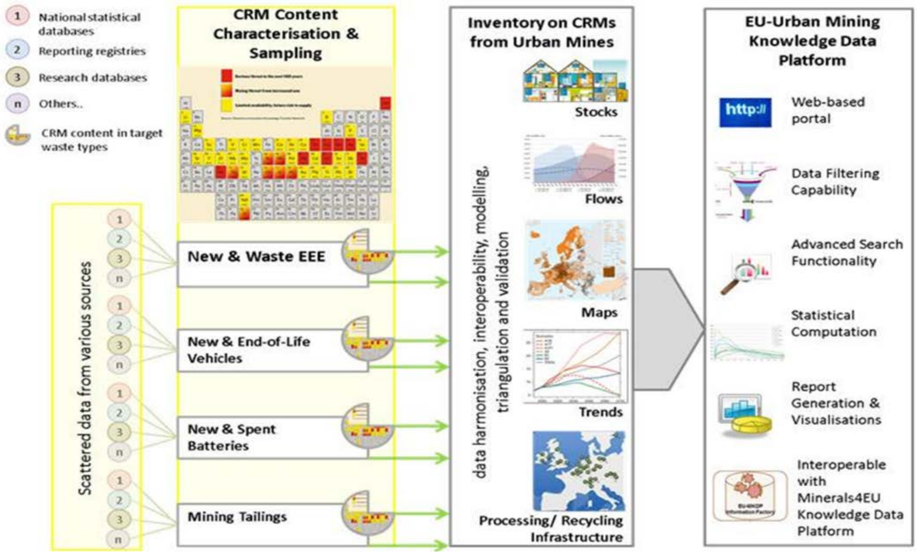
Protocols on CRM Product and Component Content and Quality Assessment
Deliverable D2.7



Project	Prospecting Secondary raw materials in the Urban mine and Mining waste
Acronym	ProSUM
Grant Agreement	641999
Funding Scheme	Horizon 2020
Webpage	http://www.prosumproject.eu/
Work Package	Work Package 2
Work Package Leader	Patrick Wäger
Deliverable Title	Protocols on CRM Product and Component Content and Quality Assessment
Deliverable Number	D2.7
Deliverable Leader	Amund N. Løvik
Version:	2.2
Status	Final
Authors	Amund N. Løvik (Empa), Claude Chanson (RECHARGE), Arthur Haarman (Empa), Jaco Huisman (UNU), Maria Ljunggren Söderman (Chalmers), Paul Mählitz (TUB), Matthias Rösslein (Empa), Patrick Wäger (Empa)
Reviewed by	Project Management Team
E-mail	amund.loevik@empa.ch
Scientific Coordinators	Dr Jaco Huisman, Scientific Advisor, UNU – IAS SCYCLE Nikolaos Arvanitidis, Chair of the EGS Mineral Resources Expert Group Huisman@unu.edu
E-mail	Nikolaos.Arvanitidis@sgu.se

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 641999.

DOCUMENT CONTROL

Project Coordinator:	WEEE Forum
Work Package Leader:	Empa
Deliverable leader	Empa
Due date:	M34 31 October 2017
Date of submission:	M34 13 October 2017
Dissemination level:	PU (Public)

Version history			
Ver. no.	**Date**	**Reason for release**	**Responsible**
1.0	22.6.2017	TOC	AL
1.1	26.6.2017	revised TOC after meeting	AL
1.2	8.9.2017	wrote text in most subchapters, including BATT text from CC	AL, CC
2.0	21.9.2017	Added text from JH, revisions based on comments from PW and MLS. Ready for PMT review.	AL, JH, MLS, PW
2.1		Updated with PMT input	SD
2.2	4.10.2017	Final revisions	AL

NOTICE

TABLE OF CONTENTS

Contents

List of figures

PURPOSE

This report documents the activities undertaken in Task 2.4 to produce update protocols and quality assessment procedures for composition data for products placed on the market.

The primary purpose of the report is to document the processes and data templates used to work with product composition data in ProSUM, to ensure that future work adopts the same harmonised approaches, and data is, therefore, comparable. A secondary purpose is to provide recommendations to the scientific community and others who publish product composition data on how to represent data in a way that facilitates sharing and inclusion in the EU Urban Mine Knowledge Data Platform (EU-UMKDP).

EXECUTIVE SUMMARY

Protocols

Product compositions change rapidly. It is therefore essential for the future relevance of the EU-UMKDP that the data on representative compositions are updated to include the best available information. It is proposed that the composition data be thoroughly updated once every 3-4 years.

Future updates to the composition data in the EU-UMKDP will need to go through the following 8 steps:

1. Initial control
2. Record raw data
3. Data quality (DQ) assessment of raw data
4. Transfer to consolidation files
5. Re-consolidation
6. DQ assessment of consolidated data
7. Transfer to portrayals
8. Harvest to UMKDP

In step 1, an initial quick assessment should be made to decide whether the data is to be further processed at all. Following this, in step 2, relevant data and metadata is recorded in the so-called "CRM parameter templates" Excel sheets. These are used to store raw data for each of the three product categories, and the original source documents are then stored together with their metadata in the UMKDP. In step 3, a data quality level is assigned to each datum based on the metadata recorded in the CRM parameter templates. Following this, in step 4, the data must be transferred to the "consolidation files", which are then in step 5 used to compare data from different sources, compute averages, calculate overall product compositions, and perform manual checking of data. After the new representative data have been generated in the consolidation files, an assessment of their data quality and uncertainty is conducted in step 6. When the consolidation is completed, the new data must be transferred to portrayals (step 7) from which they can be harvested to the UMKDP (step 8). If the parameters in question had been estimated before, and the update is merely an improved estimate, the old data in the UMKDP will be overwritten.

Recommendations for future work

A large amount of data on the composition of EEE, batteries and vehicles have been collected, harmonized and consolidated within ProSUM. For all the steps required, the general approach was to first define the necessary procedures (e.g. code lists, templates for recording data), and then implement it in practice for the available data. Each step was automated as far as possible, e.g. by writing scripts to transfer the data from one template to another. Initially, it was thought that a substantial part of the consolidation work could be automated. This proved to be difficult, amongst other things, due to the need to identify errors in the raw data and to choose a reasonable aggregation level when combining data from different sources. The consolidation of composition data is therefore a time-consuming process that requires a certain level of technical understanding and experience with the topic of critical raw materials.

