

Notes on the MEDem ATLAS Prototype

Lorenzo De Sio (Luiss University Rome)

Alexia Katsanidou (GESIS – University of Cologne)

Enrico Borghetto (University of Florence)

March 16, 2026

ABSTRACT:

This short note introduces the MEDem ATLAS prototype: an online application designed to demonstrate the potential of concept-driven dataset annotation for facilitating comparative social science research. The system implements a structured data model linking an ontology of theoretical concepts to tabular datasets containing variables and metadata. Within this framework, concepts for object properties—such as attributes of persons or political parties—are defined and organized hierarchically into dimensions, subdimensions, and indicator concepts, each supported by literature-based definitions. These conceptual elements are systematically connected to empirical operationalizations through a variables database that records how indicators are measured across datasets. By linking conceptual knowledge to dataset metadata, the tool enables interactive exploration of conceptual coverage across datasets and facilitates the identification and comparison of available operationalizations. The system also includes a Research Builder tool that allows users to select concepts relevant to a research question and automatically identify datasets containing corresponding indicators, while providing detailed information on measurement coverage. The prototype further demonstrates the possibility for cross-dataset linkage and generates automated scripts for preliminary analysis in common data science languages. Overall, MEDem ATLAS illustrates how ontology-based annotation can enhance transparency, comparability, and efficiency in secondary data analysis.

The MEDem ATLAS prototype is an online application demonstrating some of the potential of dataset annotation following the data model and procedures presented in De Sio, Katsanidou and Borghetto (under review on PS: Political Science and Politics).

Its internal data structure consists of:

- 1) An ontology document system storing, for each top-level object property concept,
 - a) a brief definition with literature references.
 - b) details about its articulation into subdimensions, and finally into indicator concepts.
- 2) Tabular Datasets: This consists of two interconnected tables:
 - a) A "datasets" table, with one row for each dataset.
 - b) A "variables" table, containing one row for every variable within each dataset. This table also incorporates entity references for variables whose values represent real-world entities.

The combination and linking of this information enables the automated, interactive generation of meaningful views (beyond the scope of harmonization tasks), as exemplified below.

The initial Concepts view (Figure 1) outlines conceptualizations by coding the properties of various objects (such as Persons and Political Parties). These conceptualizations are further structured by their division into dimensions and corresponding indicators.

The screenshot displays the 'Atlas: Concepts' web application interface. The browser address bar shows 'localhost:3000/medem/'. The application header includes the 'MEDem' logo, the title 'ATLAS', and navigation links for 'Objects', 'Concepts', 'Data', and 'Tools'. A 'Logout' button is also present.

The main content area is divided into two sections, each representing an object concept:

- Person (object concept):** This section features a 'Research builder: Person' button and a list of properties. Each property is represented by a lightbulb icon and includes its name, dimensions, and indicators.
 - Properties:**
 - Religiosity (3 dimensions, 8 indicators)
 - Social Conformism (1 dimensions, 1 indicators)
 - Party Identification (2 dimensions, 2 indicators)
 - Ideology (1 dimensions, 1 indicators)
 - Perceptions Of Political Environment (2 dimensions, 2 indicators)
 - Voting Behavior (3 dimensions, 7 indicators)
 - Democracy: Preferences (9 dimensions, 18 indicators)
 - Democracy: Evaluation (7 dimensions, 15 indicators)
 - Equality: Preferences (3 dimensions, 5 indicators)
 - Equality: Evaluation (5 dimensions, 9 indicators)
 - Freedom: Preferences (2 dimensions, 4 indicators)
 - Gender Roles: Preferences (2 dimensions, 6 indicators)
 - Gender Roles: Evaluation (1 dimensions, 3 indicators)
 - Household Composition (1 dimensions, 1 indicators)
 - Immigration: Experience (2 dimensions, 4 indicators)
 - Immigration: Perception (4 dimensions, 7 indicators)
 - Issue Positions (5 dimensions, 23 indicators)
 - Media Use (5 dimensions, 9 indicators)
 - Multiculturalism (2 dimensions, 5 indicators)
 - Personality Traits (4 dimensions, 12 indicators)
 - Political Efficacy (2 dimensions, 2 indicators)
 - Political Engagement (2 dimensions, 4 indicators)
 - Political Participation (2 dimensions, 6 indicators)
 - Political Participation: Preferences (1 dimensions, 2 indicators)
 - Political Participation: Evaluation (1 dimensions, 2 indicators)
 - Political Satisfaction (5 dimensions, 9 indicators)
 - Political Trust (3 dimensions, 14 indicators)
 - Rule Of Law (1 dimensions, 2 indicators)
 - Security (1 dimensions, 1 indicators)
 - Social Capital (1 dimensions, 2 indicators)
 - Social Trust (2 dimensions, 6 indicators)
 - Socioeconomic Status (7 dimensions, 12 indicators)
 - Biological Characteristics (2 dimensions, 2 indicators)
 - Geographic Origin (3 dimensions, 5 indicators)
 - Wellbeing (6 dimensions, 18 indicators)
 - Work (2 dimensions, 4 indicators)
 - Country Situation: Evaluation (5 dimensions, 14 indicators)
 - Evaluation Of Political Institutions (2 dimensions, 6 indicators)
 - Governance: Preferences (1 dimensions, 2 indicators)
 - Governance: Evaluation (2 dimensions, 10 indicators)
 - Attitudes Towards The European Union (5 dimensions, 19 indicators)
- Political party (object concept):** This section features a 'Research builder: Political party' button and a list of properties.
 - Properties:**
 - Political Party Identity (1 dimensions, 4 indicators)
 - Party Family (1 dimensions, 2 indicators)
 - Size (3 dimensions, 3 indicators)
 - Issue Emphasis: Economic Policies (7 dimensions, 14 indicators)
 - Issue Emphasis: Freedom And Human Rights (2 dimensions, 4 indicators)
 - Issue Emphasis: Economic Goals (3 dimensions, 5 indicators)

