

PolPaL (Political Party Links) Processes v1.1

Paul Bederke & Lea Kaftan

As of October 2026

Table of Contents

1. Introduction 5

2. Importing entries from external datasets to the PolPaL database 5

2.1 Selection criteria for datasets to be integrated into PolPaL 8

2.2 Updating metadata on external datasets 8

3. Generating new PolPaL countries 9

4. Update or delete PolPaL countries 11

4.1 Update PolPaL countries 11

4.2 Delete PolPaL countries 11

5. Generating new PolPaL entities 13

6. Update or delete PolPaL entities 15

6.1 Curate PolPaL entities 15

6.2 Update PolPaL entities 16

6.3 Delete or hide PolPaL entities 16

7. Linking external entities to PolPaL entities 18

8. Generating new user accounts 19

9. Update or delete user accounts 21

9.1 Update user accounts 21

9.2 Delete user accounts 21

10. Quality checks for PolPaL entities and links 21

10.1 Validating existing links 22

10.2 Validating curated entities 22

10.3 Generating tickets 23

10.4 Processing tickets 24

10.5 Automated quality checks using R scripts 24

11. Training new team members 26

11.1 Becoming a User 26
11.2 Training new Contributors 26
11.3 Training new Editors 28
11.4 Training new Data Managers 28
11.5 Becoming a website Admin 29
12. Publication of data-related stable releases 29
12.1 Content of the stable releases 31
13. Publishing a software-related stable releases 31
14. Publication of news 32
15. Security updates of the PolPaL software code 32
16. References 33
17. Version history 33

Figures

Figure 1: Processes to import external datasets to the PolPaL database 7
Figure 2: Updating metadata on external datasets in PolPaL 8
Figure 3. Process to generate new PolPaL countries 10
Figure 4: Updating information on PolPaL countries 11
Figure 5: Deleting a PolPaL country 12
Figure 6. Process to generate new PolPaL entities 14
Figure 7: Curating entities 15
Figure 8: Screenshot showing where to hide an entity 17
Figure 9: Screenshot showing where to delete an entity 17
Figure 10: Linking external dataset entries to entities 19
Figure 11: Assignment of linking tasks to PolPaL users 19
Figure 12: Processes to register new users 20
Figure 13: Validating links between external entries and PolPaL entities 22
Figure 14: Validating curated PolPaL entities 23
Figure 15: Screenshot showing where to find the ticket system 24
Figure 16: Standardized quality checks 25
Figure 17: Onboarding Contributors in PolPaL 27
Figure 18: Processes to publish stable release and news entry 30

1. Introduction

This document describes and visualizes the processes within PolPaL. Readers will find more information on the types of users and data gathered in the documents → **PolPaL_Codebook**, → **PolPaL_DataModel**, → **PolPaL_DataModel-extendedVersion**, and → **PolPaL_GovernanceStructure**. Please note that we ignore users with the assigned role ADMIN in most parts of this document. Unless otherwise stated, they have the right to do everything described in this document. Please note that PolPaL's governance structure is still under construction. Changes made to the governance structure will also likely affect PolPaL's processes in the near future.

2. Importing entries from external datasets to the PolPaL database

Before data can be linked to PolPaL entities, PolPaL requires the import of metadata that describes external datasets. As the import requires some sophisticated preparation of the external dataset, only trained Data Managers are allowed to import datasets (→ **PolPaL DataModel**, Section 11.4 Training new Data Managers). Technically speaking, metadata describing external data can be uploaded to PolPaL by registered users with the assigned roles DATAMANAGER.

Please note that PolPaL does not provide direct access to external datasets. It only provides the metadata necessary for linking external data to PolPaL entities and other external datasets. Users need to download the original data from its original source if they want to work with a specific dataset.

Figure 1 depicts the process by which new external datasets are prepared for integration into the PolPaL database. This process leads to the upload of an R script and a csv file (named after the → **versionKey.R**) to the PolPaL database. The R script is used to extract the relevant metadata from the external data. Each R script is available for download via the PolPaL website. The external data will be stored locally only for metadata extraction and will not be shared publicly. The generated csv file → **versionKey.csv** contains the metadata information that is required to link external entries to PolPaL entities **(→ PolPaL Data Model**, in particular the required data on projects [Projects], datasets [Datasets], and dataset versions [Versions]).

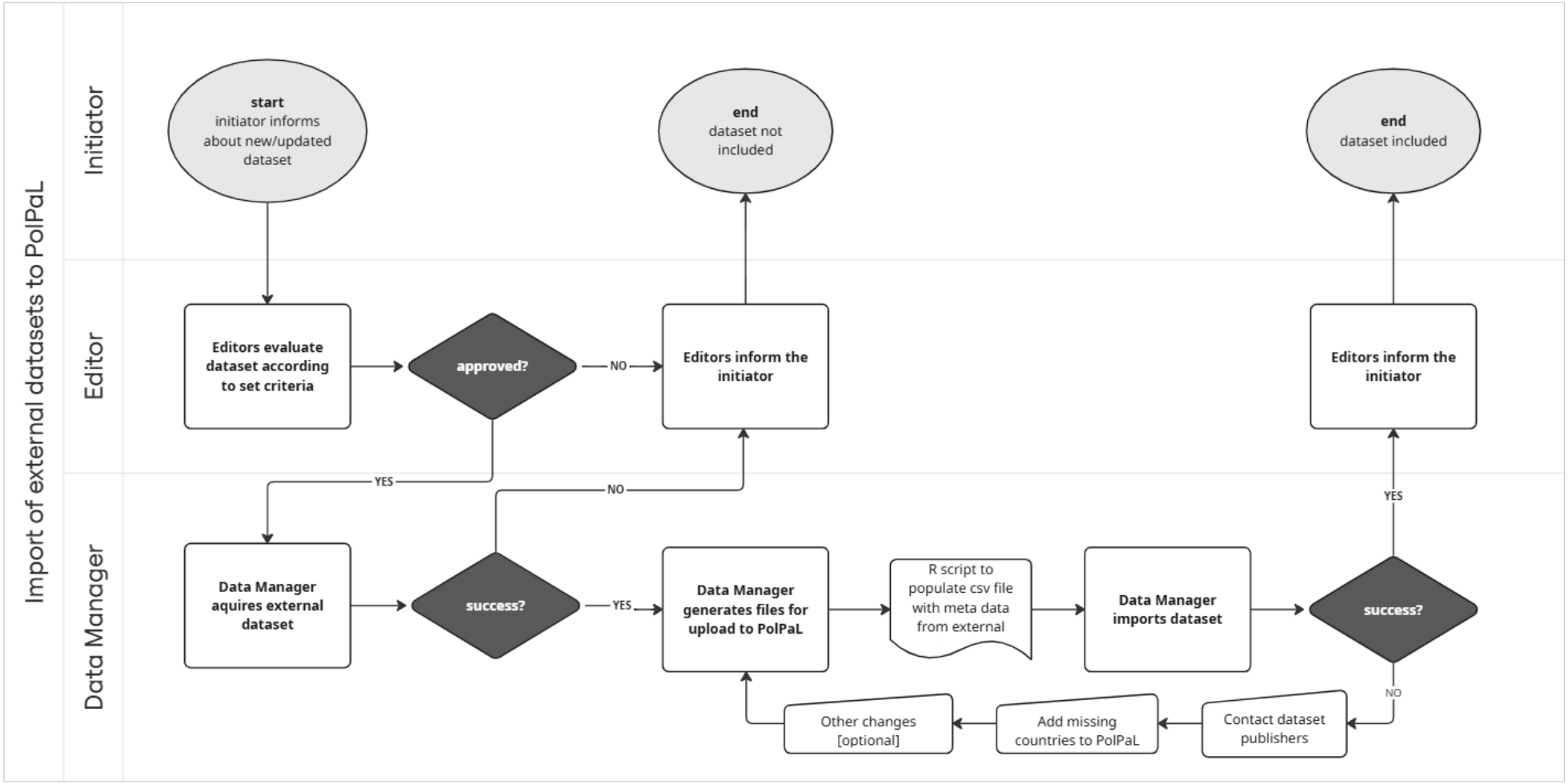
At the beginning of the process, Editors (→ **PolPaL Governance Structure**) are made aware of a new or updated external dataset by external users, by PolPaL team members, or by their own research ("initiators").[1] After an initiator has informed the PolPaL team of a new or updated dataset, the Editors evaluate it against fixed criteria (see Section 2.1 Selection criteria for datasets to be integrated into PolPaL). Should the external dataset not be integrated into the PolPaL database, the initiator will be notified, and the process will end. In case the Editors decide to integrate the external data, a Data Manager takes over.

