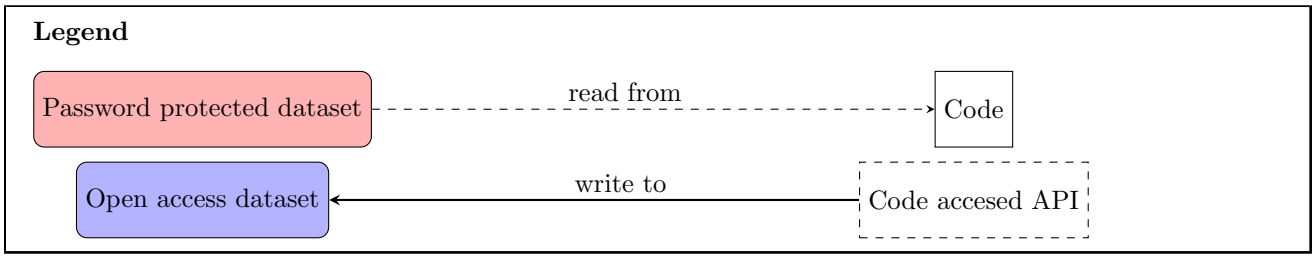
# Replication package: Data and code diagram

Francesco Bailo

February 16, 2017

Bailo, F. (2017). *The citizen-user and the crowd-mediated politics of the Five Star Movement* (Unpublished doctoral thesis). The University of Sydney.

## Diagram conventions




## Contents

**1 Data collection** **1**
1.1 Beppe Grillo's blog . . . 1
1.2 Five Star Movement (M5S) Forum . . . 2
1.3 M5S Facebook pages . . . 3
1.3.1 Beppe Grillo's Facebook page . . . 3
1.3.2 M5S' Facebook page . . . 4
1.3.3 M5S Public figures' Facebook page . . . 5
1.4 M5S meetups . . . 7
1.5 Change.org . . . 8
1.6 la Repubblica and Corriere della Sera . . . 9
1.7 Camera dei Deputati and Senato della Repubblica . . . 10
1.8 Pirate Party Forum . . . 12

**2 Data analysis** **13**
2.1 Beppe Grillo's blog . . . 13
2.2 Facebook . . . 13
2.3 Meetup . . . 14
2.4 Data integration . . . 15

# 1 Data collection

## 1.1 Beppe Grillo's blog

**url** `http://beppegrillo.it/`

**method** HTML data extraction

**period covered**

**from** 2005-01-16 16:52:00 Central European Time (CET)

**to** 2015-05-03 13:15:00 CET

**period of collection**

**from** 2014-08-05 10:07:53 Australian Eastern Standard Time (AEST)

**to** 2015-05-05 02:07:17 AEST

**data file** m5s_blog_mar2015.sqlite

**data schema**

**table** comment

**fields** comment_id, post_url, text, username, created, replied_to, timestamp

**records** 5,104,431

**table** index

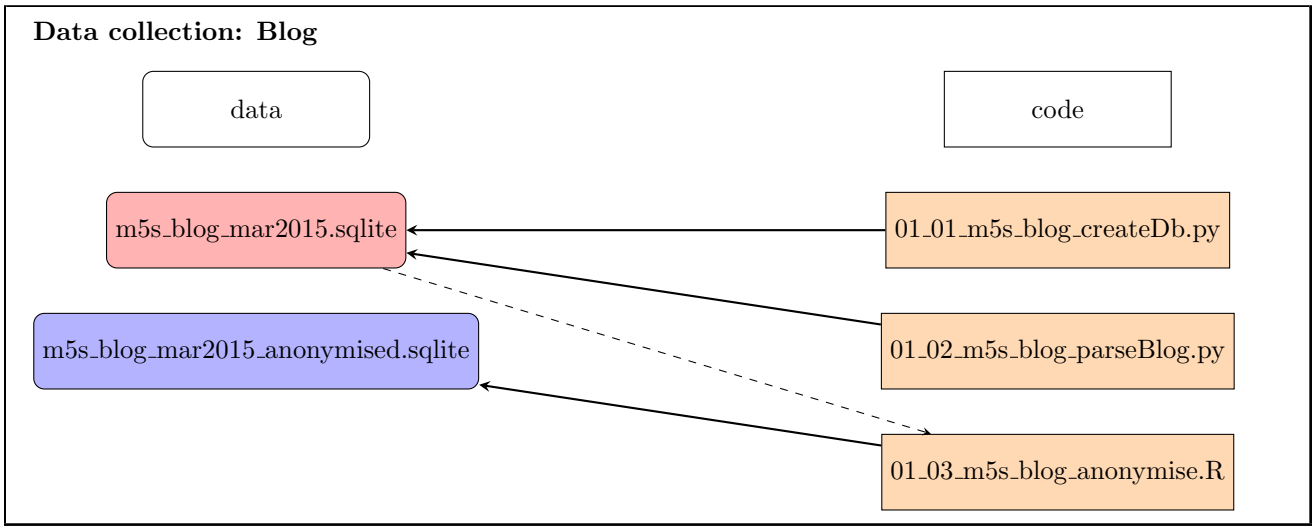
**fields** url, timestamp, status

**records** 9,172

**table** post

**fields** url, title, text, date, timestamp

**records** 9,138




## 1.2 M5S Forum

**url** `http://www.beppegrillo.it/listeciviche/forum`

**method** HTML data extraction, Application Programming Interface (API) requests

**period covered**

**from** 2009-02-11 17:53:57 CET

**to** 2015-04-08 12:07:00 CET

**period of collection**

**from** 2013-05-16 16:00:21 AEST

**to** 2015-04-09 11:30:37 AEST

**data file** m5s_forum_apr2015.sqlite

**data schema**

**table** author

**fields** authorId, name, url, timestamp

**records** 35,591

**table** comment

**fields** commentId, threadId, forum, parent, authorUrl, dislikes, rawMessage, createdAt, numReports, likes, message, timestamp

