This product is part of the RAND Corporation technical report series. Reports may include research findings on a specific topic that is limited in scope; present discussions of the methodology employed in research; provide literature reviews, survey instruments, modeling exercises, guidelines for practitioners and research professionals, and supporting documentation; or deliver preliminary findings. All RAND reports undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.

TECHNICAL REPORT

Evaluation of DG SANCO data management practices

Final report

Jan Tiessen, Claire Celia, Tom Ling, Helen Ridsdale, Maiwënn Bareaud, Christian van Stolk

Prepared for the European Commission Directorate General for Health and Consumers (DG SANCO)

RAND EUROPE

The research described in this report was prepared for the European Commission. The opinions expressed in this study are those of the authors and do not necessarily reflect the views of the European Commission.

Published 2010 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
Westbrook Centre, Milton Road, Cambridge CB4 1YG, United Kingdom
RAND URL: http://www.rand.org
RAND Europe URL: http://www.rand.org/randeurope
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

The European Commission Health and Consumer Protection Directorate-General (DG SANCO) commissioned RAND Europe to provide support in developing a comprehensive data strategy for DG SANCO that meets the needs of increasingly evidence-based policymaking in the future. This work builds on previous work by RAND Europe conducted for DG SANCO, mapping out international good practice of data management (Brutscher *et al.*, 2009).

The work described in this report had two aims:

- to assess the current data management practices within DG SANCO that relate to the four specific issues identified by DG SANCO: data needs, DG SANCO data sources, key partnerships on data, and data quality;
- to develop, on the basis of this review, recommendations for improving DG SANCO's current data management and the definition of DG SANCO's Good Practice Model for Data Strategy.

This final report presents the findings of RAND Europe's analysis. RAND Europe's work is objective and multidisciplinary, and is based upon the core value of quality. All of its products are peer-reviewed before final dissemination as part of our quality-assurance procedures. For more information about RAND's quality standards please see http://www.rand.org/standards.

For more information about this document or RAND Europe, please contact:

Jan Tiessen
RAND Europe
Westbrook Centre
Milton Road
Cambridge CB4 1YG
United Kingdom
Tel. +44 (1223) 353 329
tiessen@rand.org

Contents

List of figures

List of tables

List of acronyms and abbreviations

DG	Directorate-General
DG RTD	European Commission Research Directorate-General
DG SANCO	European Commission Health and Consumer Protection Directorate-General
DIKW	data, information, knowledge and wisdom
ECDC	European Centre for Disease Prevention and Control
ECHI	European Community Health Indicators
EFSA	European Food Safety Authority
FSDMS	Food Safety Data Management System
FVO	Food and Veterinary Office
GDP	gross domestic product
OECD	Organisation for Economic Cooperation and Development
RAPEX	Rapid Alert system for dangerous consumer products
TRACES	Trade Control and Expert System
WHO	World Heath Organisation

Executive summary

This study assesses DG SANCO's current data management practice

RAND Europe was asked by DG SANCO to assess the current data management practice within DG SANCO. Data management in this definition comprises the use of data in policy making in its wider meaning. This study maps current practice, identifies perceived strength and weaknesses and develops recommendations to address them. The context of this work is the desire of DG SANCO to develop a comprehensive data strategy to meet the increasing needs of evidence-based policy making. Such a strategy would seek to establish good practice in how data are identified, collected, stored, analysed, used and communicated.

The findings of this report were based mostly on interviews with DG SANCO staff

The analysis contained in this report relies to a large extent on key informant interviews and thus resembles a perception audit of what a subset of DG SANCO staff see as current practice, its strengths and its weaknesses. In addition internal documentation obtained from DG SANCO was reviewed, and the research team analysed a sample of documents produced during the different stages of the policy process. This report focuses mostly on areas of improvement, though it was clear from the interview findings that current practice in DG SANCO had many strengths. Therefore, DG SANCO starts from a position of strength in data management. In other areas more substantial changes are required, in the opinion of staff.

DG SANCO's current data practice is characterised by diversity between directorates and policy stages

DG SANCO staff use a relatively wide definition of data, including not only quantitative statistical data, but also more qualitative sources of data. The data needs between directorates differ substantially in terms of type and content of data; in general there appears to be little overlap in data needs between directorates, often even between units. Both (perceived) data needs and actual use of data differ substantially between policy stages. The impact assessment stage is currently the most data-heavy stage of the policy process, using the widest range of data types and sources; while enforcement and evaluation are the stages with the most comprehensive use of statistical information, often collected by DG SANCO or submitted by Member States.

DG SANCO produces some of its own data, but in the early stages of the policy process it mostly relies on the use of external data sources such as Eurostat, and the use of external contractors is widespread. For enforcement purposes, however, DG SANCO mostly relies

on self-generated data obtained through alert systems (e.g. RAPEX) or reporting requirements for Member States.

DG SANCO has access to a substantial amount of data that it requests from Member States, but these most of these data are deposited in a large number of different databases or even not stored in an electronic format. Overall, data storage is decentralised and fragmented. Data sharing and collaboration are taking place between colleagues at DG SANCO informally but there is a lack of formal knowledge about what staff are doing across units and directorates.

Data analysis tends to be descriptive and qualitative rather than statistical in nature, assessing context and problems rather than evaluating and projecting. There is a lack of quantitative data analysis and in-house specialist skills (i.e. those of economists and statisticians) to deal with data outside the specialist units, which is partially compensated for by the widespread use of contractors to analyse data.

DG SANCO staff currently conduct, for the most part, pragmatic assessments of the quality of data they use. The two main means by which staff assess data quality are (1) checking that the data source is reliable and (2) assessing the comparability of the data. There are no formal guidelines on how to assess the quality of data.

Perceived strengths and weaknesses vary by directorate and function

The perception of current strengths and weaknesses in using data for policy making differs between directorates and experience in data use, with staff from information units and Directorates D, E and F the most positive about current strengths and weaknesses. This points to pockets of good practice that should be extended to improve the overall performance of DG SANCO. Key issues mentioned both as strengths and weaknesses include the following:

- In general, **collaboration and knowledge sharing** between individuals was perceived as a strength and as working very well. There were, however, concerns that this collaboration was personalised at an individual level, which could be a barrier for new staff.
- **Specialist skills to collect, analyse and use data** are perceived to be in short supply at DG SANCO in general, but some units in Directorates D, E and F and the information units (B1, C2) have substantial skills for and experience in using quantitative data.
- While staff working on enforcement and in the information units were generally satisfied with the **availability of data** and consider it a strength, other respondents highlighted problems in access to data and knowledge about potential data sources.
- Nevertheless, respondents also felt that some of the data available to DG SANCO are **not sufficiently used** because staff are unaware of the existence of those data, or because quality and comparability make them difficult to use.
- Staff felt that there is the **importance of data for DG SANCO** is increasing, but the use of data is still very much ad hoc and data do not yet play a sufficient role in policy development;

The report includes strategic high-level and specific recommendations

High-level recommendations focus mostly on the shift in organisation and culture which will be required for the development of a comprehensive data strategy. Therefore, they are likely to represent changes that it may be difficult to implement and that require a long time horizon to embed. They focus on the following.

1. **Clarify the purpose of a data strategy**
   Prior to developing a high-level data strategy, DG SANCO should define and clarify the purpose of a data strategy and define what it wants data to be used for. Currently a wide range of sometimes competing purposes is mentioned by DG SANCO staff, ranging from monitoring enforcement action and implementation to communication to the general public. If DG SANCO wants to cover a wider range of purposes, it should make sure that the data strategy acknowledges the different uses for which data are intended.

2. **Think 'information' and 'knowledge' rather than data**
   In developing a comprehensive data strategy, DG SANCO should think about the knowledge and information it requires for policy making. Data will only become useful when structured, interpreted and understood in context. Thus DG SANCO should ensure that the capacity and skills to make use of data are available.

3. **Prioritise**
   Given the limited resources and skills available, it will be essential to prioritise the collection of specific data. This prioritisation should be based on the key questions that DG SANCO needs to answer – that is, the information and knowledge DG SANCO wants to obtain. For example, one approach that can be used is logic modelling, establishing a logic and theory of change.

4. **Create a decentralised, but coordinated, organisational structure system**
   Given the differences between directorates, both in term of subject area and type of activity, DG SANCO should consider a decentralised, but coordinated, organisational basis for its data strategy. This should be supplemented by shared resources for the whole organisation, and regular communication between these units should take place, including updates on their recent activities. Information units available should be used to a greater extent and could form the core of directorate-wide centres of expertise and support for policy units.

5. **Aim for cultural change**
   Any organisational and procedural change suggested here should be embedded in a process of cultural change. DG SANCO staff should feel that using data in their day-to-day work is relevant, valued and recognised within the wider organisation. Ultimately, DG SANCO staff will need to be aware that using data makes a substantial difference in the policy process by improving decisions taken.

Specific recommendations

Specific recommendations are closely linked to the weaknesses identified by staff, and this relationship is shown in Table 0.1 below.

Table 0.1.1: Linking recommendations to the issues identified

Weakness mentioned	Recommendation
Lack of specialist skills in-house	Produce a map of data management specific skills of DG SANCO staff
	Provide basic training on data use and management for DG SANCO staff
	Establish a central survey unit
Some of the data within DG SANCO are either not used at all or not well used	Conduct a data inventory
Harmonisation issues with data obtained from Member States	Create a central knowledge base or information hub on available data sources, both internal and external
Current data management practices are reactive and crisis driven	Develop methods to define data needs systematically
Access to data is sometimes an issue, as well as how data are used	Develop basic guidance on the use of data in day-to-day policy work
	Develop guidance on how to assess and ensure data quality
Policy is not sufficiently data led	Plan and discuss data needs as early as possible in the policy making process.
	Make data use more visible.
Timeliness is an issue in data management	Be prepared for ad-hoc data needs.
Problems with collaboration in data collection and analysis	Become more concise in contracting out research
	Strengthen external collaboration on data collection
No unified approach to data across DG SANCO	Create decentralised support structures

This report provides additional tools to implement a data strategy

The appendices to this report contain additional tools for DG SANCO to develop and implement a comprehensive data strategy, including a reflection on how to develop the high-level macroindicators that DG SANCO should use, some practical tips for staff, and indicators to help understand the effectiveness of a data management strategy. For the high-level indicators, the study proposes using a 'theory of change' approach, whereby DG SANCO examines its logics of intervention or how it produces impacts. The indicators are then developed on the basis of how DG SANCO achieves these impacts. In terms of practical tips, the report identified a number of key diagnostic questions for use by DG SANCO staff at each stage of the data management framework. These were identified on the basis of the interviews and offer a checklist for staff to help them improve data management. Finally, the impact of a data strategy has to be evaluated once changes have been made. This requires DG SANCO to establish a baseline and then look *ex-post* at how data management has become embedded and what its impact has been.

CHAPTER 1 Introduction

1.1 DG SANCO's efforts to develop a comprehensive data strategy

DG SANCO's Strategy and Analysis Unit, Unit 02, is currently working on the development of a comprehensive Good Practice Model for Data Strategy for all of DG SANCO's policy areas. To develop this strategy, DG SANCO initiated a project in two phases.

Phase 1: In an initial **mapping phase**, DG SANCO sought to analyse the current state of data handling and strategy within DG SANCO as well as outside it, in selected relevant EU and overseas organisations. For the purpose of this mapping phase, DG SANCO identified five work packages:

WP 1 – External mapping of data strategies. This comprised a review of existing data practices in other relevant administrations and, where possible, the identification of good practices that could be adapted to the specific role and needs of DG SANCO.

WP 2 – DG SANCO data needs. This work package sought to answer the following questions: 'When are data needed?'; 'Who needs these data?' and 'What are these data used for?'

WP 3 – DG SANCO data sources. This part of the mapping phase sought to identify the data to which DG SANCO already has or has had access, in order to map and review currently available data sources both in-house and outside DG SANCO.

WP 4 – Key partnerships. This work package sought to identify the main relationships on which DG SANCO relies to get the data it needs. It set out to list and study these relationships in more detail in order to identify the nature of the links and the ways to improve cooperation with key partners such as the European Food Safety Authority (EFSA) and Eurostat.

WP 5 – Data quality. This part of the mapping phase aimed to develop the main criteria/indicators of quality for DG SANCO's data and explore common practices in different policy stages such as enforcement.

Phase 2: A **strategy phase** in which the findings of the work packages are synthesised and developed into a coherent and comprehensive strategy that can then be transferred into an action plan to be implemented at DG SANCO.

As part of this process, DG SANCO commissioned RAND Europe to conduct research into Work Package 1, the external mapping of data strategies, which was completed in the summer of 2009 (Brutscher *et al.*, 2009) and subsequently on Phase 2, as detailed below.

1.2 Conducting the research

1.2.1 Research question

With the internal and external mapping nearing completion, DG SANCO commissioned RAND Europe to give support in the second phase – the strategy phase of this project – with two specific objectives:

1. to assess the current data management practices within DG SANCO which relate to the four specific issues identified by DG SANCO: data needs (WP2), DG SANCO data sources (WP3), key partnerships for data (WP4) and data quality (WP5), as well as identifying possible existing gaps;
2. to develop, on the basis of this review, recommendations for improving DG SANCO's current data management and the definition of DG SANCO's Good Practice Model for Data Strategy.