The Data Manager first acquires (typically downloads) the external data and stores it in the PolPaL OwnCloud folder, a service hosted by the European Open Science Cloud - EU Node. She proceeds by creating the required project, dataset, and version information on the PolPaL website (→ **PolPaL_DataImport**). She subsequently writes the R script for the extraction of the relevant metadata from the external data and generates the csv file containing the relevant metadata to be uploaded. The Data Manager subsequently uploads the R script to the PolPaL GitHub repository and the metadata (csv file) to the PolPaL database through the PolPaL website. A successful data import results in an upload of the metadata (csv file) and the R code used to extract the metadata from the external data. Should the data import be successful, the initiator is informed by one of the Editors, and the importing process ends.

In case the import is not successful, the Data Manager revises her data preparation steps and eventually contacts the Data Publisher for additional information/support. Typical reasons are typos or encoding errors in the external data or the R code written by the Data Manager. Another issue could be that countries are not yet part of the list of PolPaL countries and hence, need to be created first (see Section 3. Generating new PolPaL countries and → **PolPaL_DataImport** for a detailed description of steps taken by the Data Manager).

[1] As of now, PolPaL does not have the resources to systematically and continuously monitor all potential datasets with political party information.

Figure 1: Processes to import external datasets to the PolPaL database



2.1 Selection criteria for datasets to be integrated into PolPaL

To be integrated into PolPaL, external data needs to be

- machine-readable,
- citable (e.g., identifiable author, even if this is an organization), and
- relevant to political scientists in a broader sense.

2.2 Updating metadata on external datasets

In some cases, an update of the metadata describing entities in the external datasets might be necessary. An exemplary case is an issue with the encoding of the metadata, only detected after the metadata has already been integrated into PolPaL for the first time. Please note that an update occurs only when the metadata extracted from the external dataset changes. If instead the external data itself is updated, we create a new version of the dataset and link the data for this new version.

Figure 2: Updating metadata on external datasets in PolPaL

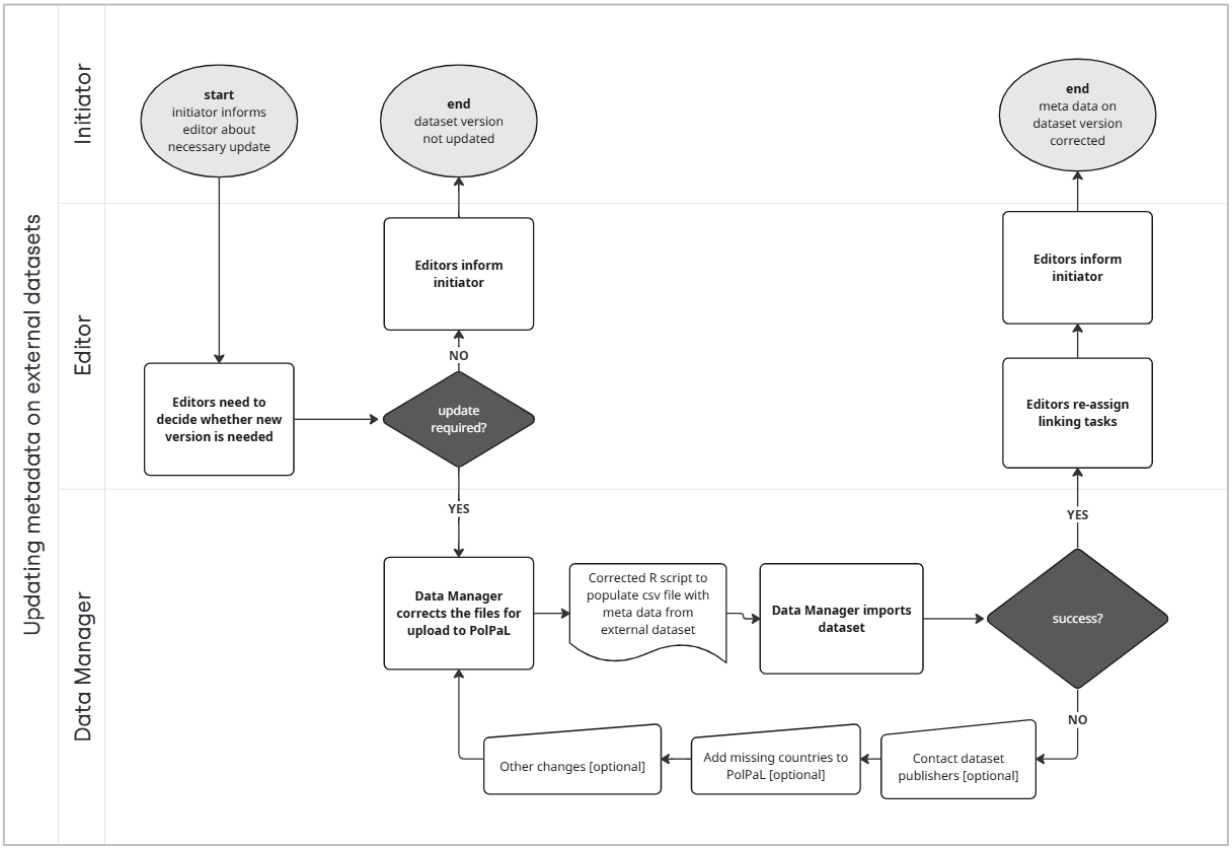


Figure 2 depicts the decision-making process behind this step. First, an initiator notifies the Editors of a reason for why a specific dataset version needs an update. Editors subsequently

discuss the issue among themselves. If they decide that indeed an update is necessary, they inform the Data Manager about the issue and instruct her to generate a new R script and a new metadata csv file for the respective dataset version. The Data Manager subsequently uploads the updated files. Once the version's metadata is updated, Editors re-assign potential linking tasks among themselves or to Contributors.

3. Generating new PolPaL countries

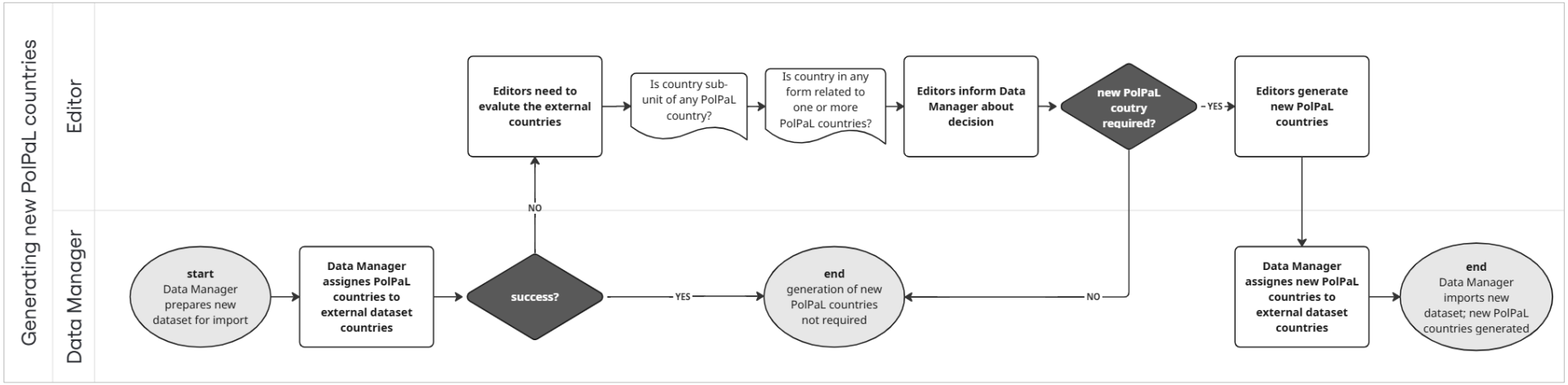
The initial list of PolPaL countries is based on the *countrycode* R package (Arel-Bundock et al., 2018). We extended the list to cover cases we are already aware of, such as the *Soviet Union*. We also extend the list whenever a dataset contains a country that is currently not included in the list of PolPaL countries (→ **PolPaL_DataModel**). New PolPaL countries can be generated by registered users with the assigned roles EDITOR or DATAMANAGER.

We follow the process described in Figure 3 to add a new PolPaL country. An initiator – typically the Data Manager while integrating external data into PolPaL – notifies the Editors of a missing country in the list of PolPaL countries. The Editors subsequently evaluate the country and decide how to proceed, based on the following questions:

- Is the country a geographical/political sub-unit of another PolPaL country?
- Is the country historically related to one or more PolPaL countries?

