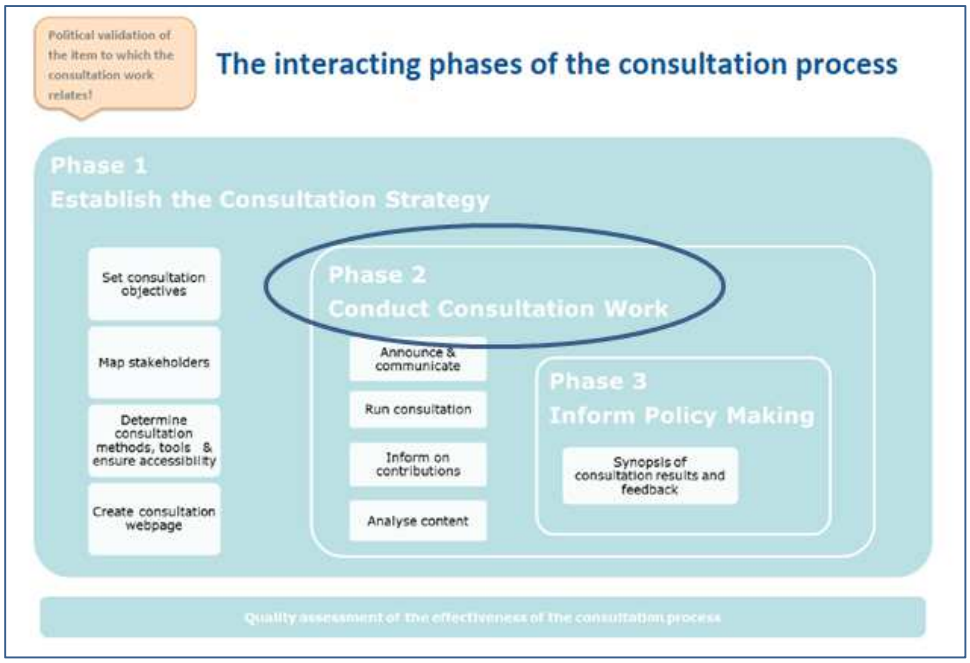
TOOL #54. CONDUCTING THE CONSULTATION ACTIVITIES AND DATA ANALYSIS



While Phases 1 and 3 of the consultation process are carried out only once – at the beginning and at the end – the four elements of Phase 2 have to be considered for each individual consultation activity linked to the specific policy initiative, evaluation or fitness check.

Box 1. Conducting consultation activities – key elements:

- Consultation activities should be conducted in line with the stakeholder consultation strategy.
- Ensure that consultation documents are explicit, clear and understandable, including for non-experts. Avoid use of technical and EU jargon.
- Questions in questionnaires should be relevant, short and simple and be designed in a neutral manner and contain the right balance between open and closed questions.
- Contributions to consultations, both public and targeted, should be published, either with personal information or anonymously, according to the option chosen of the respondent.
- Proper reference need to be made to data protection rules.
- Organisations should be urged to register in the Transparency Register. Contributions received from organisations that choose not to register will be processed as a separate category "non-registered organisations/businesses"[613] unless they are recognised as representative stakeholders via relevant Treaty provisions[614]. A public consultations

[613] See section on stakeholder categories

[614] European Social Dialogue, Art. 154-155 TFEU.

should be publicised on the relevant Europa policy webpage the same day as it is publicised on the Europa 'Consultation Portal' [615]

- Consider sufficient resources for data analysis.
- Reflect well on the questionnaire design: it determines the type of analysis that can be performed on contributions.
- Consider the target audience when deciding on type of graphs and output resulting from the analysis.
- A basic analysis should go beyond the collective results (78% of all respondents agreed that…) and should consider the responses by stakeholder group, country, area of activity etc.
- It is recommended to publish a factual summary report shortly after closing the consultation activity. This report should remain factual and neutral and therefore not contain a qualitative interpretative assessment of contributions, which should be done in the synopsis report later in the process (*see tool #55 on 'informing policymaking-the Synopsis Report'*)

1. ANNOUNCEMENT AND COMMUNICATION OF A SPECIFIC CONSULTATION ACTIVITY

Consultation activities should be prepared as early as possible, and the public - especially the targeted stakeholders – should be adequately informed about the foreseen launch of a consultation activity:

- Update the information on the specific upcoming consultation activity on the policy consultation website[616]. Add concrete dates, agenda and other relevant information. Where useful, e.g. for public internet-based consultations, create a separate subpage.

- In case of a targeted consultation activity, ensure balanced stakeholder participation, use clear and transparent criteria for selection of participants and provide information about these criteria on the policy website.

- Reach out to and invite relevant stakeholder groups to participate in the most effective way. Announce the upcoming event through various communication channels (e.g. press releases, social media[617]), and use networks and other multipliers.[618]

---

[615] https://ec.europa.eu/info/consultations_en

[616] See Tool #53 *The consultation strategy* (paragraph 5 on communication).

[617] Advert e.g. on Twitter or Facebook account of the DG; teaser question to wake interest and link directly to a consultation activity.

[618] Contact e.g. the 500 Europe Direct Centres in the Member States, Representations of the EU in Member States, umbrella organisations of stakeholder groups, SME-Panel or Network of local SMEs.

2. RUNNING A CONSULTATION ACTIVITY

2.1. How to prepare high quality consultation documents and questionnaires

When consulting stakeholders, it is essential to ensure that the documents and questionnaires used in the consultation activities are of highest quality.

There are different conceptual approaches to consulting stakeholders.

- First, one can opt for a **clearly defined and structured list of questions**. This can, for example, take the form of an (online) questionnaire or questions to be asked in person / over the phone.
- Second, one can opt for **more generic approaches**, either by simply requesting general input/views on a topic or by having stakeholders comment on a specific document such as a Commission Communication launching a consultation process or a Green Paper.
- Thirdly, it is also possible to **combine both approaches**, e.g. a generic Green Paper open for general input including embedded structured questions.

While this distinction between structured and generic approach appears similar to the distinction between closed and open questions, there is in fact only a partial overlap. While more generic approaches most often use open questions, structured approaches (from now on "questionnaires") should feature an appropriate mix of both open and closed questions.

This tool aims to provide methodological and practical support for designing both structured and generic consultation approaches. It does not describe or assess different consultation activities in detail.

2.2. Methodological and practical guidance on questionnaires

There is no 'right' answer on how to design a questionnaire. Whether a questionnaire is suitable – meaning likely to deliver the information needed – depends on a range of factors. After having decided to use a questionnaire, this implies choosing an appropriate structure for the questionnaire, designing the questions as clearly and simply as possible, and finding the most appropriate means to administer the questionnaire.

Developing a good questionnaire takes time and preparations should therefore start as early as possible. A good questionnaire increases the quality of answers and, in turn, leads to more impactful input to policymaking. The Better Regulation coordination desks can provide methodological support and procedural information. Further information can be found on GoPro[619].

2.2.1. When to use a questionnaire?

Given the many consultation activities they can be used for, questionnaires can almost always be helpful when consulting stakeholders. Much depends on *how* they are used: If little prior knowledge is available, a questionnaire consisting of mostly open questions

[619] https://webgate.ec.europa.eu/fpfis/wikis/display/REGISTRY/Stakeholders%20consultation

can help to get a better understanding of the issue. If the existing body of knowledge is substantial, closed questions can be used to rank potential solutions and open questions can be used to collect good practices or detailed information. Furthermore, the use of a clearly structured questionnaire often makes it easier to subsequently analyse the answers received.

It is worth considering whether other, more flexible approaches, or approaches that combine questionnaires with policy documents might not produce better results.

Box 2. Strengths and limitations of questionnaires	
Strengths	• Allow collecting information in a structured manner. • Facilitate the analysis of responses (e. g. descriptive statistics provided by EU Survey tool). • May be perceived as less time consuming for respondents, resulting in a high(er) number of contributions.
Limitations	• Do not allow for more detailed input from respondents, as replies to most of the questions are pre-defined. • For open questions –the number of open questions and the length of free text for replies is usually limited. • Depending on the design of the questionnaire, respondents might be pushed into a certain direction and some answers might be excluded in the first place (especially if limited range of responses is offered). • Unless Eurobarometer consultations, results from consultations are not statistically representative: Mainly the active stakeholders will contribute.

2.2.2. Questionnaire design

When designing a questionnaire, start with the **scope** of the questionnaire, as identified in the consultation strategy. What is it that you really need to know from the targeted stakeholders? Only ask those questions that are likely to provide you with the necessary information. Otherwise, try to reformulate or remove the question.

Consider how to meaningfully structure the questionnaire. Only use **sections** that are clear from the perspective of stakeholders. For example, a division into sub-themes is much more meaningful than a division into questions for an evaluation and questions for an impact assessment (in the case of a back-to-back consultation). Such a themed section could, for example, consist of several closed questions and an associated text box for further thoughts and explanations.

When targeting both **expert and non-expert stakeholders**, it might make sense to divide the questionnaire in two parts: the first part would consist of easier, more general questions to be answered by a general public, whereas the second, more detailed part would be addressed to experts. However, even if parts of the questionnaire or specific questions are particularly relevant for certain stakeholder groups, other stakeholders might have relevant input. All questions in public consultations should therefore be open to all stakeholders – also for transparency reasons. Alternatively, consider opting for two different questionnaires.

