

ADT Portfolio

Modellierung des Lobbyregisters



Noah Raupold (5022097),
David Gläsle(5022114)

Eingereicht am: November 17, 2025

Contents

1 Einleitung

Im Rahmen des FWPMs Advanced Database Techniques besteht unsere Aufgabe darin, ein realistisches Anwendungsszenario in ein funktionales Datenbankmodell zu überführen. Als Zielszenario wählten wir die Modellierung des bundesdeutschen Lobbyregisters, mit dem Ziel, die Struktur und Dynamik der Interessenvertretung in Deutschland möglichst präzise und nachvollziehbar abzubilden. Dieses Thema wurde aufgrund seiner hohen gesellschaftlichen Relevanz sowie der vielfältigen, strukturiert erfassbaren Datenaspekte ausgewählt.

Das Lobbyregister bietet eine komplexe Datenlandschaft aus Organisationen, Einzelpersonen, Tätigkeitsbereichen und Finanzflüssen, die gemeinsam ein transparentes Bild politischer Einflussnahme ermöglichen sollen. Neben den grundlegenden Entitäten wie Interessenvertreterinnen und -vertreter, Mandatsträgern und Organisationen liegt ein besonderer Schwerpunkt auf der Modellierung von Beziehungen und Aktivitäten, um Zusammenhänge und Netzwerke im Bereich der Interessenvertretung sichtbar zu machen.

Darüber hinaus stellt die Integration halbstrukturierter Daten – etwa in Form von JSON-Dokumenten oder erläuternden Textfeldern – eine besondere Herausforderung dar, die wir mit Blick auf Flexibilität und Erweiterbarkeit gezielt adressieren. Durch die Wahl von PostgreSQL als zugrunde liegender Datenbanktechnologie können sowohl relationale als auch semi-strukturierte Daten effizient verwaltet und analysiert werden.

Der Fokus dieses ersten Portfolioteils lag auf der konzeptionellen Modellierung. Im Folgenden stellen wir unseren Ansatz bei der Ideenfindung, der Definition relevanter Entitäten und Beziehungen sowie der Erstellung des Entity-Relationship-Modells (ERM) vor. Abschließend reflektieren wir die bisherigen Erkenntnisse und begründen die Entscheidung für die eingesetzte Datenbanktechnologie.

2 Ideensammlung

Zu Beginn haben wir in einer gemeinsamen Diskussionsrunde verschiedene Anwendungsszenarien abgewogen. Dabei war uns wichtig:

- Ein Szenario mit realem gesellschaftlichem Bezug.
- Ein Datensatz, der sowohl Stammdaten (Organisationen, Personen) als auch Bewegungsdaten (Lobbyaktivitäten, Treffen, Finanzen) enthält.
- Ein Umfeld, in dem wir die Vorteile einer relationalen Datenbank nutzen können – also mit vielen Relationen, Fremdschlüsselbeziehungen und ggf. halb-strukturierten Daten.

Das Thema „Lobbyregister“ überzeugte uns insbesondere, weil:

- In Deutschland mit dem Lobbyregister der Bundestagsverwaltung erstmals ein zentrales Register für Interessenvertretung vorliegt.
- Es zahlreiche Entitäten gibt: Lobbyorganisationen, Lobbyagenturen, Auftraggeber, Einzelpersonen, Finanzierung, gesetzgebende Prozesse, Ziele, Treffen etc.
- Es Potenzial für Analyse gibt: z. B. welche Branchen besonders aktiv sind, welche Verbindungen bestehen, wieviel Geld geflossen ist, welche Gesetzesinitiativen betroffen sind.
- Es auch Möglichkeiten für Datenimport und -integration gibt, z. B. Öffentliche Daten, APIs oder CSV/JSON-Quellen.

