Background document

supplementing the

“Roadmap for
Sustainability Assessment in
European Process Industries”

Challenges of cross-sectorial sustainability assessment

Authors:

Aurélie Manent¹, Lisa Mohr², Diederik Schowanek³

¹ Evonik Technology & Infrastructure
² Thyssenkrupp Steel Europe AG
³ Procter & Gamble Services Company NV, Global Product Stewardship
# Table of Content

1. **Introduction** ........................................................................................................ 4
2. **Sharing of primary data via “Black box model”** ............................................. 5
3. **Data exchange via industry averages/collaboration projects** ....................... 8
   3.1 Data collection ..................................................................................................... 8
   3.2 Representativity of industry averages ................................................................. 8
   3.3 Industry averages: horizontal vs. vertical ....................................................... 9
   3.4 LCI models ......................................................................................................... 11
4. **Data documentation in B2C collaboration** .................................................... 12
   4.1 Data quality ........................................................................................................ 12
   4.2 Methodological challenges and choices (qualitative information) .............. 13
   4.3 Quantitative (or semi quantitative) information about results .................... 15
5. **Difficulties currently not solved** ..................................................................... 17
7. **Abbreviations** ................................................................................................... 20
8. **References** ........................................................................................................ 21
9. **Annex 1** ............................................................................................................. 22
1 Introduction

This background document aims at giving an overview of the current practices observed in industrial collaborations along the value-chain or between sectors (at a B2B level only) and identifying advantages and disadvantages of the different approaches. Some recommendations will also be detailed in order to close the identified gaps.

The sharing of LCA data between companies exists but remains a difficult task because of the associated risks for the company providing data:

- LCA data can be related to costs (e.g. production cost)
- LCA data can deliver information about the process itself (yield, kind of substances emitted etc.) that are often highly confidential

Moreover, the collaboration requires much resource for both companies and the set-up of a joint-project is often required.

In order to facilitate the collaboration, a check-list is provided that summarizes the different steps that is recommended to be followed for a successful collaboration (Annex 1).
Chapter 2  5

2 Sharing of primary data via “Black box model”

Using primary data from a specific supplier for a raw material is a good option when data available in database are not representative or when a very detailed study is performed, for which a specific raw material has a decisive impact. The company requesting data is often willing to have access to full LCA results and wishes to have a model that can be used in its own LCA software. The recipient is also interested to understand the origin of the results i.e. contribution of precursors, electricity etc. However, for the reasons mentioned above, those kinds of detailed information are seldom delivered and instead of it, LCA data are exchanged via “black box” models. Black box datasets represent an aggregated form of LCA results. Black box models are a good solution for sharing data without risking of sharing confidential data about processes.

A black box dataset means that only a Life Cycle Inventory (LCI) of the whole process is provided without giving detailed data about the different process steps. This is especially useful when there are several process steps so that no data will be delivered for a single process. Black box data are very interesting because they can be shared in several formats such as gbx, Ecospold or ILCD and directly used in a software by the data recipient. An advantage is also that all categories can be assessed and a certain level of information is available i.e. a LCI gives information about the kind of flows emitted this can be useful to understand and interpret results.

On the other side, exchanging black box datasets is not always possible, especially when there are few process steps or when some inventory flows can be directly linked to a process step or precursor.

Example: Company A collaborated with company B and requires LCA data for a bio-based product P produced by mean of fermentation (Figure 1).

![Figure 1: Simplified production process of product P](image)

During the fermentation process, sugar is the main feedstock used for producing the product P. Emissions occur during the fermentation process i.e. biogenic carbon dioxide emissions. The quantity of sugar required for the fermentation process can be re-calculated based on LCI results (see Figure 2). Actually, in the LCI, the quantity of “CO2e” taken up from the atmosphere by sugar is provided (input of carbon dioxide [Renewable resources]). Knowing the quantity of “CO2” embedded in 1 kg of sugar (theoretical value), the quantity of sugar used for the process can easily be re-calculated. On the other side,
emissions occurring during the fermentation can also be read from the LCI of the whole production process because most of the emissions are occurring during the fermentation process step (output of carbon dioxide biotic [inorganic emissions to air]). Having this information, it is possible to re-calculate the yield of the process and relate the quantity of the main feedstock used to some cost information. In this case, sharing a black box dataset is not wished by the company owning data and only fully aggregated results will be provided i.e. GWP, AP, EP etc. instead of LCI results (see Table 1).