**records** 821,037

**table** commentAuthor

**fields** authorUrl, name, isAnonymous, timestamp

**records** 129,569

**table** crawledUrl

**fields** url, timestamp, codeThread, codeComments

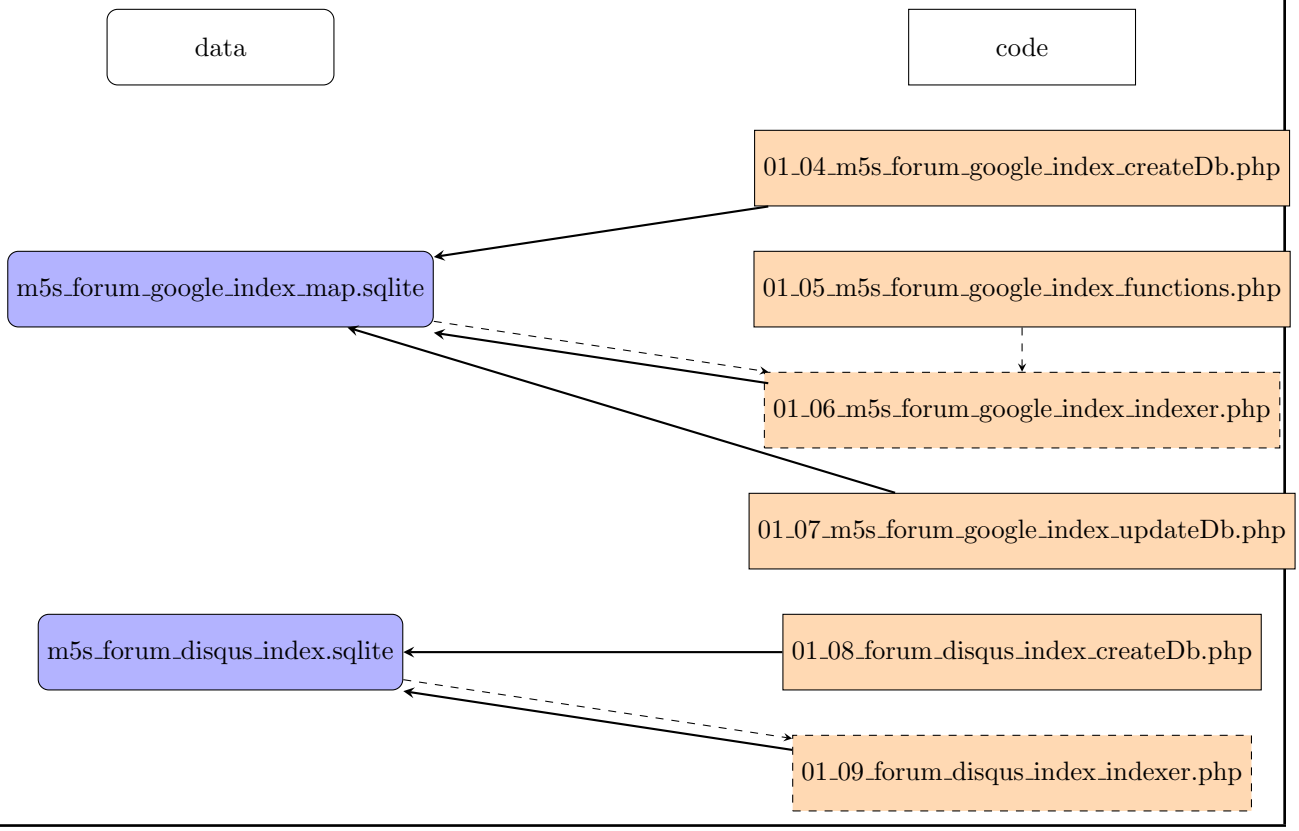
**records** 99,703

**table** thread

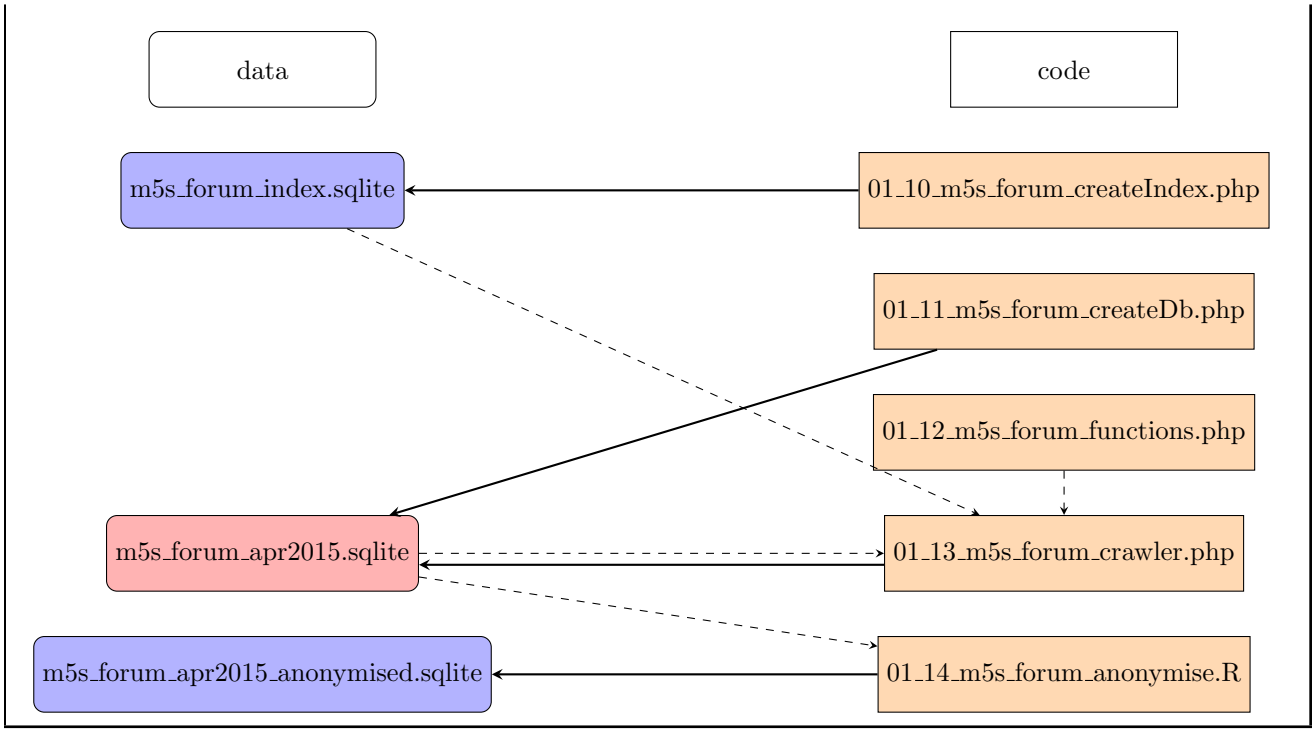
**fields** threadId, authorUrl, title, reactions, dislikes, userScore, createdAt, slug, postNumber, link, likes, message, category, score, categoryLink, timestamp

**records** 129,032

**Data collection: Forum's pages indexing with Google and Disqus APIs**




**Data collection: Forum scraping**




## 1.3 M5S Facebook pages

### 1.3.1 Beppe Grillo's Facebook page

**url** `https://www.facebook.com/beppegrillo.it/`

**method** API requests

**period covered**

**from** 2008-12-09 11:46:02 Coordinated Universal Time (UTC)

**to** 2015-01-22 21:30:00 UTC

**period of collection**

**from** 2015-01-20 21:43:53 AEST

**to** 2015-05-03 22:01:55 AEST

**data file** m5s_fb_beppegrilloit_jan2015.sqlite

**data schema**

**table** application

**fields** id, name, namespace, timestamp

**records** 13

**table** comment

**fields** id, attachment_url, can_comment, comment_count, created_time, from, message, like_count, can_remove, user_likes, post_id, parent, timestamp

**records** 5,969,620

**table** log

**fields** id, timestamp, request_string, response, error_message, error_type, error_code

**records** 2,378,353

**table** message_tags

**fields** id, post_id, profile_id, timstamp

**records** 0

**table** post

**fields** id, created_time, application, description, from, link, name, picture, status_type, story, type, updated_time, caption, icon, message, object_id, place, source, timestamp

**records** 21,041

**table** post_likes
**fields** post_id, profile_id, timestamp
**records** 51,646,517
**table** profile
**fields** id, name, category, type, timestamp
**records** 2,113,919
**table** sqlite_sequence
**fields** name, seq
**records** 2
**table** story_tags
**fields** id, post_id, profile_id, timestamp
**records** 0
**table** to
**fields** id, post_id, profile_id, timestamp
**records** 47
**table** with_tags
**fields** id, post_id, profile_id, timestamp
**records** 0

