

Zwischenbericht

ATLAS: Entwicklung eines Open-Source-Systems
für Datentreuhanderschaft.

Erkenntnisse aus der Anwendungsfallrecherche im Forschungsprojekt ATLAS – Datentreuhänder für anonymisierte Analysen in kommunalen Datenräumen



1		
Einleitung	—————	1
2		
Die ATLAS Technologien	—————	3
3		
Erkenntnisse aus der Phase der Anwendungsfallfindung	———	6
4		
Herausforderungen und Chancen	—————	17

1 Einleitung.

Zusammenfassung.

Dieser Zwischenbericht beleuchtet die Ergebnisse der Anwendungsfallrecherche im ATLAS Projekt, das innovative Technologien wie ScrambleDB und Secure Multi-Party Computation (SMPC) für die datenschutzfreundliche Nutzung kommunaler Daten entwickelt. Ziel ist es, Daten aus Silos zusammenzuführen, ohne den Datenschutz zu verletzen. So sollen faktenbasierte Entscheidungen in Verwaltung und Forschung ermöglicht werden. Die vorgestellten Anwendungsfälle decken Bereiche wie Gesundheit, Bildung und Katastrophenschutz ab, zeigen aber auch rechtliche, technische und organisatorische Herausforderungen auf. Trotz des großen Potenzials behindern datenschutzrechtliche Bedenken und Kapazitätsengpässe teils die Umsetzung der Anwendungsfälle. ATLAS sieht Chancen in der Verbesserung von Datenqualität, der Reduzierung redundanter Datenhaltung und der Entwicklung praxisnaher, sicherer Lösungen. Das Projekt will datenschutzfördernde Technologien alltagstauglich machen und so datengetriebene Entscheidungen im öffentlichen Sektor erleichtern.

Einleitung.

Die EU schätzt, dass durch eine bessere Datennutzung Zugewinne von 1% bis 2,5% des Bruttoinlandsproduktes (BIP) erreicht werden können¹. Auch in der öffentlichen Verwaltung gibt es ein enormes Potenzial für eine effektivere Verwendung von Daten. Eine faktenbasierte Entscheidungsfindung wird hier oft dadurch erschwert, dass Daten in voneinander getrennten Beständen - so genannten Silos - gehalten werden. Ein Beispiel: Vielen Behörden ist es nicht möglich, Daten zum Entwicklungsstatus der ansässigen Kinder bei Schuleintritt mit der weiteren Bildungsbiografie zusammenzubringen. Eine effektivere und effizientere Förderpolitik wird dadurch erschwert.

¹ <https://digital-strategy.ec.europa.eu/en/factpages/data>

Wie können Daten aus verschiedenen Silos zusammengebracht werden, ohne den Datenschutz zu verletzen?

Ein Ansatz für siloübergreifende Datennutzung ist das Konzept des Datentreuhänders. Damit wird eine unabhängige dritte Stelle bezeichnet, die Daten zwischen verschiedenen Parteien vermittelt und dabei kein eigenes Interesse an den Daten vertritt. Der Datentreuhänder muss dabei technische Vorkehrungen treffen, um den Schutz der Daten vor unbefugtem Zugriff zu gewährleisten - besonders dann, wenn es um personenbezogene und besonders sensible Daten geht.

Das ATLAS Forschungsprojekt - Forschung mit Anwendungsbezug.

Ziel des ATLAS Projekts ist die Entwicklung eines Open Source Systems, das verschiedene Werkzeuge zur sicheren und datenschutzfreundlichen Analyse von Daten vereint. Dabei kommen innovative kryptografische Komponenten zur Pseudonymisierung und Anonymisierung zum Einsatz.

Ein Grundgedanke des Projekts ist es, dass die Technologien nicht nur theoretisch weiterentwickelt werden, sondern auch an einem praktischen Anwendungsfall im kommunalen Bereich getestet werden. Somit wird nicht nur die Praxistauglichkeit geprüft, die entwickelten Technologien kommen gleichzeitig auch den teilnehmenden Behörden zugute.