![Table of LCI results](image)

Figure 2: Exemplary LCI results for the product P modelled in the GaBi software (thinkstep 2016)

By providing only results in some impact categories, the data recipient will have a limited usage of the data but confidentiality is ensured. However, it reduces a lot the transparency of the collaboration and the user has few possibilities to interpret results.
Table 1: Exemplary results of the product P in some selected impact categories

<table>
<thead>
<tr>
<th>Impact category</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWP incl. biog C incl. LUC</td>
<td>0.4 kg CO$_2$e/kg</td>
</tr>
<tr>
<td>AP</td>
<td>0.016 kg SO$_2$e/kg</td>
</tr>
<tr>
<td>EP</td>
<td>0.008 kg PO$_4$e/kg</td>
</tr>
<tr>
<td>POCP</td>
<td>5E-4 kg Ethene-e/kg</td>
</tr>
<tr>
<td>PED</td>
<td>90 MJ/kg</td>
</tr>
</tbody>
</table>
Chapter 3  

3 Data exchange via industry averages/collaboration projects

The provision of industry average data sets is a commonly used solution in data management, in most cases coordinated directly by industry associations. The member companies have the common goal to collect sustainability data.

Several industrial organisations have been collecting and providing LCI data already for years. The format of the data collection and provision differs but the key characteristics of industrial average data sets are in general very similar.

Industry average data sets are a good and widely used solution to overcome confidentiality issues as the company specific data is only handled by the industry association as an independent third party.

3.1 Data collection

The data collection for industrial average LCA data is in most cases done on the inventory level. So that only input and output data for process steps at gate-to-gate level are collected.

Via implementing these data in LCA models and connecting them with upstream data sets, cradle-to-gate analyses are possible. In some cases also information about other life cycle stages such as the use or end-of-life phase are collected and included in the later on published data sets.

As information of several companies is collected detailed data checks are enabled. Via simple comparison among the data of the participating companies most reporting errors can be easily identified as they are most likely not in the same range with the data of other companies. Furthermore, based on the knowledge of process experts who are active in such a data collection project, industrial associations can also analyse the technological reasonableness of data which is a huge advantage (World Steel Association 2011).

This ensures high quality data which is relevant and representative for the products.

In the case where an agreed model and final impact assessment results are already available, it is also possible to collect data directly on impact assessment level. This is usually done only for very specific case studies and not for data provision purposes.

3.2 Representativity of industry averages

The use of industry averages has many advantages as the market mix reflects the current production technology of several producers. This way it remains valid for a longer time and for a broader field of application than single producer data. Instead of using
company specific data the industry average is representative and helpful for many studies. It reflects the reality in which one material is often bought from several suppliers as otherwise a high dependency on one producer would be induced. The average LCA results allow staying representative with the real market conditions, instead of working only with one supplier in case of material and LCA supply which would require a change in the LCA every time a new supplier is used. The use of industrial averages for defined regions is seen as "the most open and consistent means of assessing improvements in “sustainability”" (FEVE 2016).

If high quality data is needed it is important to rely on process experts who know how the technology works in reality. This accurate and up-to-date knowledge is given at industrial organisations which collect data; therefore these data are technologically representative.

### 3.3 Industry averages: horizontal vs. vertical

In most cases industrial average data are aggregated to the product level so that no data for separate process steps are supplied. This is also called vertical aggregation (Figure 3). This approach is in general preferred because several companies with slightly different process step setups can collaborate and supply common data for a product that is made by all of them.

![Figure 3: Vertical aggregation and horizontal averaging of LCI data for several companies (World Steel Association 2011).](image)

However, horizontal averaging has advantages as well. It is helpful for benchmarking of process steps because the providing company can easily identify how its process performs in comparison to the industry average. With vertically aggregated data sets this is not easily done as only data for the whole process of the product is published, a process
step which performs worse than the average cannot always be identified. Therefore the contribution of each process step and possibilities for improvement are more likely to be identified by horizontal comparisons (EAA 2013).