1.2.2 A framework for a comprehensive data strategy

In the external mapping report submitted to DG SANCO in 2009 (Brutscher *et al.*, 2009), we developed a simple data strategy framework that provided a basis for the data collection across a number of case studies. In this framework we identified five key elements that represent an idealised, sequential process of developing a data strategy from scratch. The framework is shown graphically in Figure 1.1 below.

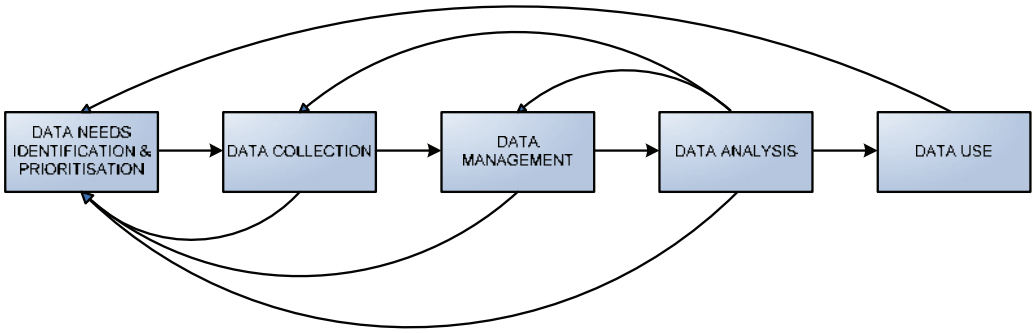


Figure 1.1: Data strategy framework

Source: RAND Europe / Brutscher *et al.*, 2009

The key elements of this framework are as follows:

1. **Data needs identification and prioritisation.** The first stage of the process is to define what data are needed, and to prioritise the data that should be collected. The definition of data needs involves consideration of the policy objectives, future uses of data and data availability.
2. **Data collection.** Once data needs and priorities are established, an organisation needs to consider how the data desired may be collected. This stage includes questions about data sources, data gathering and frequency of data collection.

3. **Data management.** Once the data have been collected, they will need to be managed within the organisation. At the most basic level, this may only involve storing the data in databases; but it may involve more advanced stages of data processing such as harmonisation and standardisation to make the collected data comparable.
4. **Data analysis.** In the subsequent stages, data are structured and analysed so they become useful information that may be used in a policy context. In the case studies we conducted, we were particularly interested in what type of analysis was performed in the organisations.
5. **Data use.** Closely related to the analysis of data is the use of data. This stage relates to what is done with them: are they produced for internal processes, for the general public or for selected stakeholders; and are they used for dissemination and communication purposes?

Our premise was that a comprehensive data strategy will need to answer the key questions 'How?' and 'Who' for each element – that is, how do we define our data needs and who will be responsible for identifying needs? This thinking takes place for the five key elements of a data strategy outlined above and also guides the analysis of this report. However, the framework was slightly adjusted and compressed to make it easier to use, to make the cross-cutting element of data quality more prominent, and to make it compatible with the internal analysis DG SANCO had already conducted. We therefore focused on five elements in this report, which are detailed in Chapter 2:

1. The identification and prioritisation of data needs.
2. A discussion of data sources and collection.
3. The management of data.
4. Data analysis, use and communication. In this step we merged the elements of the data strategy that provide an answer to the questions of what the organisation is doing with the data, and how they are used and eventually disseminated.
5. As a cross-cutting theme for analysis which affects all elements of a data strategy, we added to our conceptual framework the issue of data quality, which has been of particular interest to DG SANCO.

1.2.3 Approach

To address the specific tasks set out by DG SANCO, RAND Europe proposed a research project with ten key tasks to collect evidence, assess the current performance and develop recommendations. These tasks may be grouped into two phases. Phase 1 (Tasks 1 to 6) aimed to generate a better understanding of DG SANCO's current data management practices, while Phase 2 (Tasks 7 to 10) focused on the development of recommendations to improve the status quo. An overview of the tasks is provided in Figure 1.2 below. It should be noted that the survey (Task 4) was replaced with a second round of key informant interviews at the request of DG SANCO, so the main sources of information for this report are 24 key informant interviews with DG SANCO staff and the document review.

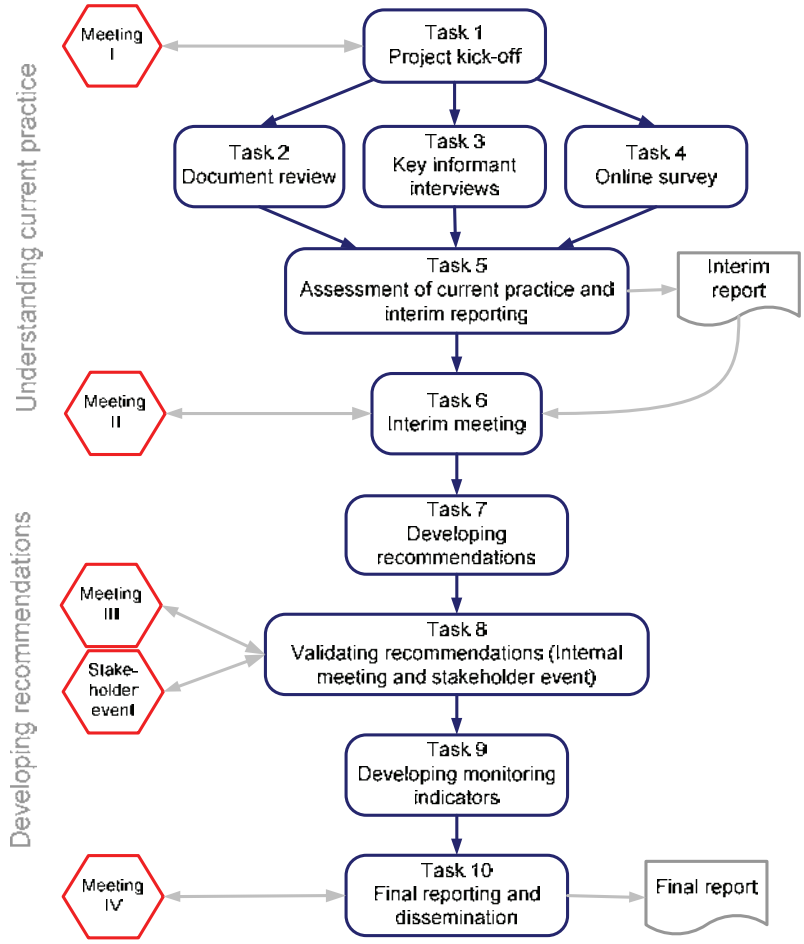


Figure 1.2: Overview of tasks

Source: RAND Europe, 2009

1.2.4 Research methods

The research approach has been built around the use of two methods of gathering evidence: a document review and key informant interviews (25 in total – see Appendix E for a list of interviewees). As stated above, an online survey of all staff who use data was planned at the outset of this project in order to strengthen the evidence base for this review of DG SANCO's current practices. However, Unit 02 later decided that it would not be appropriate to launch such a survey at this time, and a second wave of interviews was conducted to ensure a sufficient evidence base for our assessment.

Document review

To develop an initial understanding of the current state of data management in DG SANCO we conducted a document review analysing two main types of document. First, we reviewed the internal documentation on the current state of data management, in particular the interim and, where available, the final documents produced as part of the internal mapping work packages conducted by DG SANCO. The findings from these documents are integrated into our analysis throughout this report. An overview table may be found in Appendix G.

Secondly, we systematically analysed a sample of publicly available documents produced by DG SANCO during the policy making process to get a better understanding of its data use. For each stage of the policy process and each directorate, a sample of documents was chosen, as detailed in Table 1.1. These documents were analysed by a researcher according to policy stage and directorate, and each document was classified according to the following criteria:

1. the source of data used;
2. the type of data;
3. the type of analysis performed.

Table 1.1: Selection of documents for review

Policy stage	Type of document	Total number of documents
1. Policy strategy setting and policy planning	Unit management plans	45
	Public policy strategy documents	4
2. Impact assessment	Impact assessments	9
3. Policy evaluation and enforcement	Enforcement reports, evaluation, special studies	14
4. Inter-institutional process	No documents available	0
5. Communication.	Press releases	16

The classification of type of data has been based on previous work of DG SANCO on the mapping of current and future data needs (DG SANCO, 2009d). According to this classification, the following types of data may be distinguished:

- **Data type 1: economic, social and environmental context.** Data about the overall economy (e.g. GDP, size of markets), society (e.g. population projections, unemployment rate) and environment (e.g. carbon dioxide emissions).
- **Data type 2: status.** Data about the actual status in Europe of the issues that we are trying to address with health, consumer protection or food safety policies (e.g. life expectancy, consumer detriment, incidence of illness from food).
- **Data type 3: determinants driving change.** Data about the factors that affect changes in the status of the issues that we are trying to address with our policies (e.g. smoking prevalence).
- **Data type 4: public authority action.** Data about interventions by public authorities to affect the issues of concern to us. This should be broken down by:
  - **European action:** such as data on implementation of European legislation, or cases resolved through rapid alert systems, or implementation of food safety systems;
  - **Member State action:** such as provision of healthcare, vaccination campaigns, and actions against rogue traders.

- **Data type 5: cross-cutting factors.** This encompasses factors affecting all of the above which should make it possible to break down overall data, such as by gender, by geographical region, or by socio-economic group.

Key informant interviews

A total of 24 interviews with 25 individuals were conducted.[1] The selection of these interviews covered all directorates as well as Units 01 and 03. Interviewees were purposely selected by Unit 02 to cover all policy stages from policy setting and planning to communication. Furthermore, we selected interviewees according to availability and to previous involvement in data strategy activities. Interviews were conducted in two stages as it was decided to do more interviews following the cancellation of the online survey by Unit 02. The initial selection involved individuals who had already been active in developing DG SANCO's internal activities of data strategy development, either as authors of so called 'work packages' or 'pilot cases'. The second selection of interviewees involved staff from a broader range of units to ensure we had as wide a coverage as possible of DG SANCO staff in terms of directorates, units and policy stages; however the interviews were not distributed evenly across directorates or units.

Limitations of the methodology

In interpreting the results of this research project, it will be important to keep the limitations of the chosen approach in mind. The key source of evidence on which this report is based is the 24 interviews that were carried out with DG SANCO staff. These interviews were conducted with individuals selected by Unit 02 on the basis that those would be the staff best placed to speak about DG SANCO's current data management practices in their respective units. In addition to these interviews, a handful of internal documents and reports on DG SANCO's current data management practices were available as a second source of evidence. The shortage of these limited the possibility of triangulating findings with different data sources. Thus this report constitutes to a large extent an audit of the perceptions of a selected number of staff on current data management practice. With this in mind, we have included information on the number of people who mentioned a given issue or problem to give a more robust idea of how widespread that issue or problem was during the interviews. Given RAND Europe's previous wider expertise in this area and our experience in working for DG SANCO we are fairly confident that the findings of this report paint a good picture of DG SANCO's current data management practice.

1.3 This report

This report outlines the findings from this research project and is structured into four chapters. This introduction is followed by a descriptive account of DG SANCO's current data management practice (Chapter 2). Chapter 3 outlines the main strengths and weaknesses of the current data management practices, as perceived by the staff we interviewed. Finally, Chapter 4 provides a set of recommendations to address the weaknesses highlighted in Chapter 3. Furthermore, the appendices include three practical

[1] For the full list of interviewees, please refer to Appendix E.

guides to help DG SANCO: the development of top-level indicators, success indicators to evaluate the future data strategy, and practical tips for staff working with data (Appendices A, B and C respectively). Appendices D to G respectively provide a copy of the interview questionnaire, the list of interviewees and the mapping of data sources used by DG SANCO's directorates and units.

CHAPTER 2 DG SANCO's current data practice

2.1 Introduction

This section lists key observations on current practices of data use, which arose from our key informant interviews and review of DG SANCO's documents. These constitute a mapping of variations in practice and the reasons behind them. The information should inform the evidence base upon which DG SANCO may base its development of a comprehensive data strategy.

2.2 Perceptions of the term 'data' and usage of different forms of data

DG SANCO staff use different definitions of the term 'data', ranging from a narrow focus on 'hard' statistical data to fairly broad definitions, and encompassing qualitative data and information. In this section 'types' of data are used to discuss the distinction between quantitative and qualitative data, and those between data, information and knowledge – rather than the types of data outlined in the introduction, which will be explored in greater detail at a later stage in this chapter.

Almost all of the respondents (23 out of 24) referred to 'hard' or statistical data when reflecting upon the meaning of the term 'data', though a smaller number used this type of data exclusively in their work. Hard data were understood as those which could be represented numerically, for example:

- indicators such as prevalence rates of disease and import numbers;
- statistical results of opinion surveys such as Eurobarometer.

