Appendix D. Methodology of the evaluation

1. Introduction: theory-based approach

In consideration of the institutional learning and the accountability purpose of the evaluation, the team adopted a theory-based approach. By focussing not only on the "what" but particularly on the "why" dimension of the change, this approach had better help identifying lessons in terms of factors facilitating or hindering the desired change.

The reconstructed, unofficial Theory of Change (ToC) of the entire sub-programme[1] represented a series of "silo-type" areas of work, corresponding to the six thematic areas included under the sub-programme, which converge at the level of the intermediate outcomes. The intermediate outcomes are formulated in a generic form and there is no analysis of the potential for cross-fertilisation and influence among the initiatives in the six different areas. Most types of outputs are delivered for each of these areas of work, but are specific to each area of work, for example, recommendations are specific to each thematic area of the sub-programme, advocacy is done differently for each convention.

To reflect the refocus of the evaluation, two different ToC have been "extracted" from the original ToC, one for each of the two thematic areas of cybercrime and trafficking in human beings. They have been analysed, discussed and agreed upon during Inception in interaction with the Reference Group. These two Theories of Change are illustrated in the main report.

These two ToC have been taken as the point of departure for the finalisation of the evaluation questions and developing the evaluation methodology (including the evaluation matrix). The overall logic of the ToC is reflected in the data collection tools developed for this evaluation.

The two ToC are similar but present an important difference at the level of the immediate outcomes. While for cybercrime the focus is on increasing the number of signatories to the convention as a precondition for the increasing of national capacities, for trafficking in human beings, the focus is on monitoring. Notably, trafficking in human beings does not mention the increase in the number of signatory parties.

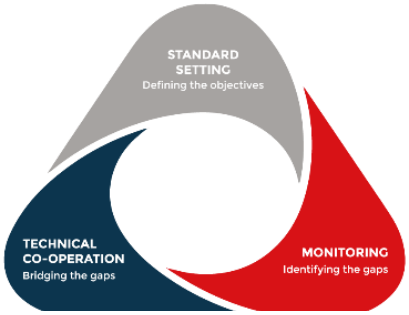
2. Sources of evidence

The evaluators based their work both on secondary and primary sources of evidence.

Secondary sources included more than 300 relevant documents such as:

- Documents at the sub-programme and at a regulatory, strategic and political level (e.g. Programme and Budget documents, the evaluation of the monitoring mechanisms, project reports, newsletters, monitoring reports, general Council of Europe documents etc.).
- Documents at the level of the two areas of work, covering the three elements of the 'dynamic triangle' of the Council of Europe.[2]
- Relevant third parties documents and literature.

Figure 1 – The Council of Europe "dynamic triangle"



Primary sources were selected in order to provide a plurality of relevant views and inputs on both areas of work. The stakeholders' groups that were identified for primary research were slightly different for each of the selected instruments and are reported in the following lines and tables. Gender balance was a constant point of attention during selection of key informants.

Three instruments were used; they were:

[1] Prepared by DIO and included in the Terms of Reference of this evaluation.

[2] During the initial part of the Inception phase, the evaluators analysed also documents related to the four thematic areas that were not retained for evaluation.

- Individual and small group semi-structured interviews.
  The evaluators had an extremely extensive interview activity and interviewed 135 individuals with a good gender balance: 48.1% women and 51.9% men; 59.3% of them were interviewed in person (in Brussels, Bucharest, Rabat, Sarajevo, Strasbourg and Vienna) while 40.7% of them were via teleconference. Table 1 represents a breakdown of interviewees, per thematic area and gender.

Table 1 – Gender of interviewees, per topic covered[3]

Topic	Interviewees: women		Interviewees: men		Interviewees: total	
	Number	%	Number	%	Number	%
Cybercrime	30	44.8 %	37	55.2 %	67	39.0 %
Trafficking Human Beings	41	54.0 %	35	46.0 %	76	44.2 %
Other topics	13	44.8 %	16	55.2 %	29	16.8 %
Total	**84**	**48.9 %**	**88**	**51.1 %**	**172**	**100 %**

Interviewees included:

- staff C-C and THB, including project and administrative staff
- chairs and members of monitoring mechanisms
- national/local partners and stakeholders (including Civil Society) during field visits
- other relevant Council of Europe staff working on C-C and THB
- external actors working on C-C and THB.

Four online surveys[4] for T-CY members, THB-CP members, relevant stakeholders and Council of Europe staff working on THB and C-C were produced. Overall, the evaluation team received 292 responses,[5] with a global 37.2% response rate, regarded as very high. The surveys were available in English (92% of the responses) and French (8% of the responses).

- Table 2

Table presents the survey response rate, disaggregated by thematic area and type of stakeholders.

