Deliverable Factsheet

Date: 29 April 2013

Deliverable No.	D3.1
Work Package	WP3
Partner responsible (Lead beneficiary)	P4 (UKS)
Other partners participating	P3, P2, P1
Nature	R
Dissemination level	PU
Delivery date according to DoW	Month 14
Actual delivery date	28 February 2013
Finalization date	29 April 2013
Relevant Task(s)	Task 3.1: Classification of existing organic market data collection methods

Brief description of the deliverable

In this report, a number of evaluation and quality criteria for data collection and compilation methods were defined. The results of an online survey on all existing organic market data collection methods in Europe were compiled and assessed. Subsequently the quality of existing data collection and processing approaches was evaluated using the following data quality dimensions: relevance, accuracy, comparability, coherence, accessibility and clarity, and timeliness and punctuality. The quality assessment was carried out exemplary to determine some good examples of data collection and processing. These cases were chosen because they delivered a very holistic and comprehensive presentation of their approaches regarding data collection methods, analyses, quality checks, and publication.

Target audience(s)

OrganicDataNetwork project partners, organic market data collectors as well as stakeholders

Publishable summary

The rapid growth of the organic sector during past years is the result of increasing demand for

organic products. Organic market actors need a solid basis for investment and purchase decisions. Up to now, data collection has been inconsistent, or data has not been comparable, because different methodologies have been used. Hence interpretations might lead to contradictory results. Moreover the organic market suffers from information-asymmetry and a lack of transparency. The potential for further market growth can best be realised by harmonising data collection and processing in addition to improving existing data sources.

This report deals with the definition of evaluation and quality criteria for data collection and compilation methods, the evaluation of existing data collection methods, and the assessment of data quality. The compatibility of methods and the general quality of data collection is analysed to establish a common basis for the collation of comprehensive European statistics. So far, data on retail sales volumes and values or import and export volumes and values of organic products is missing in the majority of European countries. More coherent data collection and thorough data analyses are needed to overcome current dispersion and fragmentation of data sources.

The online survey among organic market data collectors, carried out in the framework of this project, generated data on organic market data collection, processing, and dissemination in Europe. A number of questions on relevant issues were included in this survey to enable use of the results for the elaboration of the objective underlying this report. Thus the survey results are the basis for further data analysis and categorisation according to the data quality dimensions. An additional telephone survey was conducted to complement the responses of the online survey. The answers of the most relevant data providers were reviewed and completed. After categorising and allocating, the data sets were analysed using basic statistics. Thereby the differences in the use and processing of organic data among market actors were revealed.

After gaining an overview on all organic data collection across Europe, the most established organic data collectors were identified and described in more detail. All relevant survey questions were then allocated to the data quality dimensions. The dimension ‘relevance’ is determined by the questions on the main focus of the organisation, data sources, data uses, type of analysis and details of analysis, sample size, and start of data collection. ‘Accuracy’ includes questions on data sources, methods of data collection, details of analysis, as well as quality checks and details of quality checks. The dimension ‘comparability’ is made up by questions on methods of data collection, disaggregation of data, and sample size, while ‘coherence’ is only determined by the question on methods of data collection. The dimension ‘Accessibility/Clarity’ comprises questions on the obligation of data provision, data publication, data availability, and the format of publication. ‘Timeliness/punctuality’ is determined by questions on the frequency of data collection and publication. The allocation of survey questions to the data quality dimensions was applied to identify some good examples of data collection and processing from the underlying dataset.

Using this approach the application of the data quality dimensions for quality assessment was tested. Thereby the most consistent and elaborate data collection approaches were identified and contributed as a reference to a harmonised pan-European data collection system.

Potential Stakeholder impact(s)

Overview on data collection methods in Europe. Assessment of data quality. Application of data quality dimensions.

Interactions with other WPs deliverables / joint outputs			
WP no.	*Relevant tasks*	*Partner(s) involved*	*Context of interaction*
WP2	survey	P3 (ORC)	Survey data is collected and descriptive statistics are carried out.
WP4	survey	P2 (FiBL)	Evaluation of quality will be part of the discussion at the first workshop.
WP5	analysis	P4 (UKS)	Information on data collection methods compiled in WP3 will be used

Project no. 289376

Project acronym: **OrganicDataNetwork**

Project title:
Data network for better European organic market information

Collaborative Project
Collaborative Project targeted to a special group (such as SMEs)

SEVENTH FRAMEWORK PROGRAMME
FP7-KBBE.2011.1.4-05
Data network for better European organic market information

Title of Deliverable:

D3.1 Report on collection methods: Classification of data collection methods

Corinna Feldmann and Ulrich Hamm

Due date of delivery: Month 14
Actual submission date**: April 29th, 2013**

Start date of project: 1st January 2012 Duration: 36 months

Work package: 3

Work package Leader: UKS (P4)

Version: Final

Dissemination level: PU=Public

Contents

List of tables

List of figures

List of abbreviations

AMI	Agrarmarkt Informations-Gesellschaft mbH – Agrarian Market Information Company
CoP	Code of Practice
DG	Directorate General
EISfOM	European Information System for Organic Markets
ESS	European Statistical System
EU	European Union
EU-CEE-OFP	European Union - Central Eastern Europe - Organic Farming Policy
FiBL	Forschungsinstitut für biologischen Landbau – Research Institute of Organic Agriculture
GfK	Gesellschaft für Konsumforschung
GNP	Gross National Product
NACE	Nomenclature Générale des Activités Économiques - Statistical Classification of Economic Activities
OFCAP	Organic Farming and the Common Agricultural Policy
OMI	Organic Marketing Initiatives
OMIaRD	Organic Marketing Initiatives and Rural Development
UKS	Universität Kassel - University of Kassel
WP	Work package

1 Introduction

This report deals with the definition of a number of quality criteria for data collection and compilation methods and the evaluation of these methods. The survey results on all existing organic market data collection methods in Europe, which were collected and compiled in an online survey among European organic market data collectors, is classified according to predefined criteria. This is the prerequisite for the overall evaluation of existing data collection methods, including consistency and comprehensiveness, as well as the assessment of data quality. These results were intensively discussed and complemented by interested national data collectors in the first project workshop.

Furthermore, the compatibility of the existing data collection methods in Europe is analysed to answer the question of whether and how the different national data can be merged into European statistics. Eurostat currently compiles statistics on organic data, such as area data, operator data, livestock numbers, primary crop production, and livestock products volumes. Although Eurostat has the most elaborate system of transnational data collection related to the organic market, their current data collection does not take data on retail volumes and values, or import and export volumes and values for important organic agricultural products into account. This was one of the suggestions made by the former EU project consortium of the concerted action 'European Information System for Organic Markets' (EISfOM) (Rippin et al., 2006). In addition to the EISfOM project, several other EU projects have dealt with related issues: OFCAP (Häring and Dabbert, 2000), OMIaRD (Hamm and Gronefeld, 2004), and EU-CEEOFP (Stolze and Lampkin, 2005). Together they have published a number of reports on the development of the EU organic sector and thus have helped to develop a framework for reporting valid and reliable data. The EISfOM project suggested the introduction of legal requirements, committing member states to provide data. Since the implementation of the revised regulation on organic farming, more data has become available and can be accessed more easily (Eurostat, 2010). Nevertheless, more coherent data collection and thorough data analysis are needed to overcome current dispersion and fragmentation of data sources. So far only few countries publish consistent official statistics. Hence there is no common and holistic approach making sound decision-making in the European organic sector possible (Rippin et al., 2006). Building on this stated problem, the current EU project (OrganicDataNetwork) identifies the gaps in data bases and tries to bridge them.

The current EU project is carried out against the background of the rapid growth of the organic sector during past years resulting from an increasing demand for organic products. Today, there are many businesses in the organic sector. Besides numerous policy programmes and initiatives supporting organic farming exist throughout Europe. Due to the rapid expansion of the organic sector, a solid basis for strategic decision-making is needed. However, relevant market data is only available in a few countries and official statistics of the European market for organic farming do not exist across all countries (European Commission, DG for Agriculture and Rural Development, 2010). Until now available country data has been inconsistent or incomparable, because different methodologies and interpretations can lead to contradictory results. Furthermore the organic market suffers from information-asymmetry and a lack of transparency. The majority of potential end-users have limited access to reliable market-related information. In some cases, this can lead to incorrect entrepreneurial decisions, which in turn might result in market disturbances and the reconversion of organic farms to conventional agriculture (Sahm et al., 2012). The potential for

further market growth can best be realized by harmonising data collection and processing and by improving existing data sources (Hamm and Zanoli, 2006).

The subsequent chapters of this report combine the knowledge resulting from previous EU projects on organic market data collection and data quality assessment and explain recent developments based on the survey results. Preliminary conclusions from these findings can add to the development of a common Code of Practice for organic market data collection in Europe and hence serve as a guideline for market actors dealing with data collection, analysis, and publication.

2 Background on data quality assessment

The OrganicDataNetwork project builds on the results of the EISfOM (European Information System for Organic Markets) project which was a concerted action funded by the European Commission as part of the 5th Framework Research and Technological Development Programme. One major outcome identified within the project was the need for a harmonised nomenclature and classification system as a basis for a consistent organic data network that meets national and international demands. The EISfOM project referred to the concept of Current Best Practice (Linden and Papageorgiou, 2004), which describes the quality of statistical data as an ongoing improvement of the data production process. Quality dimensions were defined by Eurostat to establish a framework for the analysis and evaluation of the quality of statistical data and its sources: relevance, accuracy, timeliness and punctuality, accessibility and clarity, comparability, and coherence (Eurostat, 2009).

These quality dimensions are explained in more detail in the European Statistics Code of Practice (CoP), which presents the desired structure and content of a quality report to harmonise quality reporting across Member states and to facilitate comparisons (Eurostat, 2009). In addition the European Statistical System Committee prepared a Quality Assurance Framework to explain activities, methods and tools that help to implement the CoP (Eurostat, 2012). In the following paragraphs each data quality dimension is explained by a number of aspects and questions which should be taken into account when assessing data quality. The introduction of a quality report has to include a brief history of the statistical process and outputs in question, a broad statistical domain to which the outputs belong, and boundaries of the quality report with references to related reports (Hahn and Linden, 2007).