Every questionnaire should contain an **introduction** which explains – in simple terms – the background and context: What is the initiative about? What is the aim of the initiative? What is the aim and scope of this consultation? In addition, consider beginning

each section with a brief explanatory paragraph, especially when the questionnaire is addressed to non-experts.

As the Commission does not accept anonymous contributions, all questionnaires need to include a **stakeholder identification section**. This section asks for relevant information about the respondent (e. g. which stakeholder category he or she belongs to; contact details for follow-up questions). As many of these questions are used invariably for all questionnaires, the Secretariat-General has prepared a template that can be accessed via GoPro.[620] It is strongly recommended to adapt this template, as appropriate, but to use the same stakeholder identification questions for all consultation activities foreseen for an initiative. This allows comparing results.

A questionnaire is usually a **combination of closed questions** (with pre-defined answers from which the respondent has to choose) **and open-ended questions** (leaving the possibility to the respondent to formulate his/her own answer). The right balance between these closed and open questions depends on the aim of the respective questionnaire.

Closed questions are easier to answer and analyse. They should be mainly used to gather quantitative data. When used to collect opinions, the questions and range of answers should be carefully reflected upon to avoid bias. Open questions should mainly be used to gather qualitative data. They offer stakeholders the possibility to explain their view, to add individual information/concerns, and to refer to issues not yet addressed in the questionnaire. Open questions thus help to get a broader and potentially deeper picture, substantiation of responses and will improve the qualitative assessment of the contributions. A good compromise could consist of using open questions when particularly interested in the views of stakeholders on a particular issue and to cap the length of replies (character limit).

Box 3. Closed versus open questions

	Strengths	**Limitations**
Closed questions	• Suitable to collect quantitative data'. • Quick to answer & analyse • Data can be reported statistically, and answers to various questions cross-tabulated	• Force respondents to choose pre-set answer options (usually tick/circle answers) => Can exclude useful points
Open Questions	• Suitable to collect qualitative data • Allow respondents to give the answers they want in the way they want (open space). • Useful for obtaining insights into the reasons behind the responses to closed questions	• Less suitable to collect quantitative data • Difficult to carry out statistical analysis. • Can be time consuming to code and interpret, particularly if there are many responses in numerous languages.

It is usually **recommended to start a questionnaire and each individual section with simpler, more general questions**. These often take the form of closed questions. They

[620] https://webgate.ec.europa.eu/fpfis/wikis/display/REGISTRY/Stakeholders%20consultation

can then be followed up with more detailed or complex questions. These often take the form of open questions or tables containing a series of closed questions.

It is recommended to **always offer the possibility to submit separate documents** (position papers, background documents) in which the respondent can further clarify positions or views expressed in the responses to the questionnaire.

It is recommended to provide **an indication of the estimated time required to fill in the questionnaire.** It may also be useful to measure the actual time it takes to reply to the questionnaire. The longer it takes, the less likely are contributions from non-experts. In that case, try to reduce the number of questions, simplify the questionnaire or reserve one part of the questionnaire for experts.

It is equally recommended to **pilot** draft questionnaires, i.e. "test" them with selected stakeholders. Asking a few stakeholders to reply to the draft questionnaire and to identify problems – either technical or methodological – can help improving the quality and thus usefulness of the questionnaire.

2.2.3. Question design

In addition to focussing on the design of the overall questionnaire, it is important to ensure that its building blocks – the individual questions – are equally well chosen and designed.

Questions and their answer options should be **relevant, short and simple**. Note that short questions and answer options will also make it easier to present the results in tables and figures in the subsequent analysis.

The **language used should be adequate and adapted** to the stakeholders. If non-experts are targeted, abbreviations should not be used, and jargon should be avoided or, at least, explained. Ambiguous words or questions (e g. double negative) should be avoided in any case. Language should be used consistently throughout the questionnaire. For example, if several questions relate to "the period 2014-2016", all questions should use the same formulation. Deviating from formulation by referring to "the last three years" would, at best, unnecessarily confuse the respondents and might even have a slightly different meaning, thus resulting in different answers.

Questions need to be designed in a neutral manner, meaning that they should not "push" respondents to answer in any particular way. This includes using a balanced answer scale, such as a five-point scale with two positive answer options, two negative answer options and a neutral option.

Answer scales need not only to be balanced, they also need to reply to the question at hand, and need to ensure that respondents can always appropriately answer the question. The latter point is particularly important for mandatory questions as it might otherwise result in a misleading answer. If not all possible answer choices can be foreseen, respondents should be given the possibility to select "other". In addition, it is often useful to allow for an "I do not know" and/or "Not applicable" option as well as providing respondents the possibility to add further comments or explain their answers in a text box. In addition, respondents should always be given the opportunity to upload documents, to accompany the responses to the questionnaire.

2.2.4. Online questionnaire tools

While questionnaires can be used for a variety of consultation activities, many are used for internet-based stakeholder consultations. The Commission has therefore developed an online tool, EU Survey[621]. The functionalities of this tool are constantly being improved; an up-to-date overview is available online.[622]

Given that EU Survey has been developed with the needs of the Commission in mind (e. g. its interface is available in all official EU languages; it fulfils high standards of data protection; it complies with international accessibility standards) and given that the EU Survey team can be contacted to prevent or solve technical problems[623], it is recommended – though not obligatory – to use this tool. There are many commercial alternatives which offer similar functionalities[624].

2.3. Methodological and Practical guidance on generic consultation approaches

2.3.1. When to use generic approaches?

In some situations, relying on questionnaires might not produce the necessary results. More general approaches can – sometimes in combination with a short list of guiding questions – be useful alternatives. Requesting general comments from stakeholders or having stakeholders comment on a policy document such as a Green Paper or a Communication launching a consultation process might help to avoid/mitigate the bias inherent in questionnaires. They can also be useful for starting a comprehensive debate in a policy area.

2.3.2. Green Papers

Green Papers are documents published by the European Commission to stimulate discussion on given topics at European level. They invite all stakeholders to participate in a consultation process and debate on the basis of the ideas and suggestions they put forward. They are published on the Europa 'Consultation Portal'[625] and open for stakeholder input for at least 12 weeks.

2.3.3. Commission Communications launching a consultation process

A Communication launching a consultation process is a consultation document in the form of a Communication adopted by the College. The same rules apply as to Green Papers.

---

621 https://ec.europa.eu/eusurvey/

622 https://ec.europa.eu/eusurvey/home/about

623 Commission staff can also send questions to EC-HELPDESK-IT@ec.europa.eu

624 See, for example, http://survey-software-review.toptenreviews.com/

625 https://ec.europa.eu/info/consultations_en

Examples of such Communications include the Communication from the Commission to the EP and Council concerning a consultation on fishing opportunities for 2016 under the Common Fisheries Policy[626] or the Commission Communication to the Council, EP, CoR and EESC launching a public consultation on the EU Urban Agenda[627].

2.4. Further reading & references

Useful resources are available at the Commission's library and online - a brief selection:

- Creasy, Barry (2008), Effective Surveys and Questionnaires, The Consultation Institute, Biggleswade.
- Fowler, Floyd J. (2014), Survey Research Methods, Sage, Thousand Oaks.
- Fowler, Floyd J, (1995), Improving Survey Questions. Design and Evaluation, Sage, Thousand Oaks.
- Hague, Paul (1993), Questionnaire Design, Kogan, London.
- OECD (2012), Measuring Regulatory Performance. A Practitioner's Guide to Perception Surveys, OECD, Paris.[628]

3. Publication of responses, Data protection, Access to Documents and Transparency Register

3.5. Publication of responses

The Commission is committed to be open and transparent throughout the policy cycle, including in the way it consults its stakeholders. Therefore, it is strongly recommended that contributions[629] submitted in the context of the various consultation activities, public or targeted, are published on the relevant policy webpages.[630]

For all consultation activities, public or targeted, respondents should be offered the option to have their contributions published either with their personal data or anonymously. Regardless the option chosen, respondents should be required to identify themselves or the organisation on which behalf they respond. Anonymous contributions to consultations should not be accepted. The option for respondents not to have their contribution published is no longer offered by default.

For activities that collect input in writing, the options for publishing the contributions with or without personal data should be clearly mentioned in the consultation document

---

626 http://ec.europa.eu/dgs/maritimeaffairs_fisheries/consultations/fishing-opportunities-2016/doc/com_2015_239_en.pdf

627 http://ec.europa.eu/regional_policy/sources/consultation/urb_agenda/pdf/comm_act_urb_agenda_en.pdf.

628 Available online at www.oecd.org/gov/regulatory-policy/perception-surveys.htm.

629 Contributions include responses to questionnaires, position papers, background material, etc.

630 See Tool #53 on *The consultation strategy*.

(e.g. questionnaire). For oral input, such as interviews, the way the contributions[631] will be published must be made clear beforehand (e.g. for interviews, before the start of the interview)

Publication of the contribution with personal information

Contributions are published together with key personal information, including the name of the respondent and the country in which the respondent resides. In case the respondent replies on behalf of an organisation or company, only the name of the organisation/company and country of residence of the organisation/company is published together with the contribution. Any other personal data which may be collected (e-mail, phone number, address, gender, etc.) should not be made public, unless relevant.