We have produced the most extensive dataset on secondary raw materials contained in EEE, batteries and vehicles to date, which can be used to estimate the total stocks and flows of a large number of metals contained in these products. Nevertheless, there is room for improvements, firstly to make the consolidation process more efficient, and secondly to produce datasets with a higher level of detail in terms of materials and components. We therefore propose a set of recommendations for future work with composition data for the EU-UMKDP. The

upcoming Horizon 2020 project ORAMA, which is focused on procedures to generate data on primary and secondary raw materials, may offer an opportunity to develop and test some of the recommended measures.

1. *Continue improving the code lists for components and materials.* Due to the complexity of the products covered in ProSUM and the number of different representations found in data sources, it was not possible to provide complete detailed definitions and correspondence tables for all components and materials. In the future, these code lists should be improved by refining the definitions, e.g. including pictures of important components.
2. *Revise the data quality assessment procedure.* The data quality assessment procedure may be improved in the future by incorporating some of the findings from D2.6/D4.3 regarding methods for sampling, sample preparation and chemical analysis. This would enable identification of data with systematic errors e.g. due to the use of wrong digestion acid. Amendments to the data quality assessment procedure may nevertheless be limited by lack of detailed information in the original data sources.
3. *Introduce statistical data reconciliation for consolidating composition data*. The data consolidation process (calculation of representative compositions) may be improved through statistical data reconciliation. The current approach leads to loss of information as data from different sources are aggregated to a high level in terms of components and materials. Statistical data reconciliation could mitigate some of this information loss.

DELIVERABLE REPORT

1. Introduction

1.1. Aim and scope of the Deliverable

Within Task 2.2, a large number of data on the composition of products, components and materials were collected, recorded, analysed and consolidated. The result of the task was three datasets of “representative” compositions for all the product keys covered in ProSUM, one for each product category: electrical and electronic equipment (EEE), batteries (BATT) and vehicles.

Product compositions change rapidly. It is therefore essential for the future relevance of the EU-UMKDP that the data on representative compositions are updated to include the best available information. In this deliverable report, we explain how to include new data in the EU-UMKDP in the future through a detailed documentation of how to: i) record the data; ii) assess data quality; and iii) estimate representative compositions. Whilst the technical and mathematical aspects of these procedures were largely covered in D2.2, the present report focuses on the practical aspects, such as how to use the Excel templates for recording data. We also provide product-specific and general recommendations for the continued work with composition data for the EU-UMKDP, as well as recommendations for the scientific community and other data providers on how to represent composition data to facilitate their inclusion in the EU-UMKDP.

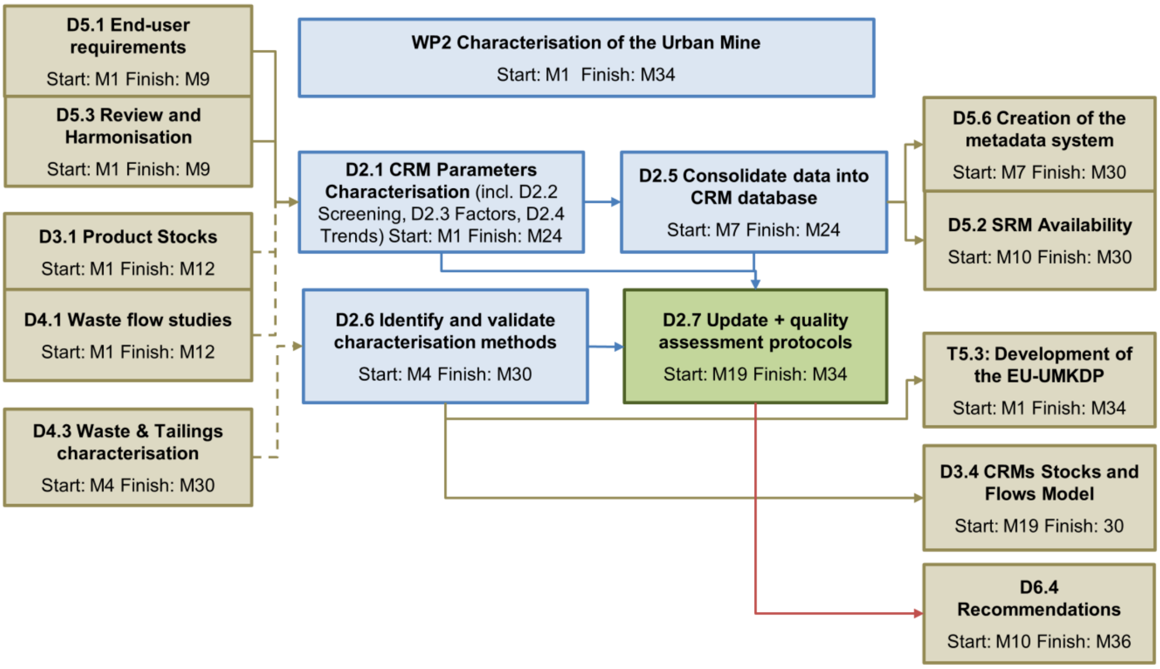


Figure 1 Pert chart showing the relationship between D2.7 and other ProSUM deliverables.

Figure 1 gives an overview of the deliverables in WP2 and their connections to other ProSUM deliverables. D2.7 is highlighted in green. The update protocols and quality assessment procedures presented here were, to a large extent, developed during the work with D2.5 (Løvik et al. 2017), with updated common project procedures updated since then, and including the assessment of data quality. D2.6 on characterization methods provides valuable recommendations on the generation of primary data from sampling and characterization of products and wastes.

1.2. Overview of update protocols and quality assessment

The UMKDP contains estimated representative compositions for a large number of EEE, batteries and vehicles placed on the market, to some extent also including changes over time. These representative compositions have been estimated by consolidation of data from many different sources. Procedures for evaluation of data quality and for calculating representative compositions were developed as part of D2.5 "consolidation of data into CRM database" (Løvik et al. 2017). Although these procedures facilitate the consolidation and ensure consistency between the product groups, they do not remove the need for manual work and decision-making during data consolidation. As an example, the components and materials used to describe the composition of vehicles were selected based on the availability of data and the importance for CRM content, and it was necessary to group all electrical and electronic components in one component group called electrical and electronic system. When more detailed data becomes available, a different set of components might provide a better representation.

Furthermore, the data quality assessment performed within ProSUM cannot capture all relevant information: closer examination of the methods used to generate the primary data may sometimes reveal a lack of representativeness or potential errors. Often, such problems are identified through comparing data from different sources. For these reasons, it is clear that the inclusion of additional data in the future cannot be automated to the level where one only needs to enter the data in the right format in a template. This is a necessary step, but future updates will also involve a substantial amount of hands-on work and critical reflection on the data and calculations.