The first quality dimension is **relevance.** It is defined as the degree to which statistical outputs meet current and potential user needs. To further describe the relevance of the statistical output it is necessary to refer to its contents and to provide the key outputs/estimates desired by different users. Therefore it is inevitable in describing relevance to identify statistical target concepts and highlight important relations. The degree of completeness in terms of the ESS regulations also belongs to the quality dimension relevance and needs to be explicitly mentioned.

The quality dimension **accuracy** implies the closeness of estimates to the true values and covers sampling as well as non-sampling errors. The methodology that is used needs to be presented in order to understand and assess specific errors. Furthermore a section on the main sources of random and systematic errors needs to be provided, as well as an assessment of bias in quantitative and qualitative terms. In a summary assessment key, estimates are highlighted and the potential for

bias of each key indicator is reflected on. Depending on the type of study, particular errors have to be defined in more detail and have to be handled individually. In addition to sampling errors, one can find coverage, measurement, nonresponse, and processing errors. Further issues which need special consideration in the context of accuracy are the definition of model assumptions, seasonal adjustment, imputation, mistakes, and revisions. These aspects should be elaborated in a final summary on the dimension accuracy.

The third dimension consists of two quality indicators: namely **timeliness and punctuality**. Timeliness is defined as the length of time between the event or phenomenon the data describe and their official availability. Punctuality means the time lag between release date and target date of data. The reasons for non-punctual releases need to be explained.

Accessibility and clarity comprise simplicity and ease with which users can access statistics. The conditions of data access depend on the following factors: media, support, pricing policies, and possible restrictions. The understanding of statistical outputs can be enhanced by the description of accompanying information. The best way to evaluate this quality dimension is a reflection on the feedback of users, which is an essential part of the quality report.

The two remaining dimensions refer to **coherence and comparability** of statistical data. The quality of statistical outputs depends on the use of the same concepts and harmonized methods. In this context comparability is defined as a special case of coherence. A lack of coherence is explained by differences in concepts and methods. Hence one part of the quality report needs to deal with the assessment of possible effects of each reported difference on the output values. To further explain this quality dimension it can be related to a variety of attributes. First of all, comparability can be regarded over time and across regions; secondly coherence can be evaluated internally, but also in comparison with national accounts or with other statistics; and finally the quality can be checked with the help of mirror statistics, which usually tackle the same topic but use a different sample or a different method (Eurostat, 2009).

Summing up the discussion on the reporting of quality dimensions, it needs to be mentioned that output quality components are not mutually exclusive and hence each trade-off and its consequences have to be explained thoroughly. Furthermore the user needs and perceptions should be handled in the final section of the report; including a description and classification of users and the different ways of using the data. Some uses need to receive special attention as they are more complex or of higher importance. In addition, the satisfaction of users concerning the statistical output and its usefulness for particular tasks needs to be discussed (Hahn and Linden, 2007).

From the EISfOM analysis of European data collection, it is possible to identify the strengths and weaknesses of organic data collection and processing methods. According to the EISfOM results, the extent and structure of farm data is better developed than other areas because of its collection as part of the regular procedure of inspection: a non-statistical motivation (Rippin et al., 2006). Currently all EU Member States as well as EFTA countries and the governments of (potential) candidate countries provide data on organic land area, land use, and conversion status of the organic land area, as well as livestock numbers. In addition some countries have information on production volumes. In most cases this data is made publicly available after having been checked by Eurostat according to the CoP. Hence primary production related data is available on a harmonised and consistent basis in almost all European countries due to the implementation of European

Council Regulation 834/2007. For market data, including data on trade, collection and processing is carried out in many different ways throughout Europe. There are different types of organisations (e.g. government bodies, private market institutes, control bodies, etc.) responsible for organic market data collection in each country; because of the diversity of data collectors there can be different data collection approaches in the same country. Data collection methods differ between countries, but also within one country and over time. In addition, organisations use different nomenclatures and product definitions, so that the collected data is not comparable. Hence, organic market data in Europe is still very limited and to some extent non-existent; it lacks consistency, reliability, and comparability. For domestic market data, international trade data, and price data, recommendations from the EISfOM project have not been implemented. The weaknesses in the European organic market database seem to be in discordance with relevance of the organic sector, which also contributes to the socio-economic welfare in terms of public goods that are not incorporated in the GNP (Rippin et al., 2006).

Based on these findings, the EISfOM group developed a suggestion for a long-term strategy to realise the tasks that were considered as primarily important in this context. According to EISfOM partners, countries which have a system that works quite well should be used as a reference system to support the development in other countries. Moreover, the nomenclature and the classification systems for production, retail sales, international trade and price data need to be adjusted and harmonised. To meet these needs, some countries have started to elaborate solutions against the national and international background. The existing attempts need to be deepened and further developed to improve coordination and availability of high quality data. Therefore an institutional framework is aimed for, which as a result of the EISfOM considerations needs to include the following aspects:

1. Consolidation of the network relationships at both national and international level
2. Increase in quantity and quality of statistical data and improvement of coverage
3. Inclusion of member states' officials responsible for agricultural statistics and food safety
4. Integration of a wide range of stakeholders, who either own, collect, or process organic data

As part of the EU research project OMIaRD (Organic Marketing Initiatives and Rural Development) a survey of data on the EU markets for organic products was carried out. The data collection was on the fundamental statement that it is 'virtually impossible to obtain accurate data about the organic market from official statistics'. The OMIaRD survey first took place in 2000; in 2002 a second survey was carried out resulting in the revised and updated study that was published in 2004. The main focus of the survey was the collection of detailed data about the organic market at the European level and on a regular basis after identifying the reasons for success or failure of Organic Marketing Initiatives (OMI). According to OMIaRD, the failure to develop an organic market data information system was costly in terms of misinvestments and lost opportunities, besides it continuously reduced the competition in this sector. Moreover the failure to collect reliable market data in the past made it difficult to predict the development of consumer demand for organic products (Hamm and Gronefeld, 2004).

The OMIaRD project involved a thorough data analysis. Consistency and plausibility of the data was checked by comparing the organic data either with conventional data or with organic data of neighboring countries. Furthermore the following measures: 'organic as a share of total production',

'organic as a share of total consumption', and the 'market share of organic products' were calculated to identify inconsistencies (Hamm and Gronefeld, 2004). The analysis of the data was subdivided into different areas of interest:

1. Organic production
2. Organic food consumption
3. Foreign trade in organic products
4. Relationships between organic supply and demand
5. Prices and price premiums for organic food

A new and very important measure in this context is 'sales of organic products as organic', because it is a very critical parameter in the creation of supply balances.

The countries investigated in OMIaRD were classified into groups depending on the number of common parameters. The results of these analyses revealed the necessity for national governments to assure an equal growth of demand and supply in the organic market by choosing an approach that combines production, demand, research, and information activities into a broad and comprehensive organic action plan. This approach appears to be especially valuable due to its horizontal and vertical integration of stakeholders and activities.

3 Methodology

The methodological approach to tackle the above-mentioned objective is a survey among stakeholders in all countries of the EU 27, candidate countries, potential candidate countries and EFTA countries, which are involved in organic market data collection, processing, or dissemination. The survey was developed and carried out in the framework of the Organic Data Network project. To reach as many respondents as possible and to facilitate the compilation of data, the survey was set up in an electronic format. With the help of all project partners, a list of potential contact persons was generated to establish a broad basis for this survey (see Gerrard et al., 2012 for further details). A number of questions on relevant issues needed to be included to enable use of the survey results for the achievement of the objective. The following categories show the content of the survey, as predefined in the project's description of work (DoW); they also serve as the basis for further data analysis and categorisation according to the data quality dimensions: relevance, accuracy, timeliness and punctuality, accessibility and clarity, comparability, coherence.

- Type of data collectors
- Type of data
- Geographical coverage
- Degree of detail of data
- Frequency of data collection
- Method of data collection (questionnaires, observations, tests)
- Sample size
- Type of quality checks
- Depth of data analysis
- Purpose for data collection
- collection of organic data only

- Obligatory or voluntary basis for data providers to deliver data
- Offer of payments or incentives to data providers
- Administrative details of organisation

Table 1: Survey topics allocated to six data quality dimensions

Relevance	Accuracy	Comparability	Coherence	Accessibility/ Clarity	Timeliness/ Punctuality
Main focus of organisation	Data sources	Methods of data collection	Methods of data collection	Voluntary or obligatory to provide data	Frequency of data collection
Data sources	Methods of data collection	Disaggregation of data		Publication of data	Frequency of publication
Data uses	Details of analysis	Sample size		Availability of data	
Type of analysis & details of analysis	Quality checks & details of quality checks			Format of publication	
Sample size					
Start of data collection					

Each question in the survey belongs to one of the content categories and was allocated to at least one of the quality dimensions as shown in Table 1. In Chapter 4.3 the procedure for quality assessment has been described and after that the quality dimensions were applied, exemplified for a number of organisations engaged in organic market data collection in Europe. After categorising and allocating the data sets, they were analysed by using basic statistics. Hence, the differences in collection and processing of organic market data between different organisations and market actors were revealed and necessary steps for data harmonisation within the European organic market were developed.

An inventory of data collectors including some basic statistics on the dataset derived from the online survey were carried out by project partners beforehand (Gerrard et al., 2012). For that purpose the dataset was divided into groups depending on their membership-status in the EU-27. Those frequency distributions served as a background for this report and gave a general idea on the numbers and figures underlying the following analysis. In contrast to the contribution by Gerrard et al. (2012), the following analysis includes additional data from a telephone survey to complement the results of the online survey and focusses on data quality issues. The telephone survey was carried out among organisations that did not fill in the online questionnaire completely or did not respond to the online survey at all, but were expected to have more information on their organisation's data collection. The telephone survey yielded 14 additional responses. The telephone survey was conducted by project partners in the respective countries, following a given interview guideline.

4 Results

4.1 Sample

The final dataset – combining the results of the online survey and the telephone interviews - consists of 126 total responses. Only the questions represented in Figures 10, and subchapters 7 and 10, which refer to the obligation of data providers to deliver data, the dependency of membership payments on the data provision, and the disaggregation of data collected by the organisation, were not covered by the additional telephone survey. Thus, the results for those questions are only based on the 112 responses to the online survey conducted in the first place. The dataset of both surveys was combined and analysed as a whole (i.e. all 126 responses) without splitting the sample into different groups of countries. As most questionnaires were not entirely completed, the number of usable responses varied among the different questions. When looking at the data output, and using results for further discussions, it should be kept in mind that multiple answers were possible for a number of questions. In addition, the overview on all participating market actors was presented in a report by Gerrard et al. (2012) and revealed the heterogeneity of data providers related to their main aim and focus of statistical work as well as to the type of data they deal with. Since a number of questions did not directly apply to each addressee, many survey questions were answered in a very heterogeneous way.