Those who used only statistical data (6 out of 24) all held roles either relating to enforcement (which in turn was related to animal health and feed – 2/24) or data management (various directorates – 4/24). This clearly reflects the nature of their work. As we shall see, those working in enforcement have a wide range of their data needs defined by legislation, and are less likely to be exposed to the same challenges of data management as units that have a stronger policy making function; for staff in enforcement there are often sets of predefined hard indicators available. Those working specifically in data management are working with databases, which are built around the storage of statistical data. It is worth noting that those who only use statistical data for their work did not mention numerically represented opinion surveys as 'hard' data.

There were far more respondents who used statistical data in conjunction with other forms of information (17 out of 24), reflecting the differing needs of those in the more creative aspects of the policy making process who have to identify a wider and more detailed picture of the problem in hand. All of these respondents worked in policy development, strategy setting, impact assessments, evaluation or communication, with a number of them mentioning more than one policy stage. Indeed, several respondents noted the 'cycle' of data used in policy making: identifying the problem (refers to types 1 and 2 as identified in the introduction – context and current status of issue); proposing legislation and assessing the possible impacts (refers to types 3 and 4, drivers of change and impacts of intervention); and evaluation and review of policy (also types 3 and 4), which often requires similar kinds of data. Compared to the other types of data highlighted for different stages of the cycle, cross-cutting information (type 5) is relevant to all of these policy stages. For example, one informant noted that this type of data was useful for conducting impact assessment since sometimes they also had to assess the current status and context as part of their work because it had not been done at the beginning of the policy setting.

As previously noted, respondents had a very broad view of what data encompass. The following is a list of what was considered as data by DG SANCO staff:

- stakeholder opinion/perceptions from broad consultation, interviews and focus groups, including opinion from Parliament, Member States, etc.;
- specialist information from stakeholders (i.e. specialist business groups);
- expert opinion from scientific advisory groups;
- public opinion (statistical surveys and otherwise, media reports);
- case studies (collected either by external contractors or projects from SANCO or other directorates, i.e. DG RTD);
- academic, grey and peer-reviewed literature;
- qualitative descriptions of processes (i.e. court cases for collective redress, management systems in Member States' competent authorities for the Food and Veterinary Office – FVO – audits);
- knowledge of relevant legislation, legal systems and best practice standards;
- reports from Member States on different sectoral developments.

DG SANCO staff's wide definition of data was nicely summed up by one informant who stated that they considered data to be 'any fact or figure that is useful, or that can be useful to define the problem'. Analysing the wide definitions of data provided pointers to the important, classical distinction between data, information and knowledge. Many responses by interviewees referred to actionable information and knowledge, rather than to unprocessed data, reflecting also their immediate needs in policy making. As a reminder Figure 2.1 below shows the data, information, knowledge and wisdom (DIKW) hierarchy.

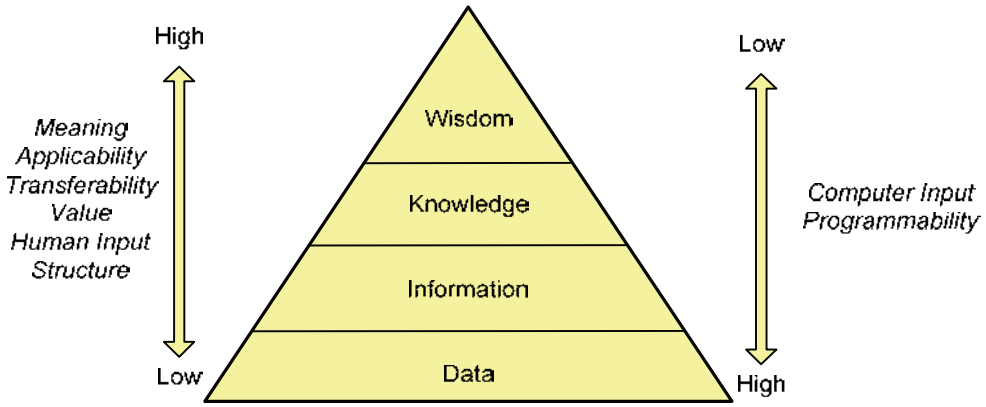


Figure 2.1: The DIKW hierarchy

Source: Rowley, 2007

The distinction between data and information was also mentioned by two interviewees, who pointed out that it is the latter rather than the former that is useful for the policy making process.

Reflecting that a majority of informants associated data with information that they could transform into knowledge rather than restricting it to hard statistical data, the remainder of the report will use 'data' in this broad sense unless clearly noted otherwise.

Key observations

- DG SANCO staff make distinctions between different forms of data.
- DG SANCO staff who worked on enforcement tended to work mostly with hard statistical data.
- Those dealing with policy making processes (the majority) were concerned with information in a more general sense – that is, knowledge, information and data which could be useful in informing their decision making. Any recommendations on data strategy should take this into account.

2.3 Data needs and prioritisation

As touched upon in the previous section, there are substantial differences both between DG SANCO's policy areas and between policy stages in how data needs are defined and prioritised. This is primarily a result of the varying subject matter and the different uses to which the data are put.

We identified five ways in which data needs were decided upon.

1. Defined by external parties:
    - by legislation;
    - by external demand.
2. Defined by the needs of the question (ad hoc):
    - policy making – status quo and drivers of change;

- European-level picture of a sector – status quo and drivers of change;
- audit – standards, guidelines and system processes.

Defined by external parties beyond informants' control

Many of our key informants did not have specific inputs into defining their data needs. They were either defined by legislative requirements or by data requests external to them.

A total of 7 out of the 24 informants stated that their data needs were defined for them. Two of these were communications officers, and three were working on policy enforcement, one in data management and one in public health.

Legislation

Data were defined by legislative requirements for four out of seven of those with predefined data needs. These were all in the field of animal health, food and feed, and three of them were in enforcement while the other was in data management. This appears to relate to both the maturity of the policy area – that is, they had developed over time, so proposing new legislation was much rarer – and their job function, which does not relate to policy making as such. In other words, the current legislation already contains mechanisms to collect most of the data necessary for the specific units, thus little time needs to be spent on what data are needed. It may nevertheless be necessary to prioritise and understand the data collected.

External demand for data – communications and the information units

One key informant in the Public Health Directorate reported that his unit primarily responds to external suggestions for data collection from Member States and other services and institutions, and DG SANCO than takes a coordinating role in collecting these data. This reflects the status of the policy area, in which primary competences lie with the Member States.

The two communications officers either had their work initiated by what policy officers desired to communicate, or by questions that journalists ask.

Key observation

- Those whose data needs are predefined tend to be those who are not involved in policy making processes (i.e. involved in enforcement and using data directly collected by DG SANCO or sent from Member States as a legislative requirement).

Defined by the needs of the question (ad hoc)

Policymakers – strategy, impact assessment and evaluation

This was by far the most common form of deciding data needs. Data needs were considered ad hoc, according to how the data were to be used; needs were discussed to develop answers to an emerging, specific policy question.

In terms of policy making and revision, this would mean identifying the status quo and the drivers for change *with a view to assessing (ex-post or ex-ante) the possibilities for intervention*, which is the key question for DG SANCO policy making staff. It is in this area that the types of data outlined in the introduction become relevant.

Most key informants who were involved in the policy making processes followed a similar pattern.

1. Establishing the status quo – collecting data of types 1 and 2.
    - Data type 1: economic, social and environmental context. Data about the overall economy (e.g. GDP, size of markets), society (e.g. population projections, unemployment rate) and the environment (e.g. carbon dioxide emissions).
    - Data type 2: status. Data about the actual status in Europe of the issues that DG SANCO is trying to address with regard to health, consumer protection or food safety policies (e.g. life expectancy, consumer detriment, incidence of illness from food).
2. Establishing the mechanisms for change – collecting data of types 3 and 4.
    - Data type 3: determinants driving change. Data about the factors that affect changes in the status of the issues that we are trying to address with our policies (e.g. smoking prevalence).
    - Data type 4: public authority action. Data about interventions by public authorities to affect the issues of concern to us. This should be broken down by:
        - European action: such as data on implementation of European legislation, cases resolved through rapid alert systems, or implementation of food safety systems;
        - Member State action: such as provision of healthcare, vaccination campaigns, or actions against rogue traders.
3. Data used for all stages of the policy process.
    - Data type 5: cross-cutting factors: factors affecting all of the above, by which it should be possible to break down overall data, such as by gender, by geographical region or by socio-economic group.

The above description represents an ideal type highlighted by our informants as we found from our document review and interviews that the impact assessments were the most data-heavy areas of work, while policy setting was fairly minimal in terms of data, and *ex-ante* evaluation only slightly more so. Many informants and DG SANCO's mapping of data use in Directorates D, E and F (DG SANCO, 2009e) noted a lack of data on the effects of interventions. Others found it difficult to extract trends as DG SANCO policy on data collection had been inconsistent and changing, so the longitudinal data necessary for observing drivers of change did not exist.

Collaboration and informal guidance from colleagues was widespread among the policy making staff – this was primarily because other colleagues held expertise which did not exist within a given unit or directorate. For example, economists within DG SANCO were widely consulted for guidance, as were legal specialists by one informant in unit C.

Some informants in this category also defined their data needs in an ad-hoc and sometimes unsystematic manner – availability being a common consideration. It was noted that there was a general lack of process and of formal guidelines and a culture of data use, this being the responsibility of individual policy officers themselves – who were often short of time and expertise. Similarly, interviewees cited a general lack of knowledge of what data were available and would be useful. While the formal impact assessment guidelines were noted

and frequently used, others said that they were not very detailed on the use of data specifically.

Information units

An alternative (though similar) question is that asked by the information units – Directorate B's consumer market scoreboard and Directorate C's health information unit. This was not specifically related to policy making, but to developing the added value of a comprehensive, European-level view of a specific sector. It involved identifying the status quo and drivers for change, but was less affected by the time pressures of policy making and involved a more systematic analytical approach to identifying data needs. In Directorate B a small number of key indicators were decided by a thorough analytical brainstorming on the nature of consumer markets, while in Directorate C a large amount of public health indicators were chosen. Availability was not a consideration in either case as they thought it a problem to be surmounted rather than worked around, however as yet they were still missing data for some of their indicators. Data needs are identified together with Member States' representatives (meeting in the Health Information Committee) as well as through confrontation with relevant European stakeholders (defining the European Community Health Indicators – ECHI). These have been set out in the European Health Information discussion paper (DG SANCO 2008b), which is itself based on the stated health policy goals of the European Commission in the health strategy.

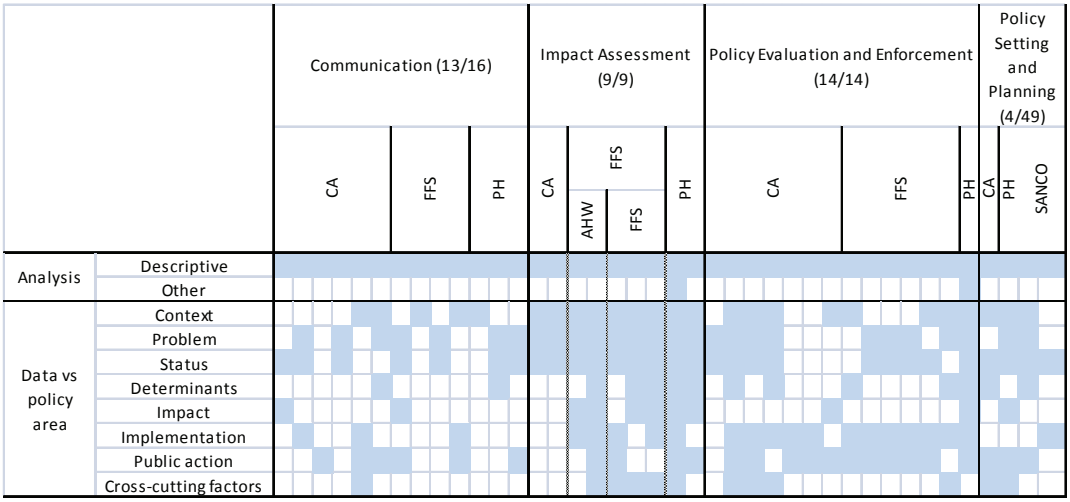
Audit

Two of our key informants were auditors working in Unit F (FVO). Their data needs were defined by their own judgement of what is required to conduct an audit. This included knowledge of international standards, the relevant legislation, and qualitative information on the processes of the country system they were auditing. Defining data needs in this sense was relatively straightforward for them.

Data needs as reflected in current documents

The extent to which data are actually used and what type of data are used in the different policy stages may also be seen in the result of the document analysis conducted (see Table 2.1 below).

Table 2.1: Data use across policy areas and stages



The top line of this table shows the different policy stages and indicates how many documents were analysed in total (e.g. 16 in Communication) and how many contained any kind of quantitative data (e.g. 13 in this case). It is particularly striking in this overview that the majority of internal policy setting and planning documents do not contain data (which are the unit management plans); the four documents listed are long-term strategies such as the animal health strategy. Secondly, this overview shows that data are mostly used in a descriptive way, and that data are not used for further, more sophisticated analysis on a regular basis. In terms of type of data, we can see that the widest range of data is used for impact assessments, including description of the context, problem and the current status. As would be expected, data on policy implementation and public action are most prominent in the policy evaluation and enforcement stage.