Table 2 – Surveys: return rate, per target group

Survey invitations sent		Survey responses	
Target respondents	**Number**	**Number**	**Response rate**
T-CY members	166	46	27.7%
THB-CP members	66	13	19.7%
Stakeholders: C-C	344	85	24.7%
Stakeholders: THB	131	78	59.5%
Council of Europe staff: C-C	46	41	89.1%
Council of Europe staff: THB	31	29	93.5%
Total	**784**	**292**	**37.2%**
Cumulative: C-C	556	172	**30.9%**
Cumulative: THB	228	120	**52.6%**

[3] Interviews with some interviewees covered more than one topic; the grand total (172) is therefore higher than the number of interviewees (head count).
[4] Surveys were realised in LimeSurvey (https://www.limesurvey.org/), a GDPR-compliant tool.
[5] 38% of them partly filled but providing answers that were relevant enough to be processed. This total reflects the situation after data cleaning.

- Field visits to Austria, Bosnia and Herzegovina, Morocco and Romania. The further planned visit (to the United Kingdom) was not undertaken because only two persons accepted the interview. During field visits, the evaluators interviewed local staff of the Council of Europe, national partners (national institutions, civil society, academia) and staff of other organisations / donors working in the two thematic areas.

3. Data analysis and evidence aggregation methods

To respond to the mandate, the evaluation team used a plurality of data analysis and evidence aggregation methods. They are briefly described in the following sub-chapters.

Results mapping

This aggregation (requested in the ToR) produced a visual illustration of the results achieved by the Council of Europe at the outcome level; it was informed by secondary research, supplemented with primary analysis.

Qualitative Comparative Analysis (QCA)

"*Qualitative Comparative Analysis (QCA) is a methodology that enables the analysis of multiple cases in complex situations. It can help explain why change happens in some cases but not others. QCA is designed for use […] in situations where there are too few cases to apply conventional statistical analysis*."[6] It was introduced in social research in 1987 to "*understand which qualitative factors are likely to influence an outcome"*[7] and since then, undergone several developments. It analyses the contribution of several factors[8] to achieving (or non-achieving) a desired outcome by using Boolean operators.[9] In spite of the adjective 'Qualitative' contained in its name and because of its mathematical basis it could be better defined -particularly in some Latin cultures- as "Quali-Quantitative Comparative Analysis".[10]

By integrating quantitative and qualitative analysis it generates "*findings that are generalisable across a wider population. QCA can help identify causal patterns when triangulated with results from other methods*."[11]

In the context of this evaluation, the QCA method was used to understand and weigh the impact of different internal and external factors on the achievement of the objective of legislative change in view of their alignment with the two conventions on cybercrime and trafficking in human beings. The QCA helped in understanding all these factors in terms of necessity and sufficiency to the achievement of the expected outcome (in our case, the legislative change to align to the relevant convention).

Charles Ragin, who is credited with developing QCA in the 1980s, specifies that the technique is based on two primary assumptions: "*change is often the result of different combinations of factors, rather than on any one individual factor; and different combinations of factors can produce similar changes."*[12]

In order to produce meaningful results, the QCA shall test both cases (in our evaluation, countries) where the outcome (in our case, legislative change) materialised and cases where it did not materialise.

[6] INTRAC, Qualitative Comparative Analysis, 2017, available here.
[7] Barbara Befani, Pathways to change: evaluating development interventions with Qualitative Comparative Analysis (QCA), EBA 2016, available here.
[8] The presence or absence of a given factor is defined as a condition.
[9] A simple definition of Boolean operators, largely used for searching the internet with browsers, can be found here.
[10] Barbara Befani, quoted.
[11] Claude Robinson, Migara Jayawardena, Ryan Waltkins, Joy Butscher, Noureddin Berrah; Qualitative Comparative Analysis, Independent Evaluation Group (World Bank Group), October 2022
[12] Claude Robinson et al., quoted.

In consideration of the many differences between C-C and THB,[13] one main QCA was ran for each of these areas;[14] each of these two analyses took into consideration the presence or absence in the three years preceding the legislative change of 10 different factors (conditions in the QCA jargon). Eight of these factors were common to both thematic areas, while two of them were specific to each field. The factors were initially identified in consultation with DIO and the Reference Group and then further refined during data gathering, based on coverage and significance of data.

For this evaluation, the team used a "*crisp set QCA*", where binary coding was used to express the presence (value 1) or the absence (value 0) of each tested condition.

The analysis took into consideration data from 30 countries in trafficking in human beings; and 41 in cybercrime. In the QCA jargon they are called "cases". Data were gathered with the use of five complementary tools:

- Surveys targeting:
    - Members of the THB Committee of the Parties.
    - Members of the T-CY.
    - External stakeholders
- Interviews.
- Background research.

