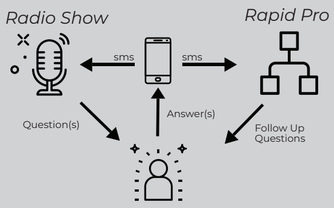
AVF Individual - An Analysis

Bachelors Project Hirschfeld 2019 / 2020



AVF conducts Radio Shows, where they conduct social-demograhic studies in a question - response manner, to understand the needs of the local inhabitants and different groups of society. After collecting responses via sms, they prepare the data with a pipeline like system for further analysis. This pipeline includes labeling

On this poster, the process of coming from a raw answer to a correctly labeled attribute value is shown via the Gender Attribute.


```
[
 {
  "Data": {
   "Source": "/rapid_pro_client.py:407:convert_runs_to_traced_data",
   "avf_phone_id": "avf-phone-uuid-ccf8e71f-d4ef-45a5-aa6a-4e12052b594a",
   "run_created_on - csap_s04e02_activation": "2019-09-03T07:01:15.998470+00:00",
   "run_exit_type - csap_s04e02_activation": "interrupted",
   "run_exited_on - csap_s04e02_activation": "2019-09-08T07:00:45.760504+00:00",
   "run_id - csap_s04e02_activation": 10910633,
   "run_modified_on - csap_s04e02_activation": "2019-09-08T07:00:45.801661+00:00"
  }
 },
 {
  "Data": {
   "operator_coded": {
    "Checked": false,
    "CodeID": "code-c7b32481",
    "DateTimeUTC": "2019-09-25T08:15:50.967015+00:00",
    "Origin": {
     "Name": "Pipeline Auto-Coder",
     "OriginID": "fetch_raw_data.py:45:label_somalia_operator"
    },
    "SchemeID": "Scheme-f86cff0b"
   },
   "operator_raw": "25261"
  }
 },
 {
  "Metadata": {
   "Source": "traced_data.py:453:update_iterable",
   "Timestamp": 1593995556.874056,
   "User": "alexander@africasvoices.org"
  },
  "NestedTracedData": {
   "survey_responses": [
    {
     "Data": {
      "Age (Category) - csap_s04_demog": "numeric",
      "Age (Name) - csap_s04_demog": "age",
      "Age (Run ID) - csap_s04_demog": 108695963,
      "Age (Text) - csap_s04_demog": "20",
      "Age (Time) - csap_s04_demog": "2019-08-25T09:56:58.914269+00:00",
      "Age (Value) - csap_s04_demog": "20",
      "District (Category) - csap_s04_demog": "Other",
      "District (Name) - csap_s04_demog": "district",
      "District (Run ID) - csap_s04_demog": 108695963,
      "District (Text) - csap_s04_demog": "asc bahda istaa AFam waadin salaamay waa ... ",
      "District (Time) - csap_s04_demog": "2019-08-25T09:52:06.640491+00:00",
      "District (Value) - csap_s04_demog": "asc bahda istaa AFam waadin salaamay waa ... ",
      "Gender (Category) - csap_s04_demog": "Male",
      "Gender (Name) - csap_s04_demog": "gender",
      "Gender (Run ID) - csap_s04_demog": 108695963,
      "Gender (Text) - csap_s04_demog": "rag",
      "Gender (Time) - csap_s04_demog": "2019-08-25T09:55:59.323039+00:00",
      "Gender (Value) - csap_s04_demog": "rag",
      "Idp_Camp (Category) - csap_s04_demog": "No",
      "Idp_Camp (Name) - csap_s04_demog": "idp_camp",
      "Idp_Camp (Run ID) - csap_s04_demog": 108695963,
      "Idp_Camp (Text) - csap_s04_demog": "MAYA",
      "Idp_Camp (Time) - csap_s04_demog": "2019-08-25T09:57:42.920366+00:00",
      "Idp_Camp (Value) - csap_s04_demog": "MAYA",
      "Recently_Displaced (Category) - csap_s04_demog": "No",
      "Recently_Displaced (Name) - csap_s04_demog": "recently_displaced",
      "Recently_Displaced (Run ID) - csap_s04_demog": 108695963,
      "Recently_Displaced (Text) - csap_s04_demog": "maya",
      "Recently_Displaced (Time) - csap_s04_demog": "2019-08-25T09:54:55.425793+00:00",
      "Recently_Displaced (Value) - csap_s04_demog": "maya",
      "avf_phone_id": "avf-phone-uuid-ccf8e71f-d4ef-45a5-aa6a-4e12052b594a",
      "run_created_on - csap_s04_demog": "2019-08-25T08:32:50.614369+00:00",
      "run_exit_type - csap_s04_demog": "completed",
      "run_exited_on - csap_s04_demog": "2019-08-25T09:57:42.920545+00:00",
      "run_id - csap_s04_demog": 108695963,
      "run_modified_on - csap_s04_demog": "2019-08-25T09:57:42.922390+00:00"
     }
    }
   ]
  }
 },
 {
  "Metadata": {
   "Source": "translate_rapid_pro_keys.py:38:set_show_ids"
  }
 },
 {
  "Data": {
   "age_raw": "20",
   "age_time": "2019-08-25T09:56:58.914269+00:00",
   "gender_raw": "rag",
   "gender_time": "2019-08-25T09:55:59.323039+00:00",
   "have_voice_raw": "HAGE ",
   "have_voice_time": "2019-09-09T08:04:04.401231+00:00",
   "in_idp_camp_raw": "MAYA",
   "in_idp_camp_time": "2019-08-25T09:57:42.920366+00:00",
   "location_raw": "asc bahda istaa AFam waadin salaamay",