Key observations

- Data needs are predefined in the case of enforcement and communication staff.
- Data needs are defined by the requirements of the question by policy making and information units.
- Data needs are defined in an ad-hoc and relatively unsystematic way by policy making staff.
- Information units such as B1 and C2 provide interesting practice in data collection and data use and could take a more pivotal role in a future data strategy.

2.4 Data collection

DG SANCO produces some of its own data, but in the early stages of the policy process it mostly relies on the use of external data sources, with Eurostat being mentioned most frequently. For enforcement purposes, however, DG SANCO mostly relies on data it generates through alert systems (e.g. RAPEX) or reporting requirements for Member States. Appendix F provides an overview of the data sources used by different unit and directorates.

Broadly and as previously highlighted, data used by DG SANCO staff are wider than those covered by the restrictive notion of numerical data and are collected in various ways. In order of importance these include the following:

- Stakeholder opinion/perceptions, including opinion from Parliament, Member States, and so on. Collected by broad consultation, interviews and focus groups, the last two frequently being done by contractors.
- Specialist information from stakeholders (i.e. specialist business groups). Collected by broad consultation, interviews and focus groups, the last two frequently being done by contractors.
- Expert opinion from scientific advisory groups.
- Public opinion. Collected from consumer groups/organisations, media reports and surveys – either by external contractors or Eurobarometer.
- Case studies. Collected either by external contractors or projects from DG SANCO or other directorates general (DGs) – for example DG RTD.
- Academic, grey and peer-reviewed literature. Collected by policy officers themselves through internet search or internal library, or through projects or tenders.
- Qualitative descriptions of processes. Collected from consumer networks (court cases for collective redress) and documents concerning management systems from Member States' competent authorities (for FVO audits).
- Knowledge of relevant legislation, legal systems and best practice standards. Collected by policy officers themselves from legal documents.
- Reports from Member States on different sectoral developments. Sent directly by Member States to units (D, E and F).

Interviewees also highlighted a number of issues related to the ways in which they sought to obtain the data they needed for their work. In terms of internet searches, some cited a lack of awareness of the proper tools and skills available to do a systematic and focused search, and spent a longer time looking for data as a result. On the other hand, while those who used external contractors were frequently positive, others found hiring procedures too rigid and slow, and felt that there was a lack of choice for finding those with the right expertise. There were also numerous management and communication problems with contractors, resulting in poor-quality data being obtained. Reliance on external data sources meant that data collected were frequently not in a similar or comparable format, and were over different time periods, used different definitions, or were of variable quality – see Section 2.5 on data storage and management for further details.

Table 2.2: Data sources by policy stage

		Communication (13/16)			Impact Assessment (9/9)				Policy Evaluation and Enforcement (14/14)			Policy Setting and Planning (4/49)	
		CA	FFS	PH	CA	FFS: AHW	FFS: FFS	PH	CA	FFS	PH	CA	PH: SANCO
Data sources	EC study												
	Other EU bodies												
	DG SANCO												
	Eurostat												
	Other DG												
	Other int. org.												
	Member States												
	Other states												
	Stakeholder												
	Research												
Data collection	Regular												
	Ad hoc												

If we look again at the document review, we see the following picture emerging. First we see a reliance on own – that is, DG SANCO – data sources for communication, evaluation and enforcement purposes. Secondly, we can see the importance of Member States' data for those purposes. Thirdly, impact assessment stands out again as the phase with the most diverse sources of data, in this case coming from a variety of sources including own data, Eurostat data, Member States' data and data from other international organisations.

Key observations

- DG SANCO primarily relies on external sources for data collection.
- Data are collected through a wide variety of methodologies.
- The use of contractors is widespread.
- The use of internet searches is widespread.
- The use of external, often several, sources means there is a lack of harmonisation in data.
- There are substantial data gaps which can currently not easily be filled (i.e. unavailable or incomplete data).

2.5 Data storage and management

Harmonisation

DG SANCO has access to a substantial amount of data externally (e.g. Eurostat), but also collects significant amounts of its own data, in particular in the field of food and feed safety and animal health. However, data are deposited in a large number of different databases and sometimes are not provided or stored in electronic format. One interviewee reported that there were at least six different databases in their directorate alone. Similarly, regarding internet searches of journal databases, one informant noted the difficulty in paying for the use of journals as instead of having a well-stocked central DG SANCO library, this was the responsibility of each policy officer, which also meant that the journals

remained with the policy officer in question – causing some difficulty in access. Harmonising and synthesising data and information from a variety of different sources, with different definitions and time periods, was frequently a large part of the work of some DG SANCO staff. This is obviously very time consuming, particularly when staff need to weigh the validity of different sources. This issue was cited across all policy stages and areas.

Harmonisation and DG SANCO's IT strategy

While many interviewees highlighted data harmonisation as a problem, it should be noted that DG SANCO's 2010 IT strategy recognises this issue and states that 'this [the different classification systems used by Member States for consumer complaints] makes the data not comparable and does not give the possibility to make any meaningful comparisons' (DG SANCO, 2009a, , p.63). The 2010 strategy is attempting to address some of these issues by enhancing electronic data collection tools. This strategy document identifies a number of projects that are currently being undertaken, some of which focus on enabling a more harmonised form of data collection and the integration of different national systems into a unified system hosted by DG SANCO (this has already been happening within ECDC, EFSA and TRACES). An additional example of such a project is the development of a data dictionary for consumer complaints (B1) to create a 'harmonised methodology for classifying and reporting consumer complaints' (DG SANCO, 2009a, p.63). Another example cited in the strategy document is that of the Trade Control and Expert System (TRACES) for Unit D1 (DG SANCO, 2009a, p.51). This system attempts to integrate the different system interfaces of Member States; however, it is difficult to do so because of the 'complexity of synchronising the data in both national and central systems', which is not an easy process for Member States (DG SANCO, 2009a, p.51).

Data sharing and access

A number of interviewees stated that they collaborated or requested information from other units or DGs. This was done in a primarily ad-hoc and personalised manner – knowing the person to ask – because of a lack of process or institutional knowledge of what data were being collected by other units, areas and DGs. One informant noted that some responses to some requests were slow as they had to go vertically through the bureaucratic hierarchy rather than horizontally between policy officers.

Key observations

- Data storage is decentralised and fragmented.
- Data are often not available in comparable formats.
- Data sharing and collaboration are taking place between colleagues at DG SANCO informally, but there is a lack of formal knowledge about what staff are doing across units and directorates.

2.6 Data analysis

Data analysis tends to be descriptive and qualitative rather than statistical in nature. In fact, only two interviewees mentioned that they produce statistical analysis with the data they collect. Examples of the types of analysis produced by DG SANCO staff included the

production of descriptive tables, descriptive reports and country profiles, as well as social and environmental impact analysis.

Use of contractors and in-house expertise/support

Many interviewees (10 out of 24) mentioned that a lot of the data analysis tended to be undertaken by outside contractors, including consultants and Eurostat. This finding was consistent across the different policy stages. Of those staff who carried out their analysis in-house, four stated that they received support from both Unit 02 and economists within DG SANCO to produce it.

Barriers to data analysis

While some staff mentioned that they could access support and expertise in-house, seven interviewees stressed that a lack of in-house specialist skills to deal with quantitative data was hindering the production of more developed and robust analysis.

Other barriers to data analysis included data gaps and issues around data harmonisation, which meant that often data were not readily available for all Member States, and the format in which information was provided needed alteration before the data could be used (because some of it was provided by Member States in PDF, Excel, etc. rather than in a readily compatible format). Data gaps sometimes required some interviewees to make assumptions in order to produce their data analysis. For example, one interviewee mentioned that staff in his unit sometimes had to rely on alternative data to build assumptions in cases where data were not available to produce estimates in food consumption trends. For example, they reported one instance in which they had to build assumptions from trends in the consumption of cow's milk to obtain estimates of trends in the consumption of goat's milk.

Time pressures and staff capacity to undertake the work were also issues that were often mentioned (8 interviewees). Sometimes these related to the fact that data are often needed quickly for policy purposes and need to be collected in an ad-hoc manner because they are not necessarily readily available for analysis. Other interviewees mentioned time pressures, in particular in relation to working with outside contractors as well as Member States in order to produce analysis.

Key observations

- There is widespread use of contractors to analyse data.
- Data analysis is largely descriptive and confined to assessing context and problem, rather than evaluation and projections.
- There is a lack of quantitative data analysis and in-house specialist skills (i.e. economists and statisticians) to deal with data.
- Staff experience time pressures in producing analysis when working with outside contractors or in-house because of the limited time available to produce the analysis.

2.7 Data quality

Collecting and using good-quality data is essential to provide a strong and credible basis for evidence-based policy making. Indeed, using poor-quality data would greatly

undermine the findings of DG SANCO's work as well as weakening the robustness of decision making based on those findings.

DG SANCO has already carried out work with regard to data quality and the design of appropriate practices in this area (DG SANCO, 2008, 2009b, 2009c).

Summary of DG SANCO's internal mapping on data quality

It is useful to reiterate here the key points that resulted from DG SANCO's initial internal mapping on data quality. This will enable us to review these findings against the main points that were made by our interviewees regarding both current practices and key issues around data quality. The points were as follows:

- The most important prerequisite is to clarify the objectives of the data collection and the intended use of the data. Regular consultations with Member States and all relevant stakeholders are key to this aspect.
- The collection of data should be harmonised as far as possible to ensure comparable data are collected.
- Caution is needed when drawing conclusions by comparing data across countries. There is a need to ensure that the data are being compared on the same basis across countries.
- Using reliable and trustworthy sources is important in collecting good-quality data. Data obtained from Eurostat or from Eurobarometer, as well as comprehensive data packages (e.g. cancer registries, Annual Pesticide Residue Report by EFSA), appear to be of very high quality. Typically, this type of data is comprehensive, well validated, representative across Member States, timely, accurate and repeatable.
- When using data from stakeholders / interest representatives, there is a need to acknowledge that the data are potentially biased; the alternative is to exclude the data altogether.
- One way to test the reliability of the data collected is to have them cross-checked by a third party (another DG SANCO service or a relevant international organisation).
- Data should always be checked against the best available data collection methodologies and data sources. Any caveats should be presented and discussed.

Current practices around data quality

Our interviews with DG SANCO staff identified three common current practices across directorates and units with regard to checking data quality, as follows.

1. **Verify and assess reliability of data source** – this was mentioned by 11 interviewees, including 7 who mentioned that they trust Eurostat data because of the high-quality standards set. Interviewees mentioned that some data, in particular those obtained from stakeholder groups, could be biased and should therefore be used with caution. In particular, one interviewee working in the field of biotechnology mentioned that this field was particularly prone to biased

information, given its highly political and ethical nature. However, interviewees also stressed that this issue was sometimes exacerbated by the fact that no other source of data existed (e.g. on the cost to industry) and that they therefore had to work around data that might be biased (e.g. by being clear about caveats when using these data).

2. **Look at comparability of data** – eight interviewees referred to checking if the definitions and indicators used to produce the data are comparable across Member States.

3. **Distinction between factual/'hard' data and perception/'soft' data –** some interviewees (5) mentioned that they distinguished between 'hard' or factual data and 'soft' or perception data. Two interviewees reported that checking the quality of perception data was a challenge, whereas another two interviewees stated that ensuring the quality of 'hard' data was straightforward because such data use less ambiguous definitions, and tend to be accurate and reliable.

Overall, interviewees mentioned that there were no guidelines on data quality in their directorate or unit, which indicates that there are currently no overarching data quality guidelines or quality-assurance processes across DG SANCO or within particular directorates or units. Only one interviewee mentioned having access to guidelines for dealing with data quality within their unit (Unit 01). Two interviewees said that they received guidance informally from more experienced colleagues on checking the quality of the data they are using.

Current issues regarding data quality

The following issues were the ones most cited when interviewees were asked about the challenges they faced in assessing the quality of the data they are using:

1. **Difficult to find timely and accurate data** (7 interviewees) – this relates to the fact that recent or longitudinal data are not always available as well as to the issue of finding accurate data. One interviewee talked of the difficulties of obtaining timely data on which to build the drafting of regulations from long-term projects as these often delivered their outputs once the drafting of regulation was in its final stages.

2. **Time pressures** (3 interviewees) – this relates to the lack of time available to check that the data collected are of adequate quality.

3. **Data gaps and comparability** (4 interviewees) – interviewees referred to the problem of receiving incomplete data from Member States and to the fact that estimates have to be used at times when complete data are not available. Another issue was that of Member States using different definitions or indicators to provide the data in the field of public health, which meant that the data obtained were often not comparable across countries. One interviewee explicitly stated that they often had to use such data if no other data were available and that they were clear about the limitations and caveats of the data when doing so.

4. **Difficult to assess quality of 'soft' or perception data** (4 interviewees) – one interviewee mentioned that he found it difficult to come by high-quality quantitative data. This interviewee said the following about data quality: 'the gold

> standard [of good-quality data] is something that is measuring a clear and well-bounded phenomenon that is counted in the same way and according to the same definitions at the same time, by the same people across all of the countries'. These criteria frequently cannot be applied to assess the quality of 'soft' data; therefore this was difficult for DG SANCO staff.