Anonymous publication

Contributions are published without any personal data provided in the context of the consultation. However, for practical reasons, documents submitted by stakeholders in the context of a consultation, such as position papers or background documents, can be published in the way they are received. Removing personal data from such documents can be cumbersome and time consuming. Therefore, **it should be clearly mentioned on the consultation webpage or in the questionnaire or feedback form that respondents should not include personal data in documents submitted in the context of consultation if they opt for anonymous publication**.

Publication of ad hoc contributions

If stakeholders provide ad hoc contributions at any point during the policy preparation or evaluation work, these contributions should also be published on the policy web page. If no information on the preferred format of publication is available, by default it should be published with the key personal information (see above).

If manageable, DGs could for courtesy reasons get back to stakeholders and ask them about their preferred form of publication (with or without personal information).

3.6. Data protection

Under EU law, personal data can only be gathered under strict conditions and for a legitimate purpose. Furthermore, persons or organisations, including the EU institutions, which collect and manage personal information, must protect it from misuse and must respect certain rights of the data owners which are guaranteed by EU law, in particular,

Regulation (EC) No 45/2001. Both apply to the processing of personal data by EU institutions and bodies within the scope of Union law.

What is understood by personal data?

According to Article 2 (a) of Regulation (EC) No 45/2001 personal data is defined as follows: "Any information relating to an identified or identifiable natural person, referred

[631] Information should also clarify how the provided responses will be published (summary or complete responses)

to as "data subject" - an identifiable person is someone who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his or her physical, physiological, mental, economic, cultural or social identity.

Privacy statement

By means of the privacy statement, respondents should be informed in a clear way on how data is collected and processed. This document describes the objective of the personal data gathering and processing, the kind of data collected, technical information on the tools or platforms used to store and process data, to whom the data can be disclosed, the way data is protected, the period data is kept as well as contact information. In practice, a specific privacy statement needs to be prepared for each consultation activity involving collection of personal data and should be published on the consultation webpage related to the initiative. Furthermore, a link to the 'protection of personal data' page needs to be provided on the consultation page. The template to be used for the privacy statement for consultations is available on GoPro[632]

Data retention Period

Personal data should be kept only for as long as follow-up actions to the Consultation are necessary with regards to the purpose(s) of the processing of personal data. All personal data should be deleted from databases 5 years after the last action in relation to the Consultation. Where necessary, personal data could be kept for a longer period as long as this is foreseen in the Privacy Statement. Consultation Reports containing personal data should be archived according to the Commission's legal framework (e.g.: SEC(2012)713 - Common Commission-Level Retention List for European Commission Files (CRL) of December 2012). Participants must be informed of the fact that they can request their personal data to be deleted."

3.7. **Access to Documents**

Contributions, including personal data provided, may be subject to a request for access to documents under Regulation (EC) No 1049/2001 regarding public access to European Parliament, Council and Commission documents ('Regulation 1049/2001')[633]. Regulation 1049/2001 provides any EU citizen and any natural or legal person residing or having its registered office in a Member State the right of access to documents of the EU institutions, subject to principles, conditions and limits defined in the Regulation. If access is requested, the request is subject to a case-by-case analysis based on Regulation 1049/2001 in order to assess the applicability of the exceptions defined in its Article 4, taking into account the legitimate interests and the justifications of non-disclosure in case provided by the author of the contribution. Where disclosure of the contribution, or parts thereof, would undermine the protection of commercial interests of a natural or legal person, the institutions shall refuse access in accordance with Article 4(2), first indent of Regulation 1049/2001.

---

632 https://webgate.ec.europa.eu/fpfis/wikis/display/REGISTRY/Stakeholders%20consultation

633 Official Journal L 345 of 29.12.2001.

3.8. Transparency Register

Organisations and businesses that wish to participate in consultation activities are asked to provide the Commission and the public at large, with information about which interests they represent and how inclusive their representation is, by subscribing to the Transparency Register. Contributions from organisations and businesses that choose not to register will be processed as a separate category "non-registered organisations/businesses"[634] unless they are recognised as representative stakeholders via relevant Treaty provisions[635].

Publishing a public consultation on the Europa 'Consultation Portal' [636]' or a roadmap or inception impact assessment on the dedicated webpage[637] will trigger an e-mail alert to registered organisations.

More info on the Transparency Register can be found on Europa.[638]

4. FACTUAL SUMMARY REPORT

For each consultation activity, it is good practice to publish factual information on the input received from stakeholders to ensure transparency. Apart from the publication of meeting agendas, consultation documents and any written contribution, this also includes a factual summary of the issues raised. This information can take different formats, e.g. workshop summary, meeting minutes, conference report, presentation of key issues. Basic statistical information on participating stakeholder groups, number of participants, geographical distribution and other basic figures relevant for an activity should be provided.

The purpose of this information on the stakeholder input is to give an overview on 'what has been said'. It should be neutral as it precedes the analysis and interpretation of consultation results. As these factual summaries may contain views and positions from stakeholders not necessarily shared by the Commission, or may refer to issues on which a decision has not yet been taken, an appropriate disclaimer[639] should be added.

It is recommended to publish this factual information soon after the closure of a consultation activity on the consultation page. No specific formal requirements apply.

---

634 See section on stakeholder categories

635 European Social Dialogue, Art. 154-155 TFEU.

636 https://ec.europa.eu/info/consultations_en

637 http://ec.europa.eu/info/law/better-regulation/initiatives

638 http://ec.europa.eu/transparencyregister/public/homePage.do

639 Disclaimer: "This document should be regarded solely as a summary of the contributions made by stakeholders [add consultation activity] on the [add title of policy initiative or evaluation or fitness check]. It cannot in any circumstances be regarded as the official position of the Commission or its services.

The factual summary should not be confused with the synopsis report, to be drafted at the end of the consultation process[640] (see tool #55 Informing policymaking - the synopsis report)

Box 4. Factual summary report	
Give a concise and balanced overview of contributions received during a specific consultation activity	
Give factual information on input received	• Who contributed? • Whom are they representing? • What aspects are addressed? • What are their views and concerns? • Which communication channels were used for contributions?
Stay neutral	• Document the input as received: • Avoid qualifying it, taking position or giving feedback
Aggregate at an appropriate level	• Cluster information
Inform on the process	• Inform on what was done so far in terms of consultation activities and on the next steps
Add Disclaimer	• Emphasise that the contributions received cannot be regarded as the official position of the Commission and its services and thus does not bind the Commission.

5. DATA ANALYSIS OF CONTRIBUTIONS TO QUESTIONNAIRES USED IN CONSULTATIONS[641]

5.1. Overview

There is no 'right' answer to the question how to best analyse data input for questionnaires, which often consists of a mix between 'open' and 'closed' questions[642]. Rather, there is a wide a range of factors to consider in order to ensure that a reasonably robust analysis can be performed within the constraints of available time and resources, and considering that the number of responses may be higher than expected.

The most efficient method is likely to involve transferring the data to a 'master' Excel spreadsheet containing responses to both 'closed' and 'open' text questions. With this in mind, it becomes obvious that a consultation which invites broad ranging submissions (including position papers) in the form of pdf documents will be difficult to process.

---

640 See Tool #55 on *Informing policymaking - the synopsis report.*

641 For further detail see also Commission study [Consultation Support and Development of Advice (Specific Contract No SG/2015/10 under Framework Contract ENTR/172/PP/20-12-FC Lot 3) : https://bookshop.europa.eu/en/consultation-support-and-development-of-advice-pbKA0217018/?CatalogCategoryID=YR4KABstrdkAAAEjLocY4e5K

642 See paragraph 2 of this tool on questionnaire design.

For the purposes of this tool, it will be primarily focused on two levels of analysis:

Basic analysis, which can be undertaken by those with a reasonably proficient knowledge of Excel.
Advanced analysis, which can be undertaken by those with specific skills to use specialised software aimed at assisting with the analysis of data and campaigns and with computer-aided analysis of open text responses.

In certain cases, it may be desirable to outsource the entire package (questionnaire design, analysis and reporting) to a ***professional*** contractor (polling or market research company.)

Note that when reporting back on the outcome of the consultation the methodologies and tools should be explained for transparency reasons.

Box 5: Overview different levels of analysis		
Approach	**Advantages**	**Disadvantages**
Basic	• Only basic spreadsheet skills required • Good for analysis of closed questions • Can be done in-house by most Commission Policy Officers with/without support from consultants	• Not efficient for high number of responses (several hundred or more), particularly when analysing campaigns and open text responses
Advanced	• Efficient means to analyse campaigns and open text responses where there are hundreds (or thousands) of responses	• Requires use of specialised software. • As such requires suitable Commission in-house staff with/without support from consultants
Professional	• Professional questionnaire design • Independent analysis • High quality presentation of results	• Potential for limited interaction with Commission Policy Officers • Approach may have to conform to a standardised 'template' with limited open text responses

5.2. Data preparation in view of the analysis

5.2.1. Data familiarisation

Once the data is on a master spreadsheet, there are two considerations to be taken into account:

- time and resources for analysis of **closed questions** do not depend on the number of responses
- time and resources for analysis of **open questions** depend on the number of responses and, to a lesser extent, to the diversity of languages in which they were submitted

Before proceeding further with the analysis, it is important to note that the **data represents the views of those that responded**. The **respondents** are self-selecting and **are not a statistical sample of the EU population**[643].