Dazu sind an dem Konsortium sowohl wissenschaftliche Partner (Hasso-Plattner-Institut und TU Dresden), Partner mit Fokus auf die softwareseitige Implementierung (SINE Foundation und KIProtect) und Partner mit Fokus auf die Anwendungsfälle (Metropolregion Rhein-Neckar und Polyteia) beteiligt.

Das Projekt geht dieses Jahr (2025) in die Implementierungsphase. Im folgenden Beitrag blicken wir auf zwei Jahre Anwendungsfallsuche zurück und analysieren Chancen, aber auch Hindernisse, für die siloübergreifende Datenauswertung in Deutschland.

2

Die ATLAS Technologien.

Welche Anwendungsfälle für das ATLAS Projekt in Frage kommen (und welche nicht), ist auch von den im Projekt genutzten Technologien abhängig. Daher folgt ein kurzer Überblick über die beiden Haupttechnologien, ScrambleDB und Secure Multi-Party Computation.

ScrambleDB.

ScrambleDB wurde von Prof. Dr. Anja Lehmann, die mit dem Hasso-Plattner Institut ebenfalls am ATLAS Projekt beteiligt ist, entwickelt². Eine nicht-technische Intuition der kryptografischen Lösung kann vereinfacht wie folgt beschrieben werden:

ScrambleDB operiert auf Grundlage von Tabellen, man kann sich z.B. eine Excel-Tabelle vorstellen. Möchte man solche Tabellen nun pseudonymisiert speichern und verarbeiten, wäre ein simpler Ansatz das Ersetzen der personenbezogenen Merkmale (z.B. Vorname, Nachname) durch zufällig generierte Zahlenfolgen (“Pseudonymisierung”). Liegen jedoch mehrere Datensätze vor, in denen Daten zu ein und derselben Person vorkommen, müsste man darauf achten, immer dasselbe Pseudonym für dieselbe Person zu verwenden.

Dieser Ansatz hat zwei Herausforderungen:

1. Eine Stelle muss die Schlüsseltabelle verwalten, um die Pseudonyme auf die betreffenden Person aufzulösen. Diese zentrale Stelle wäre ein großes Sicherheitsrisiko und gefährdet damit den gesamten Ansatz.
2. Daten können trotzdem zusammen geführt werden. Wenn es sich um besonders seltene Kombinationen handelt (wie z.B. bei personenbezogenen Daten in kleineren Gemeinden, bei seltenen Erkrankungen, Geburtsdaten usw.), können die vermeintlich pseudonymisierten Daten ungewollt doch wieder einer natürlichen Person zugeordnet werden.

² <https://petsymposium.org/2019/files/papers/issue3/popets-2019-0048.pdf>

ScrambleDB geht das Problem darum anders an: Jedes Attribut wird einzeln abgespeichert und mit einem Pseudonym bzw. "Schlüssel" versehen. Das bedeutet, das Alter einer Person wird z.B. unter dem Pseudonym "gfewEf" abgespeichert, die Nationalität der selben Person jedoch unter dem Pseudonym "O1Hdhji". Wenn eine böswillige Partei nun Zugriff auf diese Tabellen erlangen würde, so hätte sie nur sehr begrenzte Möglichkeiten, Rückschlüsse aus den Daten zu ziehen. Um die Daten für einen konkreten Anlass nutzbar zu machen, braucht es den so genannten "Converter", der unter Verwendung von kryptografischen Verfahren die Zusammenführung von Daten ermöglicht. Eine technische Besonderheit ist, dass der Converter dies "blind" macht. Das bedeutet, selbst der Converter erhält die Daten nicht in Rohform sondern die verarbeitenden Parteien erhalten ein neues temporäres Pseudonym.

Secure Multi-Party Computation.

Secure Multi-Party Computation (SMPC) befähigt mehrere Parteien, gemeinsam Daten zu verarbeiten. Eine Besonderheit dieses Ansatzes ist, dass die Daten bei der Verarbeitung verschlüsselt bleiben – damit geht SMPC über das Schutzniveau von z.B. ScrambleDB hinaus.