   "location_time": "2019-08-25T09:52:06.640491+00:00",
   "recently_displaced_raw": "maya",
   "recently_displaced_time": "2019-08-25T09:54:55.425793+00:00",
   "suggestions_raw": "But she must have a prize herself, you know,' said the Mouse.\n",
   "suggestions_time": "2019-09-09T08:10:07.481562+00:00",
   "uid": "avf-phone-uuid-ccf8e71f-d4ef-45a5-aa6a-4e12052b594a"
  },
  "Metadata": {
   "Source": "/app/src/translate_rapid_pro_keys.py:135:remap_key_names"
  }
 },
 {
  "Data": {
   "location_raw_id_WS": "b2799698ed3d8..."
  },
  "Metadata": {
   "Source": "io.py:41:compute_message_ids"
  }
 },
 {...},
 {
  "Data": {
   "gender_raw_id_WS": "a3e690053061793bb12ad4c32aace5856f57f16..."
  },
  "Metadata": {
   "Source": "io.py:41:compute_message_ids"
  }
 },
 {
  "Data": {
   "gender_raw_WS_correct_dataset": {
    "Checked": false,
    "CodeID": "code-NR-5e3eee23",
    "DateTimeUTC": "2019-09-25T08:26:15.438337+00:00",
    "Origin": {
     "Name": "Pipeline Auto-Coder",
     "OriginID": "io.py:266:import_coda_2_to_traced_data_iterable"
    },
    "SchemeID": "Scheme-fffffc6e"
   }
  }
 },
 {
  "Data": {
   "gender_coded_WS": {
    "Checked": false,
    "CodeID": "code-63dcde9a",
    "Confidence": 0,
    "DateTimeUTC": "2018-12-03T12:42:27.138Z",
    "Origin": {
     "Metadata": {},
     "Name": "Pipeline Auto-Coder",
     "OriginID": "demographic_cleaner.py:28:clean_gender"
    },
    "SchemeID": "Scheme-12cb6f95"
   }
  },
  "Metadata": {
   "Source": "io.py:259:import_coda_2_to_traced_data_iterable"
  }
 },
 {
  "Data": {
   "gender_coded_WS": {
    "Checked": true,
    "CodeID": "code-63dcde9a",
    "Confidence": 1,
    "DateTimeUTC": "2018-12-09T17:24:23.467Z",
    "Origin": {
     "Name": "Khadija Mohamed",
     "OriginID": "khadija@africasvoices.org",
     "OriginType": "Manual"
    },
    "SchemeID": "Scheme-12cb6f95"
   }
  },
  "Metadata": {
   "Source": "io.py:259:import_coda_2_to_traced_data_iterable"
  }
 },
 {
  "Data": {
   "age_raw_WS_correct_dataset": {
    "Checked": false,
    "CodeID": "SPECIAL-MANUALLY_UNCODED",
    "Confidence": 1,
    "DateTimeUTC": "2019-03-20T08:19:23.646Z",
    "Origin": {
     "Name": "Nasra Ali",
     "OriginID": "nasra@africasvoices.org",
     "OriginType": "Manual"
    },
    "SchemeID": "Scheme-fffffc6e"
   }
  },
  "Metadata": {
   "Source": "io.py:259:import_coda_2_to_traced_data_iterable"
  }
 },
 {...},
 {
  "Data": {
   "have_voice_yes_no_amb_coded_WS": {
    "Checked": true,
    "CodeID": "code-NC-42f1d983",
    "Confidence": 1,
    "DateTimeUTC": "2019-09-17T05:50:59.693Z",
    "Origin": {
     "Name": "Zakaria Sheikh",
     "OriginID": "zakaria.sheikh@africasvoices.org",
     "OriginType": "Manual"
    },
    "SchemeID": "Scheme-539ec95d"
   }
  },
  "Metadata": {
   "Source": "io.py:259:import_coda_2_to_traced_data_iterable",
   "Timestamp": "2019-09-25T08:31:46.050440+00:00",
   "User": "alexander@africasvoices.org"
  }
 },
 {
  "Data": {
   "suggestions_coded_WS": [
    {
     "Checked": true,
     "CodeID": "code-de8895af",
     "Confidence": 1,
     "DateTimeUTC": "2019-09-23T13:50:09.158Z",
     "Origin": {
      "Name": "Zakaria Sheikh",
      "OriginID": "zakaria.sheikh@africasvoices.org",
      "OriginType": "Manual"
     },
     "SchemeID": "Scheme-1c7e8dd7"
    }
   ]
  },
  "Metadata": {
   "Source": "io.py:326:import_coda_2_to_traced_data_iterable_multi_coded"
  }
 },
 {
  "Data": {
   "gender_coded": {
    "Checked": false,
    "CodeID": "code-63dcde9a",
    "Confidence": 0,
    "DateTimeUTC": "2018-12-03T12:42:27.138Z",
    "Origin": {
     "Metadata": {},
     "Name": "Pipeline Auto-Coder",
     "OriginID": "demographic_cleaner.py:28:clean_gender"
    },
    "SchemeID": "Scheme-12cb6f95"
   }
  },
  "Metadata": {
   "Source": "io.py:259:import_coda_2_to_traced_data_iterable"
  }
 },
 {
  "Data": {
   "gender_coded": {
    "Checked": true,
    "CodeID": "code-63dcde9a",
    "Confidence": 1,
    "DateTimeUTC": "2018-12-09T17:24:23.467Z",
    "Origin": {
     "Name": "Khadija Mohamed",
     "OriginID": "khadija@africasvoices.org",
     "OriginType": "Manual"
    },
    "SchemeID": "Scheme-12cb6f95"
   }
  },
  "Metadata": {
   "Source": "io.py:259:import_coda_2_to_traced_data_iterable"
  }
 },
 {
  "Data": {
   "gender_raw_id": "a3e690053061793bb12ad4c32aace5856f5..."
  },
  "Metadata": {
   "Source": "io.py:41:compute_message_ids"
  }
 }
]
```