It is questionable if the need for more data aggregation to ensure confidentiality is fulfilled by horizontal averaging as the process know-how and cost-sensitive data could still be obvious in an industry average. For companies participating in the data provision the vertical aggregation ensures the highest confidentiality as all inputs and outputs are summed up for the whole system.

In case that the setups of one process step differ heavily between the participating companies the horizontal averaging is challenging might not be fully representative. For example if two process steps are connected at one production site but split at another, the vertical aggregation would allow to calculate an average for the product made via both production routes. The calculation of a horizontal average would require an allocation and a virtual split of the process steps.

That is why using the vertical aggregation approach a data set representative for the market mix of the production can be prepared.

From a data recipient’s perspective the complexity of the horizontal averaging approach has to be kept in mind.

Even if the first impression might be to receive more and detailed data with the horizontal averaging approach, the use of these data requires high expert knowledge. In the case where no strict process boundaries exist in reality, for example for integrated processes which use by-products of each other or have a shared energy production and consumption, complex modelling approaches are required. Although solutions for a virtual separation can be found, the problem that the process steps have to be reconnected in a technically realistic and logical way remains. This connection is in case of horizontal averaging the task of the data recipient who is seldom an expert for the production technology.

A vertically aggregated data set for the whole production of a product at one production site already incorporates these linkages and connections due to the wider system boundaries.

Furthermore a model which takes into account integrated process chains is seen as the most realistic and representative solution. In most industrial situations a system consisting of many process steps is optimised as a whole and not only with the focus on especially one process step. Such a limitation could lead to unintended burden shifting. Supplying data for separate process steps shifts the holistic production cycle perspective to a single focus point.
3.4 LCI models

In most cases the collected data are imported into a specific LCA model. By using only one model with agreed underlying modelling choices, the modelling uncertainty can be effectively reduced. An agreement for using the same assumptions (such as allocation approaches) and data sets for upstream data can be found. This way the best approach is discussed within industry and not decided differently by each company.

Even if the approach chosen cannot perfectly reflect reality, it is very helpful to agree on common modelling assumptions. In this case it is ensured that all players use the same model and the respective outcomes are based on the same decisions. Although the results do not get absolutely more accurate, they get more comparable and the modelling uncertainty is reduced.

Table 2: advantages and disadvantages of industry average based LCI models

<table>
<thead>
<tr>
<th></th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data provider</strong></td>
<td>Confidentiality can be ensured via industry averages</td>
<td>Data will be collected for most important processes. These might not be the most relevant processes for each individual producer.</td>
</tr>
<tr>
<td></td>
<td>Intense data checks ensure higher data quality.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The environmental performance of a specific producer is hidden in the average. This can be advantageous or disadvantageous for a specific data provider.</td>
<td></td>
</tr>
<tr>
<td><strong>Data recipient</strong></td>
<td>Intense data checks ensure higher data quality.</td>
<td>Data interpretation depends strongly on the quality of data documentation</td>
</tr>
<tr>
<td></td>
<td>Data representative for the market mix, no dependence on one single producer.</td>
<td>No possibility to assess the individual environmental performance of specific producers.</td>
</tr>
<tr>
<td></td>
<td>Data is very representative as it is based on production experts’ knowledge.</td>
<td>No possibility to see or change modelling choices directly in the data set</td>
</tr>
<tr>
<td></td>
<td>Process steps are in most cases linked correctly, data can be easily implemented.</td>
<td>No possibility to analyse the impact of changes in process modification</td>
</tr>
<tr>
<td></td>
<td>Data is based on agreed methodology, so that the modelling uncertainty is reduced</td>
<td></td>
</tr>
</tbody>
</table>
4 Data documentation in B2C collaboration

Having a well-documented dataset is of great support when data exchanged are intended to be used for decision-making. To publish LCA data via different platforms such as the Life Cycle Data Network from the Joint Research Center, datasets have to be compliant with quality requirements aimed at guaranteeing datasets quality and coherence in terms of methodology, documentation and nomenclature. When data are exchanged between companies, no rules concerning the documentation exist and this point is often identified as a major bottleneck of the communication. Providing life cycle results without having a minimum extend of documentation presents some risks:

- Risk of misunderstanding of results and wrong communication over the value-chain
- Risk of inconsistency: when no information is provided about i.e. allocation approach or use of credits, a major risk of double counting exist.