Was zunächst unmöglich erscheint (Wie soll ich gemeinsam mit einer anderen Person etwas berechnen, wenn ich ihren Input nicht kenne?), wird seit mehr als 40 Jahren erforscht und ist durch Verbesserungen an den Verfahren (Algorithmen) sowie der Computer-Hardware praktikabel geworden.

Man kann die Technologie vereinfacht an einem Beispiel erklären:

Drei Personen wollen ihr gemeinsames Gehalt berechnen, aber keine möchte den anderen ihr individuelles Gehalt verraten. Die erste Person addiert daher eine zufällige Zahl zu ihrem tatsächlichen Gehalt und nennt der zweiten Person die Summe. Diese addiert ihr Gehalt zur Summe und nennt das Ergebnis der dritten Person, die wiederum ihr Gehalt addiert. Ziehen die drei dann die zufällig gewählte Zahl vom Ergebnis ab, so kennen sie ihr korrektes gemeinsames Gehalt. Da Person 2 und 3 die zufällige Zahl nicht kennen, kennt keine das tatsächliche Gehalt der anderen Person.

Dieses sehr simple Beispiel dient vor allem der Veranschaulichung. In der Praxis sind deutlich komplexere Analysen mit Hilfe der Technologie möglich. Ein Beispiel einer Multi-Party Computation wurde in Boston³ durchgeführt. Dabei berechneten mehrere Unternehmen gemeinsam Erkenntnisse zum Gender Pay Gap, ohne die sensiblen Gehaltsdaten ihrer Mitarbeiter:innen mit den anderen Unternehmen zu teilen. In Europa wurde SMPC in den Niederlanden genutzt, um Daten aus dem Bereich des Personenverkehrs zu analysieren⁴.

Im Rahmen des ATLAS Projekts ist die SINE Foundation federführend für die Entwicklung der SMPC Komponenten zuständig⁵.

³ https://theodi.cdn.ngo/media/documents/FINAL_PETs_in_Practice_pay_gap_analysis.pdf

⁴ <https://coe-dsc.nl/using-mpc-to-generate-relevant-insights-to-support-inclusive-travel/>

⁵ Das Open Source Package ist unter <https://github.com/sine-fdn/polytune> verfügbar

3

Erkenntnisse aus der Phase der Anwendungsfallfindung.

Die Aufgabe, einen geeigneten Anwendungsfall zu finden, stellte sich als komplexer als initial gedacht heraus. So wurden an die 20 Gespräche mit mehr als zehn potenziellen (kommunalen) Partnern geführt, um die im Folgenden vorgestellten Anwendungsfälle zu finden. Zum aktuellen Zeitpunkt (Februar 2025) rechnen wir nur bei einem Anwendungsfall mit einer tatsächlichen Umsetzung - die übrigen Anwendungsfälle scheitern an verschiedenen Gründen rechtlicher, technischer und organisatorischer Art.

Wir hoffen bei der folgenden Vorstellung der Anwendungsfälle sowohl zu zeigen, wie groß und vielfältig der Bedarf an den im ATLAS Projekt erforschten Technologien ist, als auch auf die Herausforderungen hinzuweisen, die bis zu ihrer Implementierung im Bereich des öffentlichen Sektors noch zu überwinden sind.

Anwendungsfälle nach Domänen.

Gesundheit

Erwartungsgemäß ist ein signifikanter Teil der ermittelten Anwendungsfälle der Domäne Gesundheit zuzuordnen. Die hier vorliegenden Daten, z.B. zu Erkrankungen und Entwicklung, sind besonders sensibel und schützenswert. Gleichzeitig gibt es einen großen Bedarf an Analysen und Erkenntnisgewinn, sowohl im Bereich der medizinischen Forschung, wie eine kürzlich veröffentlichte europäische Pilotstudie zur Verwendung der SMPC mit Patient:innendaten⁶ zeigt, aber auch im Bereich des Öffentlichen Gesundheitsdienstes.

