

**Grant Agreement no:** 311778

**Project acronym:** GLAMUR

**Project title:** Global and Local food chain Assessment: a Multidimensional performance-based approach

**Call identifier:** FP7-KBBE-2012-6-singlestage

**Activity code:** KBBE.2012.2.5-03: A comparative analysis of global versus local food supply chains



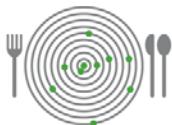
**Methodology evaluation report (deliverable 4.3)**

Jana Schwarz, Erik Mathijs (KULE)



**Due date of deliverable:** July 2015  
**Actual submission date:** August 2015  
**Start date of project:** February 2013  
**Duration:** 36 months

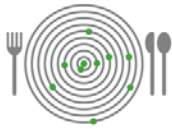
|                                                                                            |        |
|--------------------------------------------------------------------------------------------|--------|
| <b>Project co-funded by the European Commission within the Seventh Framework Programme</b> |        |
| <b>Dissemination Level</b>                                                                 |        |
| <b>PU</b>                                                                                  | Public |



GLAMUR is an EU FP7 project that aims at integrating advancement in scientific knowledge about the impact of food chains to practice, to increase food chains sustainability through public policies and private strategies. This general objective will be pursued through the following specific objectives:

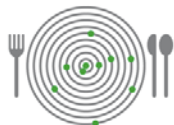
- To develop and validate a performance criteria matrix for assessment and comparison of food chains operating at a range of geographical scales through analysis of how food chain impacts are communicated in different spheres of society.
- To build a database of quantifiable indicators of impact and a set of 20 case studies aimed at understanding how impacts are generated within specific food chains.
- To advance knowledge on methodological problems and trade-offs arising when measuring and comparing the impact of food chains within and between sectors.
- To assess how performance is perceived by stakeholders in different national contexts through participatory assessment and multi-criteria analysis of the different typologies of food chains.
- To assess the actual and potential role of public and private policies addressing food chains and to turn assessment into policy recommendations.
- To build a network that turns the advancement of scientific knowledge into decision making tools for domestic and public consumers, producers, citizens, scientists, policy makers, civil society organizations.

Report to be quoted as: Schwarz J. and Mathijs E. (2015) GLAMUR WP4 – Methodology evaluation report (deliverable 4.3). KU Leuven, Belgium.



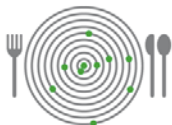
# Methodology evaluation report (Deliverable 4.3)

Jana Schwarz and Erik Mathijs – KULE



## Table of Contents

|                                         |    |
|-----------------------------------------|----|
| 1. Introduction to deliverable 4.3..... | 5  |
| 2. Overview of methodologies .....      | 6  |
| 3. Methodology evaluation.....          | 9  |
| 3.1. Participatory evaluation .....     | 10 |
| 3.2. LCA .....                          | 16 |
| 3.3. Shadow Pricing.....                | 19 |
| 3.4. Metabolic analysis .....           | 21 |
| 4. Cross-method comparison.....         | 22 |
| 5. Conclusion.....                      | 27 |
| References.....                         | 29 |



## 1. Introduction to deliverable 4.3

Deliverable 4.3 is a methodology evaluation report and belongs to task 4.2 – Comparison and evaluation of methodology. In the GLAMUR DoW the task is described as follows:

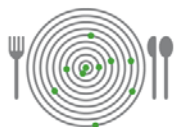
### *Task 4.2. Comparison and evaluation of methodologies*

*The challenge of the project is to create a bridge between the different methodologies employed to analyse the performance of food chains, so to complement each other and create synergies. This objective can be obtained if researchers involved in the project are able to organize a reflection on strengths and weaknesses of methodologies applied and a genuine communication with researchers applying different methodologies. The process of reflection will bring to an evaluation report of research activity of each partner, on a common discussion about self evaluation results and on a discussion with the experts of the project expert forum.*

*The task will be articulated in the following subtasks:*

- **Sub-Task 4.2.1** *Guidelines to methodology evaluation report*  
WP leader KULE will prepare a guideline for the methodology evaluation report, which will include a monitoring questionnaire to be filled by researchers during WP3 and WP4.1. The guideline will be ready by month 13 and agreed to the 3rd project meeting.
- **Sub-Task 4.2.2** *Methodology evaluation report*  
Each partner will prepare a methodological evaluation report which will identify points of strength and weaknesses of used methodologies. WP Leader KULE will write a general methodology evaluation report on the basis of partners' report and of the outcomes of the expert meeting (task 4.2.3)
- **Sub-Task 4.2.3** *Feedback from experts*  
The experts of the Expert forum will be asked to discuss the assessment report and the methodology evaluation in the 2nd expert meeting. The structure of the meeting will be planned and circulated in advance

The aim of this report is to compare and to evaluate the different methodologies that have been applied by the different partners to evaluate the performance of food chains. It is based on the results of a methodology evaluation questionnaire that has been filled in by each product group team (grains, pork, fruits and vegetables, cheese, wine) for each methodology applied by the team (participatory evaluation, LCA, metabolic analysis, shadow pricing). Table 1 summarizes the methodology evaluations that have been carried out by the different case study teams. In the row “partner”, the partner who has handed in the methodology evaluation sheet is listed. In total, we have received 11



methodology evaluations, of which 6 have evaluated the participatory checklist approach, 3 the LCA methodology, and one metabolic analysis and shadow pricing, respectively.

**Table 1:** Methodology evaluation by the different case study teams

| Partner | Product group       | Participatory evaluation | LCA | Metabolic analysis | Shadow pricing |
|---------|---------------------|--------------------------|-----|--------------------|----------------|
| City    | Grain (bread)       | X                        |     |                    |                |
| FIRAB   | Grain (bread)       | X                        |     |                    |                |
| WU      | Pork                | X                        | X   |                    |                |
| UoG     | Cheese              | X                        |     |                    |                |
| FiBL    | Wine                | X                        |     |                    |                |
| BSC     | Fruits & Vegetables | X                        |     |                    |                |
| KULE    | Fruits & Vegetables |                          | X   | X                  | X              |
| INRA    | Fruits & Vegetables |                          | X   |                    |                |

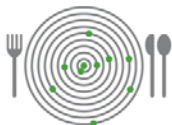
The methodology evaluation questionnaire has been developed based on two existing frameworks: (1) the criteria for indicators for sustainable development derived from Bossel (1999) and (2) the criteria for research quality as derived from Lincoln and Guba (1985) and Bryman (2004)<sup>1</sup>. In total, the questionnaire includes 19 statements covering evaluation criteria related to the usefulness of the methodology for food chain performance assessment, the quality of the obtained results and the implementation process. Respondents have been asked to rate each of the statements on a scale from 1 (completely disagree) to 10 (completely agree). Moreover, they have reflected on these statements and shared comments and insights gained while performing the food chain assessment and case-wise comparison.