4.2 Characteristics of data collection

4.2.1 Main focus of organisation

The question on the main focus of the organisation was answered by 89 out of 126 organisations. The main focus of the work carried out by the organisations was nominated as data analysis by more than 60% of the respondents, and as collection and dissemination by about 55% of the respondents (Figure 1). Some respondents, however, reported that their organisation focused only on one of these steps, depending on the type of organisation they represented. Either checking/processing, collation, or archiving is the only main focus of less than 40% of the organisations. The steps checking/processing, analysis, dissemination, and archiving were mostly named in combination with data collection and/or collation and rarely by themselves.

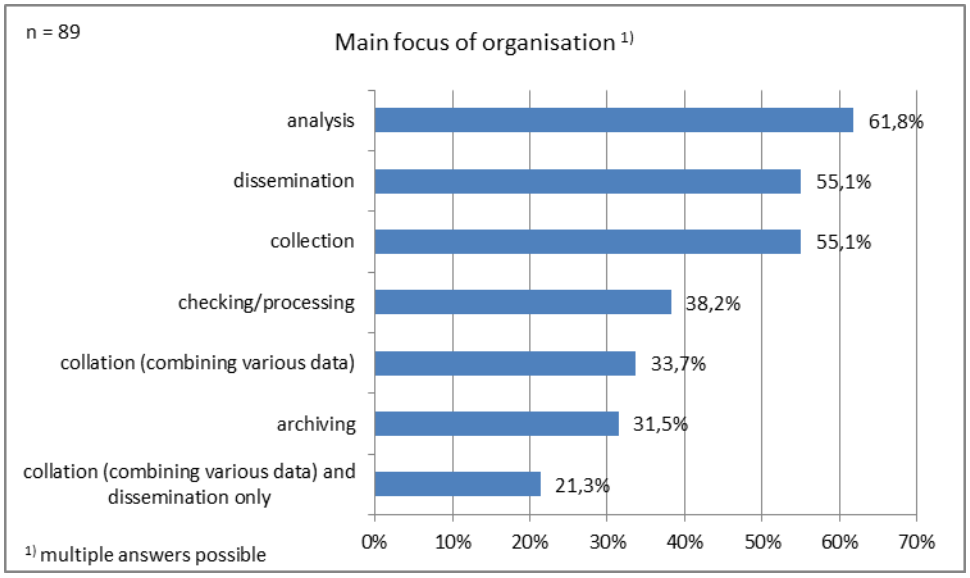

Figure 1: Main focus of organisation

4.2.2 Data sources

The question on data sources was answered by 70% of all respondents. As most of the organic market data collected by the respondents of this survey was production data, producers (59%) were the main data source for more than half of these respondents (Figure 2). This question was not answered by 30% of the organisations. The numbers on the types of data sources are reflected in the type of data recorded by the respondents and evaluated in the report on data collectors as part of this EU project (Gerrard et al., 2012).

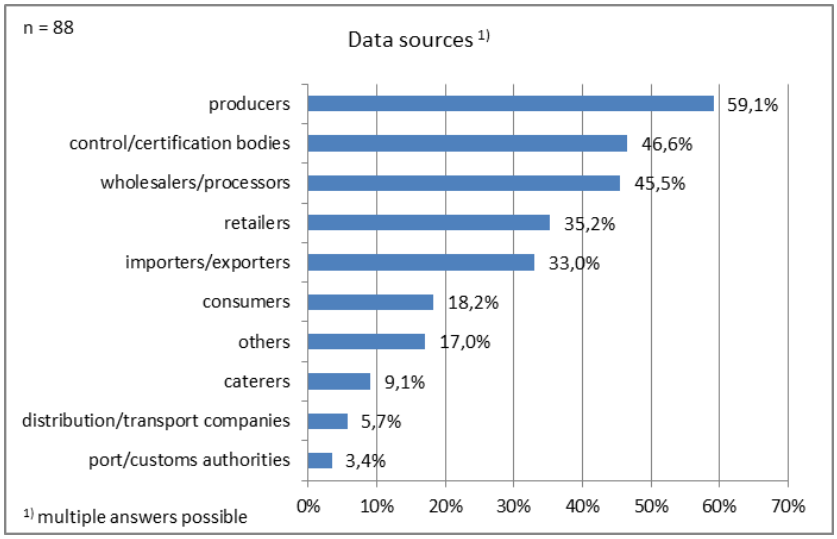

Figure 2: Sources used by data collectors

Other data sources which did not fit into any of the predefined categories mentioned by the respondents are the following:

- government bodies / national authorities
- competent authorities of autonomous regions / regional institutions
- market research institutes (e.g. Nielsen company)

- university studies
- farmer marketing cooperatives / farmer association / sector organisations

4.2.3 Data collection methods

The methods of data collection in the European organic market account for an important part of this work package. The collection methods directly influence the availability and the quality of organic market data and thus deserve special attention. Almost 60% of all organisations answered the questions on data collection methods for at least one data type. Figure 3 shows frequencies of the use of data collection methods across all data types. The most dominant methods were e-mail surveys. Panels which would allow analyses over time periods were conducted in only 20% of the organisations (Figure 3).

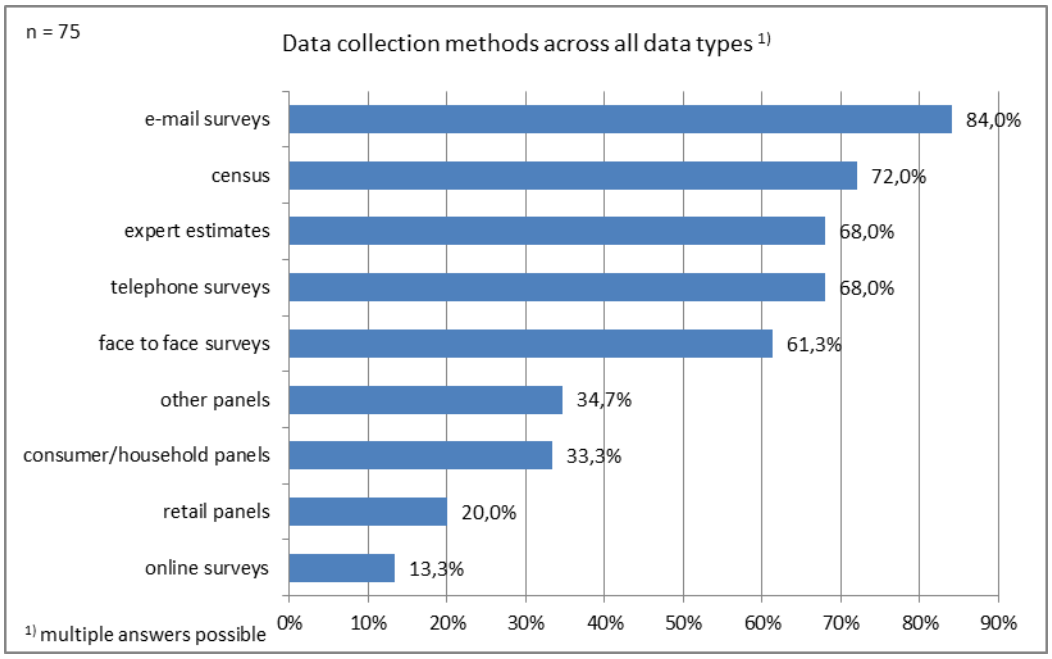

Figure 3: Methods used for data collection without reference to particular data type

The methods were analysed individually for each data type collected in the organic market sector to get a better insight into the applied data collection methods. The two subsequent diagrams (Figure 4) show the collection methods for production data: separated according to volume and value. The direct comparison reveals the different approaches depending on the type of data. With almost 50%, production volumes were predominantly assessed by censuses; expert estimates and various types of surveys (excluding online-surveys) represent the other common means of data collection. Production values, however, were most commonly determined by expert estimates (47%) as well as telephone and e-mail surveys, while censuses play a minor role.

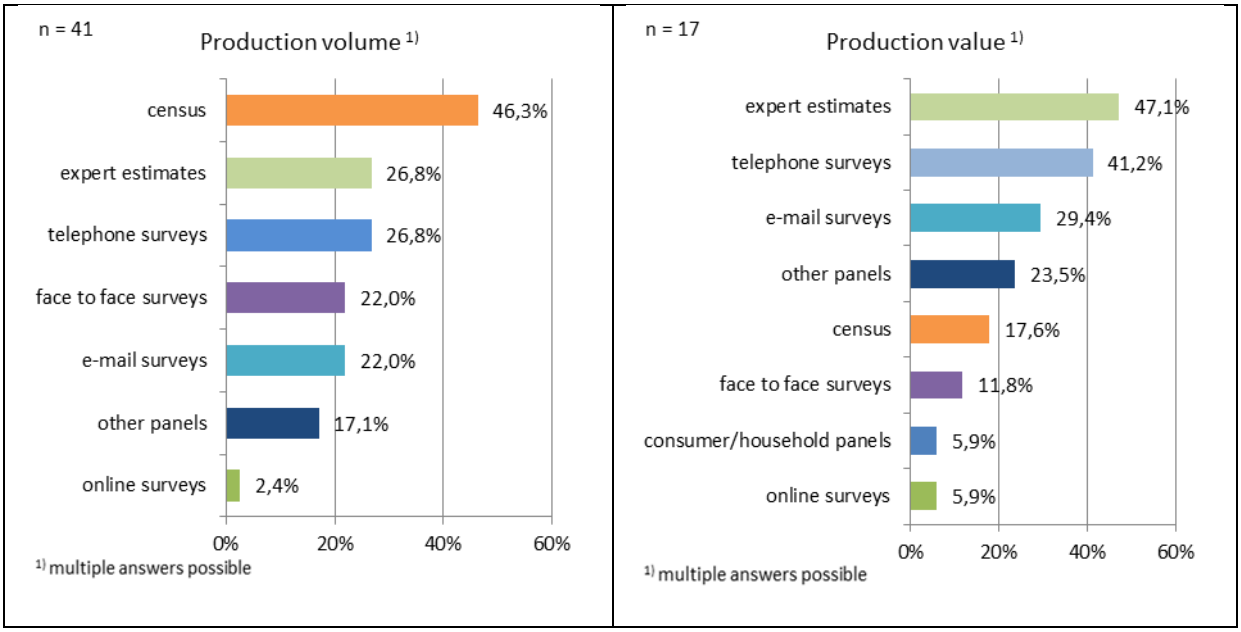

Figure 4: Methods used for collection of production data (volume and value)

Similarly, the two diagrams on the collection of retails sales volume and value data (Figure 5) reveal different distributions of the collection methods. For retail sales volumes, consumer/household panels were carried out by almost 40% of all respondents, followed by censuses with 22%. With regard to retail sales values, e-mail surveys and consumer/household panels were each used by one third of all respondents.