An overview of the procedure for including new data in the UMKDP is shown in Figure 2. New data will first be subject to an objective quality assessment and, if meeting a certain quality criterion, be recorded in the existing datasets for composition raw data that are used internally in ProSUM (grey box). After this, re-consolidation of data can be performed to update the representative compositions in the UMKDP. This re-consolidation may be as simple as adding new data in cases where there was none already existing in the UMKDP, or it may involve a reconsideration of all data used in the original consolidation together with the newly recorded raw data. An example of the former would be to include battery electronics as a component of batteries, which was not done in the original dataset. If the data becomes available, this is a relatively simple addition to make. An example of the latter would be if more detailed data becomes available for vehicle electronics: it will then be desirable to perform a complete reconsolidation with the presently existing raw data.

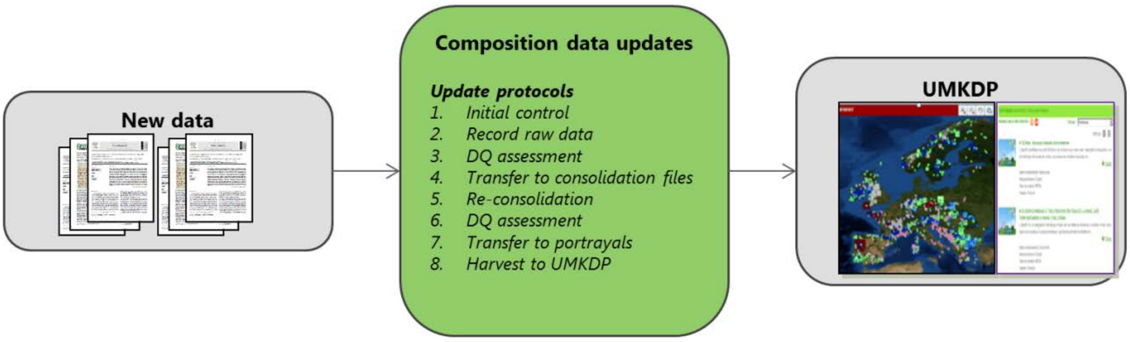


Figure 2 Overview of procedure for including new product composition data in the EU-UMKDP.

Chapter 2, explains how to include new composition data in the UMKDP, and chapter 3 presents recommendations for working with composition data based on the experiences in ProSUM. Chapter 2 will be important for anyone continuing the work of ProSUM in the future. For anyone who wishes to supply data to the UMKDP, chapter 2.2 will be important, as it explains how to record raw data in a format that facilitates consolidation. Chapter 3 is relevant for anyone working directly with composition data for batteries, EEE or vehicles, be it in connection with the UMKDP or in other projects.

2. Update protocols

2.1. Update protocols detailed process overview

A detailed overview of the procedure to include new data is shown in Figure 3. New datasets can come in many different formats and sizes. The majority of data sources used in ProSUM have been journal articles or research reports available as pdfs. In step 1, an initial quick assessment should be made to decide whether the data is to be further processed at all. Following this, in step 2, relevant data and metadata is recorded in the so-called “CRM parameter templates” (Excel sheets used to store raw data for each of the three product categories, and detail the original source documents which are then stored together with their metadata in the UMKDP). In step 3, a data quality level is assigned to each datum based on the metadata recorded in the CRM parameter templates. Following this, in step 4, the data must be transferred to the “consolidation files”, which are then in step 5 used to compare data from different sources, compute averages, calculate overall product compositions, and perform manual checking of data.

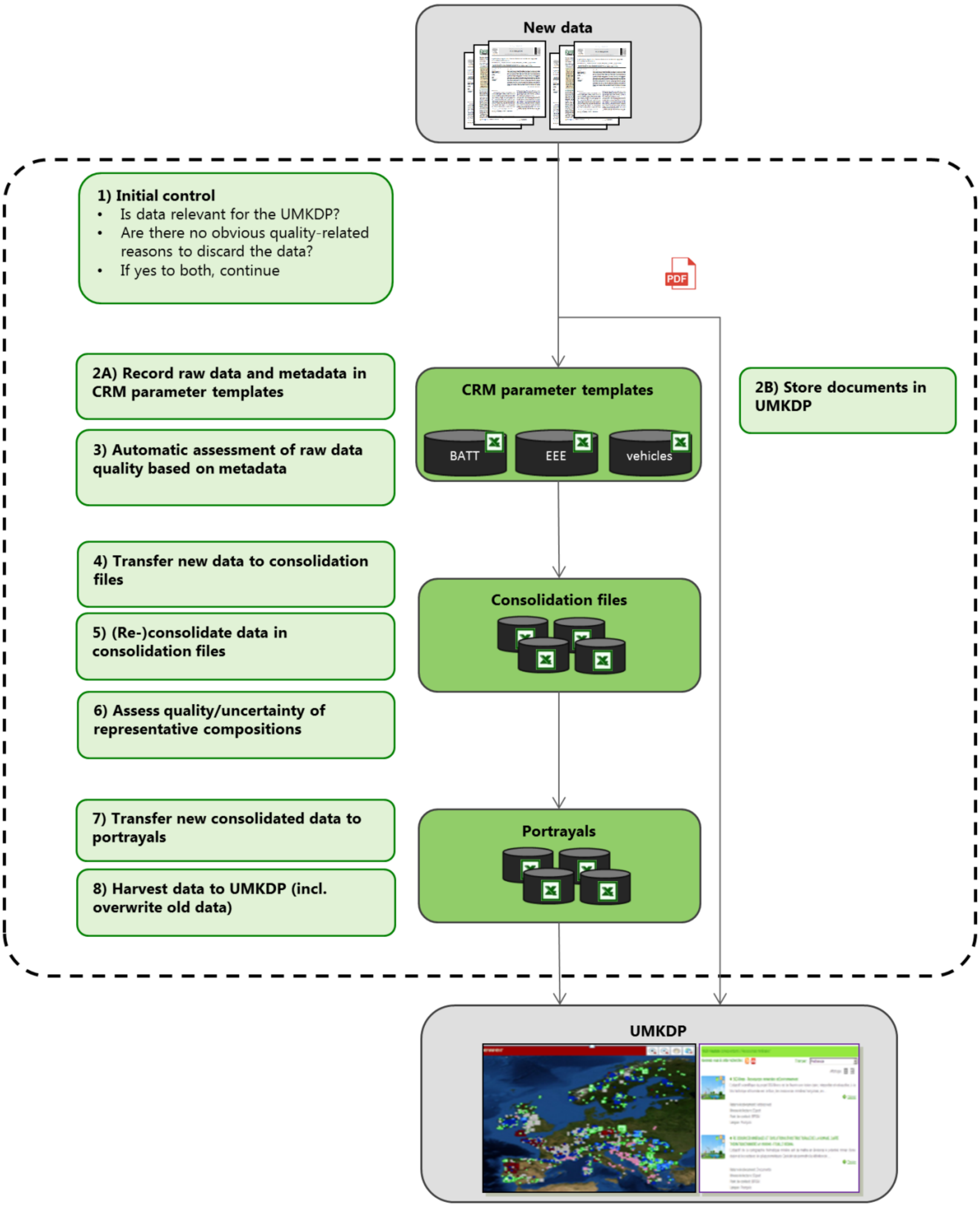


Figure 3 Detailed overview of procedures to include new data in the UMKDP.