AVF uses the software Rapid-Pro to conduct automated question response interviews (runs) with radio show listeners. Using a self written client, they extract the data from RapidPro. This data is the starting point of the pipeline. The script “convert_runs_to_traced_data” generated the internal representation and working object; the traced data object

A common - if not the main - step is to label the answers of individuals. This assignes the answer and thus the individual to a predefined group (e.g. Gender - male). This labeling is conducted with assigning a codeID, which is predefined for each project, to the key. The codeID encodes the value of the group the raw value is assigned to.

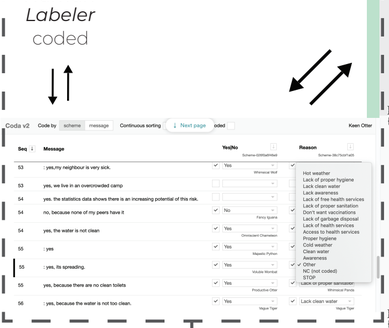
Each attribute is first preprocessed and then coded by a automated pipeline script. After this, the coding always gets checked by a human and possibly corrected.

The “import_coda_2_to_trace_data_iterable” function writes the coding that is conducted in the frontend by a human labeler to the TracedDataObject; more specifically to the corresponding key. Sample code schemes can be seen in the blue boxes

Pipeline Scripts preprocessing

Pipeline Scripts auto coding

Human Labeler coded



Process is repeated for every other attribute of an individual (e.g. age, have_voice)