In unseren Gesprächen und Workshops zur Anwendungsfallgenerierung sind uns aus dem Bereich Gesundheit unter anderem die folgenden Beispiele begegnet:

– **Wahrnehmung von Vorsorgeuntersuchungen nach sozialen Gruppen.**

Check-ups zur Vorsorge und Früherkennung von Krankheiten sind sowohl aus individueller Sicht, als auch aus Sicht der öffentlichen Gesundheit, sinnvoll und werden von den gesetzlichen Krankenkassen angeboten. Ob diese Untersuchungen auch wahrgenommen werden, und welche sozialen Faktoren die Wahrnehmungsquoten beeinflussen, sind daher aus Sicht der öffentlichen Gesundheit relevante Fragen. Ein Anwendungsfall für ATLAS könnte es sein, die Daten der Krankenkassen zur Wahrnehmung von Vorsorgeuntersuchungen mit Daten zum sozialen Status der Versicherten (z.B. über Arbeitsagenturen) oder Herkunft und Muttersprachen (z.B. über BAMF) zu korrelieren. Die Erkenntnisse könnten z.B. dabei helfen, Kampagnen für die Wahrnehmung der Untersuchungen besser auf die unterversorgten Zielgruppen zu beziehen. Aufgrund der Vielzahl der Akteure und der Sensibilität der Daten ist dies einer der ambitionierteren Anwendungsfälle. Er eignet sich daher weniger als Pilot, bietet aber interessante mittelfristige Möglichkeiten.

⁶ <https://www.nature.com/articles/s41746-024-01293-4#Tab1>

– Korrelation von Frühen Hilfen und Einschulungsuntersuchung - Der Wert von Prävention.

Die Einschulungsuntersuchung (ESU) ist ein sehr relevanter Datensatz im Bereich der Kindergesundheit, da hier umfassende Daten zu Entwicklung, Vorgeschichte und Untersuchungsergebnissen zu nahezu allen Kindern im Einschulungsalter erfasst werden. Daher ergeben sich eine Vielzahl von denkbaren Anwendungsfällen.

Ein relevanter Fall, auf den wir im Rahmen des ATLAS Projekts gestoßen sind, ist die Betrachtung der Präventionsmaßnahmen der Jugend- und Gesundheitsämter im Lichte der Ergebnisse der ESU. Die Jugendämter bieten z.B. über die Frühen Hilfen Unterstützung für Familien in den ersten Lebensjahren. Außerdem werden sie (je nach Bundesland und in Kooperation mit den Gesundheitsämtern) über nicht wahrgenommene Früherkennungsuntersuchungen (die so genannten U-Untersuchungen) informiert und halten Daten über die dazu getroffenen Maßnahmen und deren Erfolg vor.

Daraus ergibt sich die spannende, aber schwierige Frage nach dem Effekt dieser Präventivmaßnahmen. Die ATLAS Technologien können dazu genutzt werden, Fragen zu beantworten wie:

- Wie hoch ist der Anteil von Kindern mit sonderpädagogischem Förderbedarf unter Kindern, die mindestens eine U-Untersuchung versäumt haben (im Vergleich mit denen, die alle wahrgenommen haben)?
- Werden Kinder alleinerziehender Eltern weniger häufig vom Schulbesuch zurückgestellt, wenn die Eltern eine Unterstützung der Frühen Hilfen in Anspruch genommen haben?
- Gibt es eine Entwicklungsauffälligkeit, die im Rahmen der ESU untersucht wird, und die besonders stark mit einer verpassten U-Untersuchung korreliert?

Dieser Anwendungsfall ist realistisch als Pilot umzusetzen, da hier die organisatorischen Hürden eher niedrig sind. Oft werden die betroffenen Datensätze in nur einer Behörde verwaltet, teilweise sogar im selben Fachverfahren. Im Rahmen unseres Projektes scheiterte ein Versuch der Umsetzung jedoch an der Einschätzung der zuständigen Datenschutzbeauftragten, dass hier eine unzulässige Zweckänderung der erfassten Daten vorläge.

