

Was den Föderierten Datenpool zu diesem konkreten Zeitpunkt zu einer passenden Lösung macht

Der „föderierte Datenpool“ **macht Daten über die Standorte und Zugänge zu sozialen Unterstützungsangeboten in Deutschland verfügbar**, ohne ein neues Verzeichnis aufzubauen oder selbst zu pflegen. Diese Daten werden oft gebraucht: Von Wegweisern auf den Websites von Sozialämtern über KI-Chatbots zur Automatisierung von Beratung bis hin zum Beratungsstellenfinder der Sozialplattform – der Verweis auf die Regelangebote vor Ort gehört zu einem schnell handlungsfähigen digitalen Sozialstaat.

Warum jetzt? Der föderierte Datenpool verwendet semantische Technologien, wie sie im Deutschland-Stack verankert sind. Die Testphase mit realen Quellen ist erfolgreich abgeschlossen — mit dem föderierten Datenpool können diese Technologien schnell in eine reale Umsetzung gebracht werden. Eine Intensiv-Entwicklungsphase im Civic Data Lab (einem zivilgesellschaftlich getragenen Projekt, gefördert vom BMBFSFJ 07/23 - 12/26) führt zu einem aktuellen Momentum erhöhter Aufmerksamkeit bei Wohlfahrt und Sozialverbänden sowie einer Koalition der Willigen. Aufgrund der Projektfinanzierung, deren Verlängerung noch nicht sicher ist, besteht die Gefahr, dass Wissen und Netzwerke, die zu einer gelingenden Erst-Implementierung essentiell sind, zu einem späteren Zeitpunkt nicht mehr verfügbar sind. Neben der technischen Umsetzung auf den Anwendungsfall ist besonders die Bereitschaft der (verteilten) Akteure zur Beteiligung mit ihren Datenquellen ein entscheidender Erfolgsfaktor. Genau diese Aufmerksamkeit und Bereitschaft besteht jetzt.

Daher ist es wichtig, zeitnah zu entscheiden, ob der föderierte Datenpool für soziale Unterstützungsangebote in Richtung einer dauerhaften Infrastruktur entwickelt wird, die die Prozesse zur Datenpflege für die Sozialplattform (und viele andere Akteure) verschlankt, Datenqualität erhöht und die Grundlage für weitere Innovationen bilden kann.

Ausgangspunkt: Menschen finden Hilfe

Wer in einer Krise Hilfe sucht, landet heute in einer fragmentierten Landschaft aus Karten, Portalen und Verzeichnissen — veraltet, doppelt, widersprüchlich. Nicht, weil es zu wenige Angebote gäbe, sondern weil die Informationen darüber in vielen Quellen mit unterschiedlichen Strukturen liegen und keine davon das Ganze sieht. Der föderierte Datenpool setzt genau dort an: Die Daten bleiben dezentral bei ihren Quellen, werden aber zentral harmonisiert, zusammengeführt und nutzbar gemacht. Alle folgenden Argumente sind Antworten auf die Frage, warum dieser Ansatz gelingt, wo viele Anläufe zuvor gescheitert sind.

Es funktioniert bereits — öffentlich nachprüfbar

Der föderierte Datenpool ist kein Konzeptpapier. Ein lauffähiger Prototyp föderiert vier reale, heterogene Quellen — AWO, Caritas, Deutsche Hauptstelle für Suchtfragen und die Sozialplattform — zu einem Verzeichnis mit vier Entitätstypen: Träger, Einrichtungen, Angebote und Adressen. Das Ergebnis ist öffentlich einsehbar, inklusive

des Förderierungsprozesses selbst: Jeder Schritt lässt sich in der Weboberfläche nachvollziehen.

So funktioniert es technisch: Der föderierte Datenpool führt Daten über soziale Unterstützungsangebote aus dezentralen Quellen deterministisch und nachvollziehbar in einer gemeinsamen, offen nutzbaren Datenschicht zusammen — die Daten bleiben dabei bei ihren Trägern, sie werden aber zentral harmonisiert und abrufbar. Die technischen Ressourcen für den Betrieb dieser Datenschicht sind minimal. Allerdings braucht es eine stabile und vertrauensvolle Instanz, die die „Harmonisierung“ und auch die menschlichen Kontakte zu den Datenquellen pflegt - und damit das ganze System ermöglicht und am Leben hält.

Beste Grundlage für KI-Anwendungen

Je besser strukturiert und standardisiert Datenquellen sind, desto akkurater und fehlerfreier kann auch KI darauf aufbauen. Im vorliegenden Ansatz wird das föderierte Verzeichnis als Linked Data unter einem einheitlichen Schema publiziert. Eine besser strukturierte Datenquelle ist schwer vorstellbar – Linked Data ist genau dafür ausgelegt, dass Daten stets selbsterklärend sind und sich an internationalen Vokabularen und Ontologien orientieren. Die EU pflegt selbst einen wachsenden Korpus an semantischen Vokabularen, die für mehr Interoperabilität sorgen sollen – einige davon können wir nachnutzen.