#### 1.3.2 M5S' Facebook page

**url** `https://www.facebook.com/movimentocinquestelle/`

**method** API requests

**period covered**

**from** 2009-10-05 15:24:02 UTC

**to** 2015-01-19 16:01:39 UTC

**period of collection**

**from** 2015-01-19 21:00:20 AEST

**to** 2015-01-23 03:42:15 AEST

**data file** m5s_fb_movimentocinquestelle_jan2015.sqlite

**data schema**

**table** application
**fields** id, name, namespace, timestamp
**records** 9
**table** comment
**fields** id, attachment_url, can_comment, comment_count, created_time, from, message, like_count, can_remove, user_likes, post_id, parent, timestamp
**records** 649,298
**table** log
**fields** id, timestamp, request_string, response, error_message, error_type, error_code
**records** 192,057
**table** message_tags
**fields** id, post_id, profile_id, timstamp
**records** 0
**table** post
**fields** id, created_time, application, description, from, link, name, picture, status_type, story, type, updated_time, caption, icon, message, object_id, place, source, timestamp
**records** 15,084
**table** post_likes
**fields** post_id, profile_id, timestamp
**records** 4,546,176
**table** profile

**fields** id, name, category, type, timestamp
**records** 344,302

**table** sqlite_sequence
**fields** name, seq
**records** 1

**table** story_tags
**fields** id, post_id, profile_id, timestamp
**records** 0

**table** to
**fields** id, post_id, profile_id, timestamp
**records** 0

**table** with_tags
**fields** id, post_id, profile_id, timestamp
**records** 0

#### 1.3.3 M5S Public figures' Facebook page

**method** API requests

**period covered**

**from** 2009-10-21 15:05:14 UTC

**to** 2015-03-30 20:46:46 UTC

**period of collection**

**from** 2015-03-11 03:59:19 AEST

**to** 2015-03-30 22:06:34 AEST

**data file** m5s_fb_public_figures_mar2015.sqlite

**data schema**

**table** application
**fields** id, name, namespace, timestamp
**records** 106

**table** comment
**fields** id, attachment_url, can_comment, comment_count, created_time, from, message, like_count, can_remove, user_likes, post_id, parent, timestamp
**records** 2,653,715

**table** log
**fields** id, timestamp, request_string, response, error_message, error_type, error_code
**records** 1,483,122

**table** message_tags
**fields** id, post_id, profile_id, timstamp
**records** 0

**table** post
**fields** id, created_time, application, description, from, link, name, picture, status_type, story, type, updated_time, caption, icon, message, object_id, place, source, page_id, timestamp
**records** 558,687

**table** post_likes
**fields** post_id, profile_id, timestamp
**records** 24,446,197

**table** profile
**fields** id, name, category, type, timestamp
**records** 1,513,048

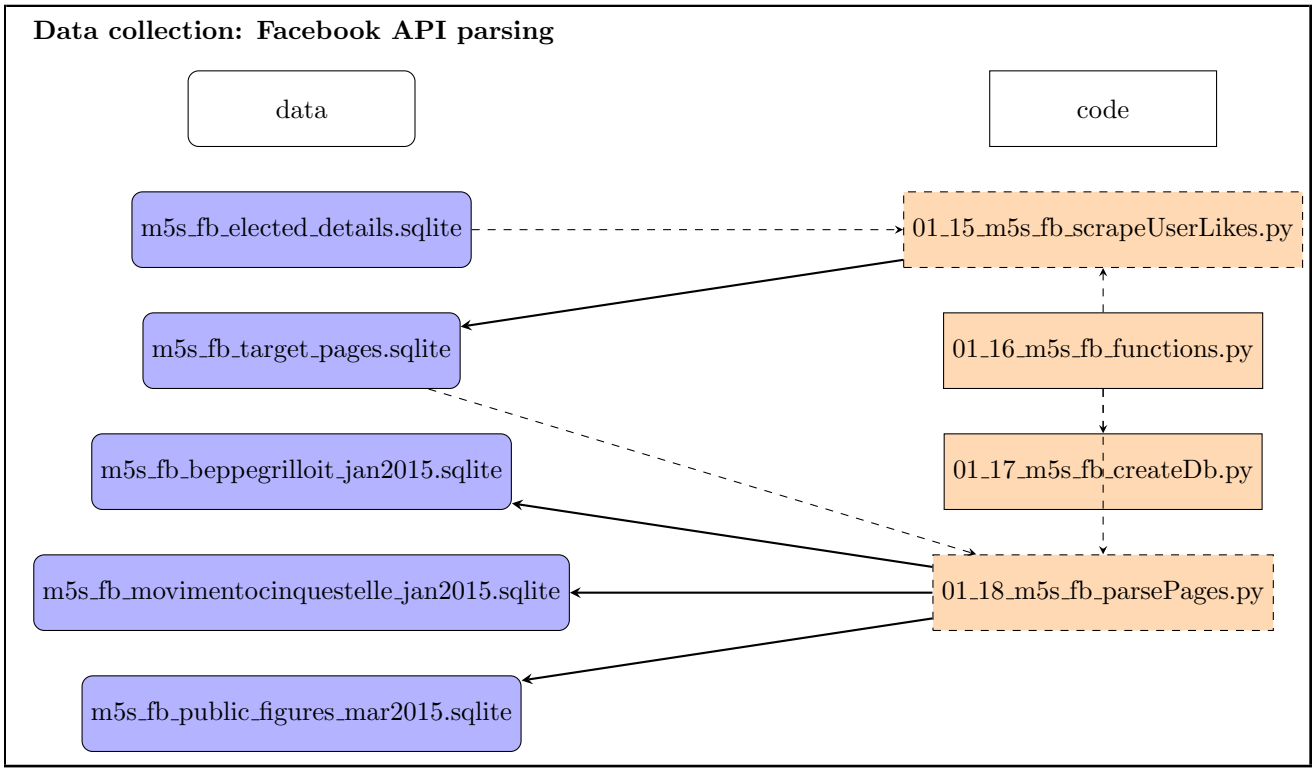
**table** sqlite_sequence
**fields** name, seq
**records** 2

**table** story_tags

**fields** id, post_id, profile_id, timestamp
**records** 0

**table** to
**fields** id, post_id, profile_id, timestamp
**records** 216,094

**table** with_tags
**fields** id, post_id, profile_id, timestamp
**records** 0




## 1.4 M5S meetups

**url** `https://www.meetup.com/`

**method** API requests

**period covered**

**from** 2005-06-10 14:45:08 UTC
**to** 2015-03-02 09:47:35 UTC

**period of collection**

**from** 2015-03-02 05:18:50 AEST
**to** 2015-03-05 12:32:28 AEST

**data file** m5s_meetup_mar2015.sqlite

**data schema**

**table** event
**fields** event_id, created, description, event_url, group_id, headcount, maybe_rsvp_count, name, rating_average, rating_count, status, time, updated, utc_offset, venue, visibility, waitlist_count, yes_rsvp_count, timestamp
**records** 128,694

**table** group
**fields** group_id, category, city, country, created, description, join_mode, lat, link, lon, members, name, organizer, state, timezone, urlname, visibility, who, timestamp
**records** 1,701

**table** group_member

**fields** group_member_id, group_id, member_id, timestamp

**records** 148,391

**table** member

**fields** member_id, bio, birthday, city, country, gender, joined, lang, lat, link, lon, name, facebook, twitter, flickr, tumblr, linkedin, timestamp