– **Masernschutzgesetz - Qualitätsmetriken zu angestoßenen Prozessen.**

Das Masernschutzgesetz verpflichtet bestimmte Personengruppen, z.B. Kinder, die in einer Kita betreut werden sollen, eine Masernimpfung vorzuweisen. Tun sie das nicht, erhalten die Gesundheitsämter eine Meldung und stoßen einen Prozess zur Nachverfolgung und ggf. Einleitung von Maßnahmen an. Da diese Nachverfolgung Ressourcen bindet, ist es wünschenswert, überflüssige Prozesse, z.B. aufgrund von Falschmeldungen, möglichst zu reduzieren. Hier kann wieder der Datensatz der ESU herangezogen werden, der ebenfalls Informationen zum Impfstatus der Kinder enthält. Es können unter der Verwendung der ATLAS Werkzeuge Datenqualitätsmetriken generiert werden, die Aufschluss darüber geben unter welchen Umständen es ggf. vermehrt zu Fehlmeldungen kommt. Diese Erkenntnisse können dann idealerweise genutzt werden, um die Prozesse zu verbessern.

Der Masernschutz-Anwendungsfall wurde mit dem Gesundheitsamt in Frankfurt am Main entwickelt und wird voraussichtlich in 2025 als Pilot umgesetzt.

Anwendungsfälle nach Domänen.

Bildung

Eine zweite fachliche Domäne, die spannende Anwendungsfälle für die ATLAS Technologien bereithält, ist der Bildungsbereich. Hier gibt es vielfältige Herausforderungen, die datengetriebenes Entscheiden erschweren: Generell liegt in den meisten Bundesländern ein eher niedriger Digitalisierungsgrad vor. Schulen übermitteln Daten für Statistiken oft nur hochgranular an übergeordnete Behörden (z.B. Gesamtzahl der Schulabschlüsse im Jahr). Darüber hinaus sind Daten von Schüler:innen, die überwiegend minderjährig sind, besonders schützenswert. Außerdem ist der Bereich Bildung stark föderal geprägt, die Prozesse unterscheiden sich von Land zu Land. Manche Bundesländer sind im Bereich Bildungsstatistik schon sehr aktiv - Hamburg z.B. führt bereits Bildungsdaten zu Forschungs- und Statistikzwecken zusammen - während andere noch an den Grundlagen arbeiten.

– **Bildungsverläufe - Der Wert von Prävention (Fortsetzung).**

Auch aus dem Blickwinkel des Bildungssektors lassen sich spannende Erkenntnisse auf Grundlage der Daten zur Einschulungsuntersuchung stellen. Verknüpft man diese mit Hilfe der ATLAS Technologien mit Daten zu den Bildungsverläufen der Schüler:innen, so lassen sich Fragen wie die folgenden betrachten:

- Wie viele Kinder mit einer bestimmten Vorerkrankung, z.B. ADHS, haben einen bestimmten Schulabschluss gemacht?
- Wie unterscheiden sich diese Werte nach Gemeinde?
- Wie wirkt es sich aus, ob den Kindern eine bestimmte Förderung empfohlen wurde?
- In welchem Bezirk ist der Anteil der Kinder, die nicht Deutsch als Muttersprache haben und Abitur machen, besonders hoch?
- Wie korreliert die Kitabesuchszeit mit dem Bildungsverlauf?

Die Umsetzung dieses Anwendungsfalls im Rahmen des ATLAS Projekts wurde in einem Bundesland diskutiert und verworfen, da noch kein digitalisierter Datensatz zu den Bildungsverläufen vorliegt. Bundesländer, die ein digitales Register zu den Bildungsverläufen aufgebaut haben, beziehen dort die Daten zur ESU sinnvollerweise schon ein, auch wenn diese normalerweise im Hoheitsbereich der Gesundheitsämter verwaltet wird. Das bedeutet, dass sie bereits eigene Lösungen zur Pseudonymisierung und Auswertung der Daten entwickelt haben.