In order to reach a decision, Editors discuss the case with each other. Due to the complexity of countries' developments, there is no uniformly applicable definition for the creation of a new PolPaL country. If the Editors decide that a new PolPaL country is not required, they inform the Data Manager about their decision. The Data Manager assigns an existing PolPaL country to the country in question in the external dataset and proceeds with importing the external dataset. If the Editors decide that a new PolPaL country is required, they create a new PolPaL country and add an explanatory note to the → **PolPaL_Codebook**. The Editors subsequently inform the Data Manager about the new PolPaL country. The Data Manager assigns the new PolPaL country to the country in question in the external data and proceeds with the import of the external dataset (see Section 2. Importing entries from external datasets to the PolPaL database).

Figure 3. Process to generate new PolPaL countries

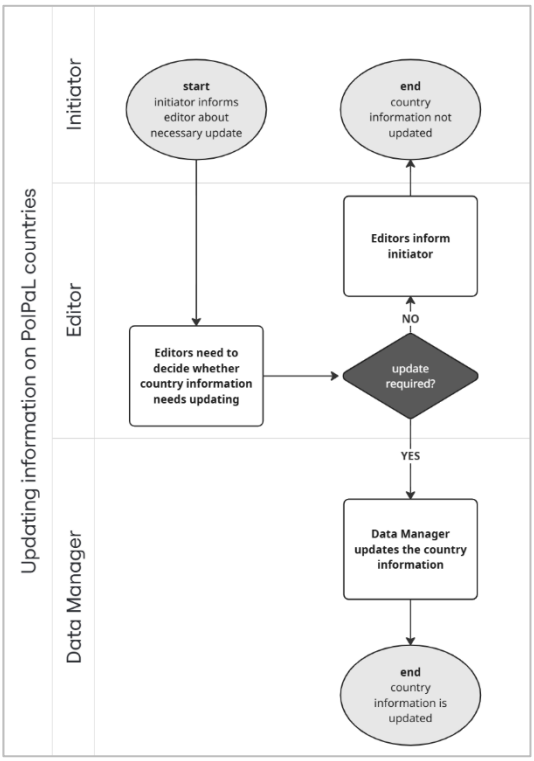


4. Update or delete PolPaL countries

In exceptional cases, it might be necessary to update or delete PolPaL countries. The following subsections explain the underlying decision-making processes.

4.1 Update PolPaL countries

Figure 4: Updating information on PolPaL countries



It is sometimes necessary to update information about a PolPaL country. This could be the case if a country's name changes (e.g., *Macedonia* to *North Macedonia*) or if a correction of the additional information is necessary (ISO codes and the United Nations geoscheme). Please note that the PolPaL countrycode can *not* be updated because it serves as an identifier. PolPaL country information can only be updated by registered users with the assigned role EDITOR or DATAMANAGER.

We follow the process described in Figure 4. First, an initiator informs editors about a necessary change to the information on a country. Editors discuss the case with one another in writing or during a meeting and subsequently inform the initiator of their decision. They also delegate the task of adjusting the information in question to the Data Manager who adjusts the respective country information.

4.2 Delete PolPaL countries

We only delete PolPaL countries in exceptional cases because this process requires an extensive re-import of external datasets previously imported and assigned to the respective country. PolPaL countries are only deleted if there is a significant reason for this, for example, if the same country appears twice in the database with different PolPaL countrycodes. Only registered users with the assigned role DATAMANAGER can delete PolPaL countries.

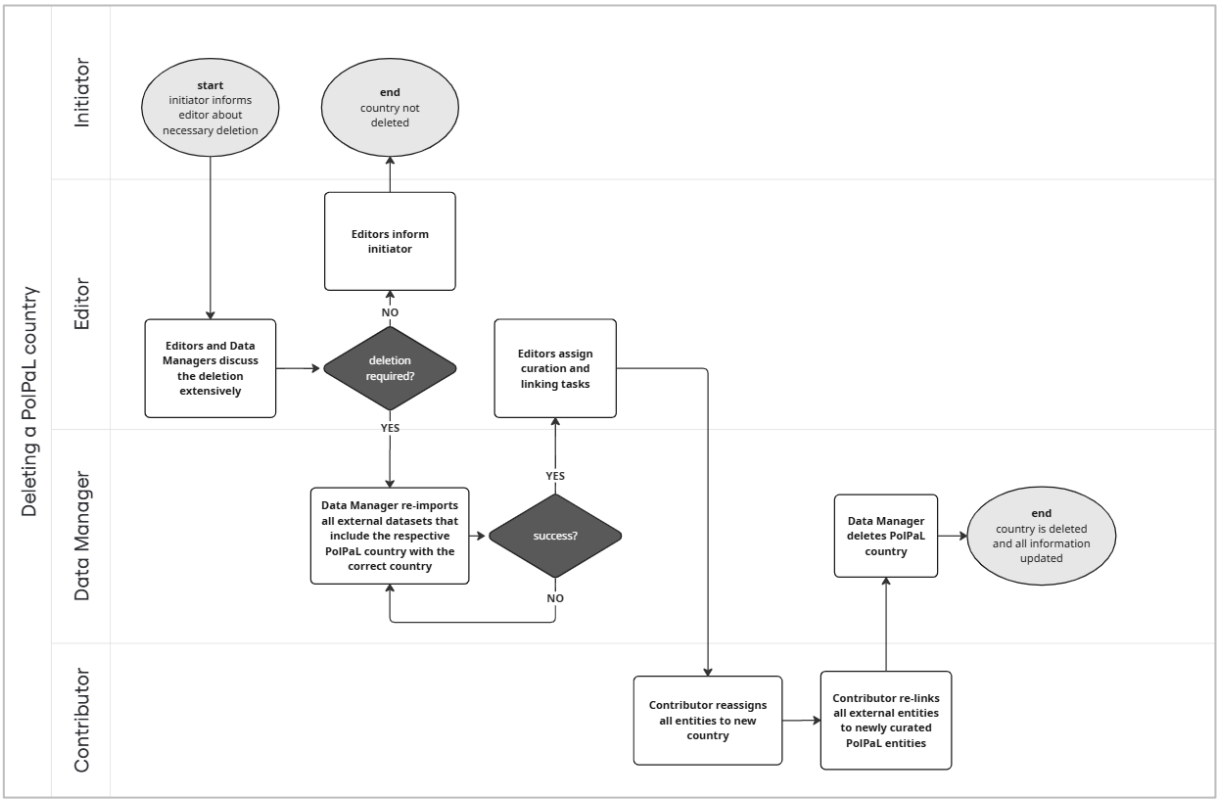
Figure 5 describes the underlying process. Once an initiator has informed the Editors of a necessary deletion of a country, Editors and Data Managers discuss this case extensively. They

decide together whether the PolPaL country needs to be deleted or not. If they believe a deletion is necessary, they initiate the re-import of all external datasets that include the PolPaL respective country. The external datasets need to be re-imported with a correctly assigned PolPaL countrycode, substituting the to-be-deleted PolPaL countrycode.

In addition, the PolPaL entities assigned to the to-be-deleted PolPaL need to be assigned to the correct PolPaL country. In this case, Editors also assign curation tasks to themselves or a Contributor (see Figure 7 and Section 6.1 Curate PolPaL entities). Should it occur that the PolPaL entities are duplicates, users follow the steps described in Section 6.3 Delete or hide PolPaL entities.

It may subsequently be necessary to re-link the external entity entries to their respective PolPaL entities. If this is the case, the Editors assign the linking tasks to themselves or to a Contributor (see Section 7. Linking external entities to PolPaL entities).