After the new representative data have been generated in the consolidation files, an assessment of their data quality and uncertainty is conducted in step 6. When the consolidation is completed, the new data must be transferred to portrayals (step 7) from which they can be harvested to the UMKDP (step 8). If the parameters in question had been estimated before, and the update is merely an improved estimate, the old data in the UMKDP will be overwritten.

2.2. Initial control of new data

An initial control of new data should always be performed to avoid unnecessary work and cluttering of the CRM parameter template with irrelevant or dubious data. The decision will be somewhat subjective, but should be guided by the following questions:

Are the data addressing products within the scope of the UMKDP?

In principle, the data should refer to a product within one of the three categories: batteries, EEE and vehicles and it should be possible to allocate it to one of the product keys in the respective code lists. There are exceptions, such as standard material compositions that may be needed for consolidation although they cannot be allocated to a specific product.

Are the data useful for calculating compositions?

For example, data may not need to be recorded if they are very aggregated compared to the existing datasets. Still, it should be considered that they may be useful for checking and controlling of results even if they are not used directly in the consolidation.

Are there obvious reasons to doubt the validity of the data?

Data that are clearly wrong should not be recorded. There may be information available that reveals mistakes in the data collection, or extreme uncertainties (e.g. when data are based on researchers' guesses or own assumptions). Since the subsequent data quality assessment only captures selected aspects of the data generation process, it is important to filter out any dubious data immediately when such issues are discovered.

2.3. Recording raw data and metadata in the CRM parameter template

After the initial control of the new data to confirm its relevance, it should be recorded together with metadata in the CRM parameter template. In parallel, the document from which the data originates should be stored, along with its metadata, in the digital library of the UMKDP, from which it can be searched and accessed (unless restricted by copyright issues). The process for storing the document and its metadata is described in chapter 3.6 in the report on Task T5.3.1 in ProSUM, part of deliverable D5.7 (Cassard et al. 2016).

The recording of data in the CRM parameter template is done to ensure that all raw data exist in a common format, so that they can easily be extracted and analysed. If somebody outside of ProSUM wishes to supply data to include in the UMKDP, the same template should be used. The empty template is supplied as Annex 1 to this report. It contains 5 sheets:

- "0. Contents" – Contents and instructions for using the template;
- "1. Data" – Where data should be recorded;
- "2. References" – The references to the data sources should be provided in full detail here;
- "Aux 1. Code lists" – Contains all code lists that are relevant for recording composition data; and
- "Aux 2. DQ_settings" – Contains settings for the automatic specification of data quality.

Only sheets 1 and 2 should be modified by the user. In practice, the data for BATT, EEE and vehicles are kept apart in three CRM parameter templates with the same structure.

The sheet "1. Data" contains 44 fields that may be filled for each datum. Each row in the template corresponds to one datum, i.e. one value, such as the mass fraction of gold in a printed circuit board. A list of all fields and instructions for how to fill them are provided in Table 1.The field names correspond to column headings in the template. The second column of Table 1 indicates what type of input is required, i.e. codes from code lists (code), free text strings (txt), or a number (#). The fields which should be completed or filled out depend on the type of data recorded,[1] and what information is available. Some fields are mandatory for any type of data (indicated by an "X" in Table 1), some are required for certain types of data (indicated by the relevant parameter subscripts in Table 1), and some are required for a complete data quality (DQ) assessment. If information for the DQ assessment is missing DQ will still be estimated, but the missing information is treated as the lowest possible category in the calculations.

Table 1 Fields used in the CRM parameter template.

field name	*input type*	*required*	*instructions*
Prod. Id #	#	X	number the separate entries: 1, 1, 1, 1, 2, 2, 2, 3, 3,… to keep track of which measurements were made for the same object. Example: The material composition of an individual car was found, divided into 5 materials. These 5 entries should have the same Prod Id #. Start the numbering from 1 for each new template. Note that all entries must have a product Id even if they do not refer to a particular product.
Product category	code	X	"EEE", "BATT" or "VEHICLE", as applicable to the recorded data
Key	code	X	product code that closest correspond to the product in question, from one of the product key code lists.
Sub-key	code		product sub-key code that closest correspond to the product in question, from one of the product sub-key code lists. Only applicable for EEE and BATT.
Sub-sub-key	code		product sub-sub-key code that closest correspond to the product in question, from one of the product sub-key code lists. Only applicable to EEE.
Description	txt	X	description of product as in the original source
Similarity between description and ProSUM key	code	X	"high", "medium" or "low". This is to account for disparities between definition of product code and object investigated in original study.
# of products in batch	#	DQ	if data originates from batch tests or similar, enter the number of products in the batch
Production year (product)	#		the year of production of the product. This is preferred over model year and design year
Model year (product)	#		the model year of product
Design year (product)	#		the design year of product
Comp. Id #	#	c, c-p, e-c, m-c	ID number for components, (see Prod. Id #). All component data must have a comp. Id #.
Component group	code		component group describing the type of component investigated
Component	code	c, c-p, e-c, m-c	component code that closest correspond to the component in question, from the component code list
Similarity between component description in reference and ProSUM component	code	c-p, e-c, m-c DQ	"high", "medium" or "low". This is to account for disparities between definition of component code and object investigated in original study.
# of components in batch	#	DQ	if data originates from batch tests or similar, enter the number of components in the batch
Production year (component)	#		the year of production of the component. This is preferred over model year and design year
Model year (component)	#		the model year of component
Design year (component)	#		the design year of component
Material type	code		material type code that closest correspond to the material in question
Material	code	e-m, m-c, m-p	material code that closest correspond to the material in question
Similarity between material description in reference and	code	e-m, m-c, m-p, DQ	high", "medium" or "low". This is to account for disparities between definition of component code and object

[1] For a description of the different types of composition data (physical quantities and parameter subscripts), we refer to D2.5.