In this report we summarize these evaluations. In the next section we summarize the four methodologies that have been used for food chain assessment within GLAMUR. The evaluation results and reflections by methodology are presented in section 3. In section 4 we make a cross-comparison of the different methods, using the average assigned scores as a starting point. Section 5 concludes.

## 2. Overview of methodologies

Four methodologies have been applied in the GLAMUR project: (1) Participatory evaluation, (2) LCA, (3) Metabolic analysis and (4) Shadow pricing.

<sup>1</sup> The *Guidelines to methodology evaluation* (Deliverable 4.2) contain a more detailed description of the two frameworks.

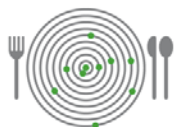


**Participatory evaluation methodologies** help the evaluation building process, allowing for the integration of different points of view, values and judgment criteria of different stakeholders (people affected by positive and/or negative supply chains effects). Different stakeholder categories are asked on the one hand to build the structure of the participatory approach, to identify areas of expected impacts (integrating scientific evidences), to define chains of causality and to choose appropriate indicators. On the other hand the stakeholders are asked to participate in providing relevant information for indicators, to interpret results and to evaluate them according to their interests and expectations. The empowerment of different stakeholder categories and the adoption of flexible and inclusive working methods are key issues in the participatory evaluation approach.

**Life Cycle Assessment (LCA)** is a method to analyse environmental issues across the life cycle of a product, from raw material inputs to waste. It can systematically identify key areas to improve environmental and economic performance. In LCA, all inputs are traced back to primary resources; for example, electricity is generated from primary fuels like coal, oil and uranium. A standardised LCA methodology provides results that are comparable between sectors and industries. LCA can be applied to agricultural systems, however, agriculture does not consume resources in a linear sense in the same way as many industrial processes and is therefore not a pure “cradle-to-grave” process. Differences between LCA studies are due to the goal of the study or the quality of the data. The goal, for instance, determines important methodological choices with regard to the system boundary, the functional unit and the type of allocation method. For example, the goal of the study may be the environmental improvement of the farm, or the whole supply chain to the final consumer. This decision will determine whether to include factors such as the emissions associated with the production of medicines, insecticides, machines, buildings and roads in the system boundary. Furthermore, a disclosure of LCA for comparability or marketing purposes requires a critical review. In GLAMUR the LCA complies with ISO14040 in order to be recognized as robust, scientifically sound and comparable with other studies, databases and software.

The **shadow price** methodology involves the pricing of impacts for which there are no markets, such as in the case of most environmental impacts. Such impacts can be monetised by computing either the cost of abating the pollution or the damage costs following the pollution.

**Metabolic analysis** changes the focus from the analysis of the value (i.e. assessing costs/benefits) of *products* – the focus adopted in economic analysis - to the value of the *functions/structures* which must be maintained and reproduced in order to be able to produce and consume the economic products within specified metabolic patterns. With this change in focus it is evident that not only the fund elements (i.e. the production



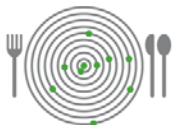
factors) of the socio-economic process are relevant for sustainability, but also that the maintenance and reproduction of fundamental elements of the ecological processes (i.e. ecosystem services guaranteeing favourable boundary conditions) has to be considered when dealing with sustainability analysis. This translates into an adoption of the strong sustainability paradigm, acknowledging the impossibility of substituting nature with human technology. An approach for developing quantitative analysis within this narrative - Multi-Scale Integrated Analysis of Societal and Ecosystem Metabolism (MuSIASEM) has been developed by the group of Societal Metabolism Studies of the Institute of Environmental Science and Technology, Universitat Autònoma de Barcelona, (Spain) in two previous EU projects - DECOIN (<http://www.decoin.eu/>) and SMILE (<http://www.smile-fp7.eu/>).

Table 2 gives a rough overview of which attributes (as identified by WP2) are covered by which method. By definition, the participatory approach is able to cover all relevant attributes of a given case study. LCA covers the environmental dimensions: resource use, pollution and biodiversity. The shadow price method combines economic information with environmental LCA data by multiplying environmental impacts with their estimated social cost and thus has the potential to cover some of the environmental dimensions (resource use and pollution) as well as some of the economic dimensions (affordability, distribution of added value, efficiency, profitability). Metabolic evaluation adds dimensions of innovation, economic development, connection and resilience.

**Table 2:** Matching methodologies and attributes

|                                        | Participatory evaluation | LCA | Shadow pricing | Metabolic evaluation |
|----------------------------------------|--------------------------|-----|----------------|----------------------|
| Affordability                          | +                        |     | +              | +                    |
| Creation & distribution of added value | +                        |     | +              | +                    |
| Contribution to economic development   | +                        |     |                | +                    |
| Technological innovation               | +                        |     |                | +                    |
| Governance                             | +                        |     |                |                      |
| Efficiency                             | +                        |     | +              | +                    |
| Profitability/competitiveness          | +                        |     | +              | +                    |
| Connection                             | +                        |     |                | +                    |
| Resilience                             | +                        |     |                | +                    |
| Food waste                             | +                        |     |                |                      |
| Information and communication          | +                        |     |                |                      |
| Food security                          | +                        |     |                |                      |
| Consumer behaviour                     | +                        |     |                |                      |
| Territoriality                         | +                        |     |                |                      |
| Labour relations                       | +                        |     |                |                      |
| Resource use                           | +                        | +   | +              | +                    |





|                |   |   |   |   |
|----------------|---|---|---|---|
| Pollution      | + | + | + | + |
| Biodiversity   | + | + |   |   |
| Nutrition      | + |   |   |   |
| Food safety    | + |   |   |   |
| Traceability   | + |   |   |   |
| Animal welfare | + |   |   |   |
| Responsibility | + |   |   |   |
| Fair trade     | + |   |   |   |

### 3. Methodology evaluation

In this section we summarize the reflections on the different methodologies. We have merged the 19 statements used in the questionnaires into the following 9 overarching topics:

a) **Facilitation of decision making**

In this question it should be assessed whether the method facilitates decision-making at different levels.