Figure 5: Deleting a PolPaL country



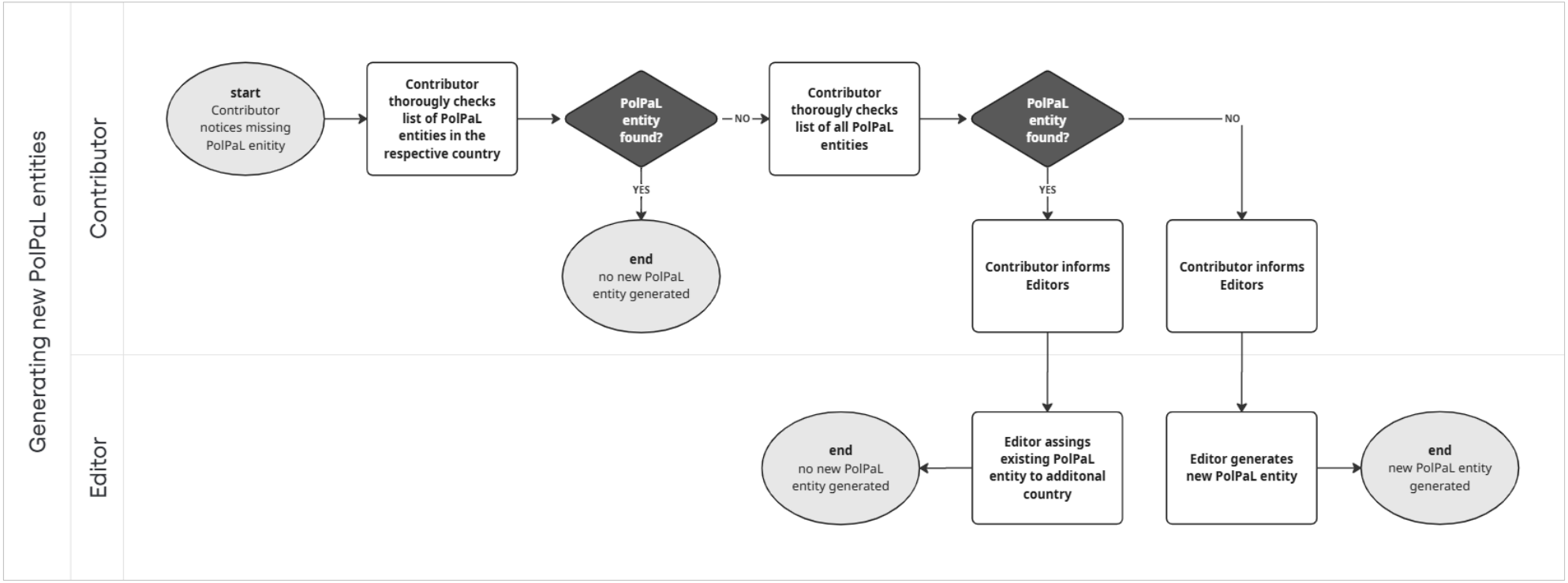
5. Generating new PolPaL entities

External data might contain entities that are not yet included in the list of PolPaL entities. New entities can be generated by registered users with roles EDITOR or DATAMANAGER.

To generate new entities, we follow the process described in Figure 6. If – while linking – a Contributor notices a missing PolPaL entity, she should first thoroughly check the list of existing PolPaL entities in the respective PolPaL country. It might be that the entity is included in the list of PolPaL entities under a different name. Hence, one should check historical names of the PolPaL entities as well as alternative spellings of the "missing" entity. Should no PolPaL entity be found, the next step is to check the list of all PolPaL entities across all PolPaL countries (available via */data/entities*). The "missing" entity might already exist in another PolPaL country. For example, the *Party for Democratic Socialism* (PDS) competed first in the German Democratic Republic and only after reunification in the Federal Republic of Germany.

If a Contributor cannot identify a suitable entity, she contacts an Editor via the PolPaL ticket system, who then creates a new PolPaL entity. Similarly, if a Contributor identifies a relevant entity assigned to a different country, she notifies an Editor via the ticket system, who subsequently assigns the existing PolPaL entity to the additional PolPaL country. The Contributor can afterwards link the entity. In contrast to Contributors, Editors and Data Managers can create new entities and assign entities to new countries by themselves.

Figure 6. Process to generate new PolPaL entities



6. Update or delete PolPaL entities

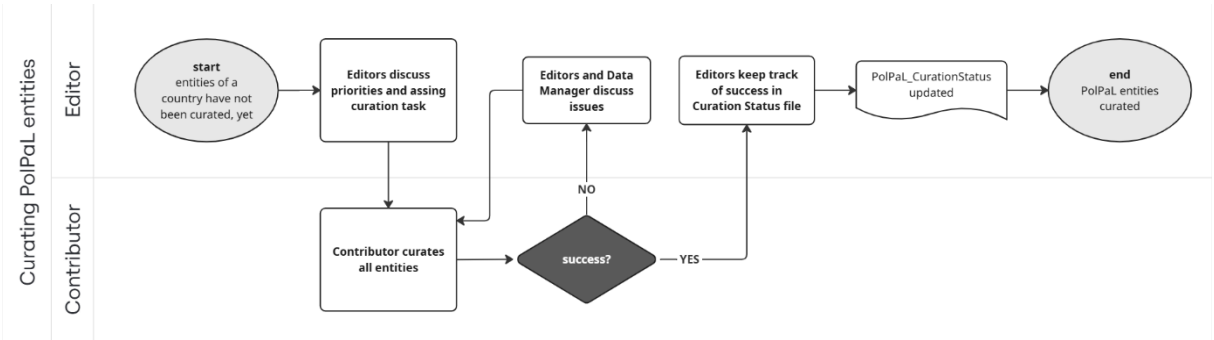
PolPaL entities are curated, updated, deleted, and hidden under specific circumstances on a regular basis. This is to keep the list of PolPaL entities up-to-date and of high quality.

6.1 Curate PolPaL entities

Since PolPaL entities are initially derived from Party Facts entities, we aim to curate all of them before linking entities for the first time. Curation includes updating and adjusting all information about the entity: their name(s), their links to Wikipedia and Wikidata, their type, and their relationships with other entities (predecessors, successors, membership in groupings). Registered users with the roles CONTRIBUTOR, EDITOR, and DATAMANAGER can curate entities.

We coordinate this process as described in Figure 7. Editors keep track of the status quo of curation in the **internal** Excel sheet → **PolPaL_CurationStatus**. Editors discuss priorities with each other, and assign curation tasks by country to themselves or Contributors, agreeing on a timeline for the curation of the respective country's entity. On or before the deadline, the curator notifies the editors of the progress made. Either the curation is noted as finished, or a new deadline is set. The curation process follows the steps described in Sections 6.2 Update PolPaL entities and 6.3 Delete or hide PolPaL entities.

Figure 7: Curating entities



6.2 Update PolPaL entities

PolPaL entities can be updated by registered users with the assigned roles CONTRIBUTER, EDITOR, and DATAMANAGER. These users can update basic information of PolPaL entities at any time. No previous discussion among the Editors is necessary (for a description of quality checks and training procedures, see Sections 10. Quality checks for PolPaL entities and links and 11. Training new team members). The documents → **PolPaL-Tutorial_Entities** and → **PolPaL_Codebook** provide more detailed information.

6.3 Delete or hide PolPaL entities

PolPaL entities can only be deleted or hidden by registered users with the assigned role EDITOR or DATAMANAGER. PolPaL entities are only deleted in a few instances:

- The PolPaL entity was created by mistake.
- The PolPaL entity is a duplicate of another PolPaL entity.
- The PolPaL entity is neither a PolPaL party nor a PolPaL grouping.

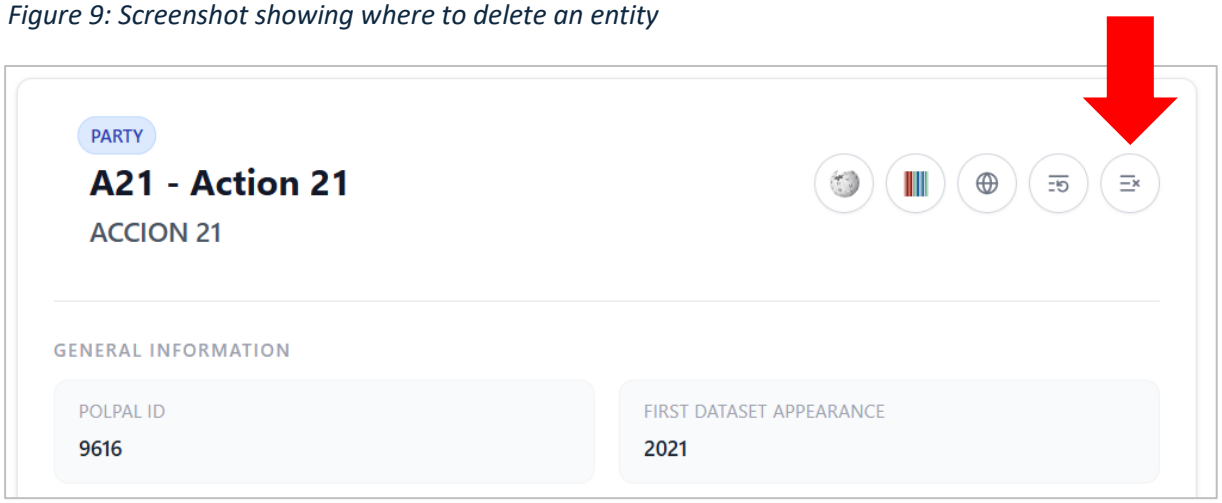
Importantly, PolPaL entities, which are based on the 2023 Party Facts stable release core parties list (Bederke, Döring, Regel, 2023) – those are PolPaL entities with an ID of 9139 and lower – are never deleted.[2] To keep the initially integrated list of Party Facts in the database, we hide these PolPaL entities (see Figure 8) instead of deleting them (see Figure 9). Deleting or hiding entities lies in the discretion of the respective user. No previous discussion among Editors is needed. See the → **PolPaL-Tutorial_Entities** for more detailed information. Registered users with the role CONTRIBUTOR can suggest a deletion by creating a ticket (see Section 10.3 Generating tickets).