– **Bildungsverlaufsregister.**

Aufgrund der Relevanz des Themas beschäftigen sich aktuell verschiedene Stellen, insbesondere die Kultusministerkonferenz, mit dem Aufbau eines Bildungsverlaufsregisters. Die Frage, zu welchem Grad dies ein zentrales Bundesregister sein wird, und zu welchem Grad die Register dezentral bei den Ländern liegen werden, wird aktuell noch ausgehandelt. Je nach Ausgestaltung könnten hier verschiedene Komponenten des ATLAS Projekts einen Beitrag leisten: Bei einer zentralisierteren Architektur könnte die Verwendung des ScrambleDB Protokolls für die Pseudonymisierung den Schutz der Daten erhöhen⁷, bei einer dezentraleren Architektur könnte eine MPC Komponente relevant werden.

Als Anwendungsfall für das ATLAS Projekt kommt das Thema aufgrund seiner Politisierung und der Zeitschienen nicht in Frage, allerdings könnten die Ergebnisse des Projekts hier mittelfristig relevant werden.

⁷ https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2023/03/konzeption-bildungsverlaufsregister-032023.pdf?__blob=publicationFile

– Katastrophenschutz.

Ein weiterer Anwendungsfall, auf den wir bei der Recherche gestoßen sind, stammt aus dem Bereich des Katastrophenschutzes. Schwerbehinderte, insbesondere Menschen mit Mobilitätseinschränkungen, sind im Katastrophenfall besonders gefährdet⁸. Für die Katastrophenschutzbehörden ist kritisch zu wissen, in welchen Straßen oder Bezirken besonders viele Menschen leben, die im Evakuierungsfall auf Unterstützung angewiesen sind. Als Indikator können dazu z.B. die vergebenen Schwerbehindertenausweise herangezogen werden, oder Daten der Krankenkassen. Die Adressen sind hier zwar meist auch als Attribut abgespeichert, allerdings ist in diesem Anwendungsfall die Aktualität der Daten sehr relevant. Ein Abgleich mit den Adressen aus dem Einwohnermelderegister erscheint daher sinnvoll. Der Katastrophenschutz könnte so für seine Ressourcenplanung mit Hilfe der ATLAS Technologien datensparsam eine aggregierte Auswertung darüber bekommen, in welchen Straßen oder Quartieren besonders viele Menschen auf Unterstützung angewiesen sind.

Dieser Anwendungsfall wurde in der Metropolregion Rhein-Neckar entwickelt und dort mit viel Engagement vorangetrieben. Allerdings konnte keine der Behörden, die Informationen zu den Schwerbehindertenausweisen hätte liefern können, von einer Zusammenarbeit überzeugt werden. Grund hierfür waren insbesondere rechtliche Bedenken.

⁸ <https://andererseits.org/doku/>

Anwendungsfälle nach Art.

Nachdem wir Beispiele für Anwendungsfälle in den verschiedenen Domänen beleuchtet haben, lohnt es sich auch noch auf die übergreifenden Muster in der Art der Anwendungsfälle einzugehen.

– Zusammenführen von Datensätzen.

Die vielleicht intuitivste Art des Anwendungsfalls ist das Zusammenführen voneinander getrennter Datensätze zu einer Person. Hier werden die erwähnten Silos aufgebrochen, da Daten aus dem Zuständigkeitsbereich einer Behörde mit Daten einer anderen Behörde verknüpft und ausgewertet werden. An dieser Stelle liegt der wohlmöglich größte Erkenntnisgewinn, da ganz neue Fragestellungen angegangen werden können. Allerdings treten auch datenschutzrechtliche Fragen auf, da laut DSGVO eine Zweckänderung in der Verwendung der (personenbezogenen) Daten im Normalfall und ohne Gesetzesgrundlage nicht zulässig ist.

– Abgleich redundanter Datensätze.