However, in some cases the respondents may represent a very high percentage of the population of particular stakeholder groups that are directly impacted by the subject of the consultation. By way of example, potential changes affecting particular industry groups may trigger responses from all the relevant manufacturers.

5.2.2. Data cleaning and duplicates

Once processed and organized, the data may be incomplete, contain duplicates, or contain errors. The need for data cleaning will arise from problems in the way that data is entered and stored. Data cleaning is the process of preventing and correcting these errors. Common tasks include record matching, identifying inaccuracy of data, overall quality of existing data, deduplication, and column segmentation. Such data problems can also be identified through a variety of analytical techniques. For example, with financial information, the totals for particular variables may be compared against separately published numbers believed to be reliable. Unusual amounts above or below pre-determined thresholds may also be reviewed. There are several types of data cleaning that depend on the type of data such as phone numbers, e-mail addresses, employers etc. Quantitative data methods for outlier detection can be used to get rid of likely incorrectly entered data. Textual data spellcheckers can be used to lessen the amount of mistyped words, but it is harder to tell if the words themselves are correct

The first step of the data analysis is to simply check the validity of the data on the master Excel sheet. Responses received before the consultation started should be deleted – as these will most likely be associated with final testing and checking of the questionnaire. Responses received a few hours after the formal closure time could be accepted if there may have been valid reasons for the delay. Clearly, responses received days/weeks after the consultation has closed can be deleted.

At this point, every response should be given a unique identifier (which could just be simple numbering 1, 2,...435).

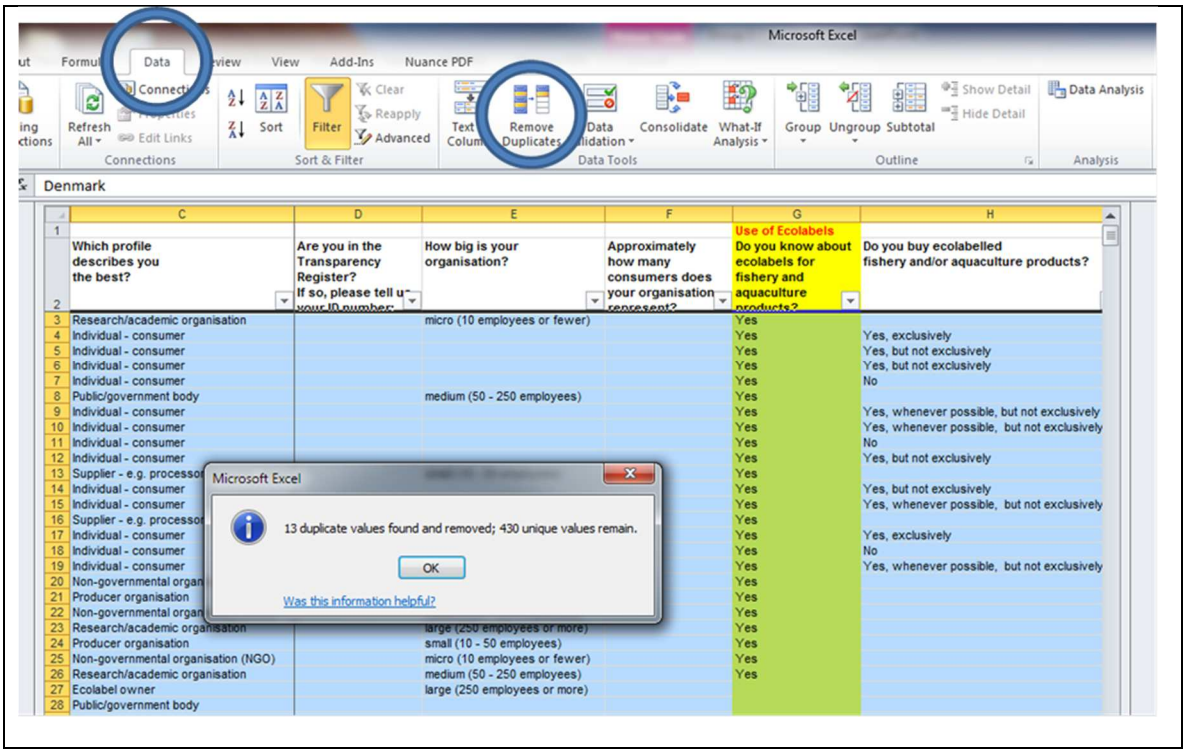
Mischievous (or mistaken) entries can be checked and validated if considered necessary (for example, individuals claiming to represent a national Government). However it is not possible to readily validate every response – particularly if there are thousands of responses to consider.

Duplicates are identical entries across all the questions (including name and location). Such entries may be entered deliberately or accidentally. The first step is to determine how many duplicates there are.

Excel has a 'remove duplicates' data tool - see example below (using Excel 2010):

---

[643] See also paragraph 5.3.1, sub para: *Interpretation of data - Weighting and representativeness of respondents and replies*



In this example, Excel found 13 duplicates in 443 submissions. Note that the Excel tool is not case sensitive (so that DAVID BROWN = David Brown) and there is a chance that more than one David Brown completed the questionnaire in the same way. Therefore, identifying the duplicates requires manual checking of the Excel sheets before/after the removal of duplicates.

A more advanced approach is to use STATA (a statistical software package used for more 'advanced analysis' – see below which readily groups the duplicates allowing you to determine quickly whether duplicate entries may be a range of common names or whether it is an obvious deliberate multiple entry as illustrated by the example below.

Box 6. Use of STATA (statistical analysis)

```
. duplicates report

Duplicates in terms of all variables

--------------------------------------
   copies | observations       surplus
----------+---------------------------
        1 |         1631             0
        2 |          264           132
        4 |            4             3
       14 |           14            13
       31 |           31            30
--------------------------------------

. duplicates drop

Duplicates in terms of all variables

(178 observations deleted)
```

In this example, there were 1,631 unique entries within the dataset and 178 duplicate ('surplus') responses representing 10% all responses. Notably, there was one response that was repeated 30 times (i.e. one 'master' and 30 copies leading to a 'surplus' of 30 entries) and another repeated 13 times. These can safely be deleted as obvious duplicates. There were also 132 pairs of identical answers as well as one with four identical entries. In this example, these duplicates were highlighted (within STATA) and the names reviewed (manually) to see if there was any possibility that these were cases of genuine duplicates (i.e. people with the same name). In this particular case, it was immediately apparent that these were duplicate entries and could be safely deleted. Removing all 178 duplicates, using STATA's *duplicates drop* command, yielded a cleaned dataset with 1,766 individual responses.

Box 7: Summary procedure for considering duplicates

- Identify the level of duplicate responses (anything over 1% is probably indicative of duplicates);
- Remove 'obvious' duplicates;
- Review and perhaps remove remaining duplicates;
- If in doubt, leave duplicate entries in place (as their overall impact on the results will be low).

5.2.3. Campaigns

Overview

Where respondents have responded to a public consultation with the same answers this may be coincidence or it may part of a co-ordinated campaign. **Campaigns are very effective in order to generate interest amongst stakeholders and to highlight key messages for policy makers. At the same time, they present a challenge for those**

analysing the responses to an public consultation. It is therefore essential to well identify campaigns, analyse them separately and present results adequately.

It is therefore necessary to consider the possible presence of campaigns, the means to identify them and how to present the results.

Presence of campaigns

Once a public consultation is launched, it should be continuously monitored. As such, occasional searches on the internet and social media may reveal the presence of organised campaigns which are suggesting answers to the questionnaire.

Once the consultation has finished, identifying campaigns through this method will become less effective as the information is changed/removed or simply overtaken by new events.

Identifying campaigns – basic analysis

Where there are only 100 or less responses to an public consultation, it is possible to sort the Excel data set by responses to successive questions and then check them by scrolling through the responses to identify rows of identical entries. Where these are the **same across all closed questions**, this suggests a campaign – particularly if the respondents represent a particular sub-group of stakeholders (by activity and/or interest and/or location).

As a rule of thumb, the minimum threshold should be 10 or more identical responses (across all the closed questions) to count as a 'campaign'. On the other hand, if there were 10 identical responses from very diverse groups of respondents to a short questionnaire with a total of 10,000 responses, this would rather be a coincidence.

Identifying campaigns – advanced analysis for closed questions

Although Excel can be used to assist with the identification of campaigns, it is more efficient to use professional statistical software such as STATA[644]. Professional software is more complex, compared to Excel, and does require someone with training or prior knowledge of the programme in order to use it. The output tables usually also require some explanation/basic understanding of statistics in order to understand the results. Furthermore, it may not be possible to easily export the outputs from this statistical software into Excel or any other programme.

Available statistical software at the Commission can be consulted on the webpage of DIGIT[645].