Wir haben eine Mindmap erstellt, in der wir zentrale Elemente des Registers identifizierten, etwa:

- Lobbyorganisationen (Name, Rechtsform, Sitz, Branche)
- Auftraggeber und Auftragnehmer von Lobbyarbeit
- Einzelpersonen (Lobbyistinnen und Lobbyisten)
- Finanzinformationen: Budget, Ausgaben, Drittfinanzierung
- Interessen- bzw. Zielkatalog: Welche Themen verfolgt die Lobbyarbeit?

Auf dieser Grundlage haben wir uns entschieden, die Datenbank so zu entwerfen, dass sie sowohl relationale Struktur (z. B. Organisationen Personen Lobbytätigkeit) abbildet, als auch flexible Datenformate unterstützen kann (z. B. JSON-Struktur für Dokumente oder Erklärungen).

Registereintrag

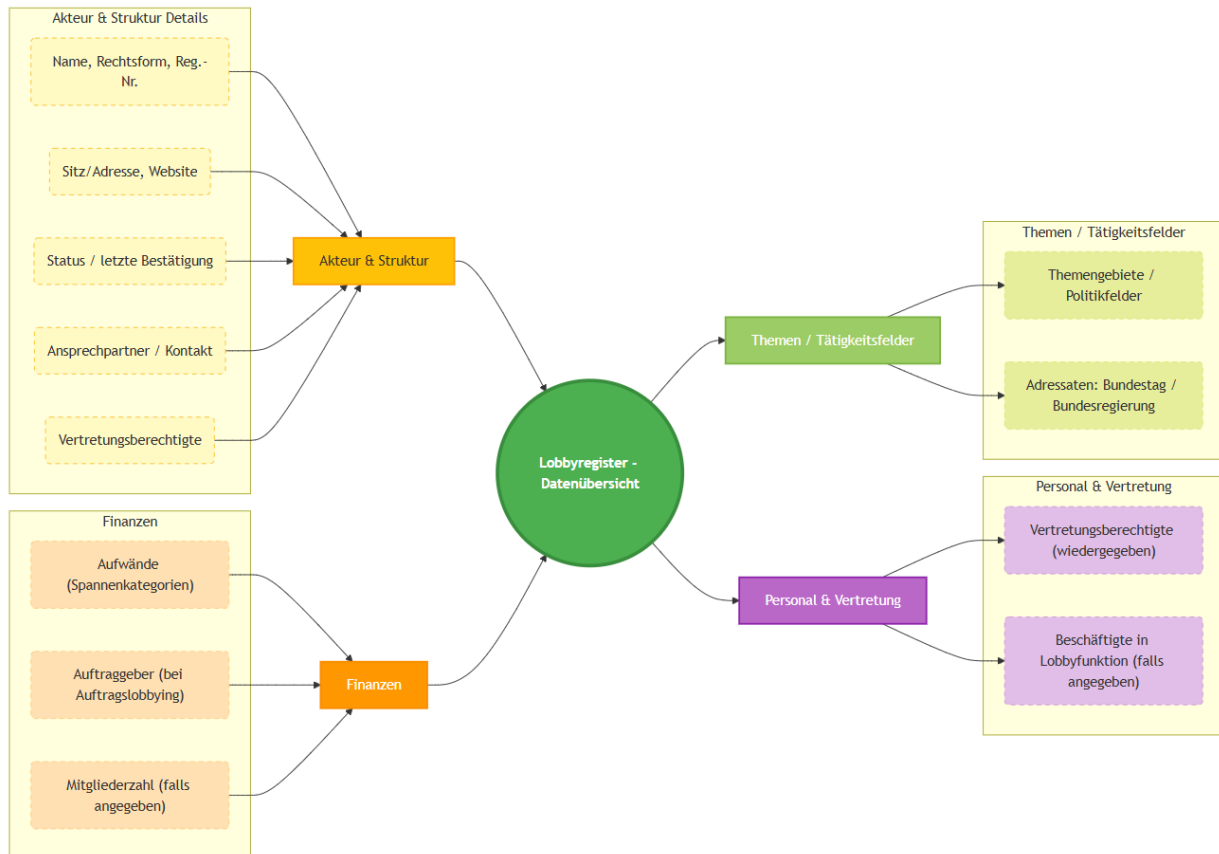


Figure 2.1: Ein Ausschnitt des Registereintrages

3 Use-Case: Analyse von Lobbyaktivitäten

Das gewählte Szenario eröffnet verschiedene spannende Use-Cases. Beispiele:

- Analyse, welche Branchen (z. B. Automobil, Energie, Versicherungen) die höchsten Ausgaben für Lobbying haben.
- Untersuchung von Verbindungen zwischen Lobbyist*innen und Wechsel in öffentliche Ämter („Revolving Door“) – welche Einzelpersonen arbeiten vorher für Verbände und später im öffentlichen Dienst oder umgekehrt.
- Darstellung von Gesetzesvorhaben und dazugehörigen Lobbyaktivitäten – z. B. welche Organisationen sich zu einem bestimmten Gesetzesentwurf gemeldet haben, mit welchem Ziel und mit welchem Budget.
- Zeitliche Analyse: Wie verändert sich die Lobbying-Aktivität über Zeit (Jahre, Wahlperioden)?

Diese Use-Cases zeigen, dass wir nicht nur eine statische Datenstruktur erstellen, sondern auch Abfragen, Auswertungen und Reports ermöglichen wollen – was Anforderungen an das Datenbankmodell und an das DBMS stellt.

4 Datenmodell

Dieses Kapitel beschreibt das zentrale Datenmodell, das zur Abbildung der Informationen aus dem Lobbyregister des Deutschen Bundestages über die bereitgestellte API dient.

Um die tief verschachtelten JSON-Daten der API in einer performanten und konsistenten relationalen Datenbank (PostgreSQL) zu speichern, wurde ein vollständig normalisiertes Schema in der 3. Normalform (3NF) entworfen. Das resultierende physische Datenmodell ist sehr detailliert und besteht aus **insgesamt 84 einzelnen Tabellen**. Diese hohe Anzahl entsteht durch die konsequente Auflösung aller Listen (Arrays), wiederverwendbarer Strukturen (wie Adressen oder Codelisten) und verschachtelter Objekte in eigene Tabellen, die über Fremdschlüssel miteinander verbunden sind.

Eine vollständige Dokumentation aller 84 Tabellen würde den Rahmen dieser Arbeit sprengen. Dieses Kapitel beschreibt daher als repräsentativen **Ausschnitt** die konzeptionelle Grundlage dieses Schemas. Es konzentriert sich auf die beiden Hauptentitäten, wie sie von der API definiert werden: `RegisterEntry` und `RegisterEntryStatistics`. Diese bilden den logischen Ausgangspunkt, aus dem das normalisierte 84-Tabellen-Schema abgeleitet wurde.

Die Basis-URL für die API-Aufrufe ist `https://api.lobbyregister.bundestag.de/rest/v2`.

4.1 Hauptentität: `RegisterEntry`

Die Entität `RegisterEntry` stellt einen vollständigen Registereintrag dar und ist die zentrale Ressource, die über die Endpunkte wie `/registerentries/{registerNumber}` abgerufen werden kann.

Die Entität beinhaltet die folgenden Hauptkomponenten:

- `source`, `sourceUrl`, `sourceDate`, `jsonDocumentationUrl`: Metadaten zur Quelle und zum Zeitpunkt der Datenabfrage.
- `registerNumber`: Die eindeutige Registernummer des Eintrags, deren Format dem Muster `^R\d{6}$` folgt (z. B. R001234).
- `accountDetails`: Enthält Informationen zum Zustand des Kontos der Interessenvertretung (IV).
- `registerEntryDetails`: Detaillierte, versionsspezifische Informationen zum Eintrag.