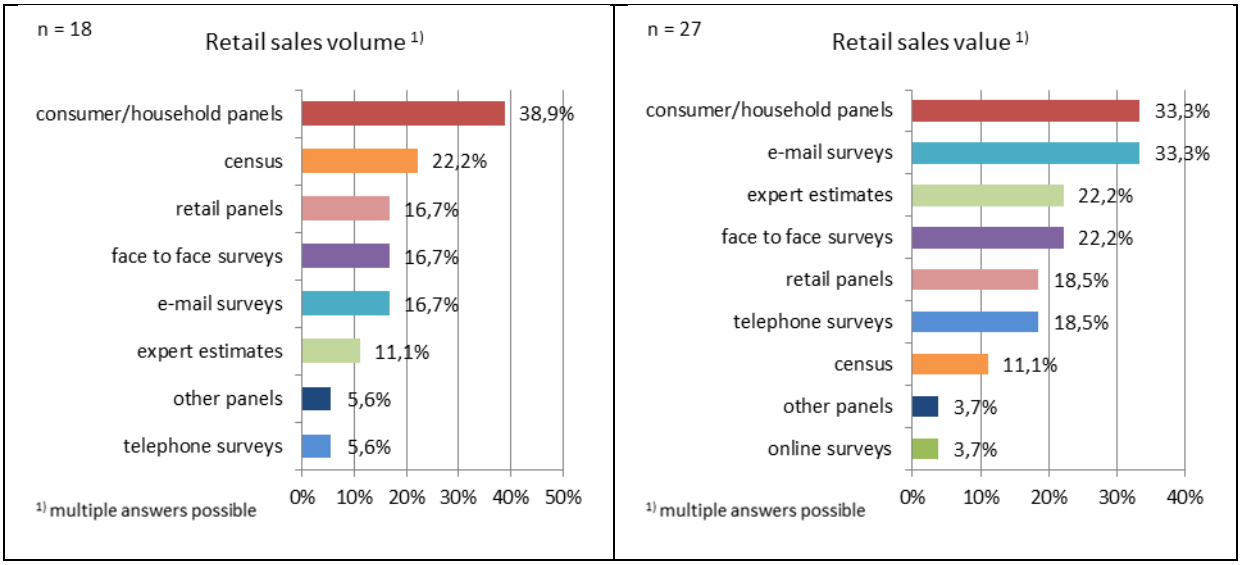

Figure 5: Methods for collection of retail sales data (volume and value)

Telephone surveys and expert estimates were most frequently used for the collection of farm level price data (Figure 6). For the collection of consumer price data, however, consumer/household panels are most important (44%). These obvious differences are attributed to the different characteristics of the data sources: farmers and consumers.

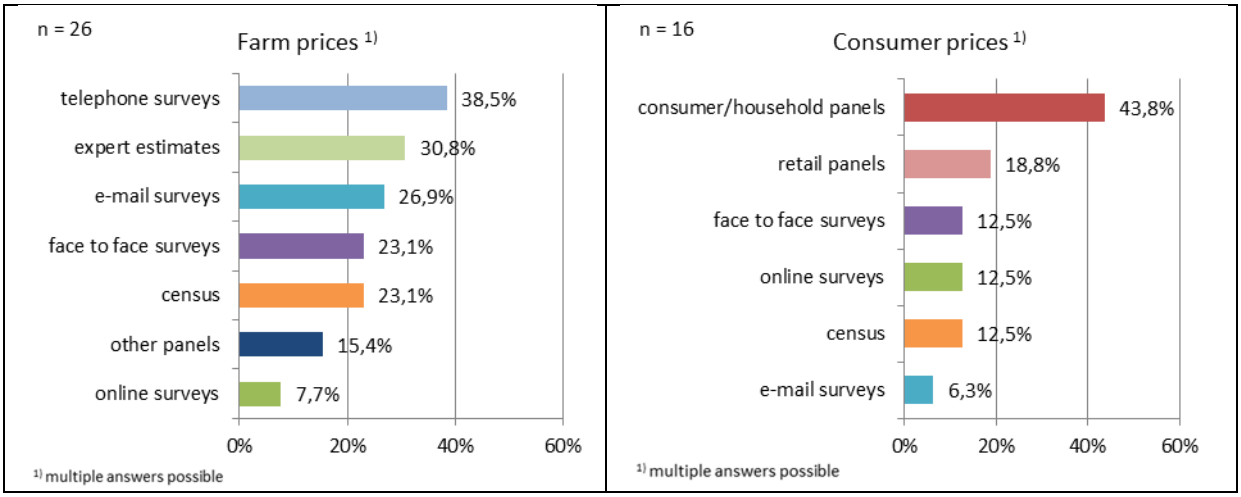

Figure 6: Methods for collection of farm level and consumer price data

The following diagrams (Figure 7) show that, for the collection of import volumes, almost 50% of the respondents used censuses, while for import value data e-mail surveys were used most often.

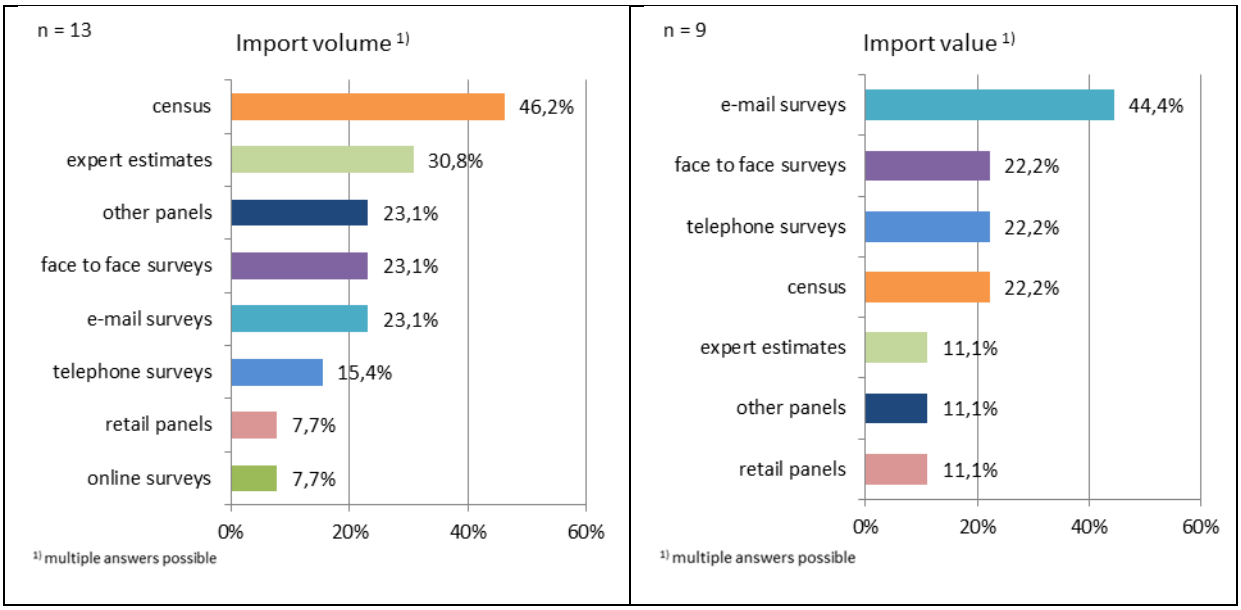

Figure 7: Methods for collection of import data (volume and value)

The frequency distributions for the collection of export data look quite similar (Figure 8). Both export volumes and values were collected most frequently by using e-mail surveys and censuses. For export value data, expert estimates, face to face surveys, and telephone surveys are used with the same frequency as censuses. It appears that consumer/household panels and online surveys are of minimal importance in export and import data collection.

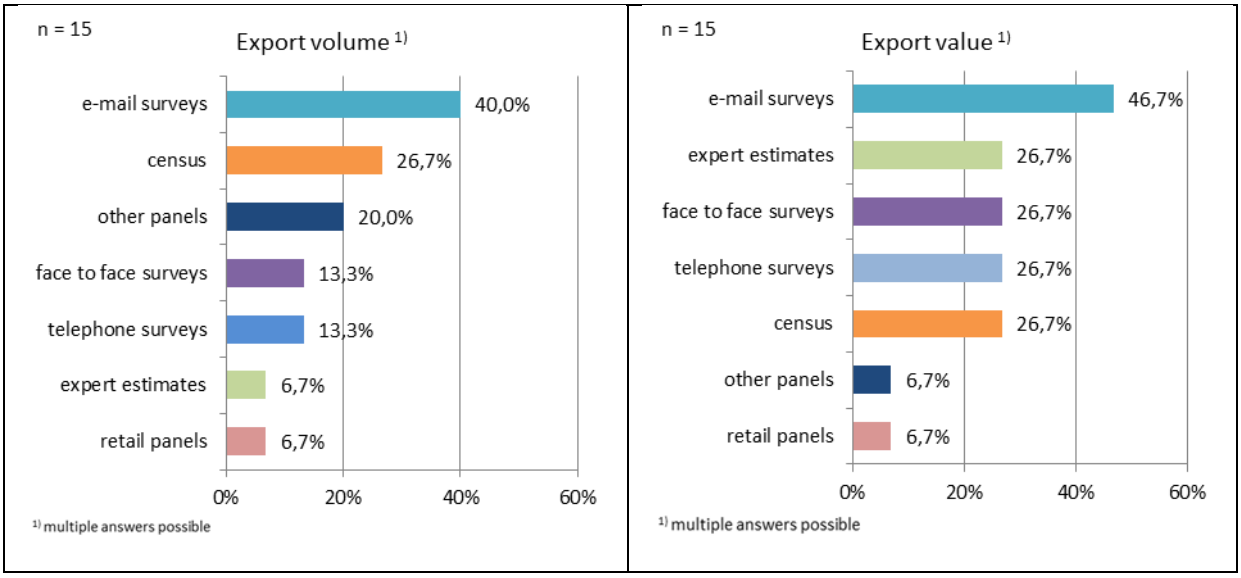

Figure 8: Methods for collection of export data (volume and value)

4.2.4 Data usage

Almost three quarters of respondents marked statistics as the main aim of their statistical work, followed by market information which was named by half of the respondents who answered the question on data usage (Figure 9). One third of all organisations did not answer this question at all, while another third of the respondents mentioned more than one data use.