The analysis may also be outsourced to a contractor that may have access to similar software.

---

644 There are various other well-known statistical packages which can provide additional functionality beyond that provided by Excel, including: R, MiniTab, SAS, SPSS, etc. A brief introduction to the 'top five' may be found here: http://www.prostatservices.com/statistical-consulting/articles-of-interest/a-review-of-the-top-five-statistical-software-systems

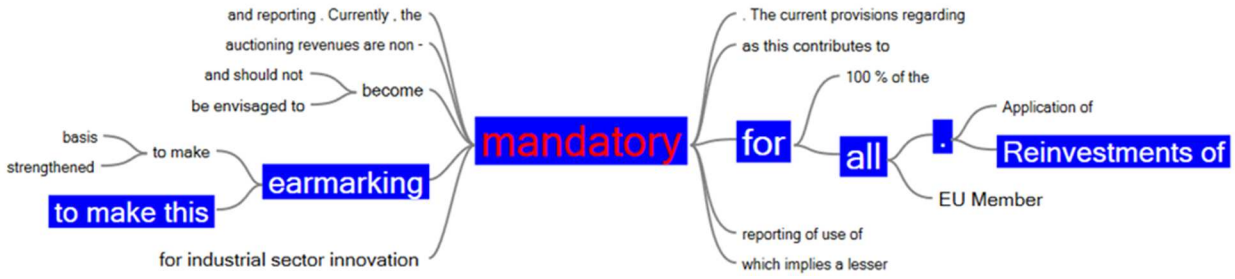
645 https://myintracomm.ec.europa.eu/dg/dgt/it_support/software/Pages/index.aspx

Identifying campaigns – advanced analysis for open questions

While some campaigns suggest a series of responses to closed questions, other campaigns may recommend that their supporters should adopt some standard text in their response. As such, the responses may not be exactly the same but some key messages will be repeated.

The most efficient way to identify the presence of campaigns in responses to open questions is to use software designed for qualitative data analysis such as the Commission tool 'Doris' or the commercial tool NVivo, but there are also other similar tools available[646].

An example of a 'word tree' around the word 'mandatory'



It is immediately apparent that is a campaign, as indicated by the larger font (reflecting greater frequency of occurrence) which includes the suggested word sequence "*…to make this earmarking mandatory for all. Reinvestments of…*". Perhaps the easiest way to identify responses from this campaign would be to search the (cleaned) dataset for "*earmarking mandatory*" and then segregate these responses.

Identifying campaigns – advanced analysis for all questions

Although professional software may be used to look for duplicates across all fields, this may not be efficient. For instance, analysis across closed questions may yield a campaign supported by a particular stakeholder. However, the wording used in the supporting comment boxes may vary slightly due to differences in use of capital letters, mistyping, etc. As such, if the search for campaigns would extend across both closed and open questions, many campaign responses may be missed.

Segregating campaigns and reporting

It is recommended to look for campaigns in both 'closed' and 'open' questions. Once campaigns have been identified, the associated responses should be segregated and analysed separately from the non-campaign responses.

If campaigns are identified, they should be referred to in the synopsis report. Reporting on campaigns should include the number of respondents supporting the campaign as well as a summary of their points of view – either in text or tabular form

---

646 http://www.predictiveanalyticstoday.com/top-qualitative-data-analysis-software/

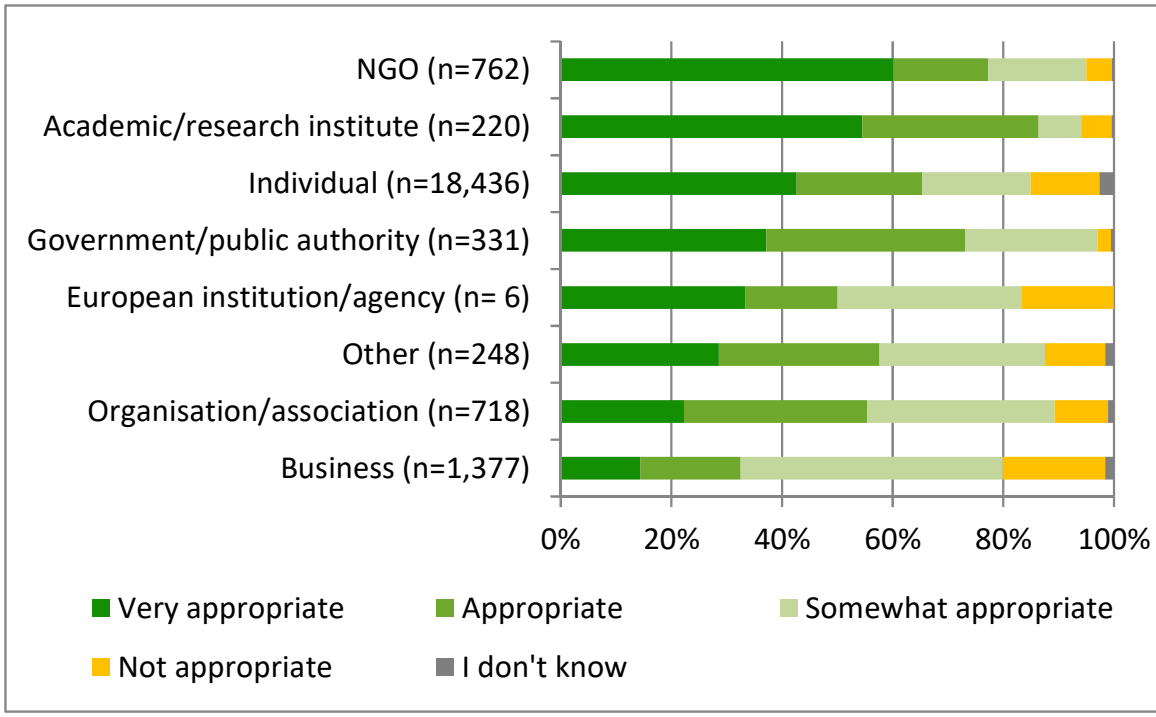
5.3. Analysis of data

5.3.1. Analysis of closed questions

Basic analysis

Basic analysis of closed questions generates information such as: "*67% of respondents considered that the legislative framework was delivering benefits."* Such information is not so helpful to policy makers unless qualified by the perspective of the stakeholder, for example: "*Although 74% of industry respondents considered that the legislative framework was delivering benefits, only 32% of citizens agreed with this view.";* or "*Less than 30% of Danish respondents considered that the legislative framework was delivering benefits, while 67% of Estonian respondents agreed with this view*." Even this information is of limited value if there were only 6 respondents from Estonia while there were 240 from Denmark.

Such quantitative information can be conveyed graphically by including the numbers of respondents as illustrated below.



Response to the question ' Is the approach set out in the Birds and Habitats Directives an appropriate way to protect species and habitats in the EU?' by stakeholder group
(based on data from the consultation on the fitness check of the Birds and Habitat Directives[647])

Basic statistical terms include:

- Mean: the total of a distribution of values divided by the number of values
- Median: the mid-point in a distribution of values

[647] http://ec.europa.eu/environment/nature/legislation/fitness_check/index_en.htm

- Mode: The value that occurs most frequently in a distribution
- Standard Deviation: a measure of dispersion around the mean
- Percentages: A rate, number, or amount in each hundred to express any proportion or share in relation to a whole
  - When to report percentages: When values are high enough for them to mean something. It is generally bad practice to report percentages if the total number of values is lower than 100, as a percentage point bigger than > 1
  - When reporting changes over time, the difference between percent and percentage points (p.p.):
    - *Percent* is used for a measure of changes in values
    - *Percentage point* is used for a measure of change in percentages

    e.g.:
    - Last year, in a workforce of 300, 30 people (10%) were smokers.
    - This year, in the same workforce of 300, 15 people (5%) are smokers
    - The number of smokers has fallen by 50% or the percentage of smokers has fallen by 5 percentage points
  - It is **good practice to calculate and report percentages *and* valid percentages** (percentages of those who answered the question) so that readers can see response rate on questions.
  - **Avoid using only percentages in the presentation of results**, make always the link with the amount of responses they correspond to.
    - Example: Q: Do you receive a disability benefit of any kind? Yes: 83 No : 256; => out of a total of 460 who returned a questionnaire (=N)

(N=460)	N	%	Valid
Yes	83	18.04	24.50
No	256	55.65	75.50
Not answered	121	26.31	-

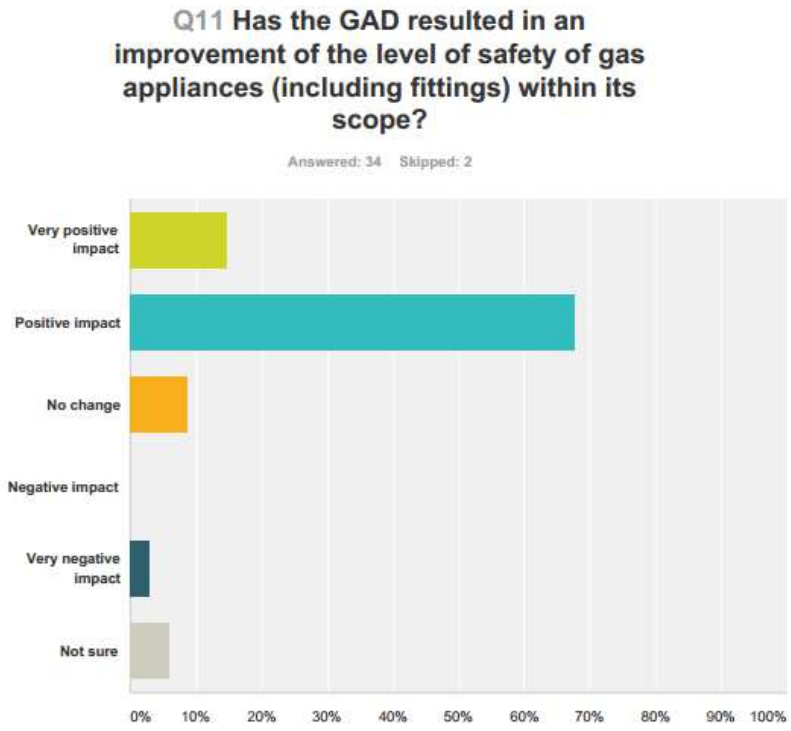
Level of analysis

The analysis of closed questions is relatively straightforward, with Excel able to generate relevant data tables and graphs. However, transferring the results into a report can be time-consuming.