Moreover, the recipient of the LCA data might face some major difficulties for interpreting its results and communicate further its own results along the value-chain when some major background information is missing.

In this section, a proposal is made for a pragmatic documentation of LCA data exchanged at the B2B level based on the experience of industrial LCA practitioners of the MEASURE project team.

4.1 Data quality

In the WBCSD guidance for the Chemical sector (WBCSD 2014), indicators are recommended for assessing the quality of the data used for the modelling (Figure 4):

- Reliability or parameter uncertainty
- Completeness
- Time representativeness
- Geographical representativeness
- Technological representativeness.

For those indicators, the quality is assessed based on a scoring system according to data quality ratings from EU PEF and the Pedigree Matrix from the ecoinvent database version 3 (from 1 to 5, with 1 representing the highest score and 5 the lowest). It is recommended to perform the assessment for each process unit with a significant contribution (>10 %) to at least one environmental impact and to aggregate results at the life cycle stage level.

This approach for assessing data quality was discussed within the MEASURE consortium and judged very suitable for the documentation of LCA data exchanged at the B2B level and adapted to the different sectors of process industries (i.e. not specific to the chemical industry).
On top of this quantitative rating, qualitative additional information concerning the quality of the impacts categories assessed might also be provided i.e. possible deviation of the data quality in the different impact categories.

Example: "The POCP impact category has a lower quality because VOC emissions could not be assessed based on primary data, literature data have been used."

4.2 Methodological challenges and choices (qualitative information)

In the analysis of sectors A, B and C, several methodological issues have been identified that are specific to each of those sectors (section 3.2 of the MEASURE Roadmap and background documents “Sector report: chemistry and FMCG”, “Sector report: metal"
and automotive” and “Sector report: waste”). When data are exchanged over the supply chain, these methodological challenges require additional information about the way they have been handled. Consequently, the data that shall be provided are sector specific. The following Table 3 provide some examples of the issues that shall be described for at least the three sectors investigated within the MEASURE project:

Table 3: Qualitative information concerning the methodological issues

<table>
<thead>
<tr>
<th>Methodological issues</th>
<th>Qualitative information to provide</th>
<th>Sector concerned</th>
</tr>
</thead>
<tbody>
<tr>
<td>By-product</td>
<td>Are by-products obtained and how have they been handled (i.e. allocation, system expansion)?</td>
<td>All</td>
</tr>
<tr>
<td>Allocation</td>
<td>At which level was an allocation applied? Which method?</td>
<td>All</td>
</tr>
<tr>
<td>Credit</td>
<td>Was any credit applied?</td>
<td>All</td>
</tr>
<tr>
<td>Land Use Change</td>
<td>How were the emissions from Land Use Change estimated (i.e. method used for averaging, data sources for carbon stock, carbon pools considered, etc.)?</td>
<td>Sector A</td>
</tr>
<tr>
<td>Biogenic carbon</td>
<td>How was the biogenic carbon considered in the analysis? Was the carbon balance checked?</td>
<td>Sector A</td>
</tr>
<tr>
<td>End of Life</td>
<td>Which method was used for accounting for the recycling at the end of life (e.g. credit, recycling rate considered, recycled content)</td>
<td>Sector B, all sector generating wastes</td>
</tr>
<tr>
<td>End of Life</td>
<td>Modelling of long term emissions from the waste treatment: which approach has been used?</td>
<td>Sector C, all sectors generating wastes</td>
</tr>
<tr>
<td>End of Life</td>
<td>Modelling of materials or energy generated during the waste treatment: which replacement options have been used?</td>
<td>Sector C, all sectors generating wastes</td>
</tr>
</tbody>
</table>
Those critical issues are often a major source of uncertainty. For this reason, a qualitative uncertainty assessment shall be given by providing information about how the choice of the approach influences the overall results. Providing a range of results can in some situation be a good approach.