b) **Representation and selection of sustainability dimensions**

This topic includes different aspects regarding the representation and selection of sustainability dimensions. More specifically, it includes comments on the following statements:

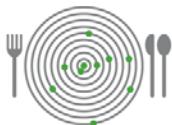
The method represents relevant dimensions of sustainability/ The method encompasses a small but sufficient amount of dimensions to be manageable and easy to communicate/ The method selects dimensions in a participatory way/ The method is sensitive to the relevant dimensions of sustainability in order to measure differences or change/ The method entails a process that includes a vision and criteria for finding dimensions and indicators.

c) **Fit of theory and indicators**

Here we summarize comments on whether the method uses indicators that correspond to what we want to measure in theory.

d) **Reliability and objectivity**

Reliability refers to the question whether the method is reliable, i.e. can the research be repeated by other researchers? We also include comments on the objectivity of the method, i.e. whether repetition by the same or other researchers leads to the



same results, and reflections on reproducibility of the results, i.e. on whether the method is standardized in order to generate reproducible and unambiguous results.

e) **Credibility**

Credibility means that interpretations are based on multiple sources (triangulation) and/or the respondent confirms the interpretation of the researcher.

f) **Transferability**

Here we summarize comments on whether the results of the method are transferable to other contexts. Moreover, we consider thoughts on whether the indicators used by the method can be generalized to other contexts.

g) **Dependability**

Are the results from the method dependable, i.e. can they be judged by peers?

h) **Confirmability**

This refers to the question whether interpretations leading to results are as free as possible from personal values or inclinations. It also includes reflections on whether the method ensures that all perspectives of all authors are represented in a fair way.

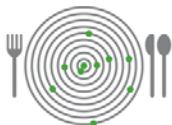
i) **Actors engagement**

This topic relates to all comments regarding the interaction with food chain actors. First, it summarizes comments on whether the method helps actors to better understand their context and to better respect other perspectives in their context. Second, it evaluates whether the method helps actors to engage in action. Third, it also includes comments on the inclusion and participation of actors in the research process.

In the following subsection we summarize the comments and reflections of the case study teams, following the structure outlined above.

## 3.1. Participatory evaluation

Six teams have assessed the participatory evaluation method that they have applied in their case studies: City University and FIRAB for the grains product group, WU for the pork case study, UoG for cheese, BSC for fruits and vegetables and FiBL for wine. In the following we summarize the comments given by the different case study teams on the participatory evaluation method.



a) **Facilitation of decision making**

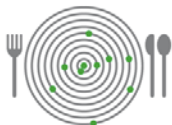
Participatory evaluation can generally facilitate decision-making at different levels. The Latvian team pointed out that it is not the method itself that causes discussion and facilitates decision making, but the way it is implemented. However, the method ensures that the questions addressed are those that really matter for the actors by involving them in the research process, i.e. to discuss research design and to feedback the results to them. The methodology has been valued as being effective for selecting key attributes and sets of indicators to evaluate specific performance levels along different food chains. These attributes and indicators integrate different viewpoints and, in turn, could facilitate decision making along the chains. The Swiss team mentioned that the method helps visualizing the results by indicators, attributes and dimensions which could help decision making at different levels. Nonetheless, the assessment through indicators must be complemented with an extended explanation to avoid biased judgements and facilitate aware decision making, as stressed by the FIRAB-team. It has also been mentioned by the FiBL team that stakeholders see the methodology more as a tool to confront their assumptions about their own performance and that of the supply chain and that the results of the participatory evaluation do not provide detailed information on how to perform better.

b) **Representation and selection of sustainability dimensions**

First, the respondents agreed that the method represents relevant sustainability dimensions. Within the frame of the GLAMUR project, the 24 attributes that have been selected in the process of WP2 have provided a large sample of criteria as a starting point for a wide and comprehensive sustainability assessment. Literature reviews and case study analysis helped to qualitatively selecting a case-specific subset of attributes, including those engaged in the process. However, the BSC-team stressed that it is important to include all relevant actors and that sometimes there might be a lack of willingness to involve in the research process, leading to a somehow incomplete picture. The stakeholder feedback that will be carried out in WP5 will verify the “goodness” of the selected attributes.

Second, the amount of dimensions covered by the method is seen as small and sufficient. However, most respondents argued that the concept was not so easy to communicate to different stakeholders and that especially for the ethical dimension it was necessary to explain it in more detail. The difficulty lies into translating outcomes into simple and univocal messages, which at the same time makes it harder to communicate the subtleties of the wider range of dimensions covered. Thus, the skills of the researcher to communicate what is required are crucially important.

Third, the dimensions/attributes have been collected in a relatively participatory way, which is seen as the essence and a main strength of the participatory evaluation



approach. However, most case study teams in reality did not only rely on participatory methods as there was a danger that this would limit the range of dimensions selected, e.g. stakeholders could focus on only one particular dimension. In the pork case studies, overall performance dimensions/attributes have been primarily selected by the research teams followed by a check among stakeholders to see if these agreed with the selection and/or missed specific performance components. The City team mentioned that the GLAMUR framework with dimensions – attributes – indicators acted as a framework, but scoring the indicators in such a way also meant that it was not possible to take full advantage of the participatory process. For example, confining the attributes and indicators within a specific sphere meant that it was not possible to take into account the full range of stakeholder views and experience, and additional perceptions were lost.

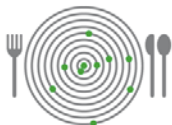
Fourth, the questionnaire respondents agreed that the participatory evaluation method is only partly able to measure differences or change in dimensions. The teams of WU and UoG pointed out that differences and change can be indicated but not measured exactly and the FiBL team mentioned that the method can nonetheless compare different chains by visualizing differences in dimensions, attributes and indicators.

Fifth, the teams acknowledged that the methodology includes a vision and criteria for finding dimensions and indicators. These were set out clearly in the WP3 guidelines which were used for navigating how to select, measure and assess the indicators within the dimensions. The pork team mentioned that the operationalisation of attributes into indicators was more subject of debate among stakeholders than the selection of relevant attributes.

#### c) **Fit of theory and indicators**

It has been mentioned that it is generally possible to *indicate* a set of indicators that measure what we want to know in theory (although there are some attributes such as resilience or animal welfare which are very difficult to operationalize in widely accepted indicators). The main problem is that it is often not possible to precisely *measure* the indicators and that researchers have to rely on rough estimates or relative indications. In this respect, data availability was mentioned as a major constraint by the case study teams. The City team pointed out that especially food systems operating at the local level are often not able to provide detailed datasets. Another problem regarding the construction and measurement of indicators was the work as 'pairs of pairs' due to which some indicators were a compromise between the two case study teams. Therefore, they did not always provide the most appropriate measure in practice.