Substantiierung des Überblicks

Die föderale Förderlandschaft aus EU, Bund, Ländern, Kommunen und privaten Förderern ermöglicht eine vielfältige soziale Infrastruktur, die einen großen Anteil am Gesamtzusammenhang des Sozialstaats in Deutschland trägt. Doch die Verteilung erschweren zugleich Überblick und Planung, die die Voraussetzung für unbürokratische, schnelle Hilfe bilden. Der Datensatz, der beim föderierten Datenpool entsteht, kann nicht nur die Grundlage für orientierende Plattformen für Bürger*innen sein, sondern auch für eine effiziente, datenbasierte Sozialplanung.

Top-down vs. Community-driven

In einigen Bereichen hat der Staat das Mandat, Standards zu verordnen. So passiert es bspw. aktuell im Kontext von NOOTS bei den anzuschließenden Registern. Daneben gibt es aber viele Bereiche, in denen eine top-down Verordnung nicht oder kaum möglich ist. Ein solcher Bereich ist die Zivilgesellschaft. 2022 erfasste das ZiviZ-Survey des Bundesverbandes Deutscher Stiftungen rund 650.000 registrierte zivilgesellschaftliche Organisationen (vgl. [ZiviZ Survey 2023](#), S.9). Gemeinsame Standards entstehen in dieser Vielzahl, wenn überhaupt, mit der Zeit durch Initiativen, die Anklang finden und sich verfestigen (, ein Problem, auf das auch kürzlich eine [Vorlaufstudie zur Digitalen Daseinsfürsorge](#) von Henning Tillmann mit Unterstützung der Mercator-Stiftung aufmerksam gemacht hatte).

Der Ansatz des föderierten Datenpools trägt genau solchen Konstellationen Rechnung. Statt eines fertigen Schemas wird ein gemeinsamer Prozess mit Daten substantiiert. Weder die Zahl der an der Förderung teilnehmenden Quellen muss zu Anfang klar sein, noch das Zielschema, auf welches die Förderung hinauslaufen soll. Dennoch

steht am Ende von Anfang an ein gemeinsamer Datensatz, der organisch wachsen kann, aber sofort für diverse Einsatzbereiche verfügbar ist.

Souveränität durch Umziehbarkeit

Der Ansatz setzt auf eine zentrale Konfigurationsdatei, die deklarativ vorliegt – unabhängig von einer Programmiersprache. Die gesamte Applikationslogik wird darin ausgedrückt. Neue Quellen hinzuzufügen oder Änderungen im Schema-Mapping – alles passiert ausschließlich dort. Somit ist es leicht, die komplette Implementierungsschicht zu wechseln. Aktuell wird Javascript verwendet, ein Wechsel zu anderen Programmiersprachen wäre vergleichsweise einfach, da quasi nur eine Spezifikation umgesetzt werden muss. Eine Dimension von digitaler Souveränität: „günstige“ Austauschbarkeit von „Execution Environments“.

Ausweg aus Harmonisierungsengpässen

So manches IT- und Standardisierungsprojekt in Deutschland scheint an sog. Harmonisierungsengpässen zu scheitern oder verzögert zu werden. Diese Gefahr besteht beim föderierten Datenpool nicht, denn er setzt auf eine deterministische und symbolische Verbindung von verschiedenen Strukturen. Dies ermöglicht eine unmittelbare Arbeitsfähigkeit inmitten von komplizierten, sich überlagernden Strukturen. Verschiedene Inputs können auf ein gemeinsames Zielschema geführt werden, das für unterschiedliche Anwendungsfälle wieder auf neue Ausgabeschemata angepasst werden kann. Langfristig kann das gemeinsame Zielschema normative Kraft entfalten, weil Quellenbetreibende es wegen der höheren Passung übernehmen. Der Datenpool kann aber schon vorher Wirkung entfalten und muss auf diese Einigung nicht warten.

Anschlussfähig an erklärte Strategie

Der Ansatz basiert durchgängig auf offenen Standards (RDF, SHACL, SPARQL, schema.org). Semantische Technologien sind Teil des Deutschland-Stacks; der IT-Planungsrat plant, den FIM-Datenfeld-Baukasten als Wissensgraph umzusetzen; die EU setzt mit SEMIC, den Core Vocabularies und den Data Spaces auf genau diese Ebene. Der föderierte Datenpool ist damit eine funktionierende Implementierung dessen, was Bund und EU strategisch bereits beschlossen haben — Rückenwind hierfür ist kein Wagnis, sondern der Beweis, dass die eigene Strategie trägt.

Kein weiteres Portal, sondern die Datenschicht darunter

Jede Förderperiode gebiert neue Karten, Portale und Chatbots, die mit dem Förderende sterben — weil jedes Projekt die Datenpflege neu erfindet. Der föderierte Datenpool ist kein weiteres Portal, sondern die Infrastrukturschicht darunter: einmal gepflegte, harmonisierte Daten, auf denen viele Anwendungen aufsetzen können.

Die Betriebskosten sind strukturell minimal (Open Source, statisches Hosting, keine Lizenzen), und das Wertvollste — teilnehmende Datenquellen, Konfiguration, kuratiertes Wissen — verbleiben. Ein gemeinsamer Betrieb durch Civic Data Lab (für das Vertrauen und die Verbindung in die Zivilgesellschaft) und der Sozialplattform (für die langfristige Sicherstellung der Bereitstellung und Nutzung) bietet sich an.