Figure 1 – The initial Concepts view

Figure 2 shows an example (mock-up) entry for religiosity (a property Concept entry: Religiosity is a property of a Person). Note the implementation of concepts-variables linkage: within each Subdimension, each Indicator concept links to actual operationalizations – “measures” – available in different datasets.

The screenshot displays the MedEM ATLAS web application interface. The browser address bar shows 'localhost:3000/medem/concepts/dbor:P...'. The application header includes the MedEM logo, 'ATLAS', and navigation links for 'Objects', 'Concepts', 'Data', and 'Tools'. A 'Logout' button is also present.

The main content area features a concept entry for 'Religiosity', which is identified as a property of a 'Person'. The entry includes a definition: "A religion is a unified system of beliefs and practices relative to sacred things, that is to say, things set apart and forbidden – beliefs and practices which unite into one single moral community called a Church, all those who adhere to them." (Durkheim 1912). It also cites Parsons (1961, 9) and provides a detailed description of religious differentiation into organizational and nonorganizational forms, referencing works by Wach (1944), Allport (1958), Mindel and Vaughan (1978), and others.

Below the definition, the 'Subdimensions and indicators' section is shown. It lists three subdimensions: 'Organizational', 'Nonorganizational: Attitudes', and 'Nonorganizational: Private Behavior'. Each subdimension is associated with specific indicators and their corresponding measures in datasets.

- Organizational**
 - Religious Attendance (2 measures in 2 datasets)
 - Religious Denomination (1 measures in 1 datasets)
- Nonorganizational: Attitudes**
 - Belief In God
 - Prayer
 - How Religious (17 measures in 17 datasets)
- Nonorganizational: Private Behavior**
 - No Premarital Sex
 - No Meat On Good Friday
 - Observe Ramadan

Figure 2 – A (property) Concept entry

A drill-down view (see Figure 3) shows the potential of this concept-centric scheme to easily organize operationalizations across multiple languages.

Concept - MedEM

localhost:3000/medem/concepts/dbo:P...

Religious Denomination (1 measures in 1 datasets)

Nonorganizational: Attitudes

- Belief In God
- Prayer
- How Religious (17 measures in 17 datasets)

Measures of **How Religious** from the Variables database:

Dataset	Variable	Label	Language	Introduction text	Question text	Values and labels
SQP Tests for: ESS Round 1 - Austria	RlgDgr	RlgDgr	German	.	Ungeachtet der Tatsache, ob Sie einer bestimmten Religion angehören, für wie religiös halten Sie sich? Bitte verwenden Sie diese Karte für Ihre Antwort. translate	0 Überhaupt nicht religiös
						1
						2
						3
						4
						5
						6
						7
						8
						9
SQP Tests for: ESS Round 1 - Czech Republic	RlgDgr	RlgDgr	Czech	.	Jak jste zbožný - bez ohledu na to, zda se hlásíte k nějakému konkrétnímu náboženství? translate	00 Vůbec nejsem zbožný
						01
						02
						03
						04
						05
						06
						07
						08
						09
SQP Tests for: ESS Round 1 - Denmark	RlgDgr	RlgDgr	Danish	.	Uanset om du tilhører en bestemt religion eller ej, hvor religiøs vil du sige, at du er? På en skala fra 0 til 10, hvor 0 betyder slet ikke religiøs og 10 betyder meget religiøs. translate	00 Slet ikke religiøs
						01
						02
						03
						04
						05
						06
						07
						08
						09

Figure 3 – Drill-down view of Indicator-linked operationalizations from different datasets

The Data view (Figure 4) shows a first important potential of the linkage between the Concepts knowledge graph and the Variables database. Due to this architecture, it's easy to implement, in the Data view, the ability to quickly review the conceptual coverage available in different datasets.

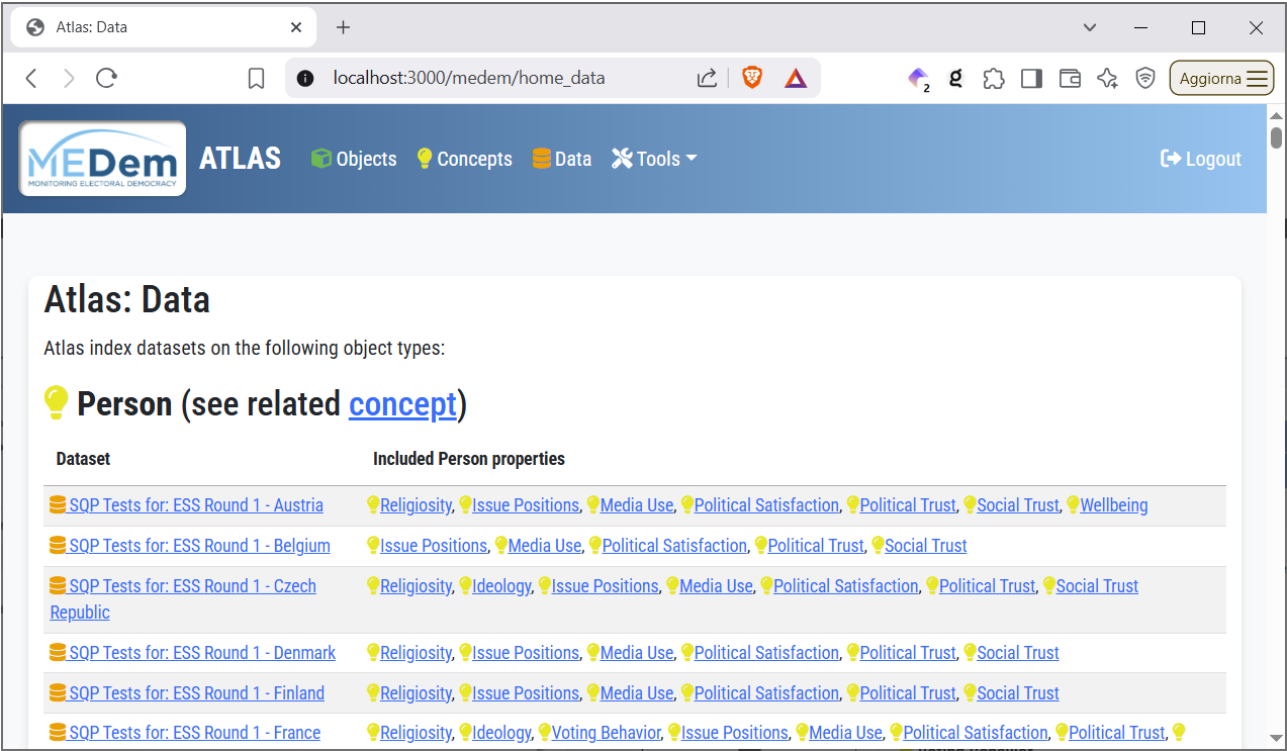


Figure 4 – The initial Data view

But perhaps this potential is made available in an even more productive fashion in the Research Builder (Figure 5). This tool makes very easy to quickly find datasets including at least one indicator for each of several concepts that are relevant for a research question.