Many respondents also referred to the SAFA guidelines. These state that "performance-based indicators are more relevant and more effective than practice-



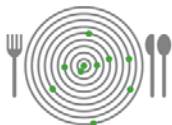
based indicators in demonstrating performance and impacts". However, the selected indicators included a mix of both types and two respondents mentioned that practice-based indicators were more suitable to the largely qualitative data collection methods of the participatory approach. For example, instead of measuring biodiversity on the field, the average frequency of practices linked to biodiversity has been measured by the FiBL team. Another problem mentioned by the City team was that data analysis only evaluated the attributes and indicators within the specific dimensions that were identified by the Multi-criteria Performance Matrix in the WP2 comparative report. For example, biodiversity within environment, nutrition within health, etc. Because of this, the evaluation had specific effects on the 'scoring' for each chain (global, regional and local in the UK bread case) and, importantly, did not directly incorporate how the attribute and indicators might 'score' under the panoply of all the performance dimensions (economic, environmental, health, social and ethical). This meant that other externalities were not considered - for example, economic or social impacts under biodiversity. Benchmarking was also constructed to reflect one specific performance dimension; for example, the indicator salt content in bread, under the attribute nutrition was benchmarked for health, using the UK's voluntary salt reduction target; benchmarking each indicator across all of the performance dimensions would have provided a more in-depth set of results, and addressed all aspects of sustainability.

d) **Reliability and objectivity**

The participatory evaluation method is extremely context dependent and does not follow a strict standardized procedure. Thus, it is only to a limited extent possible to generate reproducible and unambiguous results. This depends to a large part on the stakeholders included and on the quality of the involved researchers, e.g. on whether they are able to ensure a good communication process with the stakeholders.

Regarding reliability, i.e. the possibility of repetition of the research by other researchers, the questionnaire respondents mentioned that the method can generally be repeated by other researchers, provided there is a comprehensive explanation of the way in which the method has developed. However, most teams admitted that there is a danger to subjective interpretations of the guidelines and the results and there might be debates among stakeholders with regard to the main conclusions. The bread-team argued that the close collaboration between the two country teams counter-acted some of these issues as both teams worked closely together and learned from each other. This team-wise collaboration could be repeated by other researchers and lead to more reliability under the condition that there is enough oversight, regular reporting at every research stage and enough opportunities to discuss progress.

However, it is very likely that repetition of the research by the same or other researchers - i.e. objectivity of the method – will not lead to exactly the same results.



Food chains are very dynamic and in constant flux which would inevitably influence the research outcomes and results.

e) **Credibility**

The results of the participatory evaluation method are seen as credible because the results are based on multiple sources and in many cases, interview respondents have confirmed the interpretations of the researchers. Triangulation is one of the main strength of the participatory evaluation method and the method involved a series of 'checks and balances'. All case study teams performed interviews with a wide range of stakeholders at different stages of the research and often contacted them several times to validate the interpretations. Other sources for triangulation were literature studies, questionnaires, workshops, and focus group discussions with different groups of stakeholders. In the Latvian berry case study, triangulation led to the identification of controversial chain elements and to the detection of contradictory information from different stakeholders.

The City University team added that the data quality check was an important way to demonstrate the quality of the data, especially regarding the recentness and the origin of the data.

f) **Transferability**

There was an agreement among all questionnaire respondents that some indicators might generally be used in other contexts. However, it has been stressed that the goal of the participatory evaluation method is to grasp contextual performance and that indicators might require 'tailoring' to fit different cases. The pork team argued that only attributes might be generalized to other contexts and that the translation of the attributes into indicators is strongly context dependent.

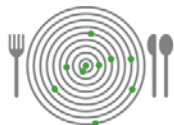
However, the transfer of research results to other contexts is very difficult. This is not mainly due to the method itself, but due to the place-specific interaction patterns and supply chain specifics. General principles might hold good, but more specific results are strongly linked to the food chain context.

g) **Dependability**

The dependability of the results of the participatory evaluation, i.e. the possibility of being judged by peers has been assessed as good. The ability of a peer reviewer to accurately assess the results of the case study depends upon there being a comprehensive description of the approach taken. The peer review system within GLAMUR has worked quite well.

h) **Confirmability**

Most questionnaire respondents agreed that the participatory evaluation method is not totally free from subjectivity. The nature of the research process inevitably



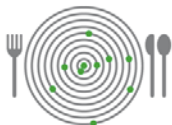
means that personal interpretations play a part in the process. For instance, the development of robust and meaningful scores to assess the qualitative data was based upon the researcher's own interpretation which was inevitably somewhat subjective. The method allows considering different viewpoints of different stakeholders but the method is rather sensitive for differences in stakeholder prioritization and interpretations. The UoG team mentioned that in this respect, having discussions among research teams with different disciplinary backgrounds led to robust discussions about the results and the interpretations.

#### i) **Actors engagement**

The engagement of actors in the research process is a main component of the participatory evaluation method.

Regarding the fair representation of all perspectives of all actors, many teams mentioned that they had difficulties to engage actors especially from the global chains in the research process. These chains are often very complex with many actors at different stages of the food chain and often, actors in global chains have been more reluctant to share and express their ideas and opinions with the research teams. The viewpoints of these actors have been indirectly assessed through secondary information from agricultural newspapers and journals, websites, foodlogs, etc. Moreover, in the Dutch pork case study, the perspectives of 'distant' actors in global pork chains are only represented indirectly through the viewpoints of social movements. But although not all actors have engaged in the research process in the same way, the research teams have endeavoured to be fair in their analyses of data. Again, this depends strongly on the ability and objectiveness of the researcher. The BSC team pointed out that the participatory evaluation method presupposes that actors have at least some level of interest in the question studied in order to participate in the research process. The method is thus better applicable to issues that are already perceived as being important.

The participatory evaluation method has only a limited capability to help actors to understand their own context. On the one hand, the method could provide actors with new perspectives, insights and critical self-reflection by using questions in the semi-structured interviews that prompted thinking 'outside the box' (City), confrontation with other views (FIRAB, FiBL, BSC), and by asking them to explain things they have previously not reflected on (BSC). The BSC-team argued that increased understanding of the context is an especially important factor in understudied food chains such as the blueberry chain. But although the method does in principle provide the opportunity for actors to reflect upon their context, in reality, it is difficult to assess the extent to which this happened. On the other hand, the pork case study team pointed out that the method contributes primarily to a better scientific understanding of chain performance and that case-study approaches should be modest in terms of stakeholder contextual 'enlightening' expectations.