field name	*input type*	*required*	*instructions*
ProSUM material			investigated in original study.
Element	code	e-m, e-c, e-p	chemical element
Parameter	code	X	indicates what physical quantity was measured, e.g. mass, mass fraction, volume. Refer to the parameter code list
Param. Subscript	code	X	indicates which entities are measured by the parameter, e.g. mass fraction of Ag in a printed circuit board would be assigned e-c while the mass fraction of lead alloy in a vehicle would be assigned m-p
Value	#	X	the observed value of the physical quantity in question
Value type	code		type of statistic: “mean” or “median”, in the case of several observations aggregated
Value lower limit	#		lower limit of confidence interval (type of interval specified later)
Value upper limit	#		upper limit of confidence interval (type of interval specified later)
Value units	code	X	units used for the recorded value
Uncertainty	#		uncertainty
Uncertainty units	code		units of the uncertainty value
Uncertainty or range type	code		type of statistic used to
Modelling method	code	DQ	modelling method used to obtain general data in original study. We mainly distinguish between “hot-spot” and “non-hot-spot”
Digestion method	code		digestion method use for chemical analysis
Measurement method	code	DQ	measurement method, including wet chemical measurement methods as well as producer data
Original measurement	code		yes/no. yes only if the measurement was conducted as part of the work reported in the given document
Measurement year	#		year of measurement
Location	code		country in which the measurement was done
Notes	txt		any other important information, esp. with respect to data quality, should be mentioned here.
Rights	code	X	rights for the document: “Confidential”, “Copyright”, “InternalUseOnly”, “Public”
Reference	#, txt	X	2 columns: reference number and reference name. All references are to be recorded in sheet 2. References. Add a new (and higher) number for each new reference.
Original data source ref.	#, txt		2 columns: reference number and reference name. All references are to be recorded in sheet 2. References. Add a new (and higher) number for each new reference.
Recorded by	txt		name of the person who did the recording

2.4. Assessment of raw data quality

An automatic assessment of data quality will be performed for each datum entered in the CRM parameter template using the information recorded. The data quality assessment takes into account a maximum of five evaluation categories: *sample size*, *temporal scope definition*, *consistency of descriptions*, *modelling method* (only for e-c, e-m, e-p) and *measurement method* (only for e-c, e-m, e-p). In each category, a score is given based on the information recorded in the raw data sheets. These scores are added up to a total data quality score, which is finally used to assign a data quality rating: *highly confident*, *confident*, *less confident*, and *dubious*. The assigned data quality can be found in column AU in sheet “1. Data” in the CRM parameter template. The columns AV to BD contain the numeric scores in the evaluation categories, and should not be edited by the user. For a detailed description of the steps behind the data quality assessment, refer to Deliverable report D2.5, Chapter 2.1.3.

The assessment of data quality is based on a limited amount of information regarding the origin of the data and the interpretation made in ProSUM. This relatively simple procedure was designed so that: 1) it does not require expertise regarding the methods used to generate the data; 2) it is free of subjective judgement, to allow for consistency when conducted by different people; and 3) it is applicable to different types of composition-relevant data. By abiding to these principles, the procedure emphasises consistency at the expense of accuracy. The procedure only allows the defined types of information to be taken into account, and treats the information the same way for all data sources. In practice, more information is nearly always available, and the same type of information may have different relevance depending on the case. Detailed investigations of sampling, sample preparation and measurement methods have been conducted within ProSUM Deliverables D2.6/D4.3 (Rotter et al. 2017, Wäger et al. 2017), for example showing the importance of element/matrix-specific methods for digestion and measurement when conducting wet chemical analysis. As an example, arsenic will tend to evaporate as AsH_3 or As_2H_4 when the sample is not allowed to cool properly after microwave digestion, thus leading to a negative error in the subsequent measurement. Such considerations could be incorporated the data quality assessment, but would require a very large amount of information (which is often not available from the primary sources), and would therefore slow down the consolidation process considerably. Despite this, it could be an option to make certain amendments to the DQ assessment based on the experiences made in D2.6/D4.3. One possible extension would be the addition of a "digestion method vs. element" indicator, as it appears to be a common mistake that elements of interest remain undissolved due to improper selection of digestion acid However, this and other possible extensions would first require a detailed mapping of such pitfalls for all elements.

2.5. Transfer new data to consolidation files

The CRM parameter templates (one per product category) should contain all relevant data for estimating product compositions. However, since this includes different types of data for many different products, components, materials and elements, it is not practical to consolidate the data directly in these files. Rather, the first step of consolidation is to extract the relevant data for the product in question. This has been done differently for each product category.

For BATT, consolidation was largely based on confidential manufacturer's data (which could not even be shared among the ProSUM project partners). Effectively, this means that there is no separate consolidation file for BATT. Depending on the openness of future BATT data, it may be necessary to create such a consolidation file.

For EEE, a very large number of data were recorded in the CRM parameter template. Detailed data was available for most UNU keys both on the m-p/c-p level and e-c level, sometimes including time-dependent data. A consolidation template was therefore created specifically for EEE. Each UNU key has its own file where consolidation is performed. This file has the same structure for all UNU keys. A MATLAB/OCTAVE script was used to extract data from the CRM parameter template to each UNU key consolidation file. The script queries the CRM parameter template (serving as a database), for all data relevant to the particular UNU key and groups this according to the data type (e-c, e-m, m-p, c-p, c-c, m-c). It should be noted that the script is not created to perform subsequent updates in the consolidation files. When new data become available for a given UNU key, there are at present two options to transfer it to the consolidation file:

1. Use the script to re-extract all data for the key in question. This will create a new consolidation file and will thus erase any manual changes that have been made during the first consolidation. This option is therefore recommended only if the new data substantially outweigh the existing data, so that a complete re-consolidation is necessary.
2. Manually transfer the new data to the consolidation file. If done in parallel with recording in the CRM parameter template and the amount of new data is relatively small, this will be quite fast. Since it requires that the new data can be easily identified, it is recommended to do it immediately. A third option to automatically update the consolidation files with new data would be useful but would require a substantial amount of programming and has therefore not been done. It is recommended to create such a routine in the future.

For vehicles, the amount of data which exists in the CRM parameter template is moderate. The consolidation of vehicles data was tailored to the data availability and involves a “hybrid” approach where the number of calculation steps to arrive at *e-p* varies between different components and materials. The approach is less structured than for EEE.[2] For these reasons, the transfer of data from the CRM parameter template to the consolidation sheets was done manually or semi-automatically in Excel. To reduce the “distance” between the raw data and the consolidation, the spreadsheets from the CRM template and the consolidation file were compiled in one Excel file. New data for vehicles must be transferred manually to the consolidation sheets after recording in the CRM parameter template.

2.6. Re-consolidate data in consolidation files

As explained in section 2.5, the structure of the consolidation was different for each product category. The procedures to re-consolidate data therefore also depend on the product category.

2.6.1. BATT

For BATT, reconciliation can be done via one of two routes:

1. Through access to the primary data that was used in the first consolidation, new data can be compared and averaged with the existing raw data. This is the preferred route but requires that the person working on the consolidation has access to the primary data.
2. If access to the original primary data is not possible, and the data reviewer can only see the consolidated (representative) battery compositions, re-consolidation can be undertaken by comparison and averaging between the existing representative battery composition and new data. For components, materials or elements that are currently not quantified, this would be as simple as adding the new parameters to the consolidated data set, given that they are compatible with the existing data. The current data only provides the mass fractions of elements in the battery cells. It is, for example, unproblematic to add more components that do not overlap with the battery cell and more elements can easily be added.

If the new data concern parameters that have already been quantified, a decision must be undertaken how to weight the new data against the existing data. The following formula is recommended:

$$\bar{x}_{new} = \frac{n_{old}\bar{x}_{old}DQ_{old} + \sum_i x_i DQ_i}{n_{old}DQ_{old} + \sum_i DQ_i}$$

Where $\bar{x}_{new}$ is the new weighted average, n_{old} is the number of data sources for the old estimate, DQ_{old} is the data quality weight of the old estimate (the representative composition), x_i are the values of the new observations, and DQ_i are the data quality weights of the new observations. The data quality weights are, as explained in Deliverable D2.5, 1, 2, 3 and 4 for *dubious*, *less confident*, *confident*, and *highly confident* data respectively. However, it is recommended to exclude data that has been graded as *dubious*, as long as there are other data available.

[2] For example, in the consolidation template for vehicles, each material and component is treated in a separate sheet. This is feasible due to the small number of components and materials (high level of aggregation). For EEE, the number of components and materials with data is too large for such a customized approach, and generally each parameter subscript (e-c/e-m, m-c/c-c, m-p/c-p) has its own sheet in the consolidation file.

2.6.2. EEE

For EEE, re-consolidation should be done in the existing UNU key consolidation files. For most parameters, new data can simply be entered into the respective sheet, and will automatically be taken into account in the computation of weighted averages. Including new data is therefore simple as long as the new data relate to parameters that have already been estimated in the first consolidation.

There are two ways in which new data may not relate directly to existing estimates: 1) data concern products, materials, components or elements for which there was no estimate (e.g. the mass fractions of elements in polymer materials have not been included yet); or 2) the data concern components or materials on a different level of detail to that included in the original estimates. If 1) is the case, re-consolidation is a simple matter of adding the new data. There is no need to compute averages with existing data. If 2) is the case, a decision has to be made whether to accommodate the new data in the existing set of components and materials, i.e. by aggregation of either the new or the old data to a smaller set of components and materials. Generally, it is desirable to keep as high a level of detail as possible but if this leads to exclusion of much of the data (because it is too aggregated), it may be a better solution to aggregate. The decision will have to be made on a case-by-case basis after a subjective evaluation of the pros and cons of aggregation. See section 2.6.4 for guidance on how to choose the right level of detail.

2.6.3. Vehicles

The vehicles dataset presently includes the following components and materials: *EESystem*, *CatalyticConverter*, *ElectricDriveMotorMagnet*, *BatteryCell*[3], *MagnesiumAlloyUnspecified*, *AluminiumCastAlloy*, *AluminiumWroughtAlloy*, *SteelStandardSteel*, *SteelHSS*, *castIronUnspecified*. A higher level of detail would be desirable both with respect to the accuracy of estimates and to relevance for recycling, however, a lack of data prevents this. Although individual sources do provide data on a higher level of detail, the lists of components and materials used are not compatible, and a substantial aggregation was necessary to allow for the inclusion of more than one data source. New data on vehicles is likely to suffer from the same problem, as vehicles are highly complex products that can be described (i.e. broken down into components and materials) in an almost unlimited number of ways. However, since the present consolidated data is at a highly aggregated level and reliant on a relatively small number of primary data with variable reliability, consideration should be given to the adoption of higher level of detail should new data become available. It is recommended to read Deliverable D2.5, chapter 2.2.3 for more a detailed description of the existing consolidation for vehicles.

Inclusion of new data without changing the level of detail can be a simple matter of adding another observation to a weighted average (e.g. mass fraction of elements in catalytic converters), but it may also involve a more careful evaluation of the relevance of the data for the parameter in question. For example, some of the data used for element mass content in EESystem are based on overall estimates for the entire vehicle (i.e. from manufacturers data to which we do not have access to the full detail) and an assumption that all of this mass resides in the EESystem[4]. Such considerations are element specific and require detailed knowledge about the methods with which the data were obtained.

The existing consolidation sheets for vehicles composition data only include those components and materials that were used in the original consolidation. Each of these sheets is customized to

[3] Note that BatteryCell data is included in both the vehicles and battery datasets. To avoid double counting, vehicle battery cells are included in the mass calculations of stock and flows of CRM of batteries and excluded from that of vehicles.
[4] Data source indicates that *most* of the mass resides in the EE system.

the availability of data in terms of representing different vehicle characteristics (which define the vehicle keys) and/or time-dependent parameters for the given material/component, and does not follow a common structure. The transfer of data from these tables to the portrayals was undertaken sheet by sheet with a MATLAB/OCTAVE script written specifically for this purpose. For these reasons, it is a somewhat cumbersome process to include new components and materials in the vehicles dataset.

If new data allows for a larger number of components and materials, it would be worthwhile to further develop the consolidation file for vehicles to harmonize the structure between the different materials and components and thereby facilitate the transfer to the portrayal. New components or materials that do not exist in the current dataset can be added by copying the structure from one of the existing sheets, selecting the one that enables representation of the desired vehicle characteristics. If new sheets are added, a corresponding update to the MATLAB/OCTAVE script for transfer to the portrayal will have to be made, or another method to transfer the new data must be devised.

It is very important to ensure that any new components or materials are not already covered by one of the (highly aggregated) components or materials in the present dataset. If this is the case, addition of the new component or material will lead to double counting.

2.6.4. General remarks

The main difficulty with including new data (as in the original consolidation) is to choose the right level of detail, i.e. the list of materials and components to include. Some product group specific aspects have been discussed above. In general, the choice depends on the amount and quality of data available at different levels of detail; the possibilities of aggregation (sometimes a common set of components cannot be found); and the usefulness of introducing more detail. The choice has to be made after subjective evaluation of all of these aspects, and can only be done case-by-case. Although we do not provide any hard rules for how to choose the right level of detail, the following questions can serve as guidance:

Does the higher level of detail for materials and components contribute to an improved estimate of total element content?

> For example, it may not be worthwhile including 15 types of polymers if no information exists on the CRM content in each of these polymers. Aggregation to 1 type may be better if it allows for including more data sources.

Is the higher level of detail relevant for recycling?

> For example, including aluminium in vehicles in separate components or component groups rather than directly at the vehicle level may provide useful information for recyclers, as separate dismantling of components can lead to separation of different alloys and thereby higher quality of the recycled material.

Does the higher level of detail lead to exclusion of most of the data, or data with high quality?

> If yes, it should be reconsidered.

2.7. Assess data quality and uncertainty of estimated representative compositions

The procedure used for data quality and uncertainty assessment of consolidated data to date should also be used in future updates. Data quality of consolidated data (estimated representative compositions) is evaluated based on the number of data sources, their estimated representativeness and the extent to which temporal changes are included. Uncertainties are estimated based on the assigned data quality and the absolute value of the physical quantity in question (in relative terms, small quantities are much more uncertain than large quantities). See Deliverable report D2.5, chapter 2.1.5 for details on the data quality assessment and uncertainty estimation for consolidated composition data.

The data quality is automatically determined by filling in a table with the required information in the consolidation files. The uncertainties are determined by running a MATLAB/OCTAVE script over the portrayals (in practice, the uncertainty assessment therefore occurs after transfer to the portrayals).

2.8. Transfer new consolidated data (representative compositions) to portrayals

Consolidated composition data is spread over a number of different sheets in the consolidation files. An automatic extraction of the data to fill portrayals was implemented in a MATLAB/OCTAVE script (one for EEE and one for vehicles). Without this or an equivalent process, the transfer would be extremely time-consuming. For EEE the same scripts can be used in the future, also if new components, materials or products are added. For vehicles the structure is less flexible, so additions to the data and re-consolidation may require modification of the transfer script.

The empty portrayal template is attached as Annex 2 to this deliverable report. Instructions for how to fill it are provided in Table 2.[5]

Table 2 Fields of the portrayal template and instructions for filling them. Not all fields are harvested.

field name	*input type*	*Instructions*
Product key	code	product key for which the data applies. Taken from code lists for EEE, BATT and vehicles products. Must be provided for all data.
productionYearStart	#	first year placed on market
productionYearEnd	#	last year placed on market (normally if data are provided over time, a new value is entered for each year, and productionYearStart is set equal to productionYearEnd.
CRM parameter	code	CRM parameter code, e.g. *MassContent* or *MassFraction*. See instructions for CRM parameter template for more information.
CRM parameter subscript (u-v OR u)	code	CRM parameter subscript, e.g. *e-c* or *m*-p. See instructions for CRM parameter template for more information.
u	code	a code that refers to the entity of the first letter in the CRM parameter subscript. For example, for *e-c* data, an element code must be entered here, and for *m-p* data a material code must be entered here. For *Mass* data, which only has one subscript letter (*c* or *p*), the relevant should be entered here.
v	code	a code that refers to the entity of the second letter in the CRM

[5] The actual data transfer may be done with a script, but it is nevertheless important to check that the fields have been filled correctly.

field name	*input type*	*Instructions*
		parameter subscript. For example, for *e-c* data, a component code must be entered here, and for *m-p* data a product key (or sub-key) must be entered here. The product key or sub-key will then be the same as in the field "Product key". For data that only has one subscript letter, this field is left empty.
Value	#	the value of the physical quantity specified by the CRM parameter and CRM parameter subscript
Units of measurement	code	units of measurement. Refer to code list in CRM parameter template.
Uncertain quantity (min)	#	minimum value of uncertainty interval (confidence interval)
Uncertain quantity (max)	#	maximum value of uncertainty interval (confidence interval)
uncertainty Type	code	type of uncertainty or confidence interval, e.g. conf90 for 90% confidence interval. Refer to code list in CRM parameter template
ID_metadata	#	an ID to identify the metadata describing the data in the ProSUM metadata catalogue
Original or estimate?	code	indicate whether the data is measured directly or estimated. For composition data, all is considered an estimate as it is not possible to observe directly the average composition of a given product
Data quality	code	data quality, as determined by the procedure described in D2.5, chapter 2.1.5
Data quality comment	txt	comments to the data quality beyond the indicated level
Data consolidation comment	txt	comments to how the data consolidation was performed. It is recommended to avoid long text, as the details of the consolidation can be explained elsewhere (i.e. in the metadata catalogue).
Person who entered data	txt	name of person who entered (is responsible) for the data

2.9. Harvest data to the UMKDP

The harvesting of data to the UMKDP is covered in chapter 3.4 of deliverable report D5 (Cassard et al. 2016).

2.10. Update frequency

A frequent update of composition data for products placed on the market is desirable, due to rapid technological development. Unfortunately, data for products recently placed on the market is for obvious reasons normally not available. Much of the available composition data is based on waste sampling and characterization conducted in the last 3-7 years. This waste provides representative compositions for products placed on the market much earlier, due to the lifespan of products.

It does not make sense to define an update frequency out of the needs for up-to-date data which cannot be met. Rather, updates should be made as more data becomes available and the frequency should reflect the work required for each step. Recording data in CRM parameter templates and consolidation sheets can be undertaken on a continuous basis as more data becomes available. Re-consolidation is a time-consuming process, even for small updates. Hence, it is better to wait for larger amounts of data to become available before re-consolidating.

We propose that the composition raw data is updated on a yearly basis (Steps 1-3 in Figure 3*), and that a thorough re-consolidation and harvesting is performed every 3-4 years (Steps 4-8 in* Figure 3*).*

3. Recommendations

3.1. BATT composition data

Availability and type of data

A rather complete description of the batteries' composition (mass fraction of elements in the battery cell) was available in different documents originating from the battery industry:

- Batteries Material Safety Data Sheets;
- Handbook of batteries, based on scientific and historical description of the batteries compositions; and
- Published data from the industry including recyclers' data.

The estimated representative values for CRM and the main material content have been based on these documents, providing the composition of the main battery designs. For each battery sub-key, a representative value and a confidence interval for each of the selected elements have been calculated based on these data sources.

The dataset derived has been used as representative batteries composition data for the ProSUM diffusion database. Although the method used requires a good knowledge of the battery technologies placed on the market, it achieved more complete and more precise information than statistical analysis (e.g. based on sampling and chemical analysis) would have provided, which is often impossible due to a lack of data.

The main limitations of the data are:

- Changes over time for defined battery sub-keys have not been taken into account. For example, it is well known that the mixture of rare earth metals used in NiMH batteries has changed over time with fluctuations in prices. Nevertheless, chemical composition normally evolves slowly within each battery technology. Major product design changes are linked to new applications: consequently, the most important change is the relative share of different technologies in the market, which is taken into account through the data and models of stocks and flows within ProSUM;
- Battery electronics are not included. Although these parts represent a small proportion of the weight of a battery, sometimes negligible or not existing at all (in the case of small portable batteries), they may be significant for large industrial batteries;
- Certain minor elements are not included due to a lack of data: again, the elements used in the battery electronics are not specifically identified. The impact of this may be mitigated by the assessment of the battery electronic as a part of the equipment or vehicle in which the battery is used. This has been done for Ag in electric vehicles, where the battery electronics are considered part of the EE system in the vehicle.

When possible, a weighted average of the product composition, according to the market share, has been applied e.g. for the usage of natural graphite (identified as a CRM) in the cathode of Li-ion batteries. Based on published market share data (Avicenne 2016), natural graphite represents 63%, and artificial graphite 37% of all the graphite used in batteries. As the different type of graphite are generally not identified in the composition or analysis data, a ratio of 63% has been applied by default to assess the quantity of natural graphite used in batteries.

Quality of data

As most of the data used were based on manufacturer information, they were assessed as reliable (DQ level "confident"). The precision of the elements composition was in general low (particularly in the “safety data sheets”). Nevertheless, the comparison of several manufacturers’ data sources allowed for the assessment of a relevant representative values. A statistical verification of the data quality was possible in the case of Lithium-ion batteries of LCO type. The result has confirmed the validity of the approach based on the confidence interval. Further acquisition of measured data may allow for more statistical comparisons in the future.

Recommendations for working with batteries data

The generation of new data (such as the composition of new battery technologies, based on chemical analysis or producer data) and their introduction into the database should be carefully reviewed with the following criteria:

- The statistical representativeness of the sample;
- The product application weight on the market, and the representativeness of this battery technology in this product application;
- The definition of the product, for compatibility with the existing datasets. For example, the electronic parts of the batteries have not been taken into account in the ProSUM Project. A quantification of these parts would certainly provide an improvement to the content quality, but would also have to be identified as a new component in the battery, and not mixed with the battery cells composition;
- In cases where physio-chemical analysis methods are used for composition assessment, a careful review of the method and the results is required. Based on the learnings of the data analysis in this project incorrect results may be published for a number of reasons:
    - Varying components or sub-set of cells may have been analysed (different perimeter);
    - Incorrect preparation of the sample (presence of electrolyte in the material analysis of electrodes);
    - Analysis of traces and impurities, not representative for the product;
    - Graphite not identified or differentiated from other types of carbon; and
    - Incoherent values (unrepresentative samples, incorrect analysis and various other reasons),
- Other improvements in the data may be achieved with an assessment of the composition change versus time. Again, compatibility with the existing data should be ensured. It is expected that such an improvement would require a much larger set of raw data, in order to achieve the expected reduction of the confidence interval per year.

3.2. EEE composition data

Availability of data:

WEEE composition data has been retrieved from a variety of sources e.g. academic and industry studies, journals, and governmental statistics agencies. Generally speaking, all lack the sophisticated data structure necessary for this work and lack harmonisation. The WEEE composition data was recorded and consolidated according to the structure described in D2.5. Despite the large number of individual data points gathered, retrieved, consolidated and stored, a substantial number of data gaps remain:

- For newer types of products, like drones or intelligent refrigerators, hardly any information is available specifically for new components. The same applies to the many types of professional appliances under the UNU keys: 0602 Large Tools; 0802; Large Medical; and 0902 Large Monitoring and Control. Here, often very specific components and functions are embedded but in low product quantities;
- For the most abundant materials (e.g. polymers and base metals) in products and those materials with a high value (e.g. gold and silver), better data is available in the literature.

There is a lack of, or insufficient information, for a wider range of elements (including many CRMs) for lamps (UNU keys 0502-0504), cooled and non-cooled dispensers (UNU keys 1001,1002), central and professional heating (UNU keys 0001 and 0101), photovoltaic panels (UNU key 0002), air conditioners (UNU key 0111), professional cooling (UNU key 0113) and professional tools (UNU key 0602);

- Only for a select number of UNU keys, a relatively complete and detailed calculation has been made from all elements to all materials to all components. For some key components, their mass fraction in the product could be determined as a function of time. However, changes over time for the element mass fraction in the component (e.g. in printed circuit boards) could usually not be determined;
- For most products there is a time dependent product weight available, but for some of the more professional and less common products the information is limited;
- Embedded electronics for communication, monitoring, etc. are being used in a much wider range of products including those that are not traditionally seen as electronics, which make it hard to quantify the composition of these (new) products.

Quality of data

- A key issue is the description of the temporal and geographical scope and descriptions of the sample size and nature of composition data. This counts especially for the academic sources. Hence the recommendation is to not only provide researchers/data reviewers with the developed code lists but also a more simple checklist/short recommendation when publishing EEE composition data. Unfortunately, the EuP studies[6] that are supposed to cover newer and rather representative products are also relatively poorly defined in this regard and thus less useful than initially expected;
- Consolidated datasets are structured, harmonised and reviewed by different members of our consortium to ensure not only transparency but also good quality. While doing this, many issues were encountered in the procedures that have now been resolved.

Recommendations for future work with composition data

- It is important for the consolidations and the relation to the stocks and flows information that the basket of goods for the composition data matches and is representative for the basket of products (sub-keys to UNU keys) in the POM and WEEE flows data. During the consolidations, products are frequently encountered that were already submitted in the CRM parameter templates that actually belong to a different UNU key. For a repeated exercise, the basket of products should be reviewed early on. Here, the developed UNU key catalogue in D2.4, the contribution of sub-keys in sampling data (p-f data) from D3.2 and D4.4 as well as the statistical data from D3.5 needs to be taken into account at the beginning of an update;
- Most data on the content of components in products originate from dismantling appliances in the return stream. Relatively recent data are rare. As components and products change rapidly with all seeing a high level of miniaturisation, an idea can be to organize dismantling sessions with recycling companies, universities, with entities like iFixit, among others, in order to receive more knowledge on the composition of new products and