**records** 92,486

**table** rsvps

**fields** rsvps_id, member_id, event_id, created, timestamp

**records** 922,337

**table** sqlite_sequence

**fields** name, seq

**records** 2

**table** topic

**fields** topic_id, name, urlkey, timestamp

**records** 6,106

**table** topic_member

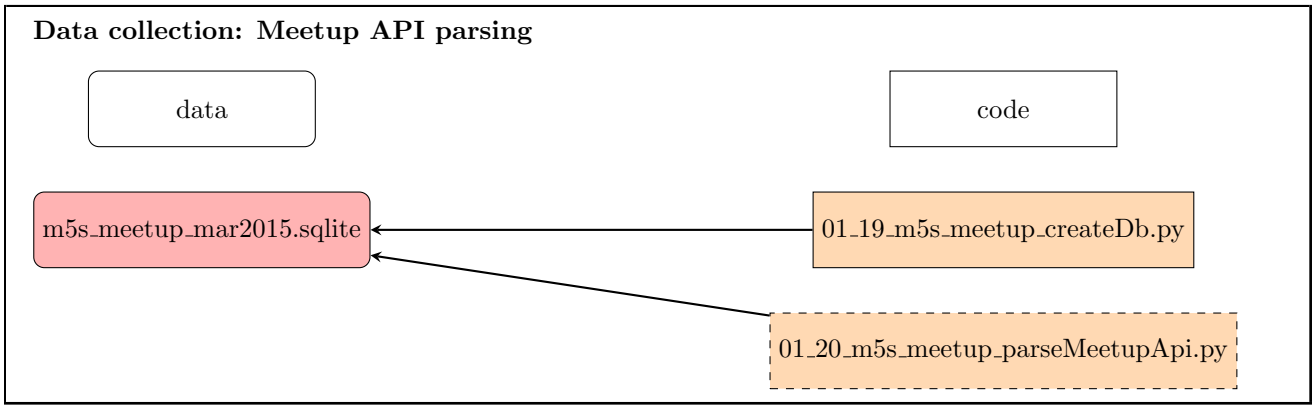
**fields** topic_member_id, topic_id, member_id, timestamp

**records** 1,178,308

**table** venue

**fields** venue_id, address_1, address_2, address_3, city, country, lat, lon, name, timestamp

**records** 30,734




## 1.5 Change.org

**url** `https://www.change.org/`

**method** API requests

**period covered**

**from** 2013-02-27 13:09:44 UTC

**to** 2014-02-19 17:58:00 UTC

**period of collection**

**from** 2015-09-11 01:33:18 AEST

**to** 2015-09-11 01:39:09 AEST

**data file** m5s_changeorg_sep2015.sqlite

**data schema**

**table** organization

**fields** organization_id, name, location, address, city, state_province, postal_code, country_name, countr_code, website, mission, organization_url, timestamp

**records** 0

**table** petition

**fields** petition_id, title, status, url, overview, letter_body, signature_count, image_url, category, goal, created_at, end_at, creator_name, creator_url, organization_name, organization_url, timestamp

**records** 1

**table** reason

**fields** petition_id, created_at, content, like_count, author_name, author_url, timestamp

**records** 10

**table** signature

**fields** petition_id, name, city, state_province, country_name, country_code, signed_at, timestamp

**records** 145,411

**table** target

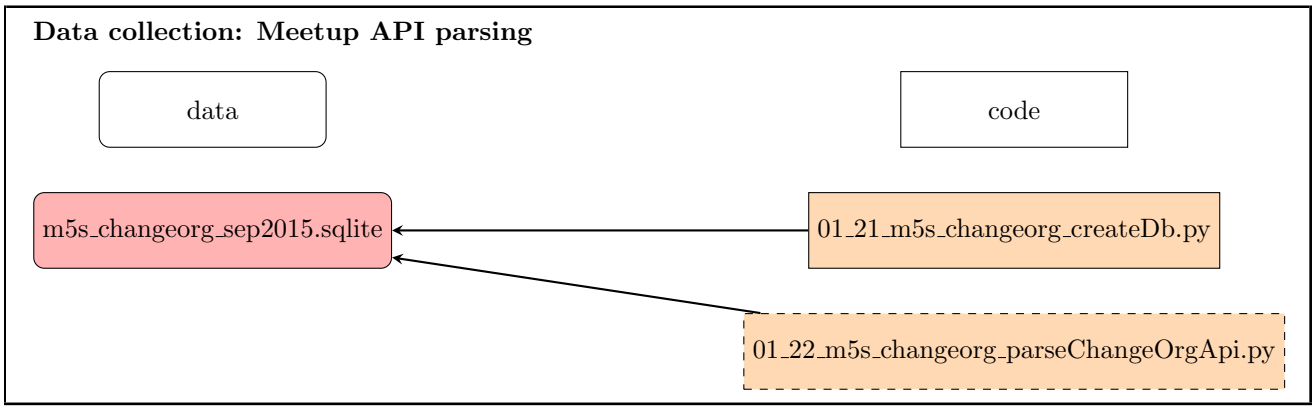
**fields** petition_id, name, title, type, target_area, timestamp