If there are 20 closed questions, then there will be a minimum of 20 tables/graphs representing the answers against another variable – usually stakeholder group (such as authorities, citizens, industry, etc.) and the related comment. There will then be further

tables/graphs if the closed questions incorporate a number of sub-questions and/or more than one possible answer.

If the analysis is to be repeated from another perspective (such as geographical location), another set of tables/graphs would be required. Through specific software (such as EU survey or SurveyMonkey) pre-filtered graphs and tables can be generated as illustrated in the example below, which represents the views of 'manufacturers'.



Answer Choices	Responses	
Very positive impact	14.71%	5
Positive impact	67.65%	23
No change	8.82%	3
Negative impact	0.00%	0
Very negative impact	2.94%	1
Not sure	5.88%	2
Total		34

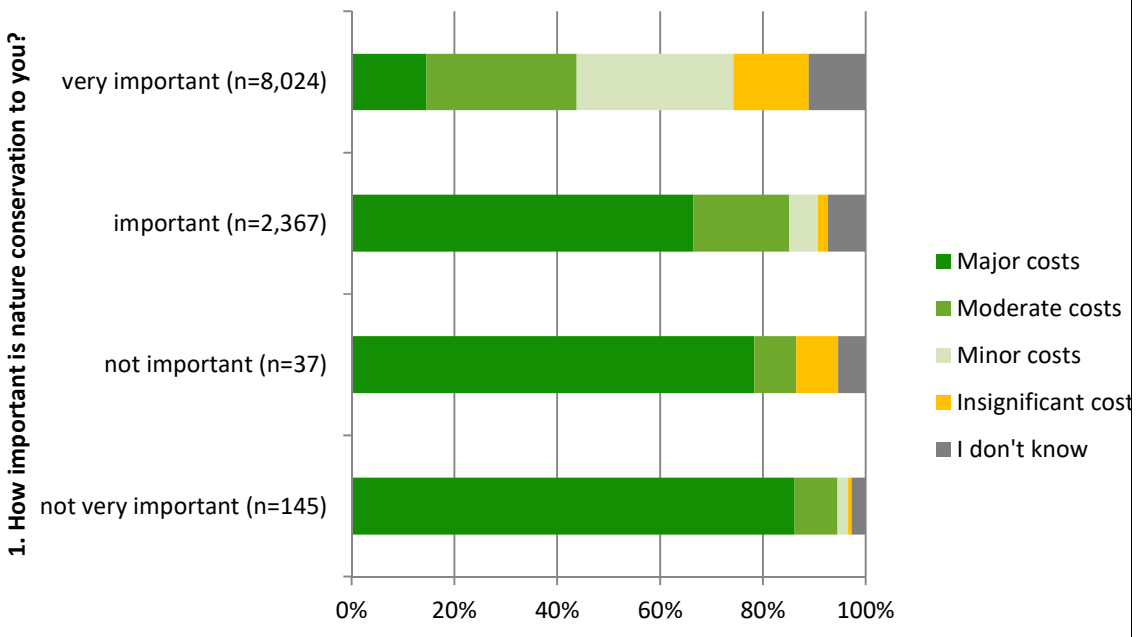
Based on data from the Public consultation on the Revision of Regulation (EU) No 994/2010 concerning measures to safeguard security of gas supply and repealing Council Directive 2004/67/EC

More advanced analysis

Professional polling companies as well as **Eurobarometer**[648] surveys are specialised in presenting the views of the general population on a great variety of issues. Where the **responses** to particular questions or survey respondents **are drawn from a representative sample**, it may be possible to apply additional analysis of the responses in order to determine the statistical significance of the conclusions – particularly if the responses are inferred to apply to a much wider population (such as that of a particular country or the EU as a whole).

This situation does not apply to the results of a **public consultation** due to the self-selection of respondents, which means that **the responses are not drawn for a representative sample**. While it is accepted that advanced statistical analysis has an important role for some types of analysis, such statistics provide limited added value to results from a self-selecting (i.e. non-randomly selected) sample and can potentially be misleading. In other words, statistics provide little additional information (of use to the policy maker) beyond that obtained by the analysis presented here.

However, this should not suggest that further analysis beyond the 'basic' analysis cannot be undertaken. By way of example, it may be useful to analyse the relationships between different questions as illustrated below.



21. How significant are the costs associated with the Directives?: Costs of protecting species other than birds

Comparison of responses to Q1 with those to Q21 ***(from the public consultation on the Birds & Habitats Directives)***

[648] http://ec.europa.eu/COMMFrontOffice/PublicOpinion/

Interpretation of data - Weighting and representativeness of respondents and replies

There is a fundamental difference between a survey, such as Eurobarometer[649], and public consultation (see former paragraph). **Data gathered through public consultation does not provide a representative view of the EU population**.

Weighting of data is a statistical technique of making answers count for more or less to ensure they are representative of the population. You can only use this technique if you have an understanding of the demographic make-up of the population and returns. It is generally very difficult to get this understanding and therefore **it is not recommended** to apply weighting techniques **for the analysis of data from public consultations.** If you need to have representative views, other tools, such as Eurobarometer, should be considered.

Note that when reporting back on the outcome of the consultation the methodologies and tools should be explained for transparency reasons, including if weighting techniques have been applied.

Box 7. Interpretation of data – key aspects

Consultations aim to gather evidence, which is used as input for policy preparation and contributes to informed decision-making. It is therefore **essential to provide the right context of the consultation when presenting the outcome**, including information on who participated and whom respondents represent:

- When analysing[650] and presenting the results, **distinction should be made between the different stakeholder categories** that contributed to the consultation. A short description should be provided about the different stakeholders (background, whom they represent, etc.)
- Do, preferably, the '**Stakeholder Credibility test**' and consider its outcome in the analysis:
  - *Longevity*: Has the stakeholder organisation been established long enough to acquire the wisdom in the policy field?
  - *Expertise*: How well does it know the subject matter?
  - *Representativeness*: Who exactly does it represent and how well does it do so?
  - *Track record*: How useful/credible has its contribution been in the past?
  - *Reputation*: How seriously do other people take this organisation?
- Contributions from **citizens** should be analysed as a separate stakeholder category
- **Campaigns should be identified** and the relevant responses should be segregated, analysed and presented separately from the non-campaign responses (see para 2.4)
- **Avoid using only percentages** when presenting results; they should be linked to the corresponding amount of responses (see para 3.1.1).

649 http://ec.europa.eu/COMMFrontOffice/publicopinion/index.cfm

650 See also better regulation Guidelines, chapter VII

5.3.2. Analysis of open questions

Overview

Textual input to open questions is considered as qualitative data, which is, compared to quantitative data rich and complex and therefore it cannot be treated statistically. However, this does not mean that systematic and rigorous analysis techniques cannot be applied. Qualitative data, more than quantitative, is extremely prone to bias, and systematic analysis helps prevent this.

Basic Analysis

Under the approach to basic analysis, responses would most commonly be grouped into broad stakeholder groups (typically citizens/NGOs, authorities, industry, others). Under the simplest approach, responses from a particular group for a particular question could then be quickly read to get an overview of the two or three most recurrent points being made.

Coding of qualitative information
Coding is a technique that allows qualitative information to be categorised/sorted/interpreted • A coding frame is constructed from the first 50 or so responses, and subsequently modified as more responses come in • A coding frame is a set of headings under which comments/texts may be placed to categorise it; the headings may be free-standing or 'nested' (with different levels, e.g. car driver -> driver of a motor vehicle -> road user • Text responses are then read and each piece of text is assigned ('coded') to one or more headings • Code frames may apply to individual questions, but (more often) apply across a whole submission

Depending on the nature of the question, one might expect **the analysis to typically yield five to ten themes from the first 20 to 50 responses**. These themes should be noted and the frequency of occurrence in subsequent response should be recorded. Thereafter, experience suggests that fewer themes will be found and, indeed, the rate of reading of responses may increase as the reader becomes familiar with the range of points being made.