4.3 Quantitative (or semi quantitative) information about results

A major bottleneck for interpreting results provided by a supplier remains in the fact that data are often provided as a black box what makes the interpretation or the comparison of these primary data with some data from database or from industrial associations almost impossible. Nevertheless, providing quantitative data is possible at an aggregated level, without risking delivering confidential information.

First of all, the MEASURE project team recommends to provide allocation factors used when the choice of the allocation approach has a significant impact on the final results (e.g. >20%).

In order to interpret results obtained, the contribution of the different process parts for the most important impact categories shall be provided (the choice of the target categories shall be agreed between the two companies collaborating):

- Contribution of precursors (Upstream)
- Contribution of energy, utilities (electricity, steam, thermal energy, process water, compressed air etc.) and own process (process emissions) (gate-to-gate)
- Contribution of the waste treatment (solid and liquid wastes, waste water)

In order to respect the confidentiality of the underlying data, the contribution shall be expressed in the following way as a range in percentage (Table 4) and aggregated for upstream, gate-to-gate and downstream processes as described above.

Table 4: Rating of the contribution

<table>
<thead>
<tr>
<th>Contribution</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fundamental</td>
<td>&gt; 80%</td>
</tr>
<tr>
<td>Major</td>
<td>50-80%</td>
</tr>
<tr>
<td>Significant</td>
<td>25-50%</td>
</tr>
<tr>
<td>Low</td>
<td>&lt; 25%</td>
</tr>
</tbody>
</table>

Table 5 is an example of a summarized data documentation that shall be provided together with LCA results.
Table 5: Example of data documentation for the LCA of the product P

<table>
<thead>
<tr>
<th>Data quality</th>
<th>Indicator</th>
<th>Score</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reliability</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Completeness</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time representativeness</td>
<td>1</td>
<td>Production data from 2014 have been used for modelling</td>
</tr>
<tr>
<td></td>
<td>Geographical representativeness</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Technological representativeness</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Methodological choices

<table>
<thead>
<tr>
<th>By-products and allocation</th>
<th>No by-products obtained in any process steps. No allocation applied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land Use Change (LUC)</td>
<td>CO₂ emissions from LUC for Soybean oil have been estimated with the “Land use Change Assessment tool” from the PAS 2050-1, for the country Brazil. For more information about method and data source, please refer to PAS 2050-1 (BSI 2012)</td>
</tr>
<tr>
<td>Biogenic carbon</td>
<td>Biogenic carbon has been considered in this analysis. The balance has been checked and the biogenic carbon content embedded in the product P corresponds to its stoichiometry (1,9 kg CO₂/kg)</td>
</tr>
</tbody>
</table>

Results (quantitative information) for GWP and Acidification Potential

<table>
<thead>
<tr>
<th>Process part</th>
<th>Contribution</th>
<th>Comment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Upstream</td>
<td>Fundamental</td>
<td>Bio-based feedstock and emissions from LUC have a fundamental impact on the GWP</td>
<td></td>
</tr>
<tr>
<td>Gate-to-Gate</td>
<td>Low</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Downstream</td>
<td>Low</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5 Difficulties currently not solved

The communication and interpretation of LCA results remains a major challenge in cross-sectorial collaboration. The results for different impact categories can have significantly different relevance and importance for each sector. In case of data sharing, an agreed approach between data provider and recipient regarding the most important impact categories should be sought to allow a correct and ideal data usage.

Apart from that, a harmonised way of communicating especially the uncertainty of results has not yet been found. During the MEASURE workshop in Mechelen, the communication of ranges of results was seen as a pragmatic solution for the near future.

This could be either done visually as in Figure 5 or in a written way. For the example shown above this would mean that instead of communicating an average result (i.e. 2.1 kg CO₂e), the whole range (-1.3 to 3.2 kg CO₂e) would be provided.

However, such results cannot be understood and interpreted easily by data users. Difficulties occur especially when a decision shall be based on the results.