- **lobbyistIdentity:** Informationen zur Identität der Interessenvertreterin/des Interessenvertreters .
- **activitiesAndInterests:** Informationen zu Tätigkeiten und Interessenbereichen .
- **clientIdentity:** Angaben zur Identität des Auftraggebers/der Auftraggeberin, falls vorhanden .
- **employeesInvolvedInLobbying:** Angaben zu den im Bereich der Interessenvertretung beschäftigten Personen .
- **financialExpenses:** Angaben zu den jährlichen finanziellen Aufwendungen für Interessenvertretung .
- **publicAllowances:** Angaben zu Zuwendungen und Zuschüssen .
- **donators:** Angaben zu Schenkungen .
- **membershipFees:** Angaben zu Mitgliedsbeiträgen .
- **annualReports:** Angaben zum Jahresabschluss/Rechenschaftsbericht .
- **regulatoryProjects:** Angaben zu Regelungsvorhaben, zu denen Interessenvertretung betrieben wird .
- **statements:** Angaben zu Stellungnahmen/Gutachten .
- **contracts:** Angaben zu Aufträgen für Interessenvertretung .
- **codeOfConduct:** Angaben zum Verhaltenskodex .

4.1.1 Detaillierte Sub-Entitäten von RegisterEntry

4.1.1.1 AccountDetails

Diese Struktur enthält Zustandsinformationen des Kontos :

- **activeLobbyist:** Gibt an, ob die IV aktuell aktive (TRUE) oder inaktive (FALSE) Interessenvertretung betreibt .
- **registerEntryVersions:** Eine Liste (array) aller Versionen, die für das Konto vorliegen . Jede Version enthält die interne ID (**registerEntryId**), die Versionsnummer (**version**), die zugrundeliegende Gesetzeslage (**legislation**, z.B. GL2022 oder GL2024), und den Gültigkeitszeitraum (**validFromDate**, **validUntilDate**) .
- **accountHasCodexViolations:** Wahr, wenn mindestens ein Verstoß gegen den Verhaltenskodex vorliegt .

- **codexViolations:** Liste der aktiven Verstöße gegen den Verhaltenskodex, mit dem Attribut **codexViolationName** .

4.1.1.2 LobbyistIdentity

Diese Struktur beschreibt die Identität des Interessenvertreters/der Interessenvertreterin :

- **identity:** Personentyp der IV, entweder **NATURAL** (natürliche Person) oder **ORGANIZATION** (Organisation) .
- Natürliche Personen nutzen Attribute wie **lastName**, **firstName** (GL2024), **commonFirstName** (GL2022) und akademische Grade (**academicDegreeBefore/After**) .
- Organisationen nutzen **companyName** , **legalFormType** und **legalForm** .
- Die Struktur enthält ferner Angaben zu Adress- und Kontaktdaten (**address**, **contactDetails**) , zur Hauptstadtrepräsentanz (**capitalCityRepresentation**) , sowie Listen von betrauten Personen (**entrustedPersons**) und gesetzlichen Vertreter/-innen (**legalRepresentatives**) .

4.1.1.3 ActivitiesAndInterests

Informationen zu den Aktivitäten der Interessenvertretung :

- **activity:** Informationen zur Tätigkeit selbst, inklusive **code**, deutscher (**de**) und englischer (**en**) Bezeichnung .
- **typesOfExercisingLobbyWork:** Liste der Arten der Ausübung der Interessenvertretung .
- **fieldsOfInterest:** Eine Liste der Interessenbereiche .
- **activityOperationType:** Gibt an, ob die Interessenvertretung **SELF_OPERATED**, **MANDATE_OPERATED** oder **BOTH** betrieben wird .
- **legislativeProjects:** Liste der Gesetzesvorhaben, inklusive Name und Drucksachenummer (**printingNumber**) .

4.1.1.4 FinancialExpenses

Angaben zu den jährlichen finanziellen Aufwendungen für Interessenvertretung :

- **refuseFinancialExpensesInformation:** Gibt an, ob Angaben verweigert wurden .
- **relatedFiscalYearFinished**, **relatedFiscalYearStart**, **relatedFiscalYearEnd:** Details zum Bezugsgeschäftsjahr (relevant für GL2024) .