The two main QCA tests ran for this evaluation were extremely wide; those discussed in literature usually consider between five and seven factors and do not include more than 15-20 cases.[15]

This had the advantage to produce highly granular results (thus, increasing their possible replicability) but at the same time was a major factor of complexity of the analysis (in terms of a very high number of possible permutations and solutions.) As a consequence, it was a decisive factor for the choice of the software that we used to support a part of our analysis.

Finally, the building of the raw data table was done in Excel while the more advanced parts of the QCA (truth table and analysis of results) were carried out with the help of the software Tosmana (v.1.61),[16] developed by prof. Lasse Cronqvist (University of Trier, Germany).[17] Visualisation of results was done in Excel, as the Venn diagrams often used to visualise the different "pathways to change" could represent a maximum of five conditions – and with difficulty of interpretation.

Quantitative analysis of survey data

This analysis was conducted to represent the main results of the surveys, disaggregated by relevant criteria.

Qualitative analysis of interviews and full-text responses to the surveys

Findings from the interviews were processed with the use of NVIVO from QSR International, which is a qualitative data analysis software package. A coding process allowed the evaluation team to analyse the unstructured texts produced during the evaluation, such as interview notes and full-text responses to the surveys. The interviews and survey responses were categorised by the individual evaluation

[13] In terms of philosophy of the Conventions, Steering and Monitoring/Assessment mechanisms, and modalities of implementation of technical co-operation projects.

[14] Some additional tests were run as well, to better understand some specific aspects of the two areas of work.

[15] For instance, the March 2022 Evaluation of the Council of Europe's work under the sub-programme 'Violence against women and domestic violence' applied QCA to a sample of 13 countries and considered six factors. See here. The World Bank's Group evaluation referred in the publication quoted in footnote 11 considered nine countries and eight factors. The example provided in Michael Baumgartner's Qualitative Comparative Analysis and robust sufficiency (Springer, Quality & Quantity (2022) 56:1939–1963, here) considers 16 cases and four conditions.

[16] This was preferred to two alternative software (QCA add-in for Excel and Kirq) in terms of robustness (the alternatives were not able to manage such complex analyses) and detail of the analysis.

[17] Source available at http://www.tosmana.net.

questions. A thematic analysis of the findings was carried out to understand the findings which were being re-enforced and could subsequently be triangulated.

Analysis of secondary sources

Information and evidence in secondary sources have been exploited to their best during the evaluation. However, as predicted in the Inception Report, several challenges emerged during this part of the analysis, such as the relevance of the evidence to the scope of the evaluation, internal coherence of data, methodologies used to gather data, reliability of the evidence, etc. The same analysis of the projects' own reports revealed indeed that the sources consulted provide helpful information, but are often inconsistent with each other, and their accuracy in reporting achievements is not homogeneous.

4. Contribution to gender and human rights

The contribution of the Council of Europe work under the two areas C-C and THB to gender equality (GE) and human rights (HR) fulfilment was observed from two concurrent perspectives, which are efficiency and effectiveness.

- Under the analysis of effectiveness, sub-question 1.d requires reporting whether any contribution to GE and HR is observed under the two areas of cybercrime and trafficking in human beings.
- Under the analysis of efficiency, the sub-question 2.c requires understanding to what extent GE and HR approaches were mainstreamed in the operations.

Findings to answer the sub-question 2.c were limited; for readers' convenience, the response to these two sub-questions were grouped under 1.d.

5. Evaluation design matrix

An evaluation design matrix (Appendix E) describes the critical elements of the methodology in accordance with each sub-question, defines the key indicators, the primary and secondary information sources, and the methods and tools for data collection and data analysis. It makes clear the attention of the team to plan for the triangulation of both information sources and methods.

6. Tools

The most relevant data collection tools are presented in Appendix F: Sample tools – semi-structured interview guidance and Appendix G: Surveys.

7. Selection of the five countries for field visit

The five countries selected for field visit (Austria, Bosnia and Herzegovina, Morocco, Romania and the United Kingdom) were determined by the analysis of all countries and territories supported by the Council of Europe during the period 2018-2021 through monitoring/assessment and technical co-operation.

It analysed four primary and four additional criteria representing the frequency of interaction with the Council of Europe in both thematic areas (C-C and THB) and both in monitoring and in co-operation. The results of this analysis were presented and discussed with the Reference Group and approved by DIO at the inception phase. The planned visit to the United Kingdom was not carried out because not justified by the low number of people (two) that accepted the interview, one of them after the end of the data gathering process.

Apart from these countries, visits were made to Strasbourg to interview staff of the Council of Europe and to Brussels (where the Team Leader is based) to interview relevant stakeholders.