While communicating different scenarios and the corresponding range of results can help to understand the impact, relevance and reliability of the data, such a wide and open communication is especially complex to interpret for people who are not deeply involved in the topic.

Nevertheless, in the workshop discussion, it was expressed that already acknowledging the fact that uncertainty exists can be the first step to a more informed interpretation of LCA data.

One possibility to assess and visualise the impact of specific value choices is to analyse the sensitivity of results so that it is at least clear that the results should not be seen as concrete and definitive numbers.
Bearing in mind the uncertainty of data, product comparisons based on LCA data are especially problematic as there is currently no generally accepted definition which differences for impact results can be called significant. If the impact assessment results of 2 products differ for example by 20 %, it could be that the uncertainty is much higher than 20 %, so that such a difference cannot be considered relevant.

More guidance is needed how to assess uncertainty and how to deal with it in case of product comparisons. The only possibility which is readily available is to agree on assumptions in, but also between sectors to reduce the uncertainty in modelling and by this also the complexity of results.

A short analysis and survey was done to evaluate the degree of penetration and adoption of the WBCSD Guidance Document ‘Life Cycle Metrics for Chemical Products’ (WBCSD 2014) amongst attendees of the MEASURE London WS and in our personal networks. There were around 10 expert respondents, distributed ~50/50 from the chemical sector versus other sectors. The results indicate that in the Chemical sector (A), the awareness of the report is reasonably good (> 50%) and that most of the guidance, examples and recommendations are deemed relevant and useful. Nevertheless, some discussion remains, mainly around choices of impacts and impact methodologies. More verbal comments came from the non-chemical sector, with some concerns around e.g. scope (cradle-to-gate vs cradle-to-grave), impact methods, allocation and end of life/recycling. Awareness in other sectors was low. It is therefore not recommended to promote the use of the WBCSD guideline outside sector A, other than ‘for information’.

Respondents who were aware of the Guideline emphasized that it should not be the intention to claim ‘compliance’ of LCA studies with this WBCSD guideline. As the name indicates, it is not a new methodology but provides guidance and recommendations. The aim was to establish a common understanding within the chemical industry on how to interpret and apply existing general LCA principles and rules. In that, the WBCSD guidance is more precise than e.g. the ISO 14040 series.
7 Abbreviations

AP  Acidification Potential
B2B  Business to Business
B2C  Business to Consumer
EP  Eutrophication Potential
GWP  Global Warming Potential
LCA  Life Cycle Assessment
LCI  Life Cycle Inventory
LUC  Land Use Change
PED  Primary Energy Demand
PEF  Product Environment Footprint
POCP  Photochemical Ozone Creation Potential
VOC  Volatile Organic Compound
WBCSD  World Business Council for Sustainable Development
8 References


9 Annex 1

Checklist: Industrial Cooperation for sustainability assessments along the value-chain

1. Preparation
   - Identify how and where connections/contact with the other companies in the supply chain can be established, i.e., can it be done via 1-1 meetings, via industry associations, via meetings with regulators, scientific events, etc.
   - Clearly define and articulate the mutual and long term category benefit, on top of interests of individual companies
   - Identify which (commercial) confidentiality and organizational conditions need to be met in order to proceed, taking into account competition law and other relevant regulations. Prepare confidentiality agreements if required.

2. Planning/organisation of the work, technical aspects
   - Define a timeline
   - Define which data are required by the recipient and which data the supplier is able to share
   - Define what the “target” categories of the study are (are other topic such as water footprint relevant?)
   - Agree on a format and the level of aggregation for data exchange:
     - Mass and energy balance
     - Life Cycle Inventory
     - Impact assessment results
     - Specific format such as gbx, Ecospold, excel.
   - Check if relevant standards and sectorial agreements are already available, based on these agree how methodological issues are intended to be handled: allocation, modelling of end of life, issues related to Land Use Change etc.

3. Data review, summarization, communication, documentation
   - Present results in a face to face meeting with both industry parties (if possible) to enable best understanding of the data and their usability:
     - Present the model and the underlying modelling choices
     - Present the major assumptions
     - Include information concerning data uncertainty
   - Exchange data in the agreed format
   - Prepare a suitable data documentation (see chapter 4)