- **financialExpensesEuro**: Zahlenbereich (**from/to**) der jährlichen Aufwendungen in Euro .

4.1.1.5 RegulatoryProjects

Angaben zu konkreten Regelungsvorhaben :

- **regulatoryProjectsPresent**: Gibt an, ob Interessenvertretung hinsichtlich eines konkreten Regelungsvorhabens betrieben wird .
- **regulatoryProjects**: Liste der Regelungsvorhaben, jeweils mit **regulatoryProjectNumber** (RVxxxxxxx) und **title** .
- Enthält Details zu vorhandenen Drucksachen (**printedMatters**) und Referentenentwürfen (**draftBill**) .

4.2 Hilfsentität: RegisterEntryStatistics

Die Entität **RegisterEntryStatistics** dient der Bereitstellung von Kennzahlen und Statistiken zu den Registereinträgen .

- Die Statistikdaten sind im Objekt **registerEntries** enthalten .
- **activeLobbyistsLegalFormType**: Aufteilung der aktiven Interessenvertreter/-innen nach Personentypen, z.B. **naturalPersons**, **juristicPersons** und **legalAssociations** (Personengesellschaften) . Jeder Typ listet **number** und **percentage** auf .
- **codexViolationOrLateUpdate**: Statistiken zu Verstößen gegen den Verhaltenskodex (**codexViolation**) oder fehlender Aktualisierung des Geschäftsjahres (**lateUpdate**) .
- **lobbyists**: Gesamtanzahl aller IVs, unterteilt in **active** und **inactive** .
- **peopleInvolvedInLobbyistWork**: Gesamtanzahl aller Personen, die Interessenvertretung betreiben (IVs und betraute Personen) .
- **recentGovernmentFunction**: Statistiken zum "Drehtüreffekt" (**existing** vs. **notExisting**), aufgeschlüsselt nach Bundesregierung, Bundestag und Bundesverwaltung .
- **fieldsOfInterest**: Listen der Obergruppen (**mainGroups**) und Untergruppen (**subgroups**) der Interessenbereiche, jeweils mit Anzahl und prozentualem Anteil .
- **activities**: Liste der Tätigkeitskategorien der IVs, ebenfalls mit Anzahl und prozentualem Anteil .

4.3 Codierte Hilfsobjekte

Viele Objekte innerhalb der Hauptentitäten verwenden ein standardisiertes Format für Bezeichnungen, die über Codes referenziert werden. Beispiele hierfür sind `naturalPersonType`, `country`, `type` des Drehtüreffekts oder `fieldsOfInterest`. Diese Objekte haben typischerweise die Attribute `code`, `de` (deutsche Bezeichnung) und `en` (englische Bezeichnung).

Nachteile, die wir in Kauf genommen haben:

- Höhere Einstiegshürde im Vergleich zu einfacheren Systemen.
- Möglicherweise höherer Ressourcenverbrauch bei großer Last.

5 Reflexion / Erkenntnisse

Zu Beginn standen wir vor Herausforderungen: Definition der Anforderungen, klare Abstimmung innerhalb der Gruppe, Auswahl der richtigen Entitäten und Attribute, Aufbau des Modells. Schnell zeigte sich, dass insbesondere die Modellierung der flexiblen Datenfelder (Themenlisten, Dokumente) sowie die Balance zwischen Normalisierung und Performanz sorgfältig bedacht werden musste.

Positiv war, dass die Wahl von PostgreSQL uns viel Freiraum gab. Wir konnten eine hybride Strategie verfolgen:

- **Relationale Normalisierung:** Für klar strukturierte 1:N- oder N:M-Beziehungen (z.B. `LobbyistIdentity` zu `LegalRepresentatives` oder die Verknüpfung von `RegisterEntry` zu `InterestsTopics`) wurde eine klassische Normalisierung mit Fremdschlüsseln gewählt. Dies ist essenziell für Joins und aggregierte Abfragen.
- **Nutzung von JSONB:** Für tief verschachtelte, variable oder selten abgefragte Listen (z.B. die `expensesBreakdown` innerhalb von `financialExpenses` oder die Liste der `statements`) wurde der JSONB-Datentyp genutzt.
- **Begründung:** Diese Balance verhindert eine "Schema-Explosion" (potenziell 50+ Tabellen bei voller Normalisierung) und hält die Komplexität beherrschbar, während die relationale Stärke für die Kernanalysen (Use-Cases) erhalten bleibt.