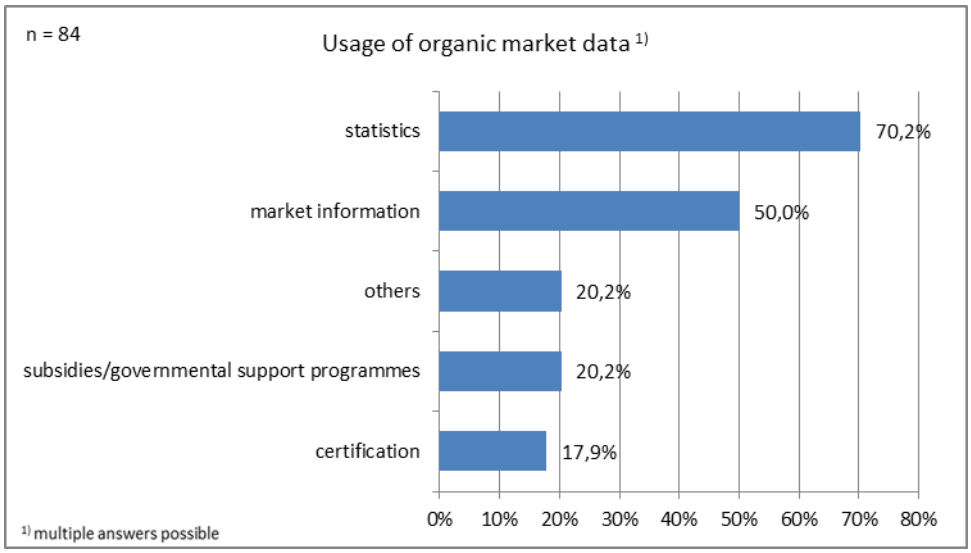

Figure 9: Data usage

In the open category for other data uses the respondents mentioned the following items:

- membership payments
- reports for Ministry of Agriculture / Eurostat
- marketing management (commercial marketing, sales improvement, fact-based consultancy)
- consumer behaviour research and specific sector analysis
- research and analysis

- governmental surveys
- project development
- lobbying

4.2.5 Data provision

Data provision was reported to be obligatory for 49% of the respondents and voluntary for 51%. However, 34% of the organisations did not answer this question (Figure 10). The following details were given on the extent of obligation:

- legal requirement / national government regulations
- Commission (Council) Regulation, obligatory according to Eurostat (834/2007 and 889/2008)
- requirement of membership / certification
- contract
- free access to database for posting prices

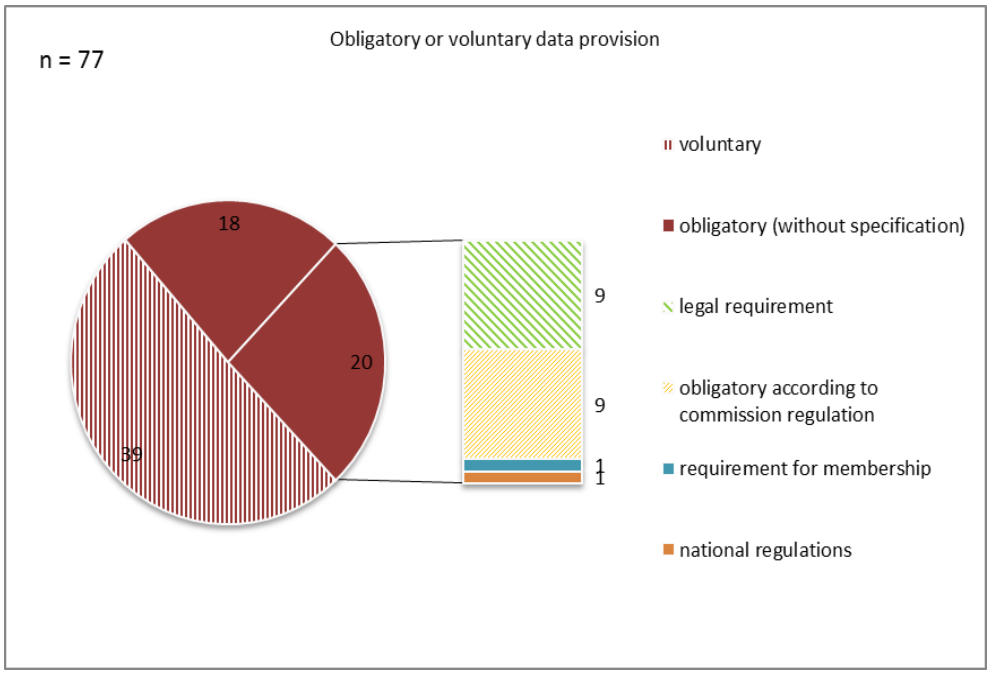

Figure 10: Obligatory or voluntary data provision

4.2.6 Payments or incentives

Less than half of the organisations answered the question on whether their organisation offers payments or incentives to data providers; with 72% of the respondents reporting that their organisation does not offer payments or incentives (**Errore. L'origine riferimento non è stata trovata.**). From the organisations that do provide incentives and payments, only ten respondents mentioned details on the type of incentive. The details are given below:

- payment for control bodies
- free analysis
- free books
- price reductions for reports (e.g. reward points)
- payments for data from certifying bodies and consultancy
- points as rewards → exchanged for vouchers

4.2.7 Membership payment

Dependency of membership payments on the data that organisations ask for can lead to underestimations by data providers to ensure lower membership fees. The question on whether the membership payment is fully or partly dependent on any of the data that the organisations ask for was answered by about half of the respondents; 62% of the respondents answered that this interdependence is not applicable and 28% responded negatively to the question. Only 10% of the respondents confirmed a link between membership payments and data request.

4.2.8 Type of analysis

More than 40% of the organisations did not answer the question on data analysis. Almost 90% of the respondents reported that their organisation compiles data. Most of the respondents to this question reported that their organisation does not apply more advanced statistics on the data collected, but about half of the respondents stated that they apply basic statistics. Others additionally mentioned that the kind of analysis depends on the type of data they collect (Figure 11).

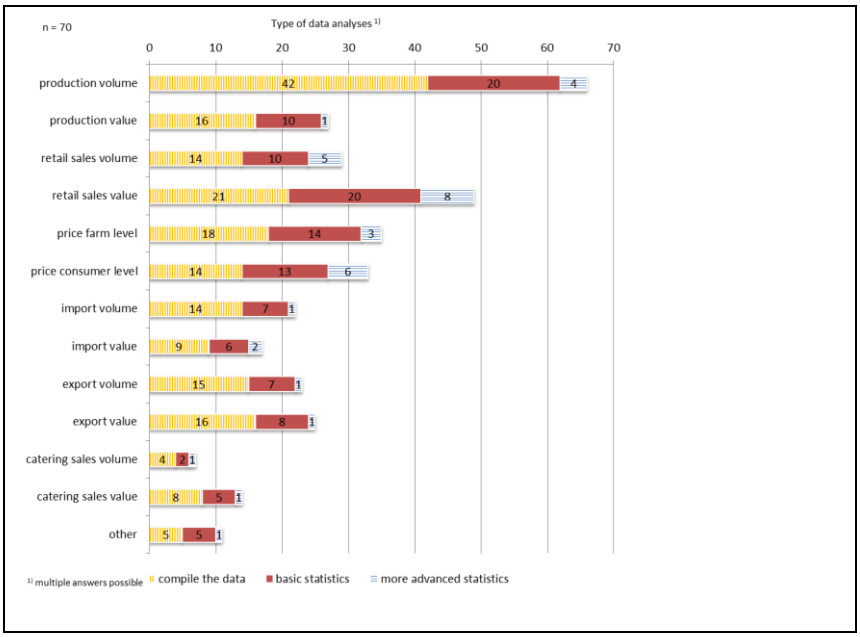

Figure 11: Type of analysis carried out by organisations

4.2.9 Data quality checks

About 60% of the organisations answered the question on data quality checks, with 69% of the 77 respondents indicating that they carry out quality checks, while 31% of the respondents answered negatively to that question (Figure 12). Of those who carry out quality checks, only 26% gave further details with most of these respondents working for government bodies. The quality checks that were mentioned most often are comparisons of organic data with conventional data or with data from previous years.

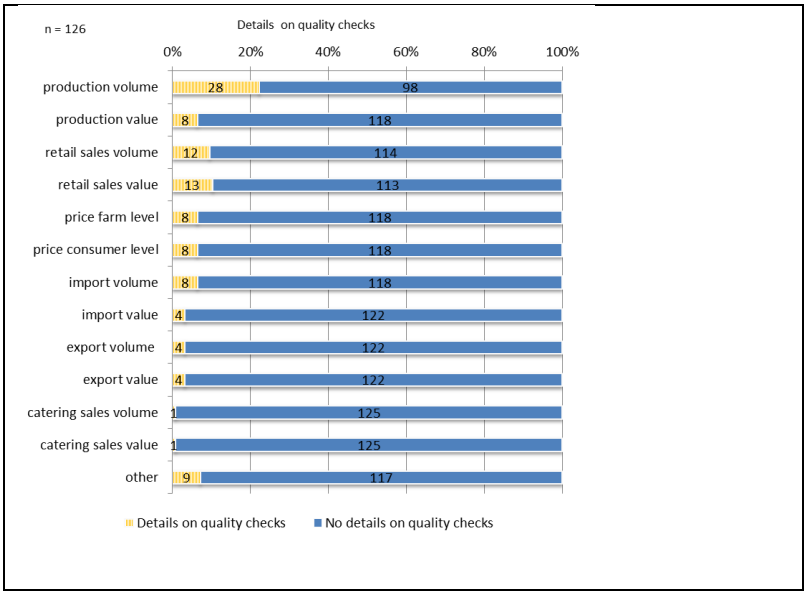

Figure 122: Implementation of data quality checks

4.2.10 Disaggregation of data

Organisations were able to give multiple responses to the question on the disaggregation of data. Out of all organisations, about one third did not answer this question. About 50% of the respondents stated that their data is collected for the whole country and 46% stated that their data is disaggregated by administrative region. The remaining 39% referred to other responses which include the following:

- type of enterprise
- code of operator
- type of registration (e.g. processors and distributors, farm processors)
- type of retailer, retail space, revenue classes
- consumer characteristics
- product categories
- production areas with different climate conditions

4.2.11 Sample sizes

The sample sizes varied greatly between different types of data, different organisations, and different countries. Due to the small number of responses to the question on sample size and the few occurrences of responses to both sample size and the country of origin, sound conclusions on the relation between country and sample size cannot be drawn. Hence, the responses for each data type need to be considered individually. The standard deviation reveals the high variation and the low informative value of a general statement on sample sizes across all respondents (Table 2).