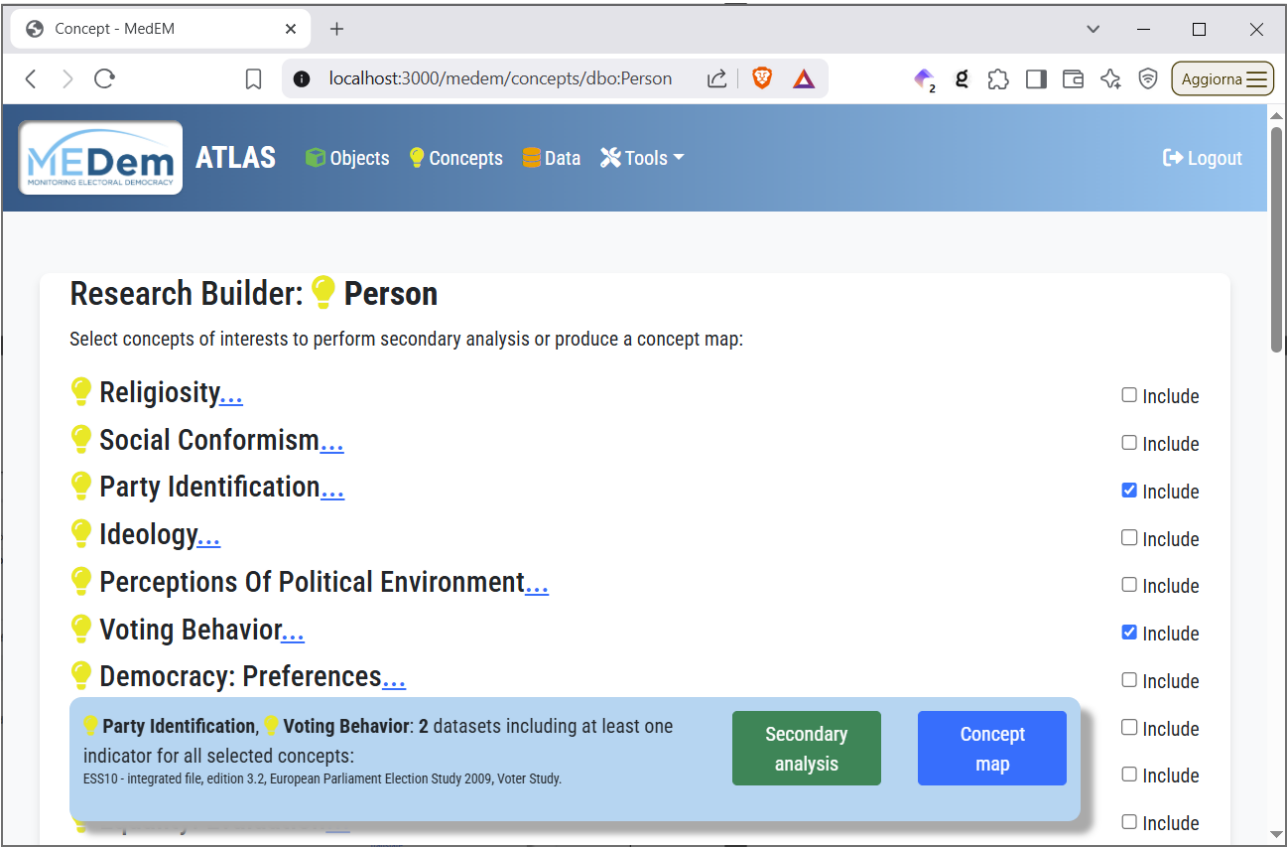


Figure 5 – The Research Builder starting page

In the second step in the Research Builder (Figure 6), Subdimension and Indicator coverage is shown for each dataset (showing the relevant variables), for all the concepts selected in the previous step. This makes easy for the researcher to identify, among candidate datasets, which offers the conceptual coverage that best suits her research needs, along with actual operationalization details.

As a result, for a set of concepts relevant for a research question, detailed information about breadth and quality of measurement coverage of each concept can be reviewed and compared across multiple datasets.

Importantly, in this step the researcher can also select options for cross-dataset linkage. In the example shown in Figure 6, a number of different party-related variables from the EES 2009 Voter study (the “main” dataset) can be used as linkage keys towards other datasets about political parties (here, CHES 1999-2019 and EES 2009 Euromanifesto appear available for linkage).