**records** 1

**table** user

**fields** user_id, name, location, city, state_province, country_name, countr_code, timestamp

**records** 0




## 1.6 la Repubblica and Corriere della Sera

**urls** `http://www.repubblica.it/` and `http://www.corriere.it/`

**method** HTML data extraction

**period covered**

**from** 1992-01-01

**to** 2016-01-05

**period of collection**

**from** 2016-03-22 16:00:45 AEST

**to** 2016-06-26 14:15:48 AEST

**data file** m5s_newsmedia_hits_mar2016.sqlite

**data schema**

**table** corriere_beppe

**fields** from_date, to_date, status, source, search_terms, hits, timestamp

**records** 49

**table** corriere_movimento

**fields** from_date, to_date, status, source, search_terms, hits, timestamp

**records** 49

**table** corriere_movimento1

**fields** from_date, to_date, status, source, search_terms, hits, timestamp

**records** 49

**table** corriere_politica

**fields** from_date, to_date, status, source, search_terms, hits, timestamp

**records** 49

**table** repubblica_beppe

**fields** from_date, to_date, status, source, search_terms, hits, timestamp

**records** 49

**table** repubblica_movimento

**fields** from_date, to_date, status, source, search_terms, hits, timestamp

**records** 49

**table** repubblica_movimento1

**fields** from_date, to_date, status, source, search_terms, hits, timestamp

**records** 49

**table** repubblica_politica

**fields** from_date, to_date, status, source, search_terms, hits, timestamp

**records** 49

**table** search_term

**fields** term_id, terms, query_mode, table_name

**records** 14

**data file** m5s_newsmedia_hits_jun2016.sqlite

**data schema**

**table** corriere_beppe

**fields** from_date, to_date, status, source, search_terms, hits, timestamp

**records** 1,204

**table** corriere_movimento

**fields** from_date, to_date, status, source, search_terms, hits, timestamp

**records** 1,204

**table** corriere_politica

**fields** from_date, to_date, status, source, search_terms, hits, timestamp

**records** 1,204

**table** repubblica_beppe

**fields** from_date, to_date, status, source, search_terms, hits, timestamp

**records** 25

**table** repubblica_movimento

**fields** from_date, to_date, status, source, search_terms, hits, timestamp

**records** 25

**table** repubblica_politica

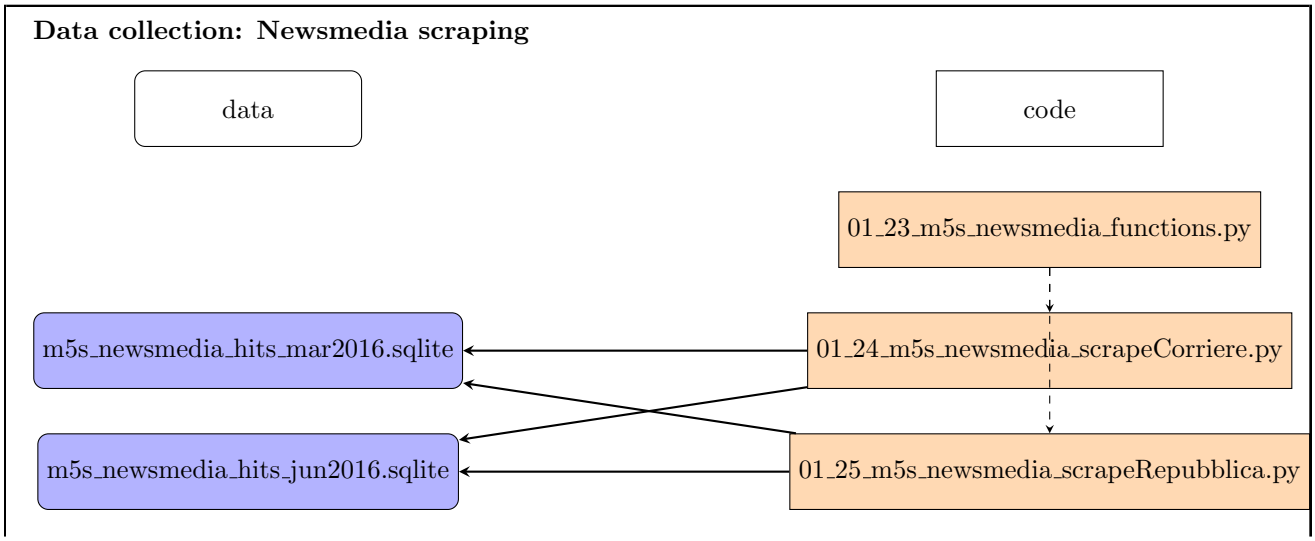
**fields** from_date, to_date, status, source, search_terms, hits, timestamp

**records** 1,204

**table** search_term

**fields** term_id, terms, query_mode, table_name

**records** 13



## 1.7 Camera dei Deputati and Senato della Repubblica

**urls** `http://dati.camera.it/` and `http://dati.senato.it/`

**method** SPARQL Protocol and RDF Query Language (SPARQL) requests

**period covered**

**from** 1993-02-17
**to** 2016-02-17

**period of collection**

**from** 2015-12-03 08:32:00 AEST
**to** 2016-02-28 02:17:00 AEST

**data file** m5s_parliament_m5s_act_details.RData

**data schema**

**table** acts_camera_m5s
**fields** act_resource, act_type, act_title, act_description, mp_resource, act_date
**records** 11,407

**table** mps_m5s
**fields** name, surname, mp_resource, group_name
**records** 109

**data file** m5s_parliament_gmi_act_details.RData

**data schema**

**table** gmi_camera_acts
**fields** act_resource, act_title, act_type, act_date, group_name
**records** 155

**table** gmi_camera_acts_cittadinanza
**fields** act_resource, act_title, act_type, act_date, group_name
**records** 89

**table** gmi_camera_acts_garantito
**fields** act_resource, act_title, act_type, act_date, group_name
**records** 79

**table** gmi_camera_bills
**fields** act_resource, act_title, act_date, group_name, act_type
**records** 30

**table** gmi_camera_bills_cittadinanza
**fields** act_resource, act_title, act_date, group_name, act_type
**records** 12

**table** gmi_camera_bills_garantito
**fields** act_resource, act_title, act_date, group_name, act_type
**records** 18

**table** gmi_senato_bills
**fields** ddl, titolo, date, nomeGruppo
**records** 75

**table** gmi_senato_bills_cittadinanza
**fields** ddl, titolo, date, nomeGruppo
**records** 36

**table** gmi_senato_bills_garantito
**fields** ddl, titolo, date, nomeGruppo
**records** 39

**data file** m5s_parliament_gmi_act_texts.RData

**data schema**

**table** gmi_camera_acts_wt_text

**fields** act_resource, act_text, act_date

**records** 161

**table** gmi_camera_bills_wt_text

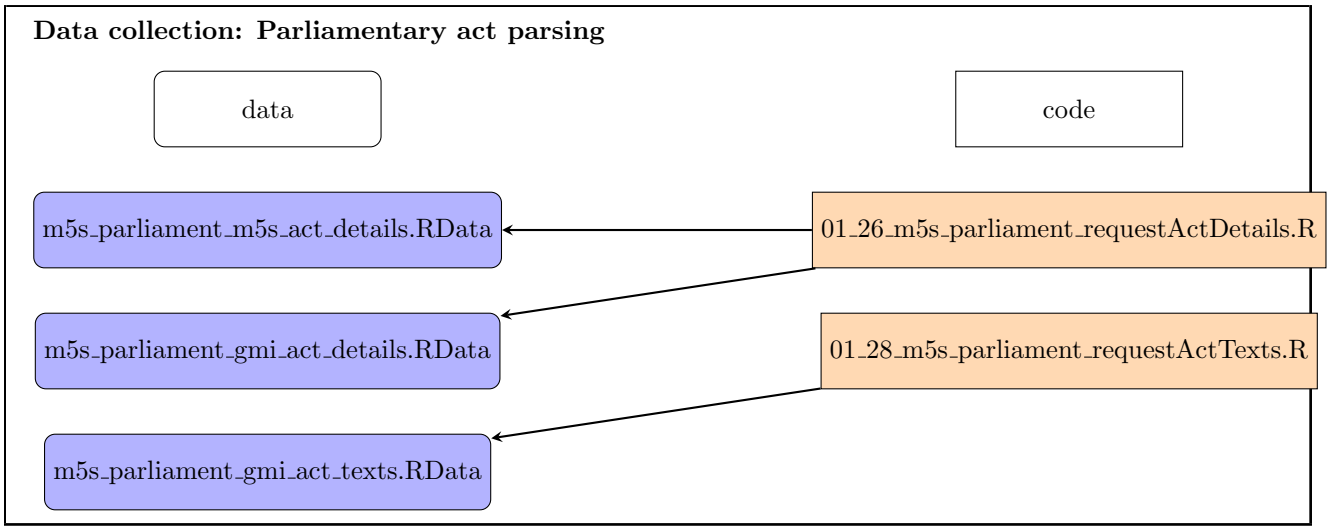
**fields** act_resource, act_text, act_title, act_date