Eine andere Art des Anwendungsfalls, die uns im Laufe der Recherche mehrfach begegnet ist, betrifft den Abgleich mehrfach geführter Datensätze miteinander. Oft sind die Zuständigkeitsbereiche der Behörden oder Fachmodule nicht scharf voneinander getrennt, was dazu führt, dass die selben Daten an mehreren Stellen verwaltet werden. Das bringt den Bedarf eines Abgleichs mit sich, der als Resultat meist eine Form von Datenqualitätsmetrik hat - also die Erkenntnis, in welchem Datenbestand die Daten möglicherweise vollständiger oder aktueller sind. Ein Beispiel ist der oben erwähnte Masernschutz-Anwendungsfall, aber auch im Bereich Impfnutzen sind wir auf Interesse für einen solchen Abgleich gestoßen. In diesem Fall werden Daten von den Gesundheitsämtern und dem Paul-Ehrlich Institut verwaltet.

– Hierarchische (föderale) Aggregation.

Ein sehr verbreitetes Szenario in unserer föderalen Verwaltungslandschaft sieht wie folgt aus: Daten werden dezentral erfasst und verarbeitet, z.B. in den Kommunen, wo auch die Datenhoheit liegt. Übergeordnete Behörden auf Landes- oder Bundesebene haben jedoch Interesse an aggregierten, statistischen Erkenntnissen aus diesen operativen Datensätzen. Oft fragen die Behörden dann die aggregierten Daten ab und alle Kommunen übermitteln die aggregierten Werte (leider immer noch oft als Excel und per Email).

Die so übermittelten aggregierten Daten sind anonymisiert, d.h. sie enthalten keinen Personenbezug mehr. Allerdings gibt es bei diesem Vorgehen auch einige Probleme. So stößt man schnell an Grenzen des Erkenntnisgewinns auf der übergeordneten Ebene, z.B. erhält man den Mittelwert für den gesamten Landkreis (normalerweise) nicht durch das Mitteln der Mittelwerte aus den Gemeinden⁹. Komplexere Statistiken - z.B. die Wahrscheinlichkeit, dass ein Kind Schulabschluss A macht unter der Bedingung, dass es Vorerkrankung B hat, ließen sich auf der aggregierten Datenbasis nur dann berechnen, wenn neben der Summenstatistik der Schulabschlüsse auch noch weitere Attribute und absolute Gruppengrößen geteilt werden. Zwar ist das oft unproblematisch, bei kleinen Kohorten und seltenen Konditionen können aber schon hier leicht Rückschlüsse auf einzelne Personen gezogen werden. In einem föderalen Kontext können die ATLAS Technologien daher auch helfen, die Privatsphäre der Betroffenen zu schützen und die Qualität der Auswertungsergebnisse zu verbessern.

⁹ Große Gemeinden mit 10.000 Einwohner:innen und kleine Gemeinden mit 100 Einwohner:innen dürfen nicht gleich gewichtet werden, wenn man den Gesamtmittelwert berechnen möchte.

– **Grenzfälle: Übermittlung statt Analyse.**

Neben den oben geschilderten Arten von Anwendungsfällen sind wir auch immer wieder auf Problemstellungen gestoßen, die einen Datentreuhänder eher im Sinne der unparteiischen und sicheren Übermittlung eines (Klar-)Datensatzes an eine andere Entität benötigen. Beispiele sind z.B. die Übergabe eines Bildungsverlauf-Datensatzes bei Umzug eines Kindes in ein anderes Bundesland oder auch die Übermittlung von Klar-Adressdaten von Menschen mit Mobilitätseinschränkung an den Katastrophenschutz im Falle einer Evakuierung. Während diese Fälle nicht weniger relevant sind, beschäftigt sich das ATLAS Projekt jedoch mit anonymisierten Analysen, weshalb diese Bedarfe nicht weiter verfolgt wurden.

4

Herausforderungen und Chancen.

Abschließend lenken wir den Blick auf die Herausforderungen, die uns bei der Suche nach einem geeigneten Anwendungsfall begegnet sind, und heben die Chancen hervor, die sich aus dem ATLAS Projekt und den erforschten Technologien für das datengetriebene Entscheiden im öffentlichen Sektor bieten.

