; zudem wird das Überprüfungsergebnis an die Personengruppe übermittelt, die ohnehin über das größte Hintergrundwissen zum Datenportal verfügt.

5 Ergebnisdiskussion und Ausblick

Das vorgeschlagene PET-unterstütze Freigabeverfahren für Offene Daten kann grundsätzlich als sinnvolle Ergänzung eines Workflows eingesetzt werden, der eine Öffnung von Datensätzen mit hoher Frequenz vorsieht. Insbesondere die versehentliche Selektion von falschen Dateien beim Upload oder die unzureichende Bearbeitung von Datensätzen mit dafür nicht optimierten Office-Werkzeugen können Ursachen für eine ungewollte Publikation personenbezogener oder -beziehbarer Daten sein, die mit einem automatisierbaren, PET-gestützten Freigabeverfahren effektiv verhindert werden kann.

[3] Projektwebseite CKAN: https://ckan.org/

[4] Projektwebseite Dataverse: https://dataverse.org/

[5] Projektwebseite DSpace: http://www.dspace.org/

Kritisch ist jedoch zu bemerken, dass eine Gefahr besteht, dass Beschäftigte im Wissen um die automatische Kontrolle der Datensätze beim Upload-Vorgang den eigenen Aufwand der Überprüfung deutlich reduzieren. Dies könnte im Extremfall einen gegenteiligen Effekt bewirken, der die Fehlerquote noch erhöht. Es muss daher Sorge dafür getragen werden, dass in Fällen, in denen OD-PET tatsächlich die Öffnung eines ungeeigneten Datensatzes verhindert, trotzdem eine Sanktionierung erfolgt, die die Sensibilisierung der betroffenen Mitarbeiter erhöht – und nicht abschwächt.

Die Lösung OD-PET befindet sich zum Zeitpunkt dieser Veröffentlichung in einem fortgeschrittenen Spezifikationszustand. Eine prototypische Implementierung ist in Verbindung mit einer Abschlussarbeit im Bachelorstudiengang E-Government der Hochschule Rhein-Waal und in Kooperation mit einer Open-Data-Kommune beabsichtigt.

Literatur

[Cin12] Fiorella De Cindio. Guidelines for Designing Deliberative Digital Habitats: Learning from e-Participation for Open Data Initiatives. The Journal of Community Informatics. Vol 8, No 2 (2012)

[DBP+16] Marcus M. Dapp, Dian Balta, Walter Palmetshofer, Helmut Krcmar. Hrsg.: Pencho Kuzev. Open Data. The Benefits. Das volkswirtschaftliche Potential für Deutschland. KAS e.V., Sankt Augustin/Berlin, 2016. ISBN 978-3-95721-202-3

[Fin16] Erica Finkle et al. San Francisco Open Data Release Toolkit. Finkle. Version 1.2 vom 3. November, 2016. Online: https://datasf.org/resources/open-data-release-toolkit/ (abgerufen: 06.04.2017)

[GCE+17] Ben Green, Gabe Cunningham, Ariel Ekblaw, Paul Kominers, Andrew Linzer, and Susan Crawford. 2017. Open Data Privacy (2017). Berkman Klein Center for Internet & Society Research Publication. Online: http://nrs.harvard.edu/urn-3:HUL.InstRepos:30340010 (abgerufen: 01.06.2017)

[Ico12] UK Information Commissioner's Office. Anonymisation: Managing Data Protection Risk Code of Practice. November 2012. Online: https://ico.org.uk/media/1061/anonymisation-code.pdf (abgerufen: 01.06.2017)

[MaKn17] Julia Manske und Tobias Knobloch. Leitfaden für Datenschutz bei Open Data. Ansätze und Instrumente für die verantwortungsvolle Öffnung von Verwaltungsdaten. Stiftung Neue Verantwortung. Policy Brief vom März 2017. Online: https://www.stiftung-nv.de/en/node/1943 (abgerufen: 07.04.2017)

[Wew12] Göttrik Wewer. Auf dem Weg zum gläsernen Staat? Privatsphäre und Geheimnis im digitalen Zeitalter. Der moderne Staat - Zeitschrift für Public Policy, Recht und Management 5.2 (2012).