**records** 12

**table** gmi_senato_bills_wt_text

**fields** act_resource, act_text, act_date

**records** 7




## 1.8 Pirate Party Forum

**urls** `https://news.piratenpartei.de/forumdisplay.php`

**method** HTML data extraction

**period covered**

**from** 2011-01-06 09:49 CET

**to** 2014-04-24 22:42 CET

**period of collection**

**from** 2014-04-15 08:53:36 AEST

**to** 2014-04-26 03:46:21 AEST

**data file** pp_forum_apr2014.sqlite

**data schema**

**table** author

**fields** authorId, authorName, authorStatus, authorNumberOfPosts, authorRegistrationDate, timestamp

**records** 10,729

**table** diff

**fields** threadId

**records** 1,106

**table** map

**fields** threadId, threadTitle, authorId, authorName, threadPosts, threadViews, threadVotes, threadRating, timestamp

**records** 1,106

**table** post

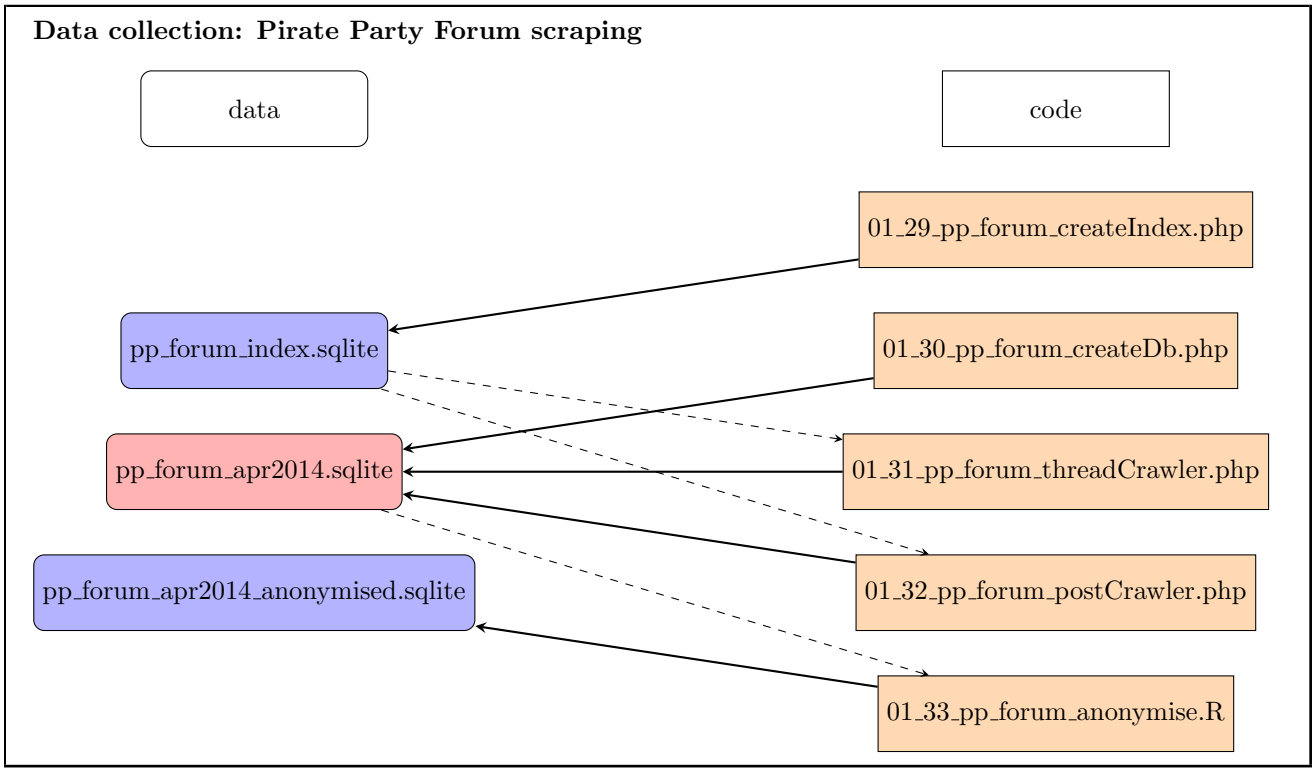
**fields** postId, postTitle, postDatetime, postOrdinalNumber, postMessage, authorId, threadId, timestamp