Brief review of current practices and key issues against guidelines of internal mapping

From the 24 interviews carried out with DG SANCO staff, it is clear that some of the key guidelines emerging from the internal mapping exercise are currently being followed: namely checking data sources and comparability of data. However, it is also apparent that some of the guidelines will be more difficult than others to integrate into DG SANCO's current practices regarding data quality. This is because of some of the key challenges that DG SANCO's staff are currently facing – such as difficulties in finding timely and accurate data, time pressures and data gaps. Indeed, if time pressures are experienced by a large number of staff when dealing with data, it becomes difficult to abide by the guideline of having the data cross-checked by a third party or checked against the best available data methodologies. Similarly, if the only data available are not complete across all Member States, there will be a need to find ways of working around this shortcoming, which may include producing estimates from the data available and being explicit about the caveats. Finally, in cases where data availability is an issue – for example, if the only data available come from a source that could be biased (such as data from stakeholders, as in the example of the biotechnology field) or if the data are limited geographically – then it is foreseeable that data quality criteria will be compromised and that the data would have to be used to produce estimates with large caveats if no other option is available. Ultimately in many cases there will be a trade-off between using and providing some kind of timely data and the comprehensiveness and accuracy of data.

Key observations

- No formal overall guidelines are currently available to assess data quality.
- Data quality is most often assessed by staff or through Eurostat (data from Eurostat are assumed to be of good quality because of the standards Eurostat applies to the data it produces).
- The two main means by which staff assess data quality are (1) by checking the data source is reliable, and (2) assessing the comparability of the data.

CHAPTER 3 Strengths and weaknesses

3.1 Introduction

This chapter lists the strengths and weaknesses of DG SANCO's current data management practices that were identified in the 24 interviews[2] we conducted across its units and directorates. Interviewees were asked specifically for their views on the strengths and weaknesses of DG SANCO's current data management practices as a whole. These issues were raised in addition to those mentioned for each data stage in Chapter 2.

Some issues, such as collaboration and skills, appear as both weaknesses and strengths. While this might appear to be inconsistent at first sight, we believe it is instead a reflection of current data management at DG SANCO. There are some units that have substantial experience and expertise in using data (such as some units within Directorates D, E, and F and the information units in B and C), while other units have more difficulties in finding, analysing and using data.

3.2 Strengths

During the interviews with staff, we asked them about the main strengths that they saw in current data management practices at DG SANCO. The following is a list of strengths that were mentioned. In order to give an idea of how widespread these findings were, we have included information on the number of interviewees who mentioned a particular strength. In addition, if an issue was mentioned by more than one interviewee whose work is linked to a particular policy stage,[3] this is recorded.

Collaboration, communication, coordination and support within DG SANCO

This was one of the two strengths most commonly cited, mentioned in 7 out of 24 interviews. Interviewees were very positive about internal collaboration and knowledge sharing within DG SANCO. Most of the collaboration and coordination took place informally, although one interviewee mentioned more formal and centralised means of support and coordination such as Unit 02 and the DG SANCO library. Three interviewees who mentioned this strength worked principally on policy evaluation and

[2] One of the interviews we conducted was with one staff member from Unit B1 and one from Unit B4, so we interviewed a total of 25 staff across 24 interviews.

[3] These policy stages were defined by Unit 02 for the purpose of this project and are as follows: policy strategy setting and planning, impact assessment, policy evaluation and enforcement, inter-institutional process and communication.

enforcement, one worked mostly on impact assessments, and the remaining three worked on more than one policy stage on a regular basis.

Availability of data

Availability of data was the other most commonly cited strength, again mentioned in 7 out of 24 interviews. Three of the interviewees who identified this strength mostly worked on policy evaluation and enforcement. They reported that data were available because Member States were required by legislation to provide it to DG SANCO, and that they had access to the right databases. One interviewee who worked in an information unit and had worked on an impact assessment stated that most of the data needed for their work were available through Eurostat and the Eurobarometer. Two interviewees who worked on more than one policy stage in the field of health mentioned that the availability of health data was generally good. The remaining interviewee said that data availability was a strength of DG SANCO because they had good access to international data sources.

Skills for processing data

Two interviewees reported that staff at DG SANCO had the right skills to process data. One of them, who worked on policy evaluation and enforcement, said that staff in his unit were well trained and had the necessary skills to work with data that were not homogenised across all Member States;[4] the other interviewee mentioned that there was a mix of skills available within DG SANCO to deal with data appropriately.

Contact with stakeholders

Contact with stakeholders was mentioned by two interviewees as beneficial to their work. They were able to obtain a wide range of data through bilateral contacts with stakeholders and consultations.

Increasing awareness of importance of data for DG SANCO

Two interviewees mentioned that data were becoming more important within DG SANCO and that this was a positive development as it would contribute to evidence-based policy. One of these interviewees worked on policy strategy setting and policy planning, and reported that policy within DG SANCO was becoming increasingly data led. The other interviewee, who worked in communication, said that the importance of data for policy making was increasingly recognised within DG SANCO.

Other issues

Further issues were individually mentioned as strengths of DG SANCO's data management practices, including:

- the availability of good guidelines for impact assessments – this was mentioned by one interviewee as a strength, although they also stated that these were, in their view, seldom implemented;
- the production of scoping papers – one interviewee found this helpful as these papers cut across policy areas and are a good source of data;

[4] In this context, 'not homogenised' meant incomplete data from Member States.

- good cooperation with Member States – the interviewee who reported this as a strength stressed that there was good data sharing between Member States and DG SANCO at the policy evaluation and enforcement stage;
- the management of external contractors – the interviewee who identified this as a strength stated that working with external contractors generally went smoothly.

3.3 Weaknesses

The following are the key issues highlighted during the interviews with DG SANCO staff. These are the underlying causes of some of the perceived difficulties DG SANCO currently has in dealing with data. It is interesting that some of the strengths listed in Section 2.2 appear as weaknesses in this section, emphasising the fact that many of the issues and findings are specific to the directorates or units in which staff work or to the policy stages on which they mostly work, rather than cutting across all policy areas and stages within DG SANCO.

Lack of specialist skills in-house

This issue was raised in seven of the interviews. It was mentioned by interviewees who worked across a wide range of policy stages. Some of these referred to the need for specialist knowledge generally in order to deal with data (4 interviewees) or the need to recruit more policy area specialists, including more economists and social scientists to deal with data more effectively (2 interviewees). Another interviewee mentioned that staff at DG SANCO 'stick to their comfort zone' and that this creates competence problems for dealing with data generally.

Some of the data available to DG SANCO staff are either not used at all or not well used

This issue was touched upon by five interviewees. Most of these interviewees (3) acknowledged that there was a lot of data available within DG SANCO but said that, for much of the time, these were either not used at all or not well used. One interviewee who worked mostly on policy evaluation and enforcement also stressed this point and referred in particular to data requested from Member States by DG SANCO, which was sometimes not useful and an unnecessary burden on Member States. The data had only been considered useful during the development of legislation, and after that they became less useful.

Current data management practices are reactive and crisis driven

Four interviewees mentioned that current data management practices were reactive rather than strategically led. As one interviewee put it, 'more action and less wishful thinking' were needed to develop data management practices. Another stressed that 'decisions are very ad-hoc and not based on real analysis', so vision was needed rather than 'short term reaction'.

Access to data is sometimes an issue

This was identified as a problem by four interviewees. These were spread across policy stages and policy areas. Two mentioned the cost of accessing data as a barrier in particular for market data and journals. The interviewee who referred to journals stressed that it was not only the financial cost but also the time involved in arranging payment for these journals that was an issue, in particular because policy officers had to pay for journal access

themselves and were reimbursed later. One interviewee mentioned that poor availability of import and export data was a problem and another said that the fact that DG SANCO did not produce its own data for many of the major policy areas it covers was an issue.

Harmonisation issues with data obtained from Member States

This problem was named by three interviewees. They referred not only to the fact that data obtained from Member States are often in different formats which are not readily compatible and require a lot of work to be made useful (2 interviewees), but also to the fact that these data submissions are not always complete (1 interviewee).

Policy is not sufficiently data led

Two interviewees reflected on the use of data in the policy process, and felt that the policy making process is not 'data led' – that is, it is not based on thorough analysis prior to initiating action. Both of these interviewees also mentioned that there was a lack of skills and competence to deal with data within DG SANCO, so it would appear that the two issues are interrelated (i.e. if the skills are not there to deal with data, then data are not put at the forefront of policy).

Timeliness is an issue

This was mentioned by three interviewees, who referred to the lack of time to deal with data; in some policy fields there is too much data to process, while in others there is a lack of time to collect adequate data as requests for data occurred at very short notice. The latter point also involved concerns about the time it takes to subcontract or collaborate with contractors or institutions such as Eurostat or EFSA to obtain data.

Other issues

Further issues that were individually mentioned as weaknesses of DG SANCO's data management practices included the following:

- Collaboration is very 'personalised' – the interviewee who mentioned this as a weakness said that collaboration between colleagues at DG SANCO is very informal and may be a barrier for some (e.g. new staff who are not familiar with colleagues in other units or directorates).
- Problems with contractors – this interviewee reported that it is sometimes difficult and time consuming to manage contractors. They also stressed that they felt limited by the choice of contractors within the framework contract, in particular because some of the specialist skills needed for their work were not available within the framework.
- Lack of reflection on future data collection needs – there is a need to define consultation strategy at an early stage.
- Overlap with other DGs (duplication of data collection) – the interviewee who mentioned this felt that DG SANCO has a lot of data available which are not currently put to the best use.
- No unified approach to data across DG SANCO – this interviewee said that different areas of DG SANCO currently have different data strategies because some policy areas have a longer history of data management than others, and that

this should be remedied to make dealing with data across DG SANCO more effective.

- Lack of awareness of what the best sources available are – one interviewee felt that not all staff were aware of the best sources of data available for their area of work.

The following chapter details some recommendations and measures that might be taken by DG SANCO to address some of the issues highlighted in this chapter.

CHAPTER 4 Recommendations and conclusions

This chapter presents some recommendations that should be considered when developing a data strategy for DG SANCO. The recommendations may be organised into two groups: (1) high-level recommendations and (2) specific recommendations to address some of the weaknesses highlighted by DG SANCO staff.

High-level recommendations set out the key principles that should guide the development of the strategy. These recommendations are based on the interviewees' view of what a comprehensive data strategy in DG SANCO should do and also on the evaluator's perception of the central themes that arose in the interviews. The recommendations focus mostly on the shift in organisation and culture that will be required for the development of a comprehensive data strategy. Therefore they probably represent changes that it may be difficult to implement and that will require a significant amount of time to embed. These recommendations are presented below.

4.1 High-level recommendations

1. **Clarify the purpose of a data strategy**
   Prior to developing a high-level data strategy, DG SANCO should define and clarify the purpose of a data strategy and define what it wants data to be used for. Currently a wide range of sometimes competing purposes is mentioned by DG SANCO staff, from monitoring enforcement action and implementation to communication to the general public. If DG SANCO wants to cover a wider range of purposes, it should make sure that the data strategy acknowledges the different uses for which data are intended.

2. **Think 'information' and 'knowledge' rather than data**
   In developing a comprehensive data strategy, DG SANCO should think about the knowledge and information it requires for policy making. Data will only become useful when they are structured, interpreted and understood in context. Thus SANCO should ensure that the capacity and skills to make use of data are available.

3. **Prioritise**
   Given the limited resources and skills available, it will be essential to prioritise the collection of specific data. This prioritisation should be based on the key questions that DG SANCO needs to answer – that is, the information and knowledge DG SANCO wants to obtain. One approach that may be used is logic modelling, which establishes a logic and theory of change.

4. **Create a decentralised but coordinated organisational structure system**
   Given the differences between directorates, both in term of subject area and type of activity, DG SANCO should consider a decentralised but coordinated organisational basis for its data strategy. This should be coordinated and supplemented by shared resources for the whole organisation, and regular communication between units should take place, including updates on their recent activities. The information units available should be used to a larger extent and could form the core of directorate-wide centres of expertise and support.

5. **Aim for cultural change**
   Any organisational and procedural change suggested here should be embedded in a process of cultural change. DG SANCO staff should feel that using data in their day-to-day work is relevant, valued and recognised within the wider organisation. Ultimately, DG SANCO staff will need to feel that using data makes a substantial difference to the policy process by improving the decisions taken.

4.2 Specific recommendations

Secondly, we arrived at a set of specific recommendations. These address the weaknesses identified by DG SANCO interviewees which are detailed in Chapter 3. We present them in Table 4.1 below before explaining them later on. We also show whether in the evaluator's judgement these changes can be implemented in the short term, medium term or long term. To make this assessment, it is important to consider the acceptability and feasibility of implementation, which involves the following:

- The extent to which changes are acceptable to key stakeholders in DG SANCO (desk officers, heads of unit, directors and wider participants in DG SANCO policy making).
- The extent to which these changes require shifts in how DG SANCO is organised and the statutory responsibilities of DG SANCO.