Pipeline Scripts cleaning


```
{
 "Data": {
  "rqa_s04e01_coded": [
   {
    "Checked": true,
    "CodeID": " ",
    "Confidence": 1,
    "DateTimeUTC": "2019-09-03T06:50:31.718Z",
    "Origin": {
     "Name": "Fatima Dahir Mohammed",
     "OriginID": "fatima@africasvoices.org",
     "OriginType": "Manual"
    },
    "SchemeID": "Scheme-eb6d805e"
   }
  ]
 },
 "Metadata": {
  "Source": "io.py:326:import_coda_2_to_traced_data_iterable_multi_coded"
 }
},
{
 "Data": {
  "rqa_s04e01_raw_correct_dataset": {
   "Checked": false,
   "CodeID": "code-NR-5e3eee23",
   "DateTimeUTC": "2019-09-25T08:38:37.861217+00:00",
   "Origin": {
    "Name": "Pipeline Auto-Coder",
    "OriginID": "io.py:266:import_coda_2_to_traced_data_iterable",
    "OriginType": "External"
   },
   "SchemeID": "Scheme-fffffc6e"
  }
 },
 "Metadata": {
  "Source": "io.py:270:import_coda_2_to_traced_data_iterable",
  "Timestamp": 1569400717.8614702,
  "User": "alexander@africasvoices.org"
 }
},
{...},
{
 "Data": {
  "gender_coded": {
   "Checked": false,
   "CodeID": "code-63dcde9a",
   "Confidence": 0,
   "DateTimeUTC": "2018-12-03T12:42:27.138Z",
   "Origin": {
    "Metadata": {},
    "Name": "Pipeline Auto-Coder",
    "OriginID": "demographic_cleaner.py:28:clean_gender",
    "OriginType": "External"
   },
   "SchemeID": "Scheme-12cb6f95"
  }
 },
 "Metadata": {
  "Source": "io.py:259:import_coda_2_to_traced_data_iterable"
 }
},
{
 "Data": {
  "gender_coded": {
   "Checked": true,
   "CodeID": "code-63dcde9a",
   "Confidence": 1,
   "DateTimeUTC": "2018-12-09T17:24:23.467Z",
   "Origin": {
    "Name": "Khadija Mohamed",
    "OriginID": "khadija@africasvoices.org",
    "OriginType": "Manual"
   },
   "SchemeID": "Scheme-12cb6f95"
  }
 },
 "Metadata": {
  "Source": "io.py:259:import_coda_2_to_traced_data_iterable"
 }
},
{
 "Data": {
  "gender_raw_correct_dataset": {
   "Checked": false,
   "CodeID": "code-NR-5e3eee23",
   "DateTimeUTC": "2019-09-25T08:40:42.240354+00:00",
   "Origin": {
    "Name": "Pipeline Auto-Coder",
    "OriginID": "io.py:266:import_coda_2_to_traced_data_iterable",
    "OriginType": "External"
   },
   "SchemeID": "Scheme-fffffc6e"
  }
 },
 "Metadata": {
  "Source": "io.py:270:import_coda_2_to_traced_data_iterable"
 }
},
{
 "Data": {
  "rqa_s04e02_coded": [
   {
    "Checked": false,
    "CodeID": "code-NA-f93d3eb7",
    "DateTimeUTC": "2019-09-25T08:43:01.126889+00:00",
    "Origin": {
     "Name": "Pipeline Auto-Coder",
     "OriginID": " apply_manual_codes.py:121:apply_manual_codes",
     "OriginType": "External"
    },
    "SchemeID": "Scheme-4a52f65b"
   }
  ]
 },
 "Metadata": {
  "Source": "apply_manual_codes"
 }
},
{
 "Data": {
  "district_coded": {
   "Checked": false,
   "CodeID": "code-97060a6b",
   "DateTimeUTC": "2019-09-25T08:43:46.386086+00:00",
   "Origin": {
    "Name": "Pipeline Auto-Coder",
    "OriginID": "...",
    "OriginType": "External"
   },
   "SchemeID": "Scheme-5292e8fb"
  },
  "mogadishu_sub_district_coded": {
   "Checked": false,
   "CodeID": "code-NC-42f1d983",
   "DateTimeUTC": "2019-09-25T08:43:46.385823+00:00",
   "Origin": {
    "Name": "Pipeline Auto-Coder",
    "OriginID": "impute_somalia_location_codes",