Redundante Datenhaltung (vermeiden).

Durch Jahrzehnte der Datenhaltung in Silos ist es Normalität, dass die Behörden die benötigten Daten im Zweifelsfall neu erfassen, statt sie aus einem anderen Silo heranzuziehen. Das führt auch dazu, dass es unseren Gesprächspartner:innen oft schwer fiel, außerhalb ihrer Domänen zu denken und Anwendungsfälle zu formulieren. Auf der anderen Seite ergeben sich so auch die bereits geschilderten Datenqualitätsprobleme zwischen duplizierten Datensätzen.

Die Verfügbarkeit von sicheren und datenschutzkonformen Mechanismen zum Verknüpfen von Datenbeständen, wie sie im ATLAS Projekt erforscht werden, bietet gleich mehrere Chancen: Könnte die redundante Datenhaltung verringert werden, so würde das nicht nur Vorteile für die Datenqualität bieten, sondern im Sinne der Datensparsamkeit auch für den Datenschutz.

Kapazitäten zur Kooperation.

Eine weitere Herausforderung, die uns vermehrt begegnet ist, war die Ausstattung der entsprechenden Behörden. Damit ist einerseits die Ausstattung mit Technologie gemeint, z.B. ob die entsprechenden Datenbestände überhaupt digitalisiert waren, aber auch die zeitlichen Kapazitäten der Mitarbeitenden. Oft wurde die Sorge vor einer zusätzlichen Belastung der bereits strapazierten zeitlichen Ressourcen als Hindernis angeführt.

Das kann aus Sicht des Projekts als nicht-funktionale Anforderung verstanden werden, die bereitgestellten Werkzeuge und ihre Implementierung möglichst niedrigschwellig zur Verfügung zu stellen. Gelingt dies, sind die Chancen enorm:

Wenn ein datenschutzkonformer Austausch von Daten für die Behörden Alltag wird, verbessert dies nicht nur die Entscheidungsgrundlagen für ihre Arbeit, sondern spart auch Ressourcen, die z.B. zum händischen Zusammentragen von Daten verwendet werden.

Innovative Methoden für den Datenschutz.

ATLAS bewegt sich, was den Datenschutz angeht, in einem Widerspruch: Einerseits werden innovative datenschutzförderliche Technologien entwickelt (engl. privacy enhancing technologies, PETs). Andererseits war oft der Datenschutz selbst das größte (wahrgenommene) Hindernis bei der Umsetzung der Anwendungsfälle.

Pseudonymisierte Daten sind - egal wie sicher die Pseudonymisierung ist - nach gängiger Lesart immer noch von der DSGVO und damit auch von der Zweckbindung betroffen. Nur wenn die Daten gänzlich anonymisiert sind (also z.B. die aggregierten Ergebnisse einer SMPC), greift die DSGVO nicht mehr. Was Anonymisierung dabei genau ausmacht, damit beschäftigt sich in diesem Jahr auch die Datenschutzkonferenz.

Eine Ausnahme besteht, wenn eine Rechtsgrundlage vorliegt. Unsere Beobachtung ist, dass in diesem Fall oft zentrale Register aufgebaut werden, die zwar DSGVO-konform einfache Pseudonymisierung und / oder Verschlüsselung zum Schutz der personenbezogenen Daten anwenden, aber darüber hinaus keine Notwendigkeit mehr für komplexere PETs wie die Multi-Party Computation sehen.

Der folgende Satz ist uns bei der Recherche begegnet: “Der beste Datenschutz ist, die Daten gar nicht anzurühren”. Das ist nicht falsch, allerdings der schlechteste Weg zum Erkenntnisgewinn aus Daten. Mit dem ATLAS Projekt verfolgen wir das Ziel, Technologien bereit zu stellen und zu testen, die Erkenntnisgewinn ermöglichen und die Privatsphäre ähnlich gut schützen als würden wir die Daten “gar nicht anrühren”.