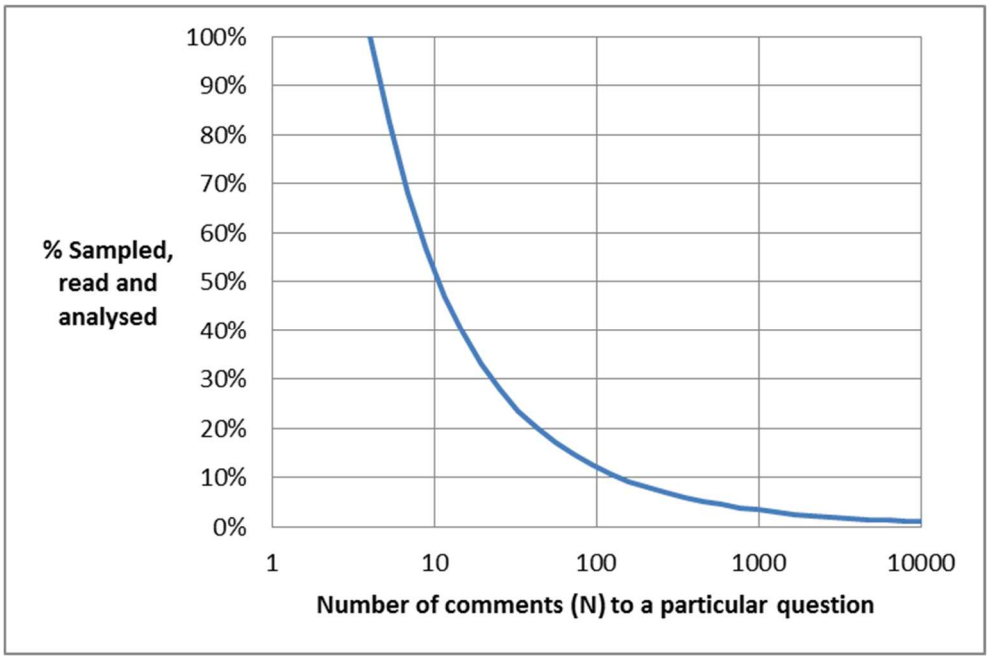
This needs to be preferably done in the languages in which the responses have been provided. In some cases, it may be easiest to translate all responses into a single language and then analyse the sample. In other cases, it may be desirable to review the responses in their native language and extract key themes. In this approach, it is preferable to run languages sequentially to avoid similar but different themes emerging for each language which can lead to confusion. In practice, the precise approach adopted will depend on the number and nature of responses and the relevant language skills.

However, as a guide, **once there are more than 500 answers to consider** (e.g. 100 responses x 5 comment boxes or 50 responses x 10 comment boxes), **it may be necessary to use a more resource-efficient method** as outlined below in the advanced analysis.

Resource estimate for consideration of open text contributions
Experience suggests that **detailed consideration of significant contributions**[651] **takes time – about five answers per hour**. For 50 responses each with 5 significant comment boxes completed will require (50 x 5)/5 = 50 hours, i.e. more than one person/ week.

Sampling for the Advanced Analysis

The more resource-efficient method involves **a combination of reading** a sample of responses **and** then using **advanced software** to analyse all responses. In developing this guidance, consideration was given to the results of sampling responses to 11 different Commission public consultation from which it appeared that **a sample size of** $\sqrt{N}+2$[652] (i.e. the square root of the number of responses, N, plus two) **would typically yield five to ten themes.** In some cases, the computer analysis would reveal a further theme. This seems a reasonable balance between the resources required for the sampling and the associated results. As illustrated in the chart below, where the number of comments being considered is relatively small, a significant percentage will be sampled, read and analysed. For 100 comments, 12 (12%) will be sampled and for 1,000 comments 34 (3.4%) will be sampled.



Sample size ($\sqrt{N}$ + 2) as a % of comments read vs number of comments (N)

[651] A 'significant contribution' has been taken to mean, typically, half a page of typed text (in any EU language) from which key themes will be identified and recorded.

[652] Empiric determined formula; See Commission study [Consultation Support and Development of Advice (Specific Contract No SG/2015/10 under Framework Contract ENTR/172/PP/20-12-FC Lot 3) : https://bookshop.europa.eu/en/consultation-support-and-development-of-advice-pbKA0217018/?CatalogCategoryID=YR4KABstrdkAAAEjLocY4e5K

Procedure for computer-aided analysis

The approach for this analysis includes the following steps:

1. **Assigning language markers**. To assist with the analysis, it is important that all responses are given a language marker. This, in turn, leads to the suggestion that if the questionnaire(s) contains comment boxes – it is recommended **to include a final question asking which language has been used for the text boxes** as this makes a significant difference to the time for subsequent analysis (particularly if there are thousands of responses)

2. **Review a sample**. For a particular question – read a sample of √N+2 responses and extract key 'themes'. By way of example, if there were 68 responses, a sample of √68 = 8 + 2 = 10 would be read and analysed.

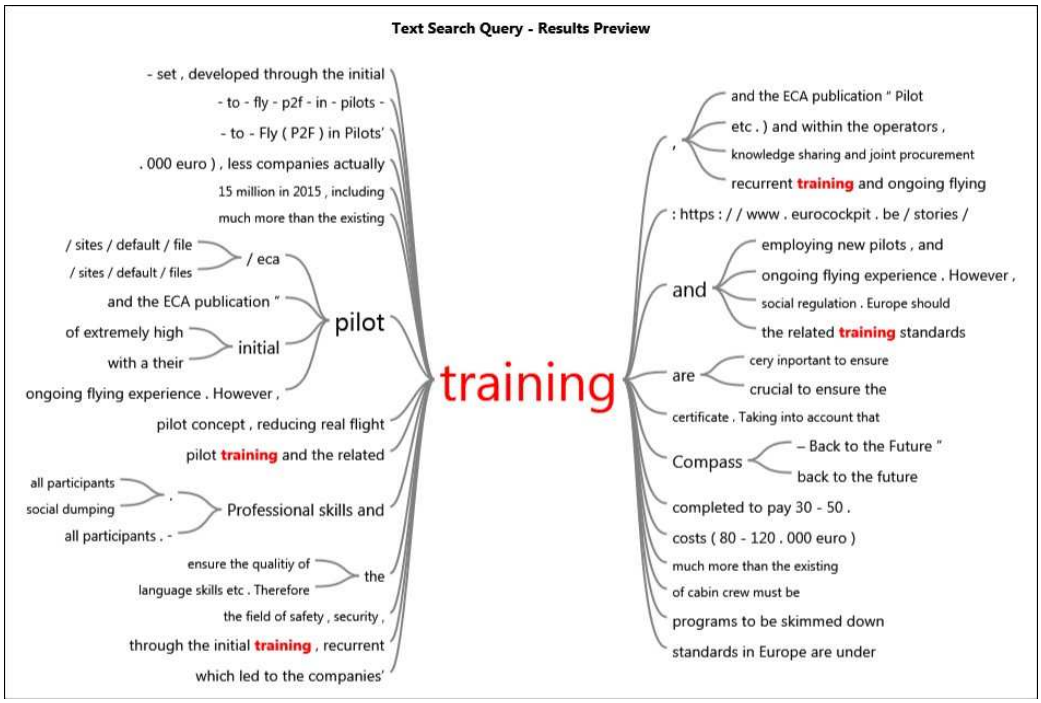
3. **Word frequency search**: Using qualitative data analysis software[653] enables word frequency searches to be rapidly undertaken to identify the top five most frequently used words in the responses. To be useful, common words (*the*, *and*, *it*, etc.) are excluded as are words from the title of the consultation. This should be done for each language for which there are more than 30 responses in multiple languages[654].

4. **Word cloud:** The 'word cloud' function of the software used could be useful to establish for instance the top 100 most frequently used words and to present the results graphically. This could help in identifying themes, particularly when combined with the 'word tree' (see below)

653 Such as Doris (Commission tool) or NVivo (commercial)

654 NVivo also has the capabilities to perform such searches using synonyms when working with some of the more common languages (EN, FR, DE, ES and PT).

5. **Word tree:** For any words identified in the above word frequency/word cloud approach, which indicated that some themes may have been missed, a 'text search' can be carried out and the results of this search displayed in a word tree[655] (see below) to quickly determine the context in which these words were used, and consequently to identify any themes missing from the sample.



Word tree showing the context of the word 'training' for question 1

5.4. ALLOCATING RESOURCES

5.4.1. Introduction

There are essentially two constraints related to the analysis of stakeholder contributions - time and availability of resources. If sufficient time and resources are available, then each and every response can be read and analysed in detail. Similarly, if the questionnaire consists entirely of closed questions, then such constraints are unlikely to pose a serious problem – irrespective of the number of responses. However, **the presence of numerous open text comment boxes can greatly increase the time and resources required.**

[655] Depending on the software used

5.4.2. Time

The figure overleaf illustrates the estimated time required for one person on a full time basis to carry out the analysis of an example consultation with a mix of open and closed questions and with a few hundred responses.