    "OriginType": "External"
   },
   "SchemeID": "Scheme-07dfc611"
  },
  "region_coded": {
   "Checked": false,
   "CodeID": "code-000c5297",
   "DateTimeUTC": "2019-09-25T08:43:46.386333+00:00",
   "Origin": {
    "Name": "Pipeline Auto-Coder",
    "OriginID": "impute_somalia_location_codes",
    "OriginType": "External"
   },
   "SchemeID": "Scheme-644e3296"
  },
  "state_coded": {
   "Checked": false,
   "CodeID": "code-0a1aaafb",
   "DateTimeUTC": "2019-09-25T08:43:46.386577+00:00",
   "Origin": {
    "Name": "Pipeline Auto-Coder",
    "OriginID": "impute_somalia_location_codes",
    "OriginType": "External"
   },
   "SchemeID": "Scheme-a390e6ce"
  },
  "zone_coded": {
   "Checked": false,
   "CodeID": "code-dc3483fa",
   "DateTimeUTC": "2019-09-25T08:43:46.386821+00:00",
   "Origin": {
    "Name": "Pipeline Auto-Coder",
    "OriginID": "impute_somalia_location_codes",
    "OriginType": "External"
   },
   "SchemeID": "Scheme-b8ce44fb"
  }
 },
 "Metadata": {
  "Source": "impute_somalia_location_codes"
 }
},
{
 "Metadata": {
  "Source": " _impute_coding_error_codes"
 }
},
{
 "Data": {
  "consent_withdrawn": "false"
 },
 "Metadata": {
  "Source": "/app/src/analysis_file.py:103:generate"
 }
},
{
 "Data": {
  "age": "20",
  "district": "belet xaawo",
  "gender": "male",
  "have_voice_yes_no_amb": "NC",
  "in_idp_camp": "no",
  "operator": "hormud",
  "recently_displaced": "no",
  "region": "gedo",
  "rqa_s04e01_CE": "0",
  "rqa_s04e01_NA": "0",
  "rqa_s04e01_NC": "1",
  "rqa_s04e01_NCOs_to_be_accountable_participatory": "0",
  "rqa_s04e01_NIC": "0",
  "rqa_s04e01_NR": "0",
  "rqa_s04e01_NS": "0",
  "rqa_s04e01_STOP": "0",
  "rqa_s04e01_WS": "0",
  "rqa_s04e01_access_to_water": "0",
  "rqa_s04e01_build_resilience_to_drought": "0",
  "rqa_s04e01_community_organisation": "0",
  "rqa_s04e01_cooperation_between_government_and_NGOs": "0",
  "rqa_s04e01_cooperation_between_government_and_people": "0",
  "rqa_s04e01_economic_development": "0",
  "rqa_s04e01_education": "0",
  "rqa_s04e01_farming_support": "0",
  "rqa_s04e01_food_nutrition": "0",
  "rqa_s04e01_good_governance": "0",
  "rqa_s04e01_greeting": "0",
  "rqa_s04e01_health_services": "0",
  "rqa_s04e01_hygiene": "0",
  "rqa_s04e01_information": "0",
  "rqa_s04e01_job_creation": "0",
  "rqa_s04e01_opt_in": "0",
  "rqa_s04e01_other": "0",
  "rqa_s04e01_peace_and_security": "0",
  "rqa_s04e01_protection_of_rights": "0",
  "rqa_s04e01_push_back": "0",
  "rqa_s04e01_question": "0",
  "rqa_s04e01_religion": "0",
  "rqa_s04e01_return_and_resettlement_of_IDPs": "0",
  "rqa_s04e01_shelter": "0",
  "rqa_s04e01_showtime_question": "0",
  "rqa_s04e01_stop_aid_dependency": "0",
  "rqa_s04e01_stop_clannism": "0",
  "rqa_s04e01_support_for_IDPs": "0",
  "rqa_s04e01_support_to_the_poor": "0",
  "rqa_s04e02_CE": "0",
  "rqa_s04e02_NA": "1",
  "rqa_s04e02_NC": "0",
  "rqa_s04e02_NGOs_to_be_accountable_and_transparent": "0",
  "rqa_s04e02_NIC": "0",
  "rqa_s04e02_NR": "0",
  "rqa_s04e02_NS": "0",
  "rqa_s04e02_STOP": "0",
  "rqa_s04e02_WS": "0",
  "rqa_s04e02_access_to_water": "0",
  "rqa_s04e02_cooperation_between_NGOs_people_government": "0",
  "rqa_s04e02_economic_development": "0",
  "rqa_s04e02_education": "0",