Table 4.1: Linking recommendations to the issues identified

Weakness mentioned	Related recommendation	Type of recommendation
Lack of specialist skills in-house	Produce a map of data management specific skills of DG SANCO staff	Short term
	Provide basic training on data use and management for DG SANCO staff	Medium term
	Establish a central survey unit	Long term
Some of the data within DG SANCO are either not used at all or not well used	Conduct a data inventory	Short term
Current data management practices are reactive and crisis-driven	Develop methods to define data needs systematically	Medium term
Access to data is sometimes an issue as well as how data are used	Develop basic guidance on the use of data in the day-to-day policy work, including where to find high-quality data sources	Short term
	Develop data quality guidance on how to assess and ensure quality	Medium term
Harmonisation issues with data obtained from Member States	Create a central knowledge base or information hub on data	Medium term
Policy is not sufficiently data led	Plan and discuss data needs as early on as possible in the policy making process (there are different data needs at different policy stages, e.g. policy setting, impact assessment)	Short term
	Make data use more visible	Long term
Timeliness is an issue in data management	Be prepared for ad-hoc data needs	Short term
Problems with collaboration in data collection and analysis	Become more concise in contracting out research	Short-term
	Strengthen external collaboration on data collection	Medium term
No unified approach to data across DG SANCO	Create decentralised support structures	Medium term

6. **Produce a map of data-specific skills of DG SANCO staff**
   To make best use of DG SANCO's current skills in collecting and using data, DG SANCO should consider mapping the data-specific skills of their staff and making this information internally available to allow for informal, collegial tutoring between staff and to improve the possibilities for internal collaboration. This would be a first step towards addressing skill imbalances between units and directorates.

7. **Provide basic training for DG SANCO staff**
   In addition, DG SANCO should consider providing basic training to interested DG SANCO staff with the aim of increasing their confidence in performing basic analysis, increasing their skills in assessing data quality and enabling them to be more confident in commissioning external data collection and analysis (i.e. working with contractors and consultants). By increasing the basic understanding of key processes and basic concepts of data management, DG SANCO would be better able to understand the nature and limitations of the work they are commissioning, which

would help improve the quality and clarity of terms of reference and ensure that the findings are useful.

8. **Establish a central survey unit**
   Surveys are often the only viable route for collecting quickly data that have not been systematically gathered previously. However, surveys are notoriously difficult to get right. DG SANCO should thus consider, in the long term, the creation of a specific central resource for conducting or contracting surveys. This would allow the build-up of expertise internally, ensure both a higher quality of data collection and comparability over time, and avoid duplication of efforts between units and directorates.

9. **Conduct data inventory**
   For all data that DG SANCO already collect on a regular basis, a systematic inventory should be established with the aim of harmonising data – that is, using consistent definitions and gathering techniques to make them consistent and comparable. Finally, DG SANCO should make internal data sources accessible to the whole of the DG where feasible, and should reconsider the collection of unused data to free up both its own and Member States' resources.

10. **Develop methods to define data needs systematically**
    DG SANCO should develop methods to define data needs systematically. This might take, for example, a logic modelling and dashboard or a systematic issue analysis approach. This would help DG SANCO to manage the time frames of data collection better, as well as being a tool for prioritising data sources.

11. **Develop basic data guidance**
    To help DG SANCO staff to increase their knowledge and confidence in dealing with data, particularly in policy units, DG SANCO should continue its efforts to develop basic 'how to' guidance for its staff.

12. **Develop data quality guidance**
    As part of a guidance package, DG SANCO staff should be provided with guidance on how to assess data quality. Given the overall availability of data, we suggest guidance following a pragmatic approach to data quality – that is, helping DG SANCO staff develop an informed understanding of the limitations of specific data sources.

13. **Create a central knowledge base or information hub on data**
    To address the problem of insufficient knowledge about internal and external data sources available, a central knowledge base in which information about data sources could be deposited should be established. In its simplest form this could be a 'wiki page', updated by the directorate's service units; in its most elaborate form this would be a central database. The key purpose of this recommendation is to induce a better understanding of what data are used within DG SANCO already.

14. **Plan and discuss data needs as early as possible in the policy making process**
    DG SANCO should consider discussing data needs at the policy setting and strategy phase, rather than only in the impact assessment phase. This would allow for more effective evidence-based policy making right from the beginning of problem

definition rather than solely in evaluations of policy. It would also ease time pressures on data collection and analysis later in the process, and economise on resources when conducting evaluations, as the baseline scenario will already be known rather than having to be reconstructed post hoc. For new or amended regulation, a new data section could outline the data needs of the specific dossier.

15. Make data use more visible

To make the use of data more visible both to internal and external audiences, DG SANCO could introduce, for example, a statistic of the month/week section on its website, linking good data to specific DG SANCO policies. This would demonstrate that good data are important for DG SANCO and reward best practice in data use, and also show DG SANCO's commitment to evidence-based policy making, thereby enhancing its legitimacy to European citizens and businesses.

16. Be prepared for ad-hoc data needs

The nature of the political process makes ad-hoc data needs unavoidable; therefore DG SANCO should be better equipped to respond to these by having a good knowledge of data currently available and should be prepared to gather additional data quickly if necessary. An in-house survey unit, for example, is a tool to react quickly to new challenges.

17. Become more precise in contracting out research

Contracting out research and data gathering is an important element in increasing skills and capacity. To maximise the value of contracting, DG SANCO should ensure that findings will be relevant when delivered and should set stronger priorities in the data it requests. DG SANCO should also explore how it might have quicker access to more specialised contractors.

18. Strengthen external collaboration on data collection

DG SANCO should further develop its external collaboration on data. Good links with Eurostat already exist within some units, and further efforts should be made to engage with other DGs and international organisations.

19. Create decentralised support structures

Create a data support unit for each directorate by developing a stronger service function in information units that already exist, such as B1 and C2. The units should not only be able to provide technical and methodological advice, but should also have high awareness of subject-specific data sources. These units should be consulted to a large extent at key stages of the policy process, such as drafting impact assessments.

REFERENCES

Reference list

Brutscher, Philipp-Bastian, Jan Tiessen, Ala'a Shehabi, Daniel Schweppenstedde, Claire Celia, Christian van Stolk and Tom Ling, *Data Strategies for Policymaking. Identifying International Good Practice*, Santa Monica, CA: RAND, 2009.

DG SANCO, 'Data Quality. Note to Mr Madelin', Brussels, European Commission Health and Consumers Directorate-General, 2008a.

——— '2010 Information Technology and Document Management "Schema Directeur",' 2009a.

———, 'Data Quality Data Strategy – First Results of Internal Mapping', Brussels, European Commission Health and Consumers Directorate-General, 2009b.

———, 'Data Strategy – First Results of Internal Mapping', Brussels, European Commission Health and Consumers Directorate-General, 2009c.

———, 'European Health Information - Objectives and Organisation,' Discussion Paper. Brussels, European Commission Health and Consumers Directorate-General. 2008b.

———, 'Sanco Data Needs – Initial Paper on Mapping', Luxemburg, European Commission Health and Consumers Directorate-General, 2009d.

———, 'Sanco Data Strategy-Mapping of Data Sources, Practices and Needs for Directorates D, E and F', Brussels, European Commission Health and Consumers Directorate-General, 2009e.

European Commission, *The Consumer Markets Scoreboard.* 2nd Edition, Luxemburg: Office for Official Publications of the European Communities, COM (2009) 25, 2009.

Ling T. and L. Villalba-van-Dijk (2009). Performance Audit Handbook. Routes to Effective Evaluation. Santa Monica, CA, RAND, pp.152-167.

Osborne, D. and T. Gaebler (1992). *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*. Reading, Massachusetts, Addison Wesley.

Rieper, Olaf, Frans L. Leeuw, and Tom Ling, eds. *The Evidence Book. Concepts, Generation and Use of Evidence,* Edited by Rist, Ray C. *Comparative Policy Evaluation*, Vol. 15. New Brunswick (USA) and London (UK): Transaction Publishers, 2009.

Rowley, J., "The Wisdom Hierarchy: Representations of the Dikw Hierarchy", *Journal of Information Science,* Vol. 33, No. 2, 2007, pp.163–180.

Tiessen, J., C. Celia, L. Villalba-van-Dijk, A. Reding, C. van Stolk and T. Ling, 'Impacts and Outcomes of Local Authority Regulatory Services', 2009.

APPENDICES

Appendix A: Developing top-level indicators for DG SANCO

Introduction

Part of DG SANCO's aim in developing a comprehensive data strategy is the development of general yet relevant top-level indicators for each policy area and sector that can be used at different stages of the policy process. Typically, such top-level indicators monitor the developments in a specific policy field in relation to the status quo or problem (e.g. life expectancy, healthy life years), and it is implicitly assumed that European policy (or the lack of it) might have contributed to the current status quo and development of the problem. Even the very elaborate and comprehensive Consumer Market Scoreboard (European Commission, 2009) focuses primarily on the functioning of several consumer markets and does not provide indicators about European consumer policy that might help to understand the underlying effects of changes. In this appendix we suggest an approach to developing top-level indicators that not only includes status indicators, but also allows for the development of indicators to monitor the effects of policy upon the status of the perceived problems.

Approach

This approach, which sits within the 'theory of change' tradition of evaluation,[5] consists of three analytical steps, each with a specific tool that could be used by DG SANCO in defining key top-level indicators:

1. mapping the intervention logic using logic models;
2. identifying and defining indicators;
3. presenting the findings in a dashboard.

Intervention logic

At the core of this approach stands the development of so-called intervention logic – that is, a representation of how DG SANCO's activities contribute to achieving the five high-level policy objectives that DG SANCO formulated:

[5] For a more detailed discussion of this approach see Tiessen *et al.*, 2009, or Rieper *et al.*, 2009.

1. Empower consumers
2. Protect and improve human health
3. Ensure food is safe and wholesome
4. Protect the health of animals and plants
5. Promote the humane treatment of animals.

The aim of this step is to identify the key causal chains that lead from a policy measure of DG SANCO to the five desired outcomes. If these causal chains are identified and 'isolated', they can be used later to structure and focus the measurement activities. One such causal chain would be DG SANCO's activities to protect European citizens from the harmful effects of second-hand smoke, which should contribute to a higher level of protection of human health in the European Union. A key tool for establishing intervention logics is logic modelling.

Logic models (Kellogg Foundation, Logic Model Development Guide)[6] enable us to produce a graphic representation of how a policy is intended to work – that is, how resources are converted into programme activities, and how those activities in turn produce the results intended. Therefore, logic models generally allow a researcher to analyse the relationship between inputs and outputs, and between inputs and outcomes. Logic models provide an opportunity within an 'accountability area' (Osborne and Gaebler, 1992) to measure results, correct problems and identify successes. It also ensures a shared understanding of the intervention and helps uncover any implicit disagreements and confusions. An abstract version of a logic model is shown in Figure A.1 below.

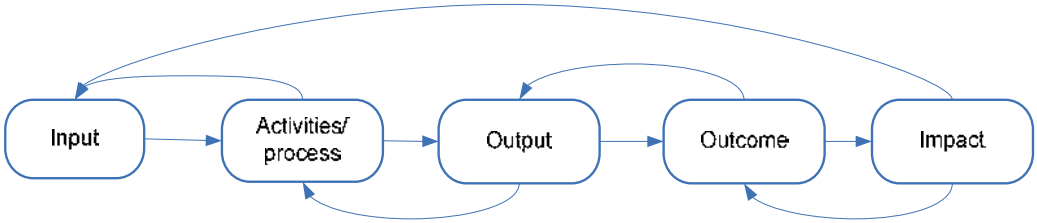


Figure A.1: Outline of a basic logic model

Source: RAND Europe, 2010

Logic models usually stop short of formulating specific links between the elements within each category. A logic model would, for example, list a number of activities as well a number of outputs, without linking the specific activity to a specific output and then a specific outcome. For the purpose of developing DG SANCO's intervention logics, we would therefore suggest supplementing the logic models with elements known from process mapping, by indicating links between the elements of the logic model and

[6] Kellogg Foundation, Logic Model Development Guide, online, available at: http://www.wkkf.org/Pubs/Tools/Evaluation/Pub3669.pdf (last accessed July 2009); see also see Ling and Villalba-van-Dijk (2009).

highlighting interactions with key partners, such as Member States, other DGs or executive agencies.

To ensure accuracy, as well as the usefulness and acceptability of such an intervention logic, it is best done in a collaborative way, such as a workshop.

Indicators and measurement

Once the intervention logic has been established and the key causal chains have been mapped out, the next stage is the identification of potential indicators that would provide a fair description of changes along the causal chain.

This list of indicators can be informed by the available, very rich knowledge of the specific policy areas within DG SANCO. The outcome of this mapping stage would be a long list of indicators that could be used to measure the elements of the causal chain, and an assessment of whether these indicators are readily available or not.

Findings from the interviews suggest that specific difficulties in identifying data along the whole causal chain may differ between policy areas and directorates. For example, we would expect Directorates D, E and F to have very good data on Member States' actions in their policy fields. That is far more difficult in public health work.

Building a dashboard

The final stage of identifying top-level indicators would be to prioritise and present indicators in a way that enables a quick but meaningful overview of how the most important policies contribute to achieving DG SANCO's high-level policy objectives. To do so, RAND Europe proposes to adapt the idea of a management 'dashboard'. Dashboards are executive information systems that present a small set of performance measures on a regular and structured basis to strategic decision makers in order to provide an overview of the organisation's performance, and thereby to identify areas of particular success or concern for more detailed examination.