**records** 227,588

**table** thread

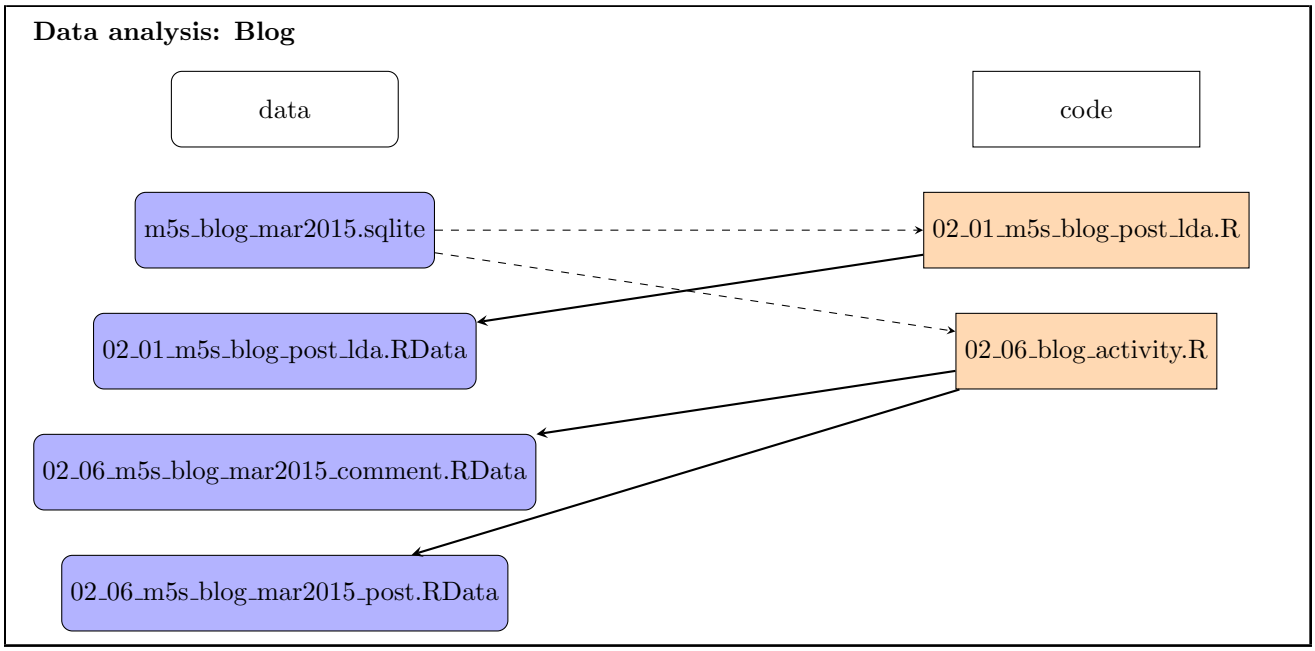
**fields** threadId, timestamp

**records** 6,591

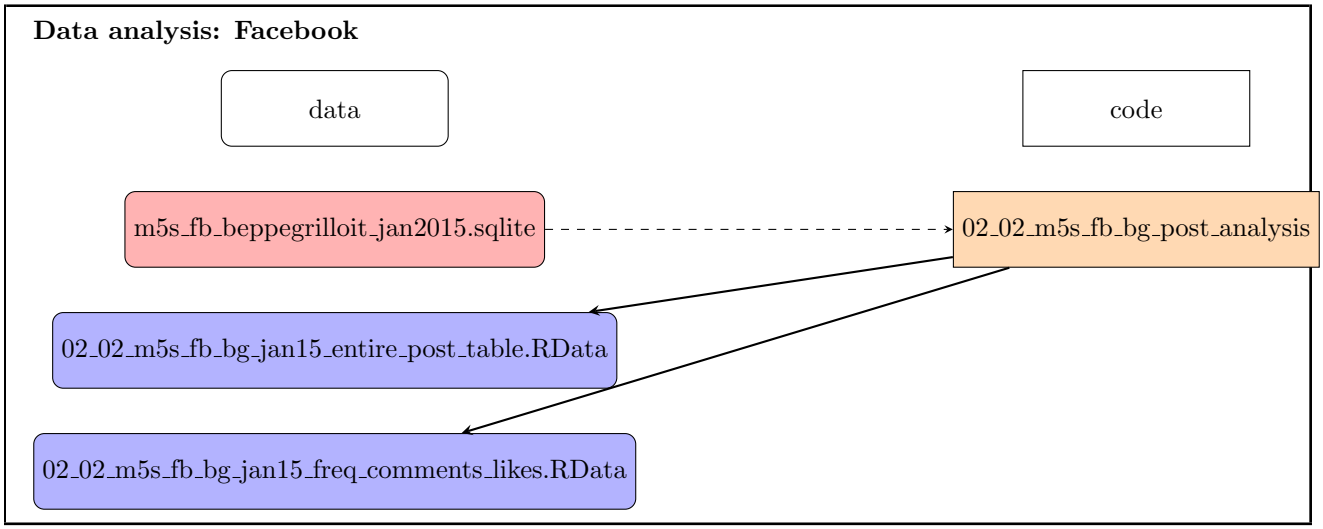



# 2 Data analysis

## 2.1 Beppe Grillo's blog

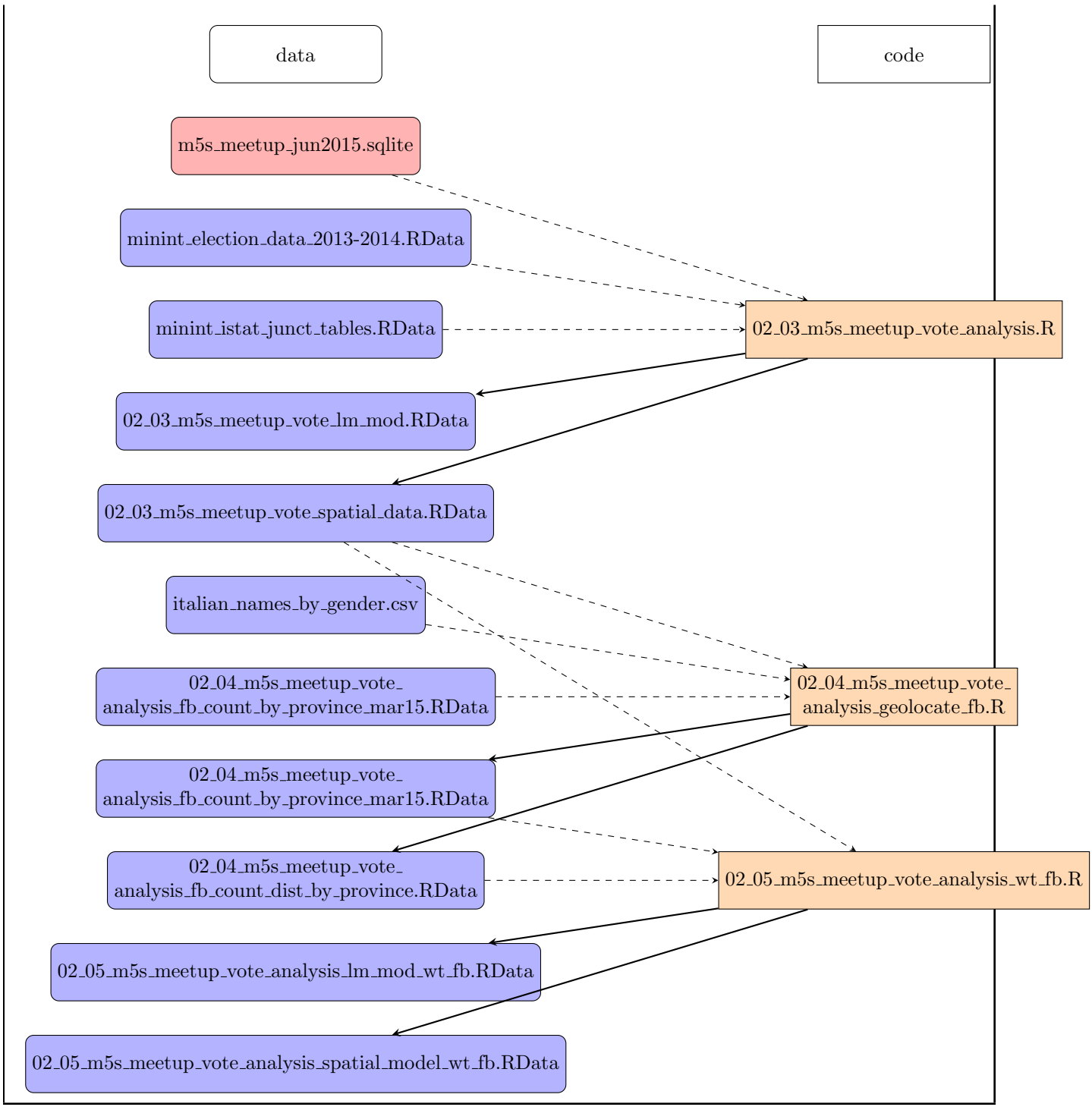



## 2.2 Facebook


Data analysis: Facebook
data
code
m5s_fb_beppegrilloit_jan2015.sqlite
02_02_m5s_fb_bg_post_analysis
02_02_m5s_fb_bg_jan15_entire_post_table.RData
02_02_m5s_fb_bg_jan15_freq_comments_likes.RData