Wir haben gelernt, dass die gesellschaftliche Relevanz eines Szenarios (wie das Lobbyregister) die Motivation stärkt und dass eine reale Datenquelle – zumindest als Beispiel – hilfreich ist. Zudem wurde klar: Dokumentation der technischen Entscheidungen (z. B. DBMS-Wechsel) gehört unbedingt ins Portfolio.

6 Fazit

Den ersten Meilenstein in der Entwicklung unserer Datenbank – die Erstellung eines konzeptionellen Datenmodells – konnten wir als Team erfolgreich abschließen. Dabei zeigte sich jedoch, dass der damit verbundene Aufwand größer war, als wir anfangs angenommen hatten. Besonders die Modellierung der vielfältigen Entitäten und Beziehungen innerhalb des Lobbyregisters erwies sich als komplex, da insbesondere die Balance zwischen struktureller Normalisierung und technischer Flexibilität sorgfältig abgestimmt werden musste.

Herausfordernd war vor allem der Umgang mit variablen und tief verschachtelten Datenstrukturen, wie sie beispielsweise bei Themenlisten oder Finanzangaben auftreten. Die Entscheidung, neben klassischen relationalen Strukturen auch JSONB-Felder einzusetzen, erwies sich dabei als sinnvoller Kompromiss zwischen Performance und Wartbarkeit.

Trotz der anfänglichen Herausforderungen empfanden wir die Arbeit mit PostgreSQL insgesamt als positiv. Die vertraute SQL-Syntax sowie die erweiterten Funktionen für semi-strukturierte Daten ermöglichten es uns, ein robustes und zugleich flexibles Datenmodell zu entwerfen. Besonders hilfreich war zudem die enge Zusammenarbeit im Team, um konzeptionelle und technische Entscheidungen abzustimmen und die Komplexität des Projekts gemeinsam zu bewältigen.

7 Ausblick / Weiteres Vorgehen

In den kommenden Arbeitsschritten werden wir das entworfene Datenmodell in PostgreSQL umsetzen und auf Basis der definierten Use-Cases weiterentwickeln. Für die Visualisierung der Daten planen wir den Einsatz von Grafana, da es umfangreiche Analysemöglichkeiten sowie eine native Unterstützung für PostgreSQL bietet.

Im Rahmen der Implementierung werden wir gezielt auf Performance-Optimierungen eingehen. So sollen beispielsweise **Materialized Views** für aggregierte Finanzanalysen erstellt und spezielle Indizes (z. B. **GIN-Indizes** auf JSONB-Feldern) eingerichtet werden, um komplexe Suchanfragen – etwa nach ehemaligen Regierungsmitgliedern oder finanziellen Verflechtungen – effizient zu gestalten.

Langfristig ist geplant, auf Basis der Datenbank interaktive Auswertungen und Visualisierungen zu realisieren. Denkbar ist etwa der Aufbau eines Dashboards, das zentrale Kennzahlen wie Branchenvergleiche oder Netzwerkstrukturen von Organisationen und Personen darstellt. Zudem könnte das Modell durch automatisierte Datentransformationen erweitert werden, um regelmäßig neue Registerdaten zu integrieren.

Damit legen wir den Grundstein für eine datenbankgestützte Analyse des Lobbyregisters, die sowohl technische als auch gesellschaftliche Aspekte sichtbar macht und zukünftige Erweiterungen – etwa durch Webanwendungen oder API-Anbindungen – ermöglicht.