In a situation that is awash with different performance measurement, indicators and targets, we suggest the use of the dashboard approach in order to make more sense of data and indicators that already exist and to provide guidance on how to link this information in a meaningful way. However, this dashboard will have a less operational perspective than management dashboards and will not need to be updated as often as a management tool (which often can be 'real time' as well). Instead, we envisage the dashboard as a more strategic tool that should be updated once or twice a year.

The key challenge of a dashboard lies in the selection of data sources and indicators. For this purpose, we suggest selecting a maximum of 16 to 20 sources and indicators. Criteria for the prioritisation of indicators and measurements should be systematic and pragmatic and should include the following:

- Does it cover a key causal chain?
- Does it cover an input, output, outcome or impact?

- Is it being collected already?
- Are data held by DG SANCO or its external partners?
- Will new data be collected?

The final step of the dashboard is to represent the indicators in a graphical interface. An example of a dashboard that structures information into inputs, outputs, outcomes and impacts may be found in Figure A.2 below. The example looks at anti-smoking interventions at the local level in the UK.

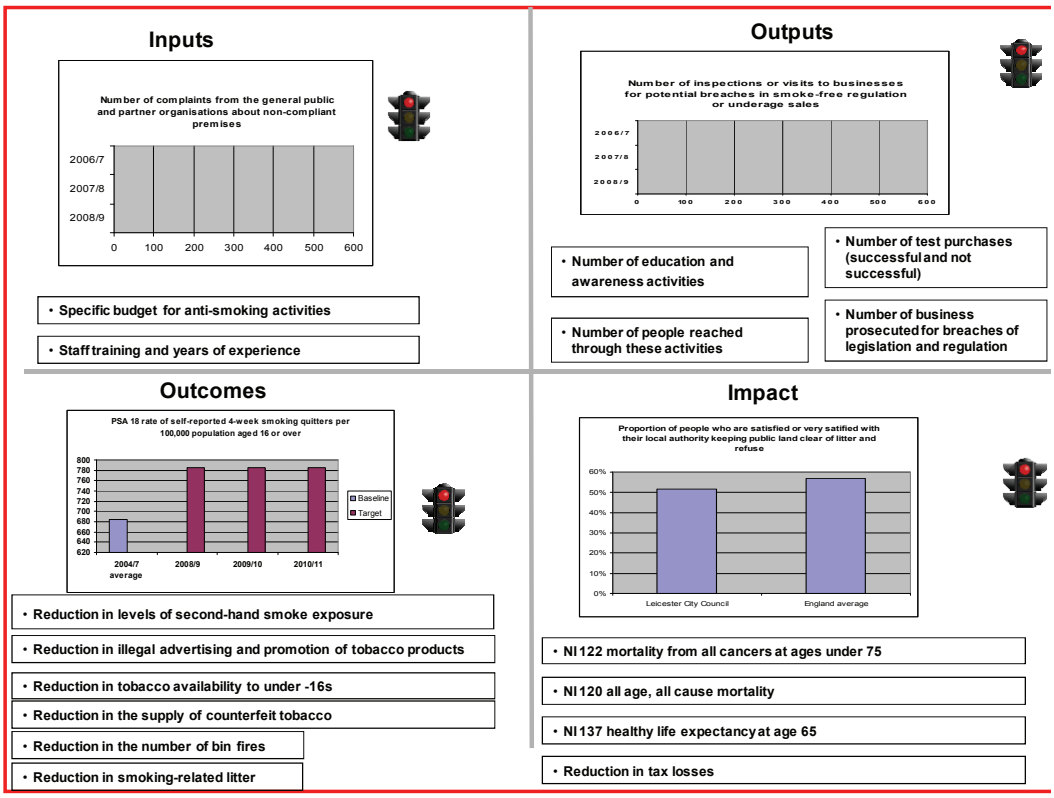


Figure A.2: Example of a dashboard for local anti-smoking interventions in the UK

Source: RAND Europe

Limitations: attribution, contribution and scalability

Three key questions need to be taken into account when analysing the impact of the activity of DG SANCO:

- How much of the impact may be attributed to the work of DG SANCO?
- To what extent has DG SANCO contributed to the achievement of specific outcomes?
- Can the approach be scaled up to cover all the activities of DG SANCO?

'Theory of change' approaches cannot completely overcome the conceptual difficulties of attribution and contribution. These relate to complexities in the dynamic of policy making

such as subsidiarity and the additionality of European Commission policy making. Nonetheless, the 'theory of change' approach allows evaluators potentially to:

- link DG SANCO action to a specific policy problem in a systematic way;
- understand the detailed causal linkages between DG SANCO action and intended and unintended impacts and outcomes;
- make sense of collected data and identify gaps for future monitoring;
- conduct further analysis by narrowing down and framing research questions.

In principle, 'theory of change' approaches would be scalable to cover all activities of DG SANCO. However, given the number of activities of DG SANCO, the approach would need to cover most of the initiatives and associated causal chains in the directorates. This has resource implications. The following section presents a worked example in the field of public health.

Improving public health: a worked example of revisions in the nutrition labelling regulations

DG SANCO in recent times proposed a number of changes to nutrition labelling in the European Union with a view to harmonising some of the information provided on nutrition labels to consumers. The background to the revision in nutrition labelling regulation was that different labelling schemes are confusing to consumers, and harmonisation of labelling requirements was seen as important in promoting better dietary choices. The theory of change in nutrition labelling relates to raising awareness and providing information and can be understood as follows:

- information provided to consumers …
- leads to increased consumer knowledge about nutrition …
- leads to better decision making on nutrition …
- leads to better health outcomes (less diet-related morbidity and mortality).

In this example, the inputs or resources are mostly provided by food manufacturers, who have to change the labels. There are some minor costs associated with Member States because they have to implement and enforce the regulations. However, the additionality of these costs on existing work was seen as minimal. In terms of process, the main activity consisted of changes in labels by manufacturers and Member State implementation and enforcement activities. The output of this would be the degree of implementation and compliance. The impact would consist of a number of potential indicators linked to the causal mechanisms:

- The number of consumers with an increased understanding of nutritional information relating to their food.
- The number of consumers changing their dietary choices on the basis of information provided.

- The related health impacts of these choices in terms of morbidity and mortality.

This list is important because it gives an understanding of some of the data that DG SANCO could be capturing to understand the impact of its policy proposal. A macro-level impact indicator would relate, for instance, to obesity or associated health conditions, with subindicators on the understanding of consumers and their dietary choices.

Appendix B: Success indicators to evaluate the future data strategy

There are two different ways in which DG SANCO could assess the effectiveness of its data strategy. First, we could assess the outputs coming from DG SANCO in a 'before and after' evaluation design. The outputs would consist of:

- communication documents;
- impact assessments;
- policy evaluation and enforcement reports;
- policy setting and planning reports.

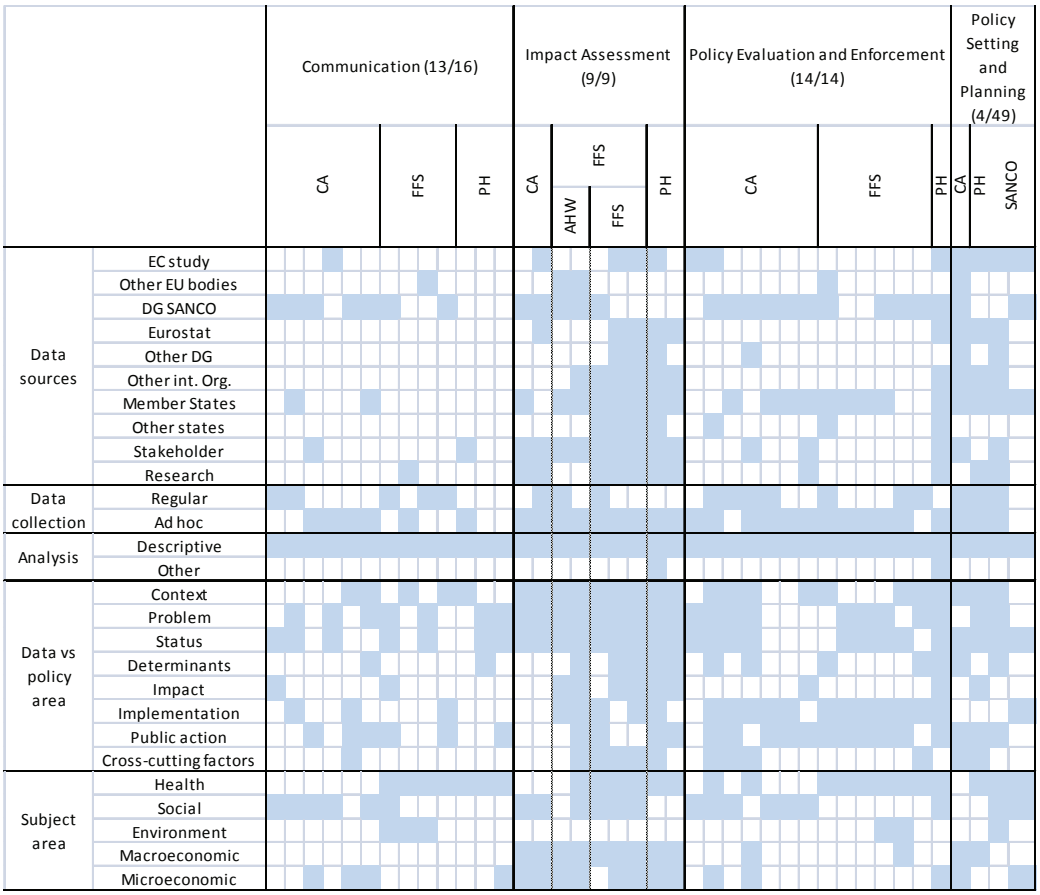
These outputs can then be assessed with regard to the type of data presented in them, the quality of the data used and the analysis provided. The last of these requires an understanding of what is meant by 'data quality' and 'quality of analysis'. Some aspects of quality assessment are provided in Chapter 3, but would need to be considered further by DG SANCO. **Error! Reference source not found.**Table B.1 gives an overview of the mapping exercise that RAND Europe conducted on DG SANCO's documents. The columns show the type of documents that we reviewed for each directorate and the rows show the source of data for each document. Certain documents did not contain data or analysis as such.

The 'before and after design' offers DG SANCO a way to assess how the quality of data and analysis in its outputs has changed over the course of the data strategy being in place. This analysis may be strengthened by looking at other external evaluations of DG SANCO outputs, such as feedback from the Impact Assessment Board.

Secondly, an evaluation could look at how the principles of the data strategy have become embedded in the work of DG SANCO. Such an evaluation would require a before and after survey, which would ask DG SANCO desk officers about their practice. The analysis of the information might lead to an assessment of the extent to which practice had changed. A template for a survey was provided by RAND Europe in the course of this study.

Of course, evaluators could also choose a combination of both approaches. It is important to note that any assessment of the success of the implementation of the strategy would need to take into account external factors that influence data management within DG SANCO.

Table B.1: Overview of mapping exercise conducted as part of this research



Source: RAND Europe

Appendix C: Practical tips – a checklist approach

In the course of conducting this study, a number of practical tips emerged. These were collected on the basis of interviews with DG SANCO staff, the review of documents, and the international comparison report that RAND Europe drafted for DG SANCO in the spring of 2009 (Brutscher *et al.*, 2009). We present these practical tips in Table C.1 as a checklist to be used by DG SANCO staff who work frequently with data. The checklist contains a number of questions with yes/no answers. A 'yes' indicates that the issue has been considered or addressed by DG SANCO staff. A 'no' may denote a problem or risk in how data are managed that requires further attention. This checklist is indicative and could be further developed. For instance, further questions and categories could be added, or existing ones could be edited. Furthermore, the checklist does not give an indication of the level of risk or the degree of the problem. More qualified or detailed questioning might allow DG SANCO to make an assessment of the level of the problem or risk. However, on the basis of our work and because of the limitations of conducting interviews, it would have been difficult to produce a full risk register that allowed for a more fine-grained analysis of where the risks and problems may exist. We have organised the questions in terms of the main areas in data strategy.

Table C.1: A checklist to inform the work of DG SANCO staff

Stage	Question	Yes/ No	Remark
Data identification	Have you mapped all relevant data sources available from established sources for your project?		'Established' in this context means data sources to which DG SANCO could reasonably expect to have access, such as those of Member States, Eurostat, DG SANCO itself, other Commission services, and the main international organisations such as the WHO and OECD. An issue that became apparent in the interviews is that sometimes DG SANCO staff are not aware of data sources in use in other parts of DG SANCO or in other parts of the Commission.
	Have you assessed how relevant these data sources will be to answer your policy or research question?		This refers to whether a data source gives direct or indirect information on the policy problem you are trying to answer.