[2] A few of these have been deleted by accident.

Figure 8: Screenshot showing where to hide an entity



Figure 9: Screenshot showing where to delete an entity



7. Linking external entities to PolPaL entities

Upon successful import of the metadata, external entities are available on the PolPaL website. Users with the roles EDITOR, DATAMANAGER, or CONTRIBUTOR (→ **PolPaL_DataModel**) can access a list of all external entities (by data version and PolPaL country) with automatically suggested links to PolPaL entities, if applicable, after log-in. The suggested link to a PolPaL entity is based on several factors:

- similarity between external entity entry name(s) and the PolPaL entity names
- similarity between external entity entry name(s) and the additional PolPaL entity names
- temporal overlap of external entity entry and the appearance of the PolPaL entity in other external datasets

Note that PolPaL uses a probabilistic linking tool to make suggestions, but does not set any links automatically.

Figure 11 depicts the decision-making process behind a regular linking task. Editors (→ **PolPaL_GovernanceStructure**) collectively assign linking tasks to each other and to Contributors. Once a linking task is assigned to an Editor or Contributor, she first reviews the external data documentation to determine whether it contains relevant information for easily identifying external entities. She subsequently either confirms the suggested link or sets one or several different links. She can also accept a suggestion for a subset of entities while choosing a different link for another subset of entities. This is only done once by one user. Users can also leave a comment to explain their decision when linking an external entity to one or more PolPaL entities. Figure 10 visualizes the actual linking task.

All links are directly saved in the PolPaL database and can be downloaded nearly immediately afterwards. Editors, Data Managers, and Contributors can easily revise any links previously set at any time by deleting a link and setting a new one. Any registered user who has not set the initial link can also validate the link after its creation (see Section 10.1 Validating existing links and the → **PolPaL-Tutorial_Linking** for further details).

Figure 11: Assignment of linking tasks to PolPaL users

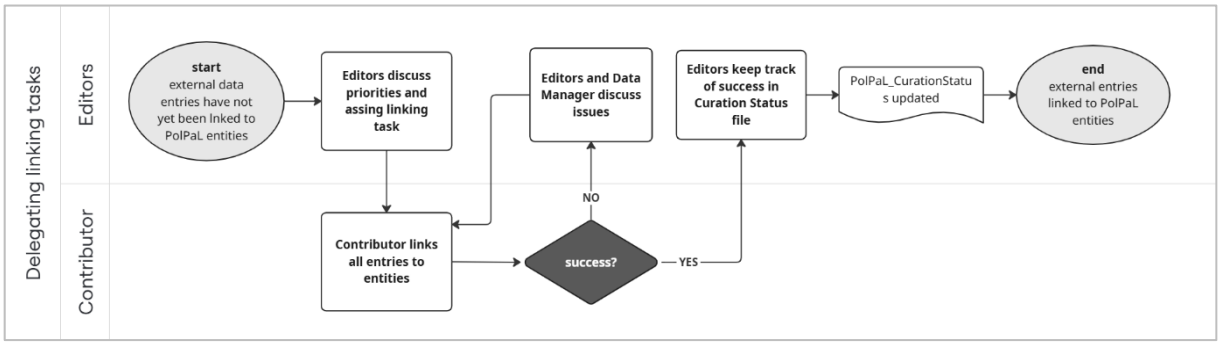
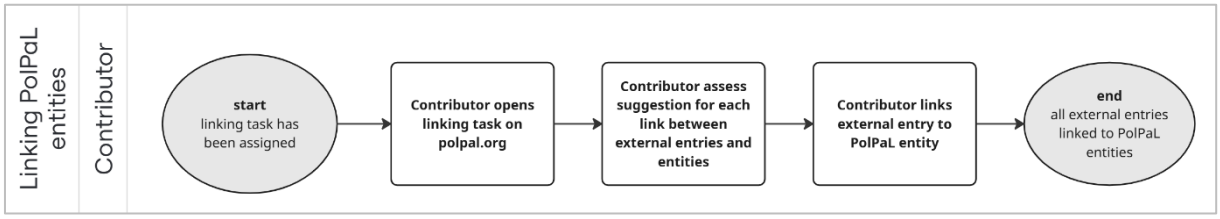


Figure 10: Linking external dataset entries to entities



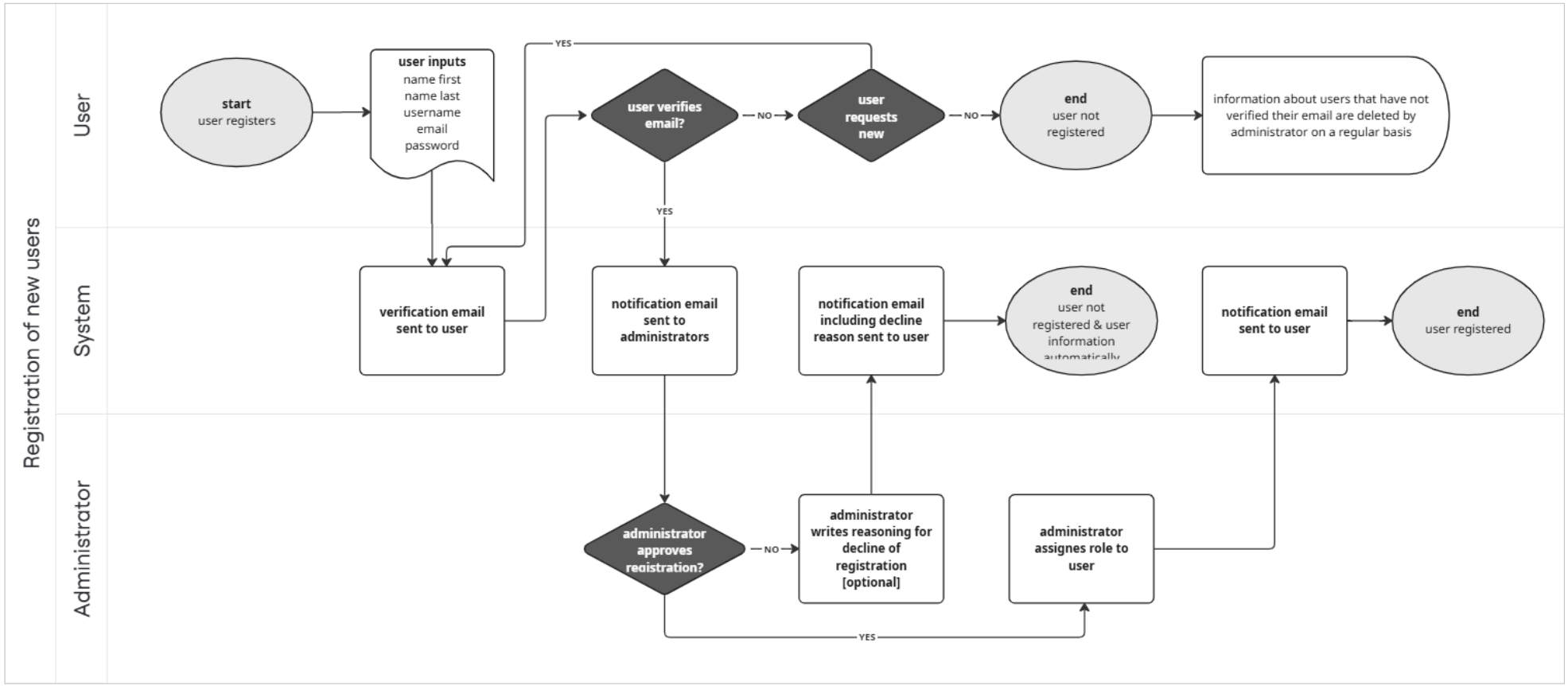
8. Generating new user accounts

Figure 12 displays the process of new users registering with PolPaL. At the start of the process, the new user provides the following information about herself/himself:

- name (first and last name of user, important to identify and judge expertise)
- username (unique username that gets displayed to other users)
- email (valid email address the user has access to for login and resetting password, also for communication)
- password (secure password for login, requires certain schema, gets hashed before stored in database)

Once the user registers, a verification email is sent to the user's email address. This email contains a welcome message as well as a verification link for completing the registration to PolPaL. The user has 24 hours' time to access the link and verify the registration. Should the user not confirm, the verification link expires. The user is required to request a new verification link. A registered user with the assigned roles ADMIN or DATAMANAGER will remove all non-verified user accounts on a regular basis. In case the user verifies her/his email via the link in the email, the field *emailVerified* is set automatically to *true*.