  "rqa_s04e02_food_nutrition": "0",
  "rqa_s04e02_government_stability": "0",
  "rqa_s04e02_greeting": "0",
  "rqa_s04e02_health_services": "0",
  "rqa_s04e02_information": "0",
  "rqa_s04e02_job_creation": "0",
  "rqa_s04e02_opt_in": "0",
  "rqa_s04e02_other": "0",
  "rqa_s04e02_peace_and_security": "0",
  "rqa_s04e02_push_back": "0",
  "rqa_s04e02_question": "0",
  "rqa_s04e02_resettlement_and_return_for_IDPs": "0",
  "rqa_s04e02_shelter": "0",
  "rqa_s04e02_showtime_question": "0",
  "rqa_s04e02_stop_aid_dependency": "0",
  "rqa_s04e02_support_for_agriculture": "0",
  "rqa_s04e02_support_the_poor_IDPs": "0",
  "state": "jubbaland",
  "suggestions_CE": "0",
  "suggestions_NA": "0",
  "suggestions_NC": "0",
  "suggestions_NIC": "0",
  "suggestions_NR": "0",
  "suggestions_NS": "0",
  "suggestions_STOP": "0",
  "suggestions_WS": "0",
  "suggestions_avoid_political_content": "0",
  "suggestions_consistent_radio_shows": "0",
  "suggestions_greeting": "0",
  "suggestions_honesty_from_radio_broadcast": "0",
  "suggestions_improve_quality_of_hage": "0",
  "suggestions_improve_radio_stations": "1",
  "suggestions_improved_participation": "0",
  "suggestions_increase_frequency_of_shows": "0",
  "suggestions_more_religious_messaging": "0",
  "suggestions_opt_in": "0",
  "suggestions_other": "0",
  "suggestions_push_back": "0",
  "suggestions_question": "0",
  "suggestions_showtime_question": "0",
  "zone": "scz"
 },
 "Metadata": {
  "Source": "/app/src/analysis_file.py:165:generate"
 }
},
{
 "Data": {
  "rqa_s04e01_NA": "0"
 },
 "Metadata": {
  "Source": "/app/src/analysis_file.py:196:generate"
 }
},
{
 "Data": {
  "rqa_s04e01_NC": "1"
 },
 "Metadata": {
  "Source": "/app/src/analysis_file.py:205:generate"
 }
},
{
 "Data": {
  "suggestions_NA": "0"
 },
 "Metadata": {
  "Source": "/app/src/analysis_file.py:196:generate"
 }
}
```


Pipeline Scripts scheme adjustment

Pipeline Scripts control flow coding

Pipeline Scripts Aggregate location object

Pipeline Scripts Check consistency

Pipeline Scripts Merging

As it can be seen, the gender_coded objects appear two times in the pipeline. The only difference is, that the suffix “_WS” is not present in this second occurence. Otherwise, the objects are exactly the same. We know that WS stands for “wrong scheme” and therefore assume that by default every attribute coding is conducted with an assumption of applying the wrong scheme. Later on the scheme is checked again, which can be seen in step #44, and the correct dataset is coded like every other attribute.

It appears that this step was conducted one year after the initial coding as can be seen in the timestamps that have been logged within the objects.

The individuals pipeline as it is unfolded here, is created in the last step of the pipeline. Where previously during processing steps, every flow is treated and labeled separately, the folding algorithm in the analysis file generat function folds the analysis of every flow that belongs to a single individuum and creates this long individuals object.

The data object analysed on this sheet can be found in the individuals_data json file that is generated with the AnalysisFile.generate() function.