	Have you assessed the quality of these data sources?		There is no central guidance in DG SANCO on how to assess the quality of data, but other organisations such as the Organisation for Economic Cooperation and Development (OECD) have such guidance. Guidance normally takes into account whether data are longitudinal and how they are collected.
	Have you mapped the limitations of the data sources in answering your policy or research question?		An important aspect of using data following from the previous question is to outline the associated problems. Knowing what the data will not tell you makes it easier to understand what the data will tell you.
	Do you have more than one data source available to answer the research question?		Triangulation of data is important in answering policy questions. Overreliance on one data source may indicate a problem for analysis. By comparing different data sources the robustness of the findings will be increased and the potential shortcomings of one data source will be identified.
	Have you assessed the gaps in data availability?		By making a careful assessment of data availability and the quality of this data, an assessment of the gaps between the data available and the questions that need answering can be made.
	Do you have an approach to overcome these data gaps?		The approach here may mean the collection of primary data, the use of assumptive modelling or the use of proxy variables/data. The use of outside contractors was mentioned in interviews, but may not be a substitute for understanding the data gaps upfront.
Data collection	Have you considered the limitations of collecting data?		This refers to problems a study team might have in terms of resources available or lack of skills in collecting data.
	Have you thought about the limitations of using one approach compared to others?		Merely collecting data is not a guarantee that the data are of good quality. A critical assessment of the collection approach gives a good context for interpreting the data.
	Have you mapped the problems in data collection?		Understanding limitations is one aspect, but identifying the specific problems is more likely to lead to the solution.
	Have you looked at an approach to overcome problems in data collection?		The approach here may mean the use of external consultants, other relevant experts in DG SANCO (e.g. Unit 02), obtaining more money to do primary research, collaboration with partners, and

			consultation with other parts of the Commission. The interviews showed that often interviewees had lofty expectations of external contractors and did not consider working with others to collect primary data.
	Have you thought about how and where data would be stored?		An issue that came up in interviews is that data often are not stored in a standardised and user-friendly way in DG SANCO. Working with partners such as Member States and Eurostat may make it easier to store data on DG SANCO systems in order to facilitate wider and easier access to data. This also indicates a need for more dialogue between DG SANCO desk officers and those managing the databases.
Data analysis	Have you considered your limitations in analysing data?		Data analysis requires specific skills. It is clear from the interviews that these skills are not always present in all units of DG SANCO.
	Do you fully understand the options available to you in data analysis?		Knowledge of research methods leads to a better assessment of what type of analysis is applicable and also likely to be best.
	Have you looked at an approach to overcome problems in data analysis?		The approach here may mean the use of external consultants, other relevant experts in DG SANCO, collaboration with partners, and cooperation with other parts of the Commission such as Eurostat.
	Do you have a plan in place to check the quality of the analysis?		It is important for data analysis to be peer-reviewed or checked by parties external to the project.
Dissemination of findings	Have you considered your audience in the study?		The audience is relevant in how results are communicated and also in terms of what data are collected and how they are analysed.
	Have you devised a plan to communicate the findings clearly to your audience?		Communication is important in informing policy debates and ensuring the relevance of the study undertaken.

Appendix D: Interview questionnaire

Introduction to the project

DG SANCO is currently reviewing and developing their data strategy. As part of those efforts Unit 02 commissioned RAND Europe to conduct an evaluation of the current data management exercise. This evaluation will be built on three streams of evidence: an online survey, a document review and a number of key informant interviews, of which this interview will form an essential part. The ultimate aim of the evaluation will be to assess the strengths and weaknesses of the current data management practices and to suggest improvements for DG SANCO.

Your experience of using data for policy making

In this first section we want to learn more about your experience of using data for policy making. To make our discussion more tangible, we would like you to consider a recent example where you used data for policy making to answer questions 1.3 to 1.6.

Introduction

Could you please describe your position and role within your unit?

Recent experience in data use

Do you have recent experience in using data in one of the policy stages listed below? If so, could you please describe to us what your activity involved?

a. Policy strategy setting / policy planning

b. Impact assessment

c. Policy evaluation and enforcement

d. Inter-institutional process

e. Communication

Data needs and prioritisation

- How did you identify and define the data needs?
- Was there a formal decision about which data to use for your purposes? Were there guidelines you could use in choosing the right data?

- How did you proceed to choose one data source rather than another (most recent data, most complete time series, most focused on EU data, most recognised organisation, regularly used, well-known source, etc.)?
- Did you receive support and guidance on how to define data needs? If so, who provided it?
- Did you collaborate with any other organisations in defining data needs? If so, which ones?
- In summary, what were the key difficulties you encountered in identifying data?
    - Capacity?
    - Resources?
    - Skills?
    - Collaborations?

Data collection

- How did you collect the data?
- Did you receive support and guidance on how to collect these data? If so, from whom?
- Which tools did you use to collect the data (directorate internal data sources, EC information systems, Eurostat, Google, external contract)?
- Which data sources were available to you? Were any of the data you needed not found / not available or not accessible to you (e.g. require payment or a subscription)?
- Did you have to collect new data (primary data)?
- Did you collaborate with other organisations and partners in the process of collecting these data? If so, which ones? Which partnerships worked best, and for what reasons? What tools have been put in place to facilitate the cooperation on data?
- In summary, what key difficulties did you encounter in collecting data?
    - Capacity?
    - Resources?
    - Skills?
    - Collaborations?

Data analysis

- How did you use the data? What kind of analysis did you perform with the data collected?
- Did you collaborate with other organisations and partners in the process of analysing the data? If so, which ones?

- Did you receive support and guidance on how to analyse data? If so, from whom?
- In summary, what key difficulties were there in analysing data?
    - Capacity?
    - Resources?
    - Skills?
    - Collaborations?

Data quality

- How did you assess the quality of the data you used and the analysis you performed?
- What criteria do you use for ensuring data quality in your directorate?
- What is your best-quality data set? Why is it so good?
- Do you have any examples of poor-quality data you have commissioned? With hindsight, what was/were the problem(s) with these data?

Strengths and weaknesses in DG SANCO's current data management system

In this section we should like to learn more about your overall assessment of the current state of data management at DG SANCO and ask about the strengths a future data strategy can build on as well as the weaknesses it should address.

Strengths

- From your experience of using data at DG SANCO, which would you say are the elements you think work very well?
- In your view, what is the single most helpful feature of the current data management practice at DG SANCO?

Weaknesses

- What are the elements of current DG SANCO's data management strategy that need improvement the most?
- In your view, what is the single weakest element of DG SANCO's data management practice?

Recommendations

- What would you improve in the current data management practice? How would you improve it?

Your work package

You have been recommended to us for a key informant interview because you have been involved in one of the work packages for the internal mapping exercises. In this section the

research team wants to learn more about your work package and the key findings emerging from it.

- Could you please provide a brief overview of the work package on data quality you were responsible for?
- What are the key findings in terms of data quality?
- What are the key challenges DG SANCO needs to address to improve/maintain quality?
- What would be your recommendations to develop further data quality at DG SANCO?

Appendix E: List of interviewees

Table E.1: Mapping of interviewees by directorate or central unit

Directorate or central Unit	Unit within directorate
Unit 01 Audit and Evaluation	Mentor Murtezi
Unit 03 Science and Stakeholder Relations	Jeannie Vergnettes
A General Affairs	A1 Institutional relations and communication – Jan Marek Ziolkowski
B Consumer Affairs	B1 Consumer markets – David Mair, Luca Protti B4 Financial services and redress – Isabelle Rouveure, Rosella Delfino, Sebastian Bohr B5 Enforcement and European consumer centres – Paolo Catalini
C Public Health and Risk Assessment	C5 Health strategy and health systems – Maya Matthews C4 Health determinants– Charles Price C7 Risk assessment – Takis Daskaleros C3 Health threats – Paolo Guglielmetti C6 Health law and international – Anna Jassem-Staniecka C2 Health information – Nick Fahy, Federico Paoli
D Animal Health and Welfare	D1 Animal health and standing committees – Francisco Reviriego Gordejo D2 Feed – Ana-Marie Cucuteanu
E Safety of the Food Chain	E5 Enforcement – Joaquim Odeig Vila, Veronika Jeszo E1 Biotechnology and plant health – Marco Valleta E4 Food law, nutrition and labelling – Helen Lee E2 Food hygiene, alert system and training – Jose de Felipe
F Food and Veterinary Office (FVO)	F1 Country profiles, coordination of follow-up – Aidan O'Connor F7 Quality and planning development – Juha Juntilla

Appendix F: Mapping of data sources by directorates and units

Table F.1: Data sources produced by DG SANCO

MS: Member State(s)

Database	Description	Used by	Legal basis
EUROPHYT – European Network of Plant Health Information Systems	Notifications from MS on interceptions of harmful organisms or the detection of non-compliance with requirements listed in EU plant health legislation in consignments from third countries (import controls)	FVO – F4	Council Directive 2000/29/EC
RASSF – Rapid Alert System for Food and Feed	Network tool to exchange information about measures taken in response to serious risks detected in relation to food or feed	Managed by E2 Directorate E (1) Directorate F (1)	Regulation (EC) 178/2992 (art. 50)
TRACES – Trade and Control Expert System	Web-based system networking veterinary authorities in all MS and economic operators	Managed by D1 Directorate F (1)	Commission Decision 2003/623/EC
ADNS – Animal Disease Notification System	Notification system that has as its main purpose the registration and documentation of certain important infectious animal diseases	Directorate D (1) Directorate F (1)	Council Directive 82/894/EEC
TSE-BSE – transmissible spongiform encephalopathy - bovine spongiform encephalopathy database	Database on TSE monitoring programmes from MS	Managed by E2	Voluntary and Regulation (EC) 999/2001
Residues of veterinary	Database on residues	Managed by E5	Council Directive

medicinal products	control plans in MS		96/23/EC
EU pesticides database	Database on pesticide maximum residue levels (MRLs)		Reg. (EC) no. 396/2005
Zoonosis database	Database on zoonosis monitoring reports from MS	Managed by EFSA	Directive 2003/99/EC
Avian influenza surveillance online reporting database	Database on avian influenza surveillance programmes in MS, quarterly reports		Decision 2008/897/EC
BT-Net – Bluetongue	Daily, monthly and biannual reporting		Regulation (EC) 1266/2008
MisDoc – Management Information System on Documentum Platform	Repository warehouse on planning, reporting and follow-up of FVO missions	Directorate F (1)	
FVO (Food and Veterinary Office) inspection report search engine	Publication tool and search engine for FVO inspection reports on Europa website		
FSDMS – Food Safety Data Management System	Data warehouse of ten existing data sources (including EFSA databases)		
RAPEX – rapid alert system for dangerous consumer products		Directorate C (1)	

Source DG SANCO, 2009a; RAND Europe; informant interviews with DG SANCO staff

Table F.2: Data from other DGs, agencies or EU institutions and other international organisations

Data source	Description	Used/managed by
COMEX	Council of Europe Committee of experts	Directorate E (1) Directorate F (1)
ECDC	European Centre for Disease Prevention	Directorate C (3) Directorate E (1) Directorate F (1)
EFSA – European Food Safety Agency	databases and scientific opinions	Directorate C (2) Directorate F (1) Unit 01 (1) Unit 03 (1)

Eurostat		Directorate B (3) Directorate C (3) Directorate E (3) Unit 01 (1) Unit 03 (1)
Eurobarometer		Directorate B (2) Directorate C (1)
EFSA and Community Reference Laboratory database	For validation reports on GMO events	
European Chemicals Agency		Directorate C (1)
European Medical Agency		Directorate C (1)
EEAC – Equal Employment Advisory Council		Directorate C (1)
Eur-Lex database of community legislation		Directorate F (1)
DG AGRI – European Commission Agriculture Directorate-General	Reports and studies	
DG RTD – European Commission Research Directorate-General	Reports on research projects	Directorate E (1)
EMEA – European Medicines Agency	Opinions and reports	
WHO – World Health Organisation	(including INFOSAN – International Food Safety Authorities New work – on health threats)	Directorate C (2) Directorate E (1)
OECD – Organisation for Economic Cooperation and Development		Directorate C (3)
WTO – World Trade Organisation	On international audit standards	Directorate F (1)
OIE – World Organisation for Animal Health	On international audit standards	Directorate F (1)

Source: DG SANCO, 2009a; RAND Europe; informant interviews with DG SANCO staff

Table F.3: Data from other external sources

Data source	Description	Used by
Member States	Various	Directorate B (2) Directorate C (1) Directorate D (1) Directorate E (1) Directorate F (2)
Third countries		Directorate E (1)

National statistics organisations in MS	Consumer market statistics	Directorate B (1)
Consumer organisations in MS	Consumer market statistics	Directorate B (2)
Consumer policy network	Civil court data	Directorate B (1)
External contractors	Ad-hoc primary data collection	Directorate B (5) Directorate C (3) Directorate D (1) Directorate E (1) Unit 01 (1)
Internet search	Google, PubMed, Medline, etc.	Directorate B (3) Directorate C (2) Directorate E (1) Directorate F (1)
Industry stakeholders		Directorate B (3) Directorate C (1) Directorate E (3)
Scientific committees	Give opinions based on peer-reviewed literature	Directorate C (2)

Source: DG SANCO, 2009a; RAND Europe; informant interviews with DG SANCO staff

Appendix G: Document analysis