Table 2: Overview on sample sizes with reference to the particular data type

	Number of responses	Minimum sample size	Maximum sample size	Mean value	Standard deviation
Production volume	23	2	2,900,000	148,085.20	647,770.10
Production value	6	34	1,500	710.80	678.67
Retail sales volume	12	10	30,000	11001.11	12,434.74
Retail sales value	17	2	30,000	6,305.08	11,120.56
Price farm level	10	20	3,680	705.50	1,276.05
Price consumer level	9	6	30,000	11,555.14	13,327.57
Import volume	5	2	350	101.75	165.96
Import value	3	350	540	445.00	134.35
Export volume	8	17	906	263.83	328.11
Export value	8	17	906	340.50	330.30
Catering sales volume	3	13	2,277	1,145.00	1,600.89
Catering sales value	5	5	540	317.00	278.38
Other	3	20	995	507.50	689.43

4.2.12 Choice of respondents

The main criteria mentioned for the choice of respondents are listed below. The wording of the open responses on this question indicates the attempt of the data collectors to achieve a comprehensive sample, which covers as many respondents as possible. This question was asked for each data type individually; on average (across all data types) 21 respondents answered this question.

- all
- as many as possible
- representative households
- geographical representativeness
- existing supermarkets trading organic food / specialist shops
- importance
- Farm Accountancy Data Network (FADN) system

4.2.13 Frequency of data collection

Altogether 235 responses were given on the frequency of data collection for all data types. Data is collected annually in 70% of these cases. Weekly data collection is very uncommon and was only mentioned in relation to retail sales data (volume and value), price data on producer and consumer level, and for catering (volume and value). In 21% of the cases, data is collected monthly; monthly collection occurs for all types of data, but altogether in far fewer cases than does annual collection (Figure 13).

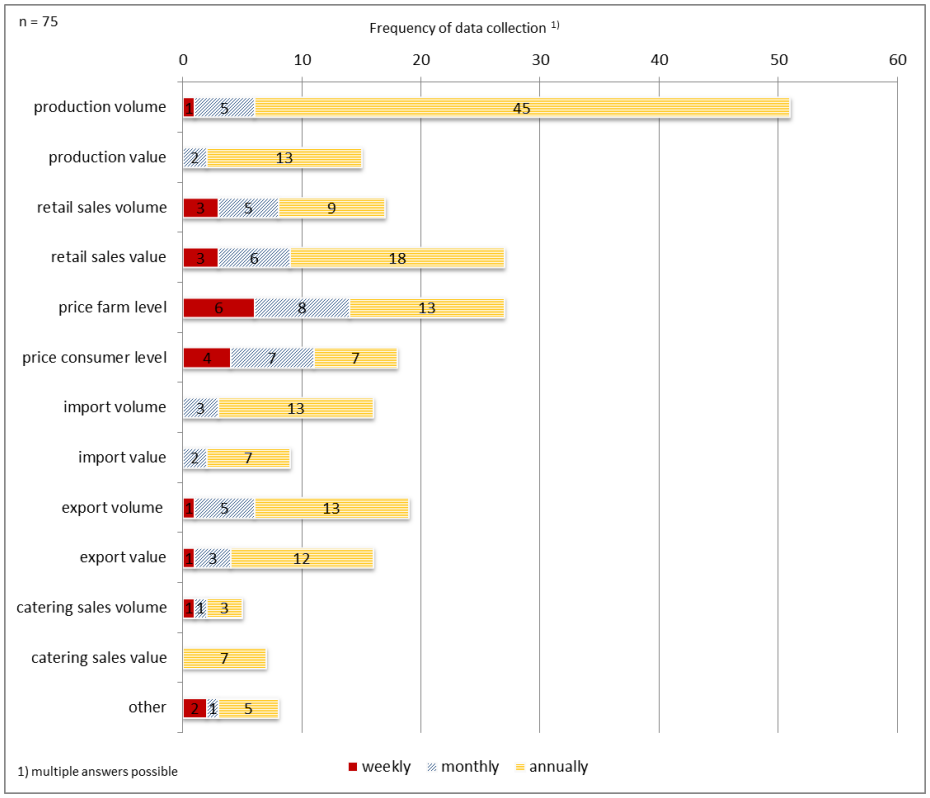

Figure 13: Frequency of data collection depending on different data types

4.2.14 Start of data collection

Collection of organic market data started after 1990 for all organisations. Germany and the UK have the longest history of data collection. Some organisations from Finland, Turkey and Lithuania also referred to some data collection which started before 2000. Catering sales data has not yet been collected for a very long time, because organic products in catering companies are still not widely-used.

4.2.15 Publication

About 60% of all organisations answered the question on whether the data is published or not, with 71% indicating that they publish the data, and 29% indicating that their organisation does not. The respondents who stated that their organisation publishes the data were most often government bodies.

4.2.16 Availability of publications

Due to the option of giving multiple responses to the question on the availability of publications, the number of responses is larger than the number of respondents. Most data is publicly available with 87% of respondents answering that their organisation offers at least one data type that is publicly available. It is less common to offer data at a cost or exclusively to the members of the organisation. An agreement implying the availability of data at a lower cost is made by less than 10% of the organisations; such an agreement is only made for retail sales, price, and import data and only by three different organisations responding to the survey (Figure 14).

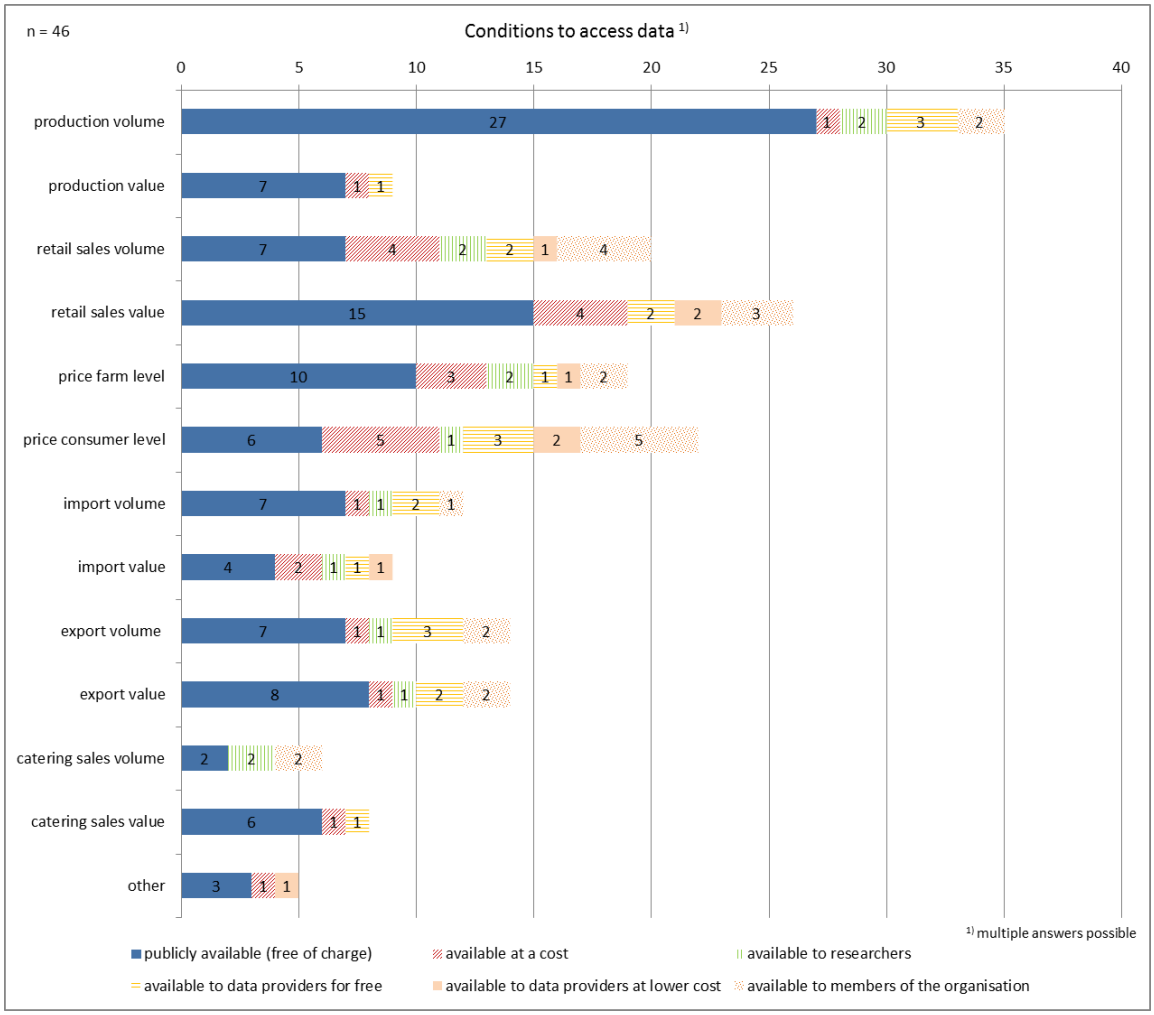

Figure 14: Conditions to access data depending on different data types

4.2.17 Frequency of publication

Similarly to the frequency of data collection, the publication mostly takes place annually as well; in 128 of 159 cases data is published annually. Weekly publications most frequently occur for price data, while publications in a monthly interval rather occur for retail sales and price data (Figure 15).