Figure 12: Processes to register new users



Once the user has verified her email, all Admins get notified about a newly registered user. If the Admins do **not** approve the registration, the user gets notified by email about the declined request (and the reason if provided) and the related user information is deleted from the database. In case an Admin approves the request, she also sets a role for the user. The field approvedByAdmin is automatically set to *true*. This ends the registration process; the user is successfully registered and can access the PolPaL database according to the assigned role.

9. Update or delete user accounts

User accounts can, with very few exceptions, only be updated by Admins. Admins can also delete user accounts upon request. We explain the underlying process in greater detail below.

9.1 Update user accounts

Registered users can update their password via their profile page. They receive an email with a link to the password reset page. On the forwarded page, the user can choose a new password. Any other information cannot be updated by registered users. Only Administrators can change user details via the admin interface. Registered users should request this via email.

9.2 Delete user accounts

User accounts can be deleted upon request from the list of registered users. A registered user should send an email; the Admin subsequently deletes the user account via the admin interface as soon as possible. Complete deletion of the user data from the database takes seven (7) days after deletion via the admin interface due to the retention time of security backups.

10. Quality checks for PolPaL entities and links

To ensure high quality of the data generated and provided by PolPaL, we conduct several quality checks at different points in time. These checks are outlined in the following subsections.

10.1 Validating existing links

Figure 13 shows the validation process for previously set links. Editors first select a dataset version and a country for which links shall be validated. Based on the Editors' and Contributors' country expertise, they assign the validation task to a specific Editor or Contributor, agreeing on a timeline for the validation. These decisions are tracked in the **internal** Excel document → **PolPaL_CurationStatus**. Before or on the deadline, the validator communicates the progress of the validation, and the validation is either marked as completed in the **internal → PolPaL_CurationStatus**, or a new deadline is agreed upon.

While validating, the respective Editor or Contributor goes through each link for each external observation for the country and the dataset version. If the Editor or Contributor would have made the same linking decision, she validates the link by clicking on the thumbs up. If the Editor or Contributor would have taken a different decision, she flags the linking decision by clicking on the thumbs down symbol (see also the → **PolPaL-Tutorial_Validating**).

This "thumbs down" button triggers a warning to PolPaL Editors. PolPaL Editors, on a regular basis during their meetings, discuss these warnings and decide whether the link needs to be deleted and corrected or is valid. If linking decisions need to be revisited, the Editors delegate this task to one among themselves or to another Contributor.

10.2 Validating curated entities

Figure 13: Validating links between external entries and PolPaL entities

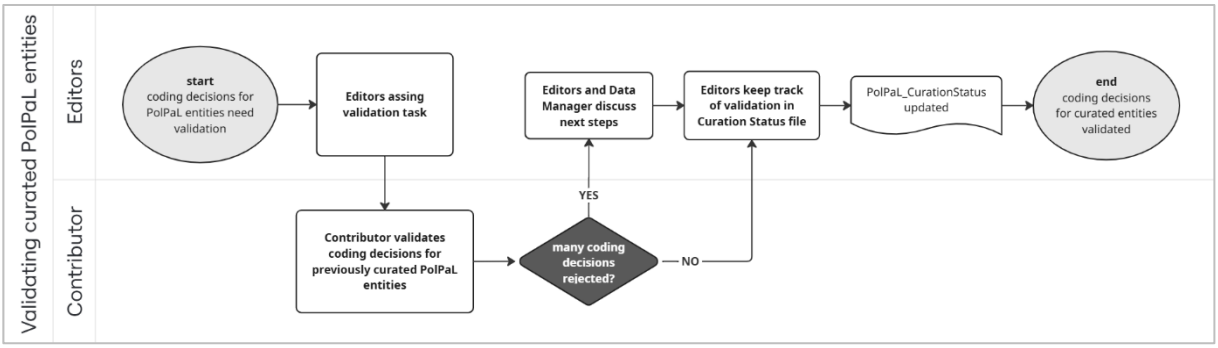
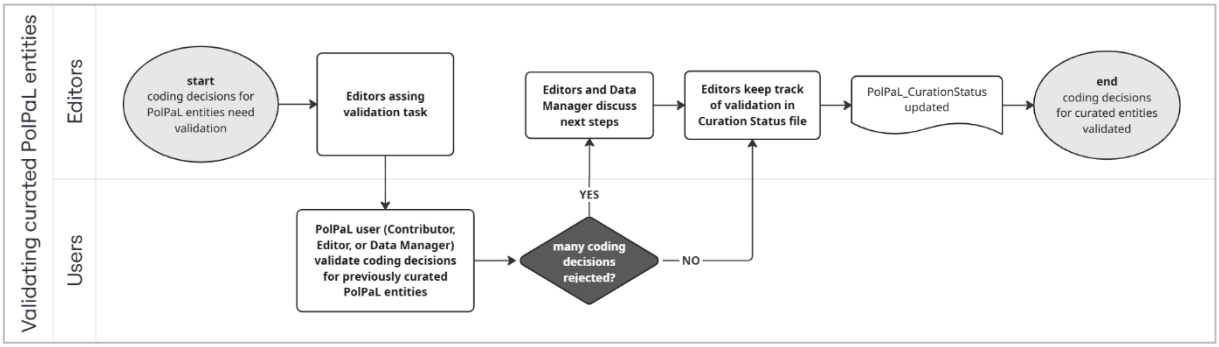


Figure 14 shows the validation procedure for curated entities. Editors first select a country whose entities need further validation. They subsequently select some of the country's entities *randomly* and assign the validation task to an Editor or Contributor based on the person's country knowledge and time. Together with the validator, the Editor sets a timeline for the validation. Before or on the deadline, the validator communicates the progress she has made. Either, the validation task is marked as completed, or a new timeline is agreed upon. Please note that we only assign validation tasks to experienced Contributors. Decisions

regarding the country assignment and the timeline are tracked in the Excel sheet → **PolPaL_CurationStatus**.

While validating curated entities, the respective Editor or Contributor checks the following information: the names of the selected entities and the decisions made regarding the entity types. The validator tracks how many of these data points she would have coded differently *at the point of the initial curation* (note that a party might have changed its name afterwards) and *without actually adjusting the information*, yet. She reports the numbers to the Editors for documentation in the PolPaL quality checks. Editors decide whether the validator should adjust the identified issues or whether the initial decision is more valid.

Figure 14: Validating curated PolPaL entities



10.3 Generating tickets

Registered users (all roles) can create tickets in PolPaL's internal ticket system. The ticket system is intended to allow logged-in users to flag any issues, inconsistencies, and questions within the PolPaL website. All tickets can be viewed by users with the assigned roles ADMIN, DATAMANAGER, or EDITOR. These users are also responsible for processing outstanding tickets.

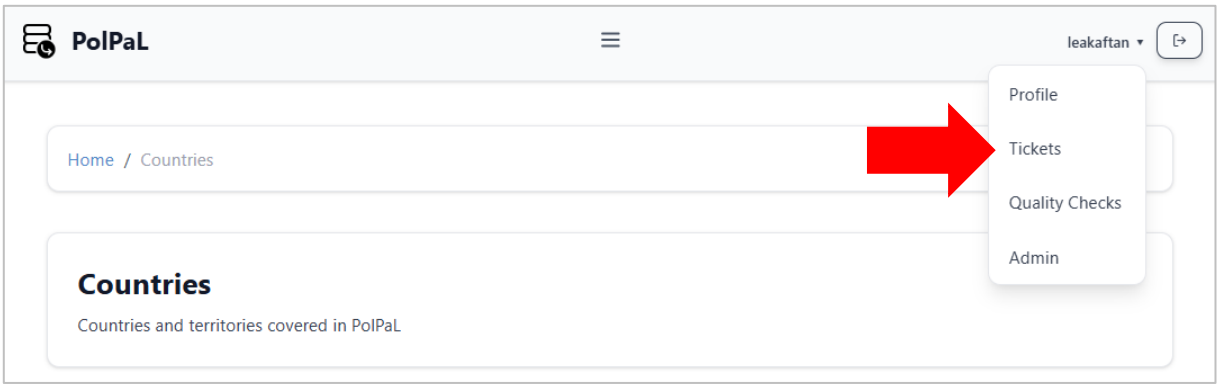
The ticket system is intended for the following cases, but not limited to those:

- Request revision of PolPaL entities (create/update/delete)
- Request assignment of PolPaL entity to another country
- Request clarification on external dataset entries and links

To create a new ticket, the logged-in user navigates to the tickets page on the PolPaL website. On that page, one can review open tickets and create new ones. Registered users with the assigned roles USER and CONTRIBUTOR can only review their own tickets. A title and

a description are required for the ticket. If feasible, a direct link to the relevant page can be added.