Dataset selection

These are the datasets that include at least one indicator for all the 4 concepts you selected (Party Identification [7], Voting Behavior [15], Issue Positions [65], Media Use [71]):

Show 10 entries

Showing 1 to 1 of 1 entries

Search:

Previous 1 Next

Dataset	Available external links	Party Identification	Voting Behavior	Issue Positions	Media Use	
European Parliament Election Study 2009, Voter Study dbo:Person	Select	q25 (q25 Which party voted for in EP elections) ☐ dbo:PoliticalParty: CHES1999-2019 ☐ dbo:PoliticalParty: EES2009Euromanifesto q26 (q26 Which party, if R had voted in EP elections) ☐ dbo:PoliticalParty: CHES1999-2019 ☐ dbo:PoliticalParty: EES2009Euromanifesto q27 (q27 Actual vote in last general election) ☐ dbo:PoliticalParty: CHES1999-2019 ☐ dbo:PoliticalParty: EES2009Euromanifesto q28 (q28 Vote intention in general elections) ☐ dbo:PoliticalParty: CHES1999-2019 ☐ dbo:PoliticalParty: EES2009Euromanifesto q87 (q87 close to a party [pid]) ☐ dbo:PoliticalParty: CHES1999-2019 ☐ dbo:PoliticalParty: EES2009Euromanifesto q90 (q90 not close but closer to party: which?) ☐ dbo:PoliticalParty:	Direction Which Party Feel Closer To: q87 (q87 close to a party [pid]) q89 (q89 not close to a party - but closer to a party?) q90 (q90 not close but closer to party: which?) Intensity How Close To Each Political Party: q88 (q88 close or sympathizer)	Party Vote Party Voted For In Last National Election: q27 (q27 Actual vote in last general election) Party Voted For In Last Eu Election: q25 (q25 Which party voted for in EP elections) q26 (q26 Which party, if R had voted in EP elections) Party Voted: Turnout Voted: Voted In The Eu Election: q24 (q24 Did or did not vote in EP elections) Vote Intention First Vote Preference: q28 (q28 Vote intention in general elections) Second Vote Preference:	Civil Rights Homosexuals Should Be Free To Be With Whom They Want: Stance On Same-sex Marriages: q58 (q58 agree/disagree: Same-sex marriages should be prohibited by law) Stance On Abortion: q60 (q60 agree/disagree: Women should be free to decide on matters of abortion) Economy Stance On Income Equality: q63 (q63 agree/disagree: Income and wealth should be redistributed towards ordinary people) Fair Level Of Gross Pay: Stance On Government Intervention Economy: q57 (q57 agree/disagree: Private enterprise best to solve [country's] economic problems) q61 (q61 agree/disagree: Politics should abstain from intervening in the economy) Stance On Role Trade Unions: Stance On Private Enterprise Best Solution: Stance On State Ownership: q59 (q59 agree/disagree: public services and industries should be in state ownership) Immigration Evaluation Of Allowing Immigrants From Poorer Non-european Countries: Evaluation Of Allowing Immigrants Of Different Race:	Internet Use Use Of Internet: Use Of Internet On Politics: q20 (q20 how often look into website concerned with election?) Newspaper Use Total Use Of Newspaper: q13 (q13 Read other paper more frequently?) q15a (q15a days a week read other paper 1st) q15b (q15b days a week read other paper 2nd) q15c (q15c days a week read other paper 3rd) q15d (q15d days a week read other paper 4th) q15e (q15e days a week read other paper 5th) Use Of Newspaper On Politics: q17 (q17 how often read about election in newspaper?) Radio Use Total Radio Use: Use Of Radio On Politics: Tv Use Total Tv Use: Use Of Tv On Politics:

Figure 6 – The Research Builder dataset selection step

Finally, the Research Builder, currently in its pilot phase, will provide the researcher with an automatically generated kick-start script in the most common data science languages, such as Stata, R, and Python. The script will include:

- information on how to obtain the dataset(s) from the appropriate archives;
- automatically-generated syntax performing linkage among the selected datasets (if selected: not shown in the example);
- automatically-generated syntax for descriptive analyses of all variables linked to the originally selected concepts, along with extensive comments detailing the role of the variable in the concept articulation, and actual operationalization information (e.g., question wording for survey data).

This kick-start script, with its inclusion of conceptual and operationalization information, is deliberately designed to encourage concept- and measurement-aware data analysis.