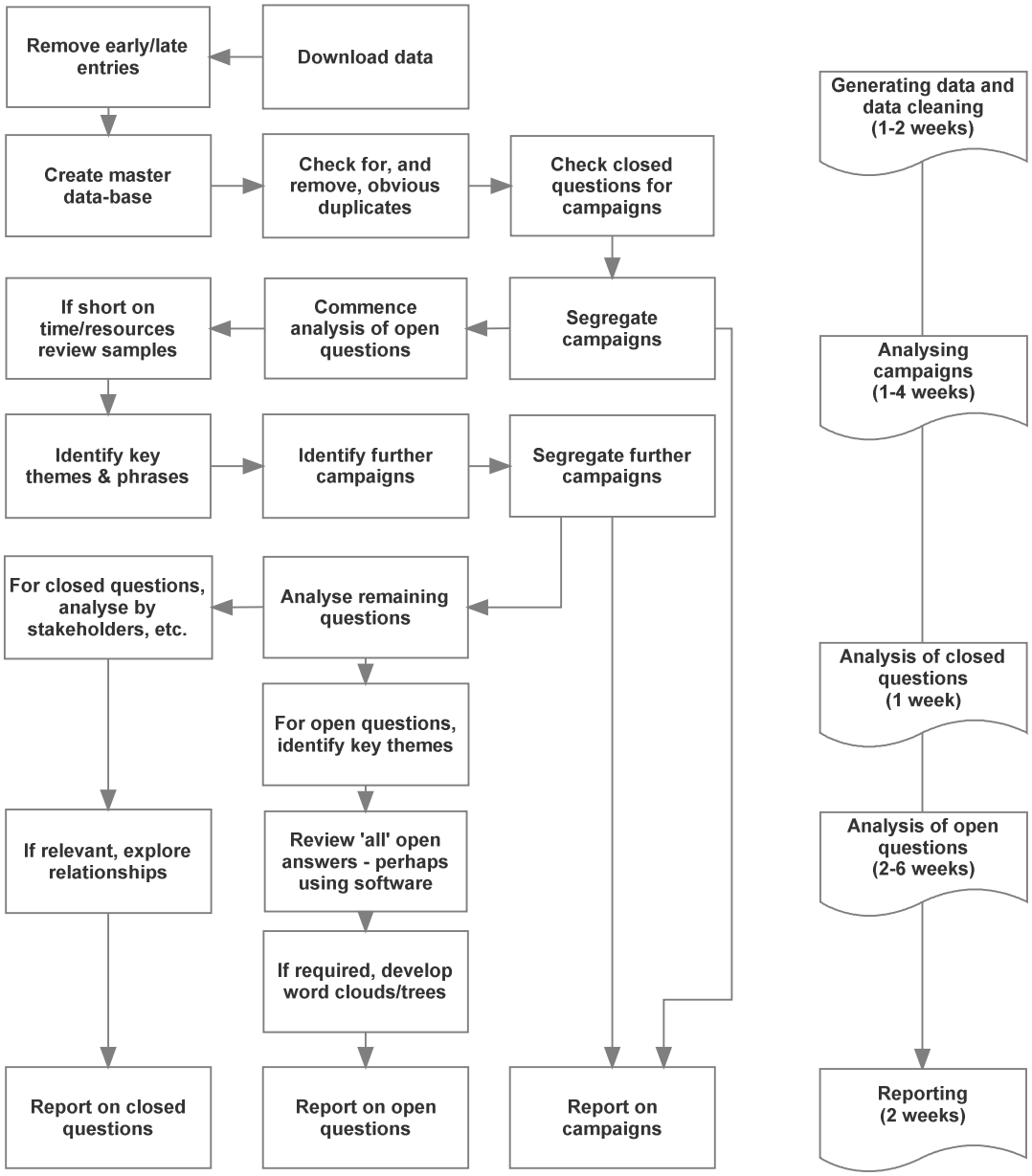
5.4.3. Resources

Based on the time taken to analyse responses to a range of public consultations, it is possible to provide indicative guidance as to the resources required (in person-days) for a thorough analysis of a public consultation which attracts a degree of interest from stakeholders.

Clearly, the resources required will increase with:

- The number of closed questions (little impact)
- The number of open questions (big impact)
- The number of responses (mainly for open questions)
- The number of stakeholder groups to be analysed

The estimated time required to analyse a consultation



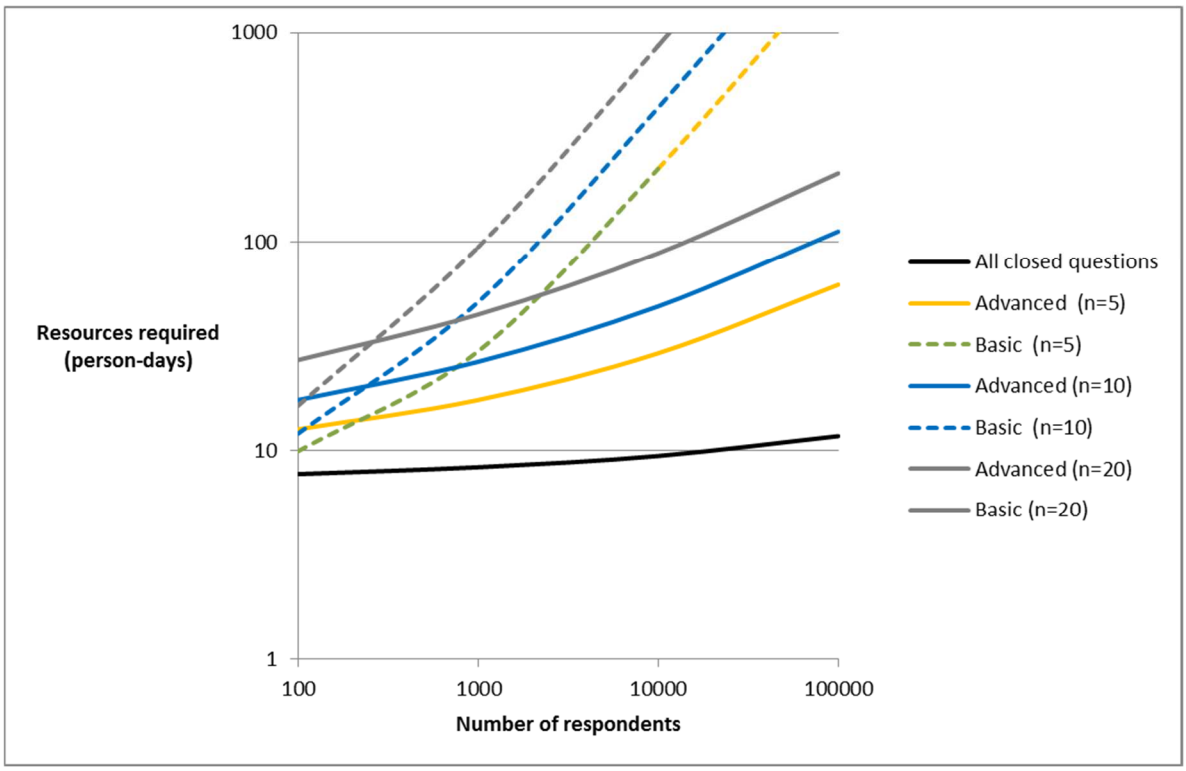
The illustrative example below shows the resources required to analyse the data and provide a summary report based on a mid-sized questionnaire with 30 substantive closed questions divided into five themes and one open comment for each section (five in total). The table below illustrates how the resources required will vary depending on whether the questionnaire elicits 200 or 2000 responses. For this example, two further assumptions have been made:

- The analysis of open and closed questions will focus on three broad stakeholder groups
- The response rate to the open questions is 30% (i.e. they are not mandatory)

Parameter	200 responses	2000 responses
Download data to Excel and analysis of	8 days	9 days

closed questions (including reporting) (*Note advanced analysis using STATA for campaigns does not add extra time)*		
Basic analysis of open questions (i.e. all comments read and analysed at a rate of 10/hour)	4 days	43 days
Advanced analysis of open questions	6 days (3 days for sampling (@ 5/hour) and 3 days for computer aided analysis)	11 days (7 days for sampling (@ 5/hour) and 4 days for computer aided analysis)
Total (basic analysis of open questions)	**12 man days**	**52 man days**
Total (advanced analysis of open questions)	**14 man days**	**20 man days**

Further examples are provided in the figure below:



Illustrative guide to the resources required to analyse and report on an OPC questionnaire with 30 closed questions and a number (n) of open text comment boxes using both basic analysis and advanced analysis

5.5. FURTHER READING & REFERENCES

- Commission Study 'Consultation Support and Development of Advice' (RPA, 2016): https://bookshop.europa.eu/en/consultation-support-and-development-of-advice-pbKA0217018/?CatalogCategoryID=YR4KABstrdkAAAEjLocY4e5K

- The Consultation Institute: https://www.consultationinstitute.org/
- Commission data analysis tool Doris: http://doris.cnect.cec.eu.int/dorisBoard
- Available data analysis software, including request procedure: https://myintracomm.ec.europa.eu/dg/dgt/it_support/software/Pages/index.aspx