Figure 15: Screenshot showing where to find the ticket system



10.4 Processing tickets

The processing of tickets is the responsibility of Editors, Data Managers, and Admins. They can change the status of tickets between 'open', 'in progress', 'resolved', and 'closed'. Registered users with the assigned roles USER and CONTRIBUTOR can only review and delete their own tickets.

Usually, the processing of tickets should not require deliberation between Editors, Data Managers, and Admins. However, they are encouraged to deliberate tickets if deemed necessary. In cases where more information on the open ticket is required, the user who created the ticket might be contacted via email. Contact is established only by Admins; email addresses are not accessible to any other user group.

10.5 Automated quality checks using R scripts

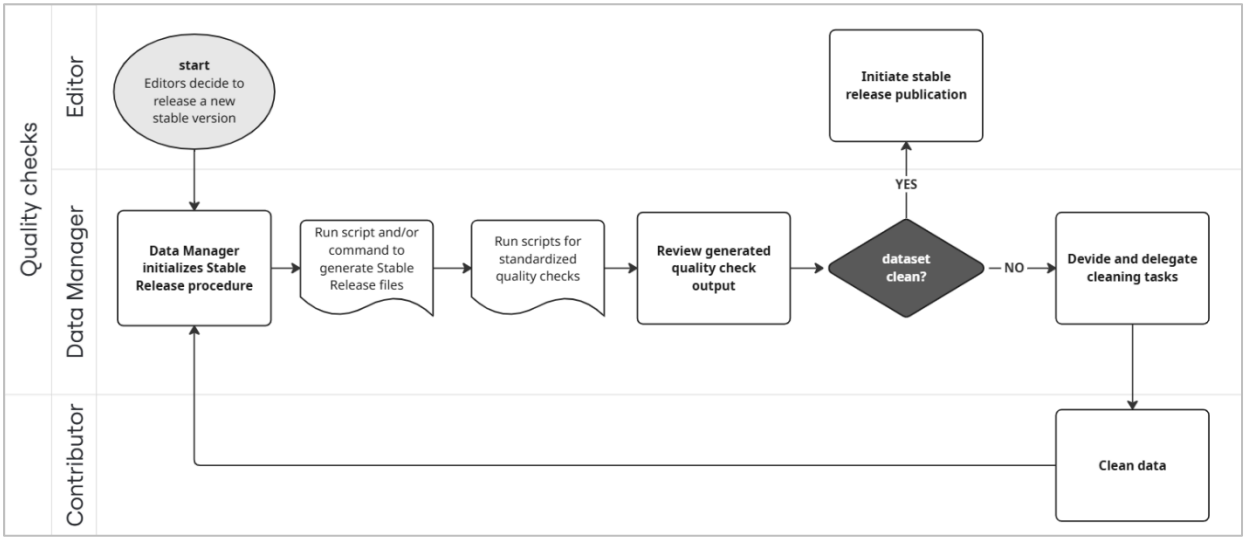
Before generating and publishing a new stable release (see Section 12. Publication of data-related stable releases), the Data Manager conducts further quality checks. These control whether

- any dates are out of range,
- citations follow the agreed upon citation style,
- there are logical inconsistencies in the assignments of members to groupings,
- there are mistakes in the assignment of external entities to the PolPaL entities unknown, others, and residuals,

- external IDs are consistently assigned to PolPaL IDs across datasets within projects, e.g., whether any entity ID from V-Dem and V-Party is assigned to the same PolPaL entity, independent of the dataset,
- if links to Party Facts entities are missing; we aim at full coverage here to ensure comparability and compatibility.
- Further quality checks might be agreed upon by the Editor in the future.

Figure 16 visualizes the underlying process. When Editors initiate a new stable release, they take the time needed for quality checks into account and instruct the Data Manager(s) to initiate these quality checks. Data Managers can divide the tasks among themselves. They run the scripts on the data available for download on the website and on the data that would be integrated into the stable release (see Section 12. Publication of data-related stable releases), keeping track of any issues on the observation level. They report the test results to the Editors, who decide whether adjustments are needed before generating a new stable release. If adjustments are necessary, they assign specific tasks to Editors or Contributors and agree upon a timeline. Once the adjustments are implemented, the Data Manager(s) run the respective quality checks again, reporting the results again back to the Editors. This process continues until the quality of the data to be integrated into the stable release is deemed of sufficient quality by the Editors.

Figure 16: Standardized quality checks



11. Training new team members

Onboarding is an important process to guarantee the high quality of user-generated data. Depending on the access rights and responsibilities a user receives, she needs to be trained in different ways. In the following subsections, we explain the onboarding process for the different user categories. We also explain who makes decisions regarding user rights and responsibilities. See the → **PolPaL_Codebook,** the → **PolPaL_DataModel-extendedVersion** and the → **PolPaL_GovernanceStructure** for further details on the rights and responsibilities of the different user categories.

11.1 Becoming a User

Following the process described in Section 8. Generating new user accounts, people can register as users on the PolPaL website. If they only need the rights and responsibilities of a USER, no training is conducted.

11.2 Training new Contributors

Before a user can do the tasks of a Contributor, she needs thorough onboarding for entity curation and entity linking tasks. Figure 17 describes the process that an Editor should follow when onboarding a new Contributor. The Editor selects a country for the Contributor to work on first, while, if possible, taking the preferences of the Contributor into account. The Editor subsequently explains the entity curation process (see Section 6.1 Curate PolPaL entities) and the respective → **PolPaL-Tutorial_Entities**. The Contributor then curates 10 entities of the selected country, which are discussed in-depth with the Editor before proceeding with curating the remaining entities. The Contributor should reach out to the Editor or any other PolPaL team member when she has questions. Once all entities are curated, the Contributor can be onboarded for linking tasks.

Figure 17: Onboarding Contributors in PolPaL

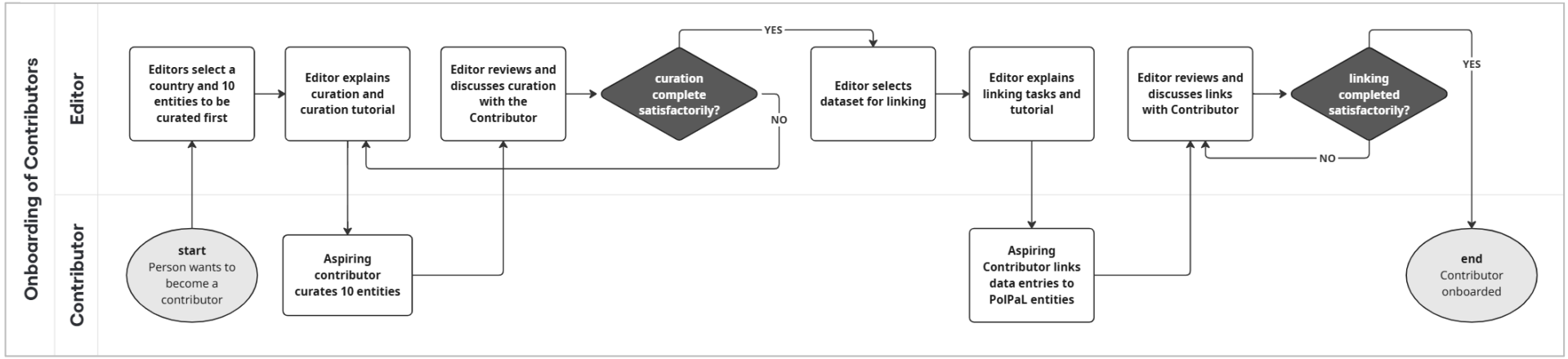


Figure 17 also explains the onboarding process for Contributors before they fulfil their first linking tasks. Similar to the process above, the Editor selects a dataset version to be linked by the Contributor first, if possible, taking the Contributor's preferences into account. The Editor subsequently explains the linking procedures and introduces the respective → **PolPaL-Tutorial_Linking.** After having read the material, the Contributor links the entities for the selected dataset version and country (the country the Contributor has curated before) while discussing potentially difficult cases with the Editor. The Editor controls the linking decisions made by the Contributor on the first dataset version before the Contributor is assigned new dataset versions to link. Once all dataset versions are linked for data on a respective country, the Contributor can be assigned other countries.

In a few exceptions, Contributors are allowed to link data without previously curating entities. They should, however, be onboarded before linking for the first time, as laid out in Figure 17. These are the exceptional cases in which a Contributor is allowed to link data without curating entities for the respective country, first:

- The Contributor is an expert on a specific data project, dataset, or dataset version and only wants to link this to the PolPaL entities.
- Entities of the countries that the Contributor wants to curate are already curated. In this case, the Contributor could validate the existing curation (see Section 10.2 Validating curated entities).