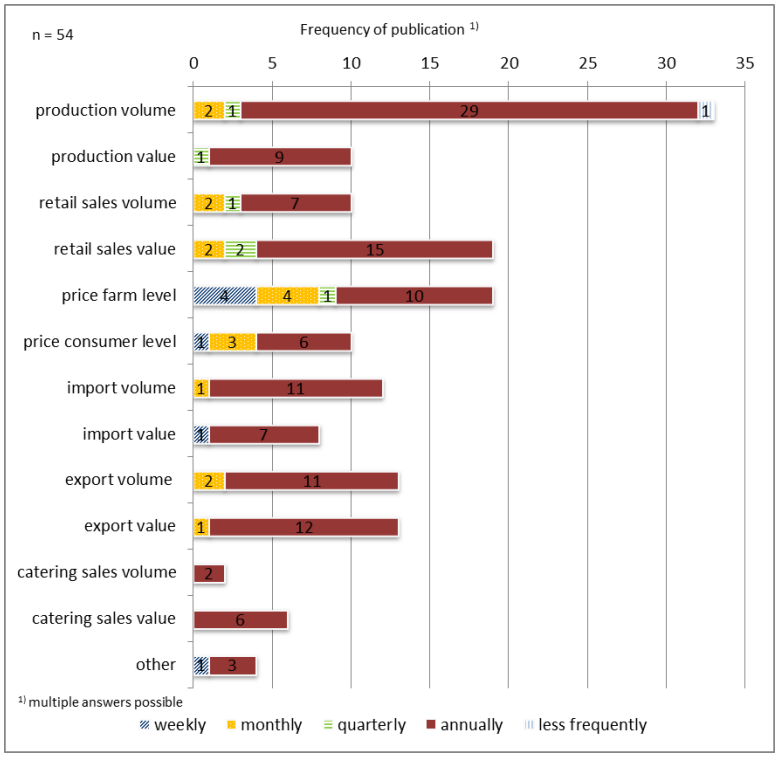

Figure 15: Frequency of publication depending on different data types

The following responses were given in addition to the predefined categories:

- real time publication
- trimester

4.2.18 Format of publication

Less than half of the respondents answered the question on the format in which data is published. The web page is the most common format, as it is used by the majority of respondents. More than half of the respondents publish their data in online and paper reports as well as statistical tables. Less than 20% of the respondents use other formats, including scientific journals, press conferences / press releases, public conferences, and (tele-)fax (Figure).

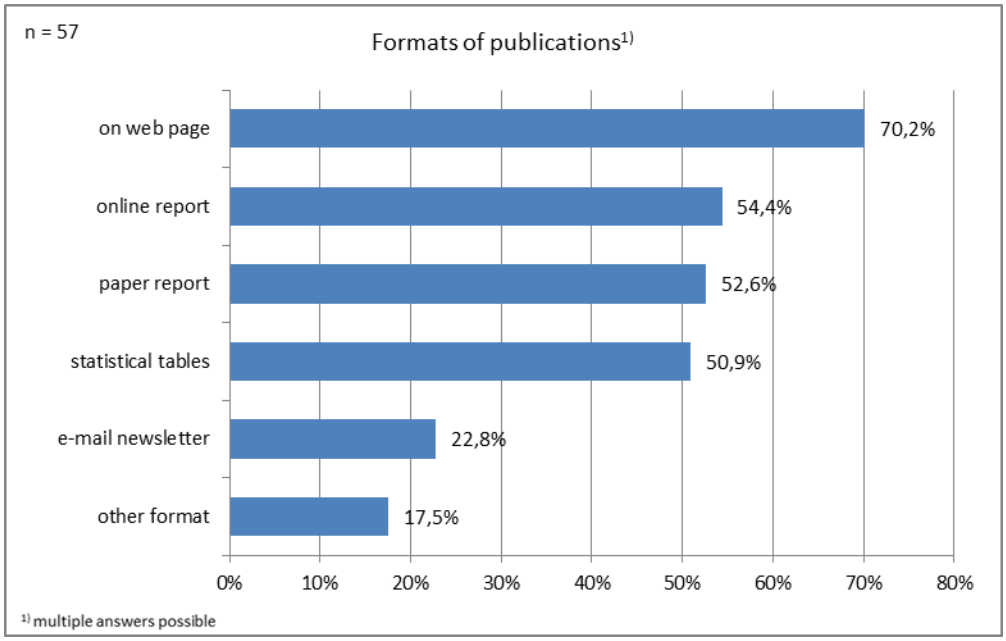

Figure 16: Formats of publications on data

4.2.19 Countries covered by data collectors

The following table (Table 3) indicates how well one country is covered by organic market data collection. While, in many countries, only one organisation is responsible for organic market data collection, or only one organisation replied to the survey, in a few countries more organisations collect organic market data. These organisations often deal with varying types of data and collect data for different purposes and hence they might use different data sources and data collection methods. Nine organisations responsible for data collection from Italy and Germany answered the survey (Table 3).

Table 3: List of countries showing the number of times a country is covered by data collection

Number of times a country was mentioned	For which country/region is the data collected?
0	MT, YU
1	AL, AX, BA, BG, CZ, EG, FI, IS, LU, LT, MK, NL, PL, RO, RS, SF, SI, SK, SY, MOAN, West & North Africa
2	HR, HU, LB, LI, LV, MA, ME, NO, TN, England, worldwide
3	ES, SE, CH, EE, Wales, Scotland
4	AT, BE, IE, FR, PT, Europe
5	UK, TR
6	DK
9	DE IT
19	“BLANK”

4.3 Quality assessment

Classification of the methodological approaches applied by data collectors in Europe

In the context of organic market data the concept of data quality dimensions introduced through the ESS was used as the basis for the assessment of data quality in the European organic market. The data quality dimensions were adopted to the new context and applied for the subsequent steps: sample generation, data collection, data processing, data analysis, and publication. Interrelations between these steps need to be taken into account, as they might influence each other and consistency in the whole approach is very important, e.g. type of data influences the type of collection method and the samples size as well, moreover the size of the final dataset determines the type of analysis, etc.

The dimensions described in Chapter 2 are the underlying concept for assessing data quality. Each quality dimension consists of a few parameters, which each reveal an individual step in data processing and were inquired in the survey among organic market data collectors (Table 1: Survey topics allocated to six data quality dimensions). Through exploration and evaluation of these parameters the quality of the organisations' approaches to data collection can be assessed. Therefore, step by step, the data collection procedure of each organisation (based on their response to the survey) has been analysed to determine the quality and to identify obvious drawbacks and improvement possibilities in the current performance. A possible approach on how an organisation can rank its own performance on each individual parameter is presented in the following table (

Table 4).

Table 4: Parameters underlying data quality assessment

+: basic implementation (or non-existent) ++: good +++: excellent

1. Main focus of organisation is closely related with data uses:
   → sound combination of responses to both questions (2-3 data uses and 4-6 steps as main focus): +++
2. Data source is closely related to data type:
   → responses to both questions have to be consistent
   → more data sources are generally rated slightly better (higher reliability of data)
3. Sample size: large/representative sample (+++) and very small sample/no information (+)
4. Data collection method (also closely related to data type):
   → responses to both questions have to be consistent
   → more data collection methods are generally rated slightly better (higher reliability of data)
5. Frequency of collection and frequency of publication:
   → same interval (+++), different interval (++) and no publication/no information (+)
6. Start of data collection:
   → long experience (+++) and less experience (+)
7. Type of analysis:
   → more advanced/adapted to data type (+++), basic (++), and no analysis, respectively compilation (+)
8. Quality checks:
   → quality checks and details given (+++), simple quality checks (++), no checks (+)
9. Disaggregation of data:
   → disaggregation into comparable groups (+++), no disaggregation (++), meaningless disaggregation (+)
10. Publication:
    → publication (+++) and no publication (+)
11. Frequency of publication → frequency of collection
12. Format of publication: accessibility depends on data user
    → more formats (+++) and only one or no response (+)
13. Availability:
    → publicly available, free of charge (+++) and not available (+)

The data quality dimensions were applied on a number of datasets resulting from the online and telephone-survey. The performance was assessed for each parameter and ranked according to the schema outlined in the table above (Table 4). Thereby some examples for “best practice” approaches were identified for the performance in each data quality dimension. These examples can function as a reference system for those organisations that see the need to improve their data collection and processing approaches. To present an overview the “best practice” examples are summarised in the following table (Table 5).

Table 5: Examples for "best practice" identified from the survey among organic market data collectors

	Examples for "best practice"	
Relevance	Agence Bio and AMI	• Correspondence of the focus of statistical work and the purpose of data collection • Consistency of data sources and data type • Large ample sizes • Long-time experience • Advanced statistics
Accuracy	Agence Bio and AMI	• Consistency of data sources and data type • Compliance of data sources with data collection methods • Advanced analyses • Application of data quality checks
Comparability	Agence Bio and AMI	• Compliance of data collection methods with sample choice and sample size • Disaggregation of data by regions
Coherence	Soil Association, Agence Bio and AMI	• Choice of a data collection method which meets the requirements of the data sources, the sample, the data use, and the final analyses
Accessibility/Clarity	Eurostat, Statistics Denmark, Soil Association and Agence Bio	• Availability of data: easy to access and user-friendly • Different formats of data publication • Publicly available and mostly free of charge
Timeliness/Punctuality	AMI and Bio Suisse	• Timely and punctual publication • Same time interval for data collection and publication

The quality dimension **relevance** is made up of the parameters 'main focus of the organisation', 'data source', 'data uses', 'type of analysis', 'sample size', and 'start of data collection'. Both, the AMI and Agence Bio, follow an approach in which the focus of the organisation and its statistical work correspond to the purpose the data is collected for, respectively the data use. The sources from which they receive their data are consistent with regard to the type of data the organisations are interested in; besides, the data sources are appropriate for the data uses. The samples sizes are large enough to ensure a solid basis for the type of analysis applied by the organisations. Furthermore, the AMI has got a long-time experience in organic data collection and thus gathered more knowhow in terms of interaction among all the parameters relevant for consistent data collection and processing. Both organisations applied advanced statistics to achieve results that are useful for the end users of their data. Altogether their approaches rank highly on the parameters associated to relevance.

The quality dimension **accuracy** is made up by the parameters 'data source', 'data collection method', 'type of analysis', and 'quality checks'. As explained above the approaches followed by AMI and Agence Bio are both very consistent with regard to 'data source' and 'type of analysis'. The data collection methods used comply with the data sources used by AMI and Agence Bio and lay the foundation for accurate analyses. Due to the successful interplay of data sources, collection methods, and the types of data analysis the performance of both organisations is ranked high for the quality dimension accuracy. Furthermore both, AMI and Agence Bio, apply quality checks to search

for inconsistencies and errors in their datasets. Thereby they ensure accurate results for the end users of their statistical outputs.

Concerning the quality dimension **comparability,** which consists of the parameters 'data collection method', 'disaggregation of data', and 'sample size', again the performances of AMI and Agence Bio offer examples of best practice. While the data collection methods comply with the choice and size of the sample, the data of these two organisations are also disaggregated by certain regions within the country to allow for comparisons on a national level.

Soil association, Agence Bio, and AMI reveal a good performance in the dimension **coherence**, which is made up of the parameter data collection and which is very closely related to the dimension comparability. A coherent approach is characterised by the use of a data collection method which is wisely chosen to meet the requirements of the data sources, the sample, the data use, and the final analyses. Thereby the organisations achieve a comprehensive approach, which leads to comparable and consistent results.

Eurostat, Statistics Denmark, Soil Association and Agence Bio all represent "best practice" examples for the data quality dimension **accessibility/clarity**, which comprises the parameters 'voluntary or obligatory to provide data', 'publication of data', 'availability of data', and 'format of publication'. All these parameters relate to the ease of accessing and using the data produced by the organisations. The "best practice" examples are characterised by a good availability of data which is easy to understand and user-friendly. Due to the different formats of publication offered by the organisations the accessibility is enhanced for the data users. Furthermore, most of the data is publicly available without any charges and the conditions of data provision may reflect in the clarity of the data which is available to the end user.

The quality dimension **timeliness/punctuality** consists of the parameters 'frequency of collection' and 'frequency of publication', which both need to be coordinated to ensure the prompt publication of results close to the aspired release date of data. The organisations, AMI and Bio Suisse, both follow approaches in which the frequency of publication and the frequency of collection happen in the same time intervals. The timely and punctual collection and publication of data improve its quality and the satisfaction of the end user concerning the statistical output.

5 Conclusions

The current EU project "OrganicDataNetwork: Data network for better European organic market information" follows previous Europe-wide research on organic markets and aims at identifying and closing knowledge gaps as well as closing gaps identified in prior research projects. Since the final EISfOM project report with detailed recommendations (Rippin et al., 2006), some improvements in organic data collection have been achieved, but methods and data quality remain very heterogeneous. Especially data on consumption volumes or import and export volumes of organic products for important organic agricultural assets are missing in the majority of European countries. Improvements in the organic data basis refer mainly to primary production data and are likely to be due to the official regulations implemented on the European level. Data collection does not take place evenly across Europe; most responses from the survey conducted among organic data collectors came from countries with better developed organic markets (Gerrard et al., 2012). That

makes it difficult to analyse the constraints regarding organic data collection in countries with less developed organic markets. In most European countries, organic market data collection involves the collection of area and operator data, i.e. this data is already collected and published in a fairly comprehensive way across Europe. However, organic market data collection in all other fields is inconsistent and incomplete in nearly all European countries. The heterogeneity of the responses within this survey and the uneven return of responses across all European countries mirror the inconsistency and the incompleteness of organic market data.

The results of the frequency distributions from the online and telephone survey (Chapter 4.2) served as an overview on the current situation of data collectors and were meant to introduce the reader to the underlying dataset used for the quality assessment. The results reveal that most organic market data was collected from producers, closely followed by control/certification bodies, and wholesalers/processors. Data from caterers, distribution/transport companies, and port/customs authorities was collected by less than 10% of the respondents.

Most of the organisations used their data for statistics and market information. However, about 90% of the respondents claimed that they compile data, while only 55% also conducted basic statistics. Very rarely was more advanced statistical analysis was carried out, mainly for retails sales data. Data quality checks were applied by about 70% of the organisations. Unfortunately, only few respondents provided information about the nature of their quality checks. These quality checks were mostly applied to production volume data. Similarly, comprehensive conclusions from the question on sample sizes cannot be drawn, because only few organisations responded to that question. If the number of responses is broken down to the different countries and the different types of data, the resulting picture would be very heterogeneous and not meaningful. Therefore a general conclusion on sample sizes for organic market data collection in Europe cannot be drawn.

Most of the organic market data is collected annually and also published annually. Consumer and farm level price data are more often collected and published on a weekly basis than other data types. The most common format for publications is the web page. About half of the respondents also named online and paper reports as well as statistical tables as the formats in which they published their data. Most of the rather limited organic market data collected in Europe is publicly available, but given that mainly production data are collected this is not very helpful to market actors. Hence, as far as data exists the access should not be a barrier to data availability.

This report has specially focused on the data collection methods as they form an important basis for harmonising the organic market data situation in Europe. The collection methods were analysed individually for each data type to investigate the compliance of data type and collection methods and thereby detect inconsistencies in the methodological approaches. As mentioned in the previous paragraph figures have to be handled with care, as the underlying database was very heterogeneous and incomplete. Nevertheless, for most data types there is one data collection method carried out most frequently, revealing the conformance of these approaches; e.g. for export volume and value data, most organic market data collectors use e-mail surveys to obtain their information (see Figure 8). It is striking that many organisations use expert estimates for data collection, as expert estimates are not an acknowledged statistical method, but rather a source for additional validation of information. This finding reflects the current situation in the organic market sector, in which a viable information infrastructure has not been established throughout all European countries yet. Although

it should be avoided to compile data by only using expert estimates, they can be a valuable addition for data quality assessment as they add to the accuracy of data.

For the assessment of data quality the quality dimensions were explained and applied to a limited number of datasets. The procedure is outlined and the ranking of the performance for each parameter of the data quality dimensions is described in detail (see Table 4). Thereby the reader gets an insight into the data quality evaluation approach chosen in this study. Furthermore organisations interested in the evaluation of their data quality learn how to apply the concept in order to reveal inconsistencies and improvement possibilities in their own data collection, processing, and publication approach. The results show a few "best practice" examples which can also serve as a reference system for other data collectors throughout Europe and for the implementation in the case studies, which will be conducted in the framework of the Organic Data Network project.

6 References

Agence Bio, http://www.agencebio.org/la-bio-en-france

AMI, http://www.ami-informiert.de/ami-english/ami-about-us.html

Bio Suisse, http://www.bio-suisse.ch/en/home.php

European Commission, DG for Agriculture and Rural Development (2010): An Analysis of the EU organic sector, available online at: http://ec.europa.eu/agriculture/analysis/markets/organic_2010_en.pdf

Eurostat, http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM&StrGroupCode=LEX_MANUAL&StrLanguageCode=EN

Eurostat (2009): ESS Standard for quality report, published by European Communities, Luxembourg, ISBN: 978-92-79-07854-5, available online at: http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/ESQR_FINAL.pdf

Eurostat (2010): Organic farming statistics — setting higher standards, Sigma – The Bulletin of European Statistics 2010/1, available online at: http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Organic_farming_statistics_%E2%80%94_setting_higher_standards

Eurostat (2012): Quality Assurance Framework of the European Statistical System, Version 1.1, available online at: http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/QAF_2012/EN/QAF_2012-EN.PDF

Gerrard, C.L., Vieweger, A. and Padel, S. (2012): Deliverable 2.1 Report on data collectors: Inventory of data collecting and publishing institutions, Organic Data Network project report, Organic Research Centre Newbury, online available at: http://projects.fibl.org/organicdatanetwork/SitePages/Deliberables.aspx

Hahn, M. and Linden, H. (2007): The European statistics Code of Practice for a high quality and European Statistical System, published by European Commission, Luxembourg, available online at: http://panda.hyperlink.cz/cestapdf/pdf07c1/hahn.pdf

Hamm, U. and Zanoli, R. (2006): The need for a long-term strategy for data collection on organic markets, paper at: Joint Organic Congress, Denmark (May 30-31, 2006), available online at: http://orgprints.org/7140/

Hamm, U. and Gronefeld, F. (2004): The European Market for Organic Food. Revised and Updated Analysis, OMIaRD Vol. 5, published by The University of Wales, Aberystwyth UK, ISBN 0-9543279-4-7

Häring, A.M. and Dabbert, S. (2000): The economic impact of the CAP reform and potential future policy developments on typical organic farms in the EU, contribution to the 13th International IFOAM Scientific Conference in Basel, Switzerland, archived at: http://orgprints.org/00002365

Linden, H. and Papageorgiou (2004): Standard Quality Indicators, paper at: European Conference on Quality and Methodology in Official Statistics, Germany (May 24-26,2004) available online at: http://213.133.108.158/rdb_db/PDF/bev_bevoelkerung_demographie/bew_Bewegung_Entwicklung/EuEurope_Conference_on_Quality_and_Methodology_in_Official_Statistics_(Q2004)/Content/Session_11/19_01_Linden.pdf

Rippin, M., Vitulano, S., Zanoli, R. and Lampkin, N. (2006): Synthesis and final recommendations on the development of a European Information System for Organic Markets (Deliverable 6, final report), archived at: http://www.orgprints.org/8961, available online at: http://www.agromilagro.de/downloads/rippin-etal-2006-eisfom-final-recommendations.pdf

Sahm, H., Sanders, J., Nieberg, H., Behrens, G., Kuhnert, H., Hamm, U. (2012): Reversion from organic to conventional agriculture: A review, Renewable Agriculture and Food Systems (*First View* Article), published online in May 2012, available at: http://journals.cambridge.org/download.php?file=%2FRAF%2FS1742170512000117a.pdf&coco=098e5b46f5eb9b7fad870197ef071dd8

Soil Association, http://www.soilassociation.org/marketreport

Statistics Denmark, http://dst.dk/en/Statistik.aspx

Stolze, M. and Lampkin, N. (2005): Einbettung des ökologischen Landbaus in die EU-Agrarpolitik - Der EU-Aktionsplan für den ökologischen Landbau, Agrarwirtschaft und Agrarsoziologe 01/05, available online at: http://ageconsearch.umn.edu/bitstream/32006/1/05010057.pdf