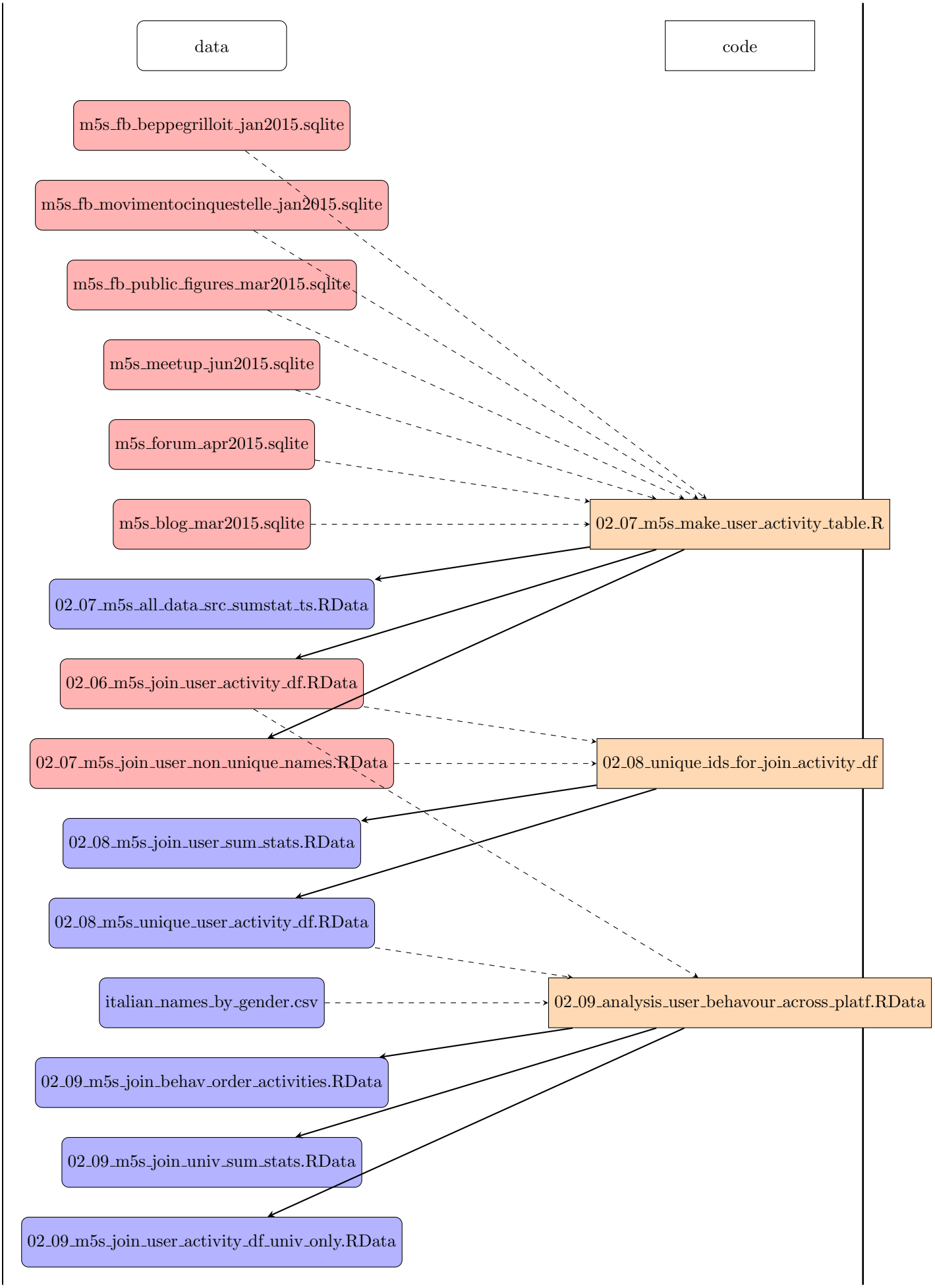
## 2.3 Meetup

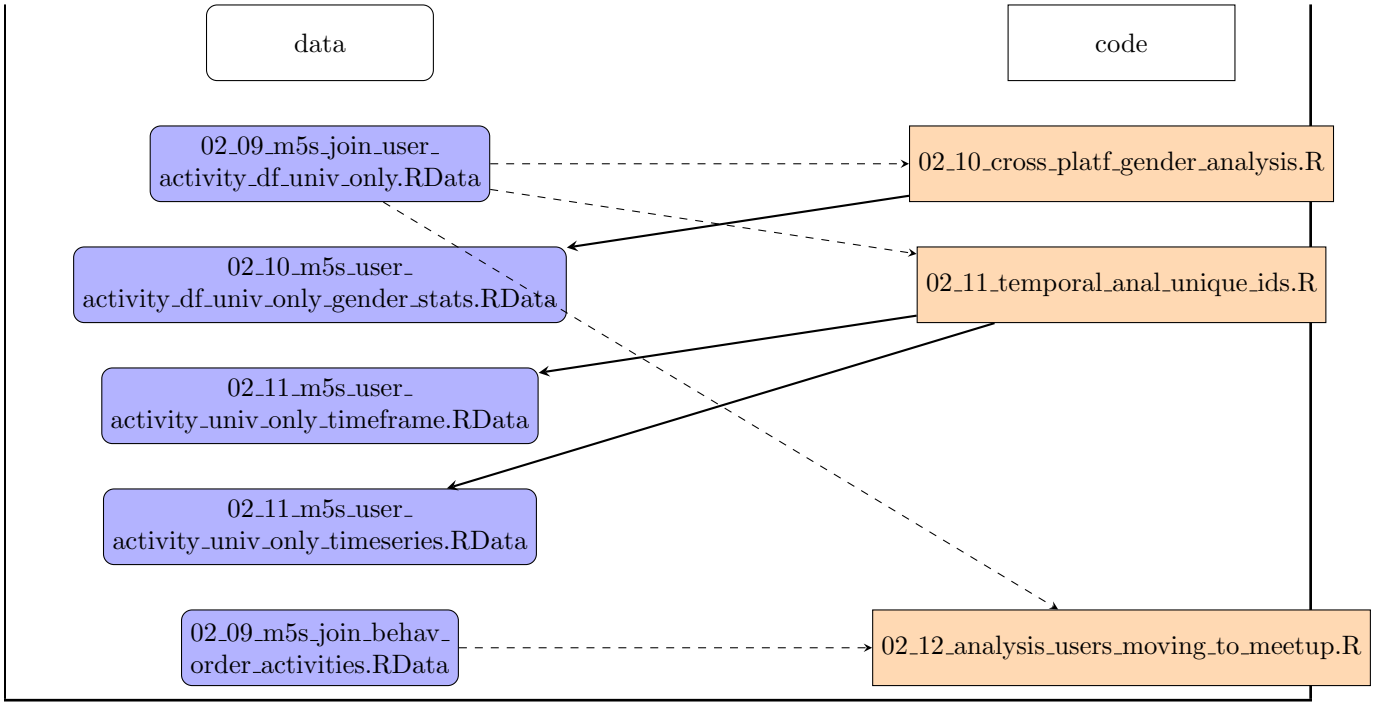
Data analysis: Meetup




## 2.4 Data integration

**Data analysis: Meetup**


data
code
m5s_fb_beppegrilloit_jan2015.sqlite
m5s_fb_movimentocinquestelle_jan2015.sqlite
m5s_fb_public_figures_mar2015.sqlite
m5s_meetup_jun2015.sqlite
m5s_forum_apr2015.sqlite
m5s_blog_mar2015.sqlite
02_07_m5s_make_user_activity_table.R
02_07_m5s_all_data_src_sumstat_ts.RData
02_06_m5s_join_user_activity_df.RData
02_07_m5s_join_user_non_unique_names.RData
02_08_unique_ids_for_join_activity_df
02_08_m5s_join_user_sum_stats.RData
02_08_m5s_unique_user_activity_df.RData
italian_names_by_gender.csv
02_09_analysis_user_behavour_across_platf.RData
02_09_m5s_join_behav_order_activities.RData
02_09_m5s_join_univ_sum_stats.RData
02_09_m5s_join_user_activity_df_univ_only.RData


data
code
02_09_m5s_join_user_
activity_df_univ_only.RData
02_10_cross_platf_gender_analysis.R
02_10_m5s_user_
activity_df_univ_only_gender_stats.RData
02_11_temporal_anal_unique_ids.R
02_11_m5s_user_
activity_univ_only_timeframe.RData
02_11_m5s_user_
activity_univ_only_timeseries.RData
02_09_m5s_join_behav_
order_activities.RData
02_12_analysis_users_moving_to_meetup.R