11.3 Training new Editors

Since Editors have more rights and responsibilities, they need to be better familiar with the PolPaL database than Contributors. Contributors can become Editors only if they have gained sufficient experience with curating *and* linking data for PolPaL. "Enough" means that the respective Contributor has curated entities for at least one country and linked the respective entities for at least 5 versions of *different* datasets. Editors can subsequently decide to promote a Contributor to an Editor. If they do so, the Contributor receives an introduction to the additional rights and responsibilities.

11.4 Training new Data Managers

Before a user can perform the tasks of a Data Manager, she needs thorough onboarding for the data ingestion task. The Editor first selects an external dataset that has already been ingested into PolPaL for training purposes. The Data Manager generates the respective

metadata in an Excel sheet (the metadata that one would have to generate within PolPaL before uploading the metadata, see Section 2. Importing entries from external datasets to the PolPaL database). The Data Manager also writes the R code to extract metadata from the external data, as if she were ingesting it. Once finished, the Editor and the aspiring Data Manager discuss the work. If the Data Manager has no further questions and understands the data ingestion process, the Editor assigns a dataset version to be ingested by the newly instated Data Manager.

11.5 Becoming a website Admin

The Board of Directors has full discretion over those who receive admin rights on the PolPaL website. These people should have been Editors beforehand, and they should be willing to dedicate part of their work time to PolPaL in a reliable manner. The Board of Directors, with a unanimous decision, can withdraw admin rights (→ **PoLPaL_GovernanceStructure**).

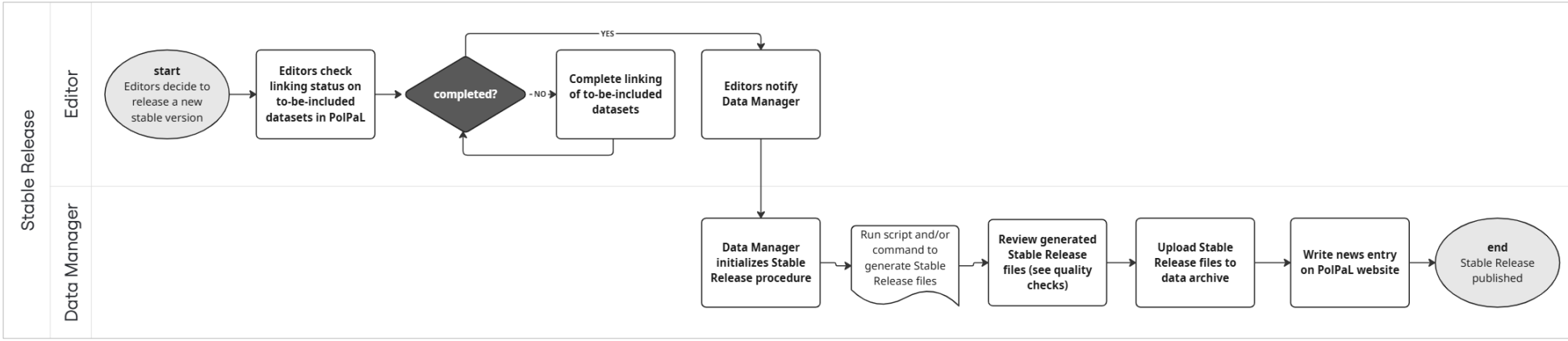
12. Publication of data-related stable releases

Figure 18 shows the process of publishing a stable release and a related news entry. A new stable release is published at least annually, and more often if deemed necessary. The Editors initiate a new stable release. The Editors also decide which new external dataset versions, in addition to the previously included ones, are included in the stable release. The included dataset versions need to meet the following criteria:

- The data version needs to be already public on the PolPaL website.
- For each country, at least 90% of external entity entries are linked to PolPaL entities.
- Links must have been created for all countries in an external dataset.

The Editors subsequently instruct the Data Manager to prepare the stable release files. This is done via standardized R scripts, which are available in the PolPaL GitHub repository. After generating the stable release files, the Data Manager reviews the files (see Section 10.5 Automated quality checks using R scripts). The stable release files get uploaded by the Data Manager to the PolPaL data archive [ARCHIVE(S) TO BE DECIDED]. At the end, the Data Manager writes a news entry on the PolPaL website announcing the stable release (see Section 14. Publication of news), providing information about the differences between the new and the previous stable release.

Figure 18: Processes to publish stable release and news entry



12.1 Content of the stable releases

All stable releases shall include the following files:

- Anonymized PostgreSQL database
- PolPaL countries
- PolPaL entities
- PolPaL party changes
- PolPaL grouping memberships
- PolPaL additional names
- External data version metadata
- External entities metadata

13. Publishing a software-related stable releases

Software-related stable releases concern the public release of the underlying software code of PolPaL. The software code is made openly accessible for other scholars with an interest to adopt the PolPaL linkage key database structure to their own linkage key projects or to run a localized version of PolPaL.

The most current software code of PolPaL will be publicly available via the PolPaL GitHub Repository, the stable release of the software code will be archived in the selected long-term archive [ARCHIVE TO BE DECIDED]. An initial stable release of the PolPaL software code is published along the full launch of PolPaL in November 2026. Afterwards, stable releases of the software code are only published after major changes to the software code. Major changes could be the implementation of new features to PolPaL (e.g., adding politicians to PolPaL) or significant re-work of the existing code (e.g., changing the technical stack of PolPaL). This is expected to happen in connection to secured funding for new extensions to PolPaL. Updates to the current software stack (e.g., latest version of node.js) and security patches are not considered qualifying for a new stable release of the PolPaL software code. The latest software code is always available at the PolPaL GitHub repository.

The publication of a software-related stable release will be done by the Technical Operator in coordination with the Administrator(s). In general, changes to the software code are first tested locally before deployed to the production website. Hence, the software code should always be working fine. For software-related stable releases, the technical operator tests the

software code thoroughly and ensures, that all the relevant metadata and versioning is updated. In coordination with the Administrator(s), the software code will then be archived.

14. Publication of news

Registered users with the roles EDITOR or DATAMANAGER can generate new news. Several events can trigger a news publication:

- New data version completely linked and/or set public
- New stable release published
- Major events (conferences, presentations, meetings etc.)
- Major website-related updates (new features/tools, security updates, etc.)

For these events, Editors or Data Managers can publish news on the PolPaL website without further discussion among editors. Other events, such as the publication of research using PolPaL, will need agreement among Editors before related news is published on the PolPaL website.

15. Security updates of the PolPaL software code

In order to ensure the security of the personal data stored in the PolPaL database, it is required to perform security updates of the PolPaL software code. The security updates are performed by the Technical Operator.

Security updates are performed if:

- Relevant vulnerabilities are detected (e.g., GitHub Dependabot, BSI Report, etc.) and deemed relevant to the PolPaL software.
- Major updates within the tech stack are released (e.g., Node.js releases the new active LTS version *Node.js 26* in November 2026).

Once relevant vulnerabilities are detected or a major release within the tech stack are released, the Technical Operator starts updating the PolPaL software code accordingly. Once the update of the software code is done, the Technical Operator tests the functionality of PolPaL within a localized version of PolPaL. Only if the localized version of PolPaL runs without issues and the code linter (automated process of checking source code for programming errors, bugs, stylistic mistakes) plus all required tests are passing, the updated PolPaL software code is pushed into the production environment.

16. References

Arel-Bundock, V., Enevoldsen, N., & Yetman, C. (2018). countrycode: An R package to convert country names and country codes. The Journal of Open Source Software, 3(28), 848. https://doi.org/10.21105/joss.00848

Bederke, P., Döring, H., & Regel, S. (2023). Party Facts – Version 2023 [Dataset]. In Harvard Dataverse. https://doi.org/10.7910/dvn/tjinlq

Rovny, J., Polk, J., Bakker, R., Hooghe, L., Jolly, S., Marks, G., Steenbergen, M., & Vachudova, M. A. (2025). The 2024 Chapel Hill Expert Survey on political party positioning in Europe: Twenty-five years of party positional data. Electoral Studies, 97, 102981. https://doi.org/10.1016/j.electstud.2025.102981

17. Version history

- Adding Section 13 - Publishing a software-related stable releases
- Adding descriptions and visualizations of processes previously undefined:
    - Updating metadata on external datasets
    - Curating PolPaL entities
    - Linking external entities to PolPaL entities
    - Quality checks for PolPaL entities and links
    - Training new team members
    - Publication of software-related stable releases
- Adjusting all previously described processes to provide more detail.
- Changing the order in which processes are described in this document
- Language editing
- Increasing the size of Figures and adding new Figures describing the processes
- Adding a note that this document is likely subject to change in the near future, depending on pending decisions regarding the governance structure.