Similarly, the ability of the participatory evaluation method to help stakeholders to better respect other perspectives in their context is not clear. Whereas some teams agreed that stakeholders started thinking more critical about other perspectives (City) and that this is a main aim of the method (BSC), others were more cautious about this statement. The UoG team mentioned that the question whether the method helped actors to better respect other perspectives is perhaps a moot point; it certainly enabled them to at least consider the perspectives of others in their context. Other teams agreed that this depends very much on the stakeholders' interests and the intensity of actors' involvement.

The capability of the methodology to help actors to engage in action remains to be seen. Stakeholders have been provided with an assessment that incorporates a range of perceptions and viewpoints which might lead to further understanding and might be translated into actions. This, however, depends strongly on the actors themselves and the kind of action. In the pork case study, global pork chain actors might not have noticed much from the research activities and are not expected to engage in action. Local actors, on the other hand, might use case study findings helpful for future attempts to mobilize institutional and societal support and to convince consumers about the societal benefits of their initiatives and practices. The FiBL team mentioned that stakeholders expressed the feeling that the research did neither provide them with clear ways of how to perform better nor with reasons for change. In summary, the translation of research results into actors' awareness and engagement into action is very context dependent and remains to be seen in the future.

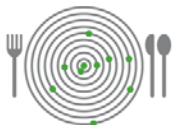
## 3.2. LCA

The LCA methodology has been evaluated by three case study teams: WU for pork and both KULE and INRA for the fruits and vegetables teams.

### a) Decision making

The respondents agreed that the LCA method provides valuable information for decision makers but only on environmental and health aspects. There is a risk of missing important aspects if only LCA results are used for decision making. Additional information on social and economic aspects for instance is needed. Moreover, LCA focuses on impact efficiency, i.e. impacts per food product, but does not take into account the total impact at higher levels, e.g. the whole agricultural sector. This should also be taken into account in decision making.





#### b) Sustainability dimensions

The traditional LCA method does not represent all relevant dimensions of sustainability as it focuses only on environmental aspects and some health issues. However, on these issues the method provides clear quantitative results and it is widely accepted among policy makers and professionals. For non-specialists the method is a 'black box' and for them it is not possible to assess the reliability of the results. In the communication process it is important to clearly stress the strength, weaknesses and significance of the analysis. The INRA team argues that LCA should be considered in combination with other quantitative and qualitative methods in order to cover other dimensions of sustainability and chain performance. One attempt in this respect is the inclusion of social impacts into LCA approaches, i.e. Social Life Cycle Analysis. However, this has not been applied in the GLAMUR case studies.

Generally, the LCA method is specialist based and does not select dimensions in a participatory way. But feedback from farmers and stakeholders can be used for fine-tuning, e.g. the choice of the functional unit and input variables, or the weighting of impacts.

Regarding the sensitivity of the method to the relevant dimensions of sustainability in order to measure differences or change the questionnaire respondents agreed that LCA is a static approach and does not pay explicit attention to changes in time. But results of the method can be used to measuring differences between different chains at the same scale, if the method is implemented in exactly the same way. Moreover, different scenarios can be tested, e.g. changing environmental impacts due to a reduction of fertilizer use or a switch in transportation means. This has however not been done in the GLAMUR project.

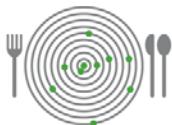
As LCA is an existing method that uses existing models, the method does generally not entail a process that includes a vision and criteria for finding dimensions and indicators. It is however possible to choose among different indicators within the environmental and the health dimensions.

#### c) Fit of theory and indicators

The LCA method includes only a limited number of indicators for few attributes. But the questionnaire respondents find that these correspond to what we want to measure in theory.

#### d) Reliability and objectivity

A standardized framework exists for carrying out LCAs. ISO standards (14040 and 14044) shall guaranty data quality and the implementation of the methodology. But when comparing LCA results of different case studies it is important to compare the basic hypotheses, the analytical model, the system boundaries and the software



chosen to make the LCA inventory. Without checking these criteria a throughout comparison of LCA results is not possible.

The LCA method is very reliable and objective, i.e. the research can be repeated by other researchers and repetition by the same or other researchers are likely to lead to the same results. But all the assumptions and methodological choices need to be clearly explained and some choices and assumptions, such as the choice of the functional unit, are always debatable. An example has been given for the pork case study by the WU team: They chose not to consider different parts of the pig or different meat qualities and assumed all non-processed meat to have the same value for the consumer.

e) **Credibility**

Results of the LCA analysis need to be understood considering the context. The method itself does not include triangulation of sources but the method and results can be explained to stakeholders to enable them to give feedback. This can increase the credibility of the results.

f) **Transferability**

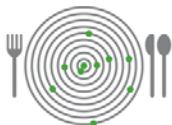
LCA indicators can generally be generalized to other contexts. However, the pork case study team pointed out that the meaning of an indicator can be different depending on the context. For example, in the Netherlands the functional unit non-processed carcass weight is meaningful. This is not the case in Italy where the quality of the processed meat is far more important for instance. So, it could be argued that only attributes might be generalized to other contexts and that the translation of attributes into indicators is strongly context dependent, as in different ways confirmed by the pairwise pork chain comparison between the Netherlands and Italy. Regarding the transferability of results to other contexts, the respondents agree that this is low for the LCA method. Results are very context specific, especially because of differences at the farm stage such as farm management, practices, etc.

g) **Dependability**

The respondents agreed that the results of the method are dependable, i.e. they can be judged by peers.

h) **Confirmability**

Results of an LCA are also seen as confirmable, i.e. interpretations are as free as possible from personal values or inclinations. However, some important choices, e.g. regarding the functional unit and some other assumptions might be influenced by personal inclinations.



i) **Actors engagement**

The fair representation of all perspectives of all actors is thought to be low because LCA is a specialist-based method. The WU team argues that the method contributes primarily to a better scientific understanding of chain performance. In building upon available sets of indicators the method only to a lesser degree helps actors to substantiate claims on distinctive performances. When used for comparing different food chains (i.e. the pairwise comparison between different case studies), the LCA method might contribute to better respect other perspectives in the same context. But LCA only considers the environmental dimension which limits the number of different perspectives.

### 3.3. Shadow Pricing

The shadow pricing methodology has been applied by the apples case study team and the methodology evaluation questionnaire has been filled in by KULE.

a) **Decision making**

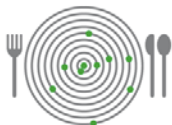
Shadow pricing is only used to evaluate costs of externalities. But not all externalities can be quantified which leads to the risk, that decisions are made based on partial information. Additional information and descriptions are necessary to inform decision makers.

b) **Sustainability dimensions**

In theory, shadow pricing can be used to quantify externalities related to different sustainability dimensions, e.g. the social or the environmental dimension. But the degree to which these costs can be monetised differs. The questionnaire respondent argued that costs related to the environmental dimension are the easiest to quantify by monetizing impacts that have been calculated with LCA. Generally, impacts and the way of monetising them are selected by the researcher and it depends on the quality of the researcher to do this in a scientifically sound way. The method can be used to quantify impacts at different points of time, but this is not the focus of the shadow pricing methodology.

c) **Fit of theory and indicators**

The fit of what we want to measure in theory and the actually used indicators is rather low in the shadow price methodology. The process of monetising externalities already entails an abstraction from the original impact.



**d) Reliability and objectivity**

There are different methods that can be used to convert impacts into costs, and the selection of impacts can differ in different studies. Thus, the way of monetising impacts always needs to be explained by the researcher.

**e) Credibility**

Results of the shadow price methodology are generally not triangulated, as the cost of environmental externalities is a hypothetical concept which is hardly comparable with other information. The inference of interpretations based on shadow prices is not debatable; it is rather the way of monetising that could lead to discussions about the credibility of the results.

**f) Transferability**

Results of the shadow price methodology are context specific and not easy to generalize to other context in the area of food chain assessment. Similar to the LCA methodology, there are often major differences at impacts at the farm stage. The questionnaire respondent argued that impacts linked to the processing and transport stage are easier to generalize. Moreover, the monetisation of impacts is often based on consumers' willingness to pay for external costs which might be very different in different regions and thus lead to different results.

**g) Dependability**

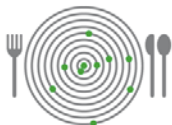
Yes, if the way in which shadow prices have been calculated is clearly explained to other researchers, they can easily interpret the results.

**h) Confirmability**

Personal values and inclinations do not influence the results derived from the shadow price methodology. The methodology is thus very confirmable.

**i) Actors engagement**

Regarding the extent to which the method helps actors to better understand their own context the questionnaire respondent stressed that shadow pricing only relates to estimated costs of externalities and that a big part of context information is not part of the methodology. With its focus on environmental aspects, the method helps actors to better understand environmental impacts but does not include much information on other aspects, such as social or ethical issues. In this way, shadow pricing can help actors to engage into action and to improve the environmental performance of their food chain.



## 3.4. Metabolic analysis

The metabolic analysis has been carried out by the fruits and vegetables team, specifically for the case studies on tomatoes and apples. The methodology evaluation questionnaire has been filled in by KULE.

### a) **Decision making**

The method might facilitate decision making as it takes into account different levels of the system in which the chain is embedded. This can help decision makers to consider the different impacts their decisions can have. The method can be used to study the resilience of food systems by quantifying internal (the socioeconomic system) and external (interactions with the ecosystem) constraints and it analyses the main drivers that put pressure on the system to evaluate the state, possible impacts and responses. Moreover, the methodology enables the evaluation of different scenarios which might help decision makers to make choices. However, the focus is on bioeconomic criteria, social and especially ethical aspects are less studied.

### b) **Sustainability dimensions**

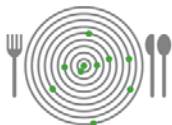
Metabolic analysis bridges the societal and environmental dimension by establishing links between the social system and the ecosystem. Dimensions are usually not selected in a participatory way, but additional indicators can be integrated based on participatory methods. The number of indicators can be chosen by the researcher, but generally the number of different indicators is relatively small and easy to present and communicate. MuSIASEM has the potential to be used to analyse changes in time, but this has not been done for GLAMUR.

### c) **Fit of theory and indicators**

Mostly, there is a good fit between what we want to measure in theory and the actually used indicators. However, some issues such as biodiversity are very difficult to measure directly.

### d) **Reliability and objectivity**

The metabolic analysis (especially MuSIASEM) is not a standardized method such as LCA. The methodological procedure needs to be well explained to other researchers who want to carry out and interpret such an analysis. However, there is a general process that can be followed to generate unambiguous results and it is very likely that experienced researchers would get the same results when applying the methodology.



e) **Credibility**

Results of the metabolic analysis are very credible. In order to build the models, data from different sources have to be compared and there are checks at different analytical levels to confirm the fit of the data. Moreover in theory, the metabolic analysis should build on a participatory approach and include stakeholders in the research process.

f) **Transferability**

Results of the metabolic analysis are context specific and thus difficult to generalize to other contexts.

g) **Dependability**

If the fellow researcher knows the methodology, there is no problem in doing a peer review of the result. Peers can relatively easily judge the results and interpret them.

h) **Confirmability**

MuSIASEM uses quantitative indicators to analyse the metabolism of systems. Thus, the personal values of the researcher do not influence the model results. However, the choice of indicators and indicator definition is mainly done by the researcher. But interpretations of the results are based on the comparison of numbers and do not depend on the researcher.

i) **Actors engagement**

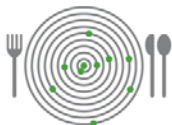
One of the main strength of the method is that it helps actors to better understand their own context and the functioning of the system they work in. It can be seen as a basis on which different actors can discuss and compare their perspectives and narratives. As it is especially directed towards policy makers, the method is very suitable for helping actors to engage into action.

## 4. Cross-method comparison

In this section, we present the average scores given for the different statements related to the performance of the four methodologies. Scores range from 1 (completely disagree) to 10 (completely agree). In detail, the statements given were:

### Sustainable development indicators framework

- 1) The method facilitates decision-making at different levels
- 2) The method represents relevant dimensions of sustainability

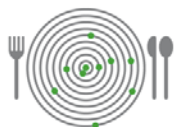


- 3) The method encompasses a small but sufficient amount of dimensions to be manageable and easy to communicate
- 4) The method selects dimensions in a participatory way
- 5) The method is standardized in order to generate reproducible and unambiguous results
- 6) The method is sensitive to the relevant dimensions of sustainability in order to measure differences or change
- 7) The method entails a process that includes a vision and criteria for finding dimensions and indicators

### Research quality framework

- 1) The method uses indicators that correspond to what we want to measure in theory
- 2) The method uses indicators that can be generalized to other contexts
- 3) The method is reliable, i.e., research can be repeated by other researchers
- 4) The method is objective, i.e., repetition by the same or other researchers leads to the same results
- 5) Results from the method are credible, i.e., interpretations are based on multiple sources (triangulation) and/or the respondent confirms the interpretation of the researcher.
- 6) Results from the method are transferable to other contexts
- 7) Results from the method are dependable, i.e., results can be judged by peers.
- 8) Results from the method are confirmable, i.e. interpretations are as free as possible from personal values or inclinations.
- 9) The method ensures that all perspectives of all actors represented in a fair way.
- 10) The method helps the actors to better understand their context.
- 11) The method helps the actors to better respect other perspectives in their context.
- 12) The method helps the actors to engage into action.

In Table 1 we summarize the results. The method with the highest score is highlighted in green. Standard deviations for the average values of the participatory evaluation and LCA are indicated in brackets. These scores are based on very few evaluations (especially for the metabolic analysis and the shadow pricing methodology). Nonetheless, they give some indications on the strengths and weaknesses of the different methodologies.



**Table 3:** Average evaluation scores of the four methodologies

|                                                                                      | Participatory Evaluation | LCA         | Shadow Pricing | Metabolic Analysis |
|--------------------------------------------------------------------------------------|--------------------------|-------------|----------------|--------------------|
| Number of observations                                                               | 6                        | 3           | 1              | 1                  |
| <b>Sustainable development indicators framework</b>                                  |                          |             |                |                    |
| 1. Facilitates decision making                                                       | 7.5 (1.5)                | 4.0 (0)     | 3              | 6                  |
| 2. Represents relevant sustainability dimensions                                     | 8.5 (0.76)               | 5.7 (0.47)  | 4              | 8                  |
| 3. Small but sufficient amount of dimensions (easy to communicate)                   | 7.0 (1.29)               | 6.3 (1.7)   | 7              | 7                  |
| 4. Selects dimensions in participatory way                                           | 7.2 (2.14)               | 3.3 (0.47)  | 3              | 5                  |
| 5. Reproducible and unambiguous results                                              | 5.3 (1.49)               | 8.3 (0.94)  | 4              | 5                  |
| 6. Measure sustainability differences or change                                      | 6.5 (1.58)               | 3.3 (0.47)  | 3              | 5                  |
| 7. Process that includes a vision and criteria for finding dimensions and indicators | 7.8 (1.21)               | 3.7 (0.47)  | 3              | 3                  |
| <b>Research quality framework</b>                                                    |                          |             |                |                    |
| 1. Indicators correspond to what we want to measure in theory                        | 5.8 (1.57)               | 6.7 (1.89)  | 4              | 7                  |
| 2. Generalization                                                                    | 6.0 (1.29)               | 7.7 (0.47)  | 7              | 7                  |
| 3. Reliability                                                                       | 7.0 (1.15)               | 8.3 (0.47)  | 8              | 8                  |
| 4. Objective                                                                         | 5.5 (2.06)               | 8.3 (0.47)  | 8              | 6                  |
| 5. Credible                                                                          | 7.5 (0.96)               | 7.3 (2.7)   | 8              | 8                  |
| 6. Transferable                                                                      | 4.7 (1.8)                | 3.3 (0.47)  | 4              | 4                  |
| 7. Dependable                                                                        | 7.8 (0.69)               | 8.5 (0.5)*  | 8              | 6                  |
| 8. Confirmable                                                                       | 6.5 (1.12)               | 8.0 (0)*    | 8              | 8                  |
| 9. Fair representation of perspectives                                               | 7.0 (1.73)               | 3.3 (0.94)  | 4              | 4                  |
| 10. Actors understand context                                                        | 7.83 (1.34)              | 5.3 (1.25)  | 5              | 9                  |
| 11. Respect other perspectives                                                       | 7.2 (0.69)               | 4.3 (0.47)  | 5              | 5                  |
| 12. Engage into action                                                               | 5.8 (1.34)               | 4.5 (1.5)** | 6              | 6                  |

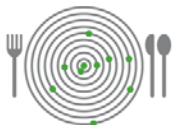
Note: Scores range from 1 (completely disagree) to 10 (completely agree).

\* These results are only based on two observations as one respondent did not understand the question correctly.

\*\* Based on two observations because one respondent did not rate this statement.

Regarding the sustainable development indicators framework we observe that the participatory evaluation method received the highest average scores for nearly all statements. This is mainly due to the fact that the participatory selection of sustainability attributes and indicators is inherent in the participatory evaluation method and that it has the broadest scope with regards to different aspects of sustainability. The other methods focus mainly on environmental and economic aspects and are more science-

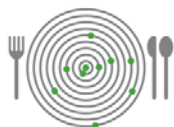




based, i.e. indicators are mainly selected by the researchers themselves and do not require a participatory process. However, the LCA methods scores by far highest for the statement “The method is standardized in order to generate reproducible and unambiguous results”. This is mainly due to the standardization of the LCA procedure as part of the ISO 14000 environmental management standards and the use of well-known software to calculate the LCA impacts. The other methodologies are much more flexible and require the researcher to make more decisions, e.g. choosing between a wide variety of possible indicators, valuing different stakeholder statements, benchmarks, the choice of the monetary valuation method, etc. Thus, a repetition of the same research by other teams is more likely to lead to other results and needs a lot of ‘thick’ description to explain the methodology applied. This is also the reason why LCA scores very well in the statements of the research quality framework. Especially the generalization of indicators, reliability and objectivity of the method, and dependability and confirmability of the results have been assessed as being best in the LCA methodology.

Generally, the transferability of results has been rated very low by the research teams for all methodologies, with average scores between 3 and 4.67. This is due to the location- and context-specificity of food chains: No two food chains have exactly the same configuration, actors are different and many research results on the sustainability performance are very specific for the food chain under study. All results need a throughout description of the chain context and the approach taken for defining and calculating the indicators. Moreover, the combination of different methodologies can be a valuable tool. This has for instance been done by the pork team who has used a methodological combination of participatory evaluation and LCA analysis.

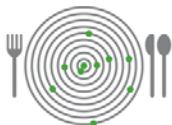
We summarize the main potentials and limitations of the methodologies regarding food chain assessment in Table 4. The main strength of the participatory evaluation is its flexibility and the close cooperation with food chain stakeholders that leads to an integration of different viewpoints and opinions. The drawback of this is the reliance on the interest of stakeholders to be involved in the research process. Moreover, the skills of the researcher to communicate with chain actors and to objectively interpret research findings influence the quality of the research results and are a potential limitation of the methodology as compared to the other methods. A potential benefit of LCA and shadow pricing is the ease of communicating the results. Moreover, LCA is a standardized method which improves reliability and comparability of the results. But LCA and shadow pricing focus mainly on the assessment of environmental impacts which represents an important limitation for the use of these methodologies for food chain assessment. Additional analysis is needed to include other aspects such as social and ethical ones into the sustainability assessment. Similarly, the main focus of the metabolic analysis is on biophysical aspects. However, the strength of this approach is the consideration of different hierarchical levels of analysis and interactions between the social system and the ecosystem.



**Table 4:** Main strengths and limitations of the methodologies

|                                 | <b>Strengths</b>                                                                                                                                                                                                                                                                    | <b>Limitations</b>                                                                                                                                                                                                                                                                     |
|---------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Participatory evaluation</b> | <ul style="list-style-type: none"><li>• Integration of many different viewpoints and aspects</li><li>• Very flexible approach</li><li>• Triangulation of different sources</li></ul>                                                                                                | <ul style="list-style-type: none"><li>• Relies heavily on stakeholder participation</li><li>• (Communication) skills of the researcher are very important</li><li>• Not standardized and difficult to replicate</li><li>• Does not provide a vision on how to perform better</li></ul> |
| <b>LCA</b>                      | <ul style="list-style-type: none"><li>• Easy communication of results</li><li>• Standardized method that helps to make results comparable</li></ul>                                                                                                                                 | <ul style="list-style-type: none"><li>• Limited to environmental and health aspects</li><li>• Only considers impacts per food product, does not take into account interactions at higher levels</li><li>• Static approach</li></ul>                                                    |
| <b>Shadow pricing</b>           | <ul style="list-style-type: none"><li>• Easy communication</li><li>• Inclusion of issues in accounting that would otherwise not be considered</li></ul>                                                                                                                             | <ul style="list-style-type: none"><li>• Limited capability to assess food chain performance, focus on environmental externalities</li><li>• Abstraction from real impacts through monetisation</li></ul>                                                                               |
| <b>Metabolic analysis</b>       | <ul style="list-style-type: none"><li>• Consideration of different levels of analysis</li><li>• Interactions between the social system and the ecosystem</li><li>• Flexible approach</li><li>• Enables comparison and discussion of different narratives and perspectives</li></ul> | <ul style="list-style-type: none"><li>• Focus on biophysical aspects</li><li>• Not standardized</li></ul>                                                                                                                                                                              |

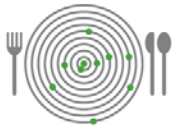
Differences in the methodologies regarding the assessment of local versus global food chains are not pronounced. All methodologies are able to analyze and compare the performance of different kinds of food chains. However, there might be differences in data availability between local and global chains, but there is no clear pattern: On the one hand the City team argued that the local bread chain was not able to provide detailed datasets. On the other hand, the WU team mentioned that stakeholders of the global chain are more reluctant to share and express their ideas and opinions with researchers. Thus, the methodology evaluation unravels important differences in the potential of different methodologies for the assessment of food chain performance. The differences among the methodology in the ability to assess local versus global food chain performance is, however, limited.



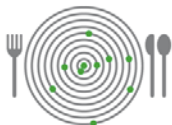
## 5. Conclusion

In the GLAMUR project, four different methodologies have been applied by different case study teams to assess the sustainability of food chain performance: participatory evaluation, LCA, metabolic analysis and shadow pricing. In this report we have summarized the methodology evaluations of the teams who have reflected on the suitability of the applied methodologies for food chain sustainability assessments and on the research quality. Most teams have applied the participatory evaluation, which is based on a strong involvement of stakeholders in the research process. Three partners have evaluated LCA analysis which focuses on the environmental performance of food chains. Metabolic analysis and shadow pricing have been applied by the fruits and vegetables product teams. Overall the methodology evaluation shows that the participatory approach is the most open and flexible one, as attributes and indicators are selected in close collaboration with food chain stakeholders. It can incorporate all relevant dimensions of sustainability. However, the approach is highly context dependent and each food chain assessment needs to include a 'thick' context description in order to put the research results into a broader framework. Moreover, participatory evaluation relies heavily on the abilities of the involved researchers and on the interest of the involved actors. Research teams mentioned the risk of subjectivity of interpretations. These problems are less pronounced in the methodologies that rely on quantitative information. However, LCA and shadow pricing are focusing on the environmental dimension of food chain assessment and do not include other important aspects such as social and ethical ones. In these cases it is important to add additional information to the research results in order to provide a complete picture for decision makers and stakeholders. The metabolic analysis does also focus on the environmental dimension but it is more flexible and has the capability to bridge environmental, economic and some social aspects.

Regarding the applicability of the four methodologies to the assessment of local and global food chain assessments, respectively, we do not find significant differences between the methods. Theoretically, all methods can be applied for the analysis of different food chains and can be used to identify differences in performance. The choice of the method depends on the focus and the goal of the analysis, e.g. LCA or metabolic analysis are the preferred options for assessing environmental impacts whereas participatory evaluation has a broader scope but does not necessarily lead to quantifiable results. A tool that might be used for the selection of the most appropriate methodology for sustainability assessments has been developed by Zijp et al. (2015). It proposes an identification key that proposed methodologies based on a set of key questions the researcher has to answer. However, one point that has been raised in by the Dutch team is that the combination of different methodologies – in their case LCA and participatory evaluation – can be a strong tool for food chain assessments. As all methodologies have



their strengths and drawbacks, a combination of methodologies can help to increase the quality of research results. This has already partly been done in the GLAMUR project, where research teams have applied different methodologies and where stakeholders have been included in the research process not only for the participatory evaluation. This will be further done in WP5 where stakeholder workshops will be carried out to validate the 'goodness' of the research results. The stakeholder feedback will help to assess the quality of the choice of indicators for food chain performance and to evaluate the credibility of the research process from the view of the stakeholders.



## References

Bossel H. (1999). *Indicators for sustainable development: theory, method, applications*. A report to the Balaton Group, International Institute for Sustainable Development, Winnipeg, Canada.

Bryman A. (2004). *Social Research Methods*. Oxford: Oxford University Press.

Lincoln Y.S., Guba E.G. (1985). *Naturalistic Inquiry*. Newbury Park, CA: Sage Publications.

Zijp M., Heijungs R., van der Voet E., van de Meet D., Huijbregts M., Hollander A., Posthuma, L. (2015). An identification key for selecting methods for sustainability assessments. In: *Sustainability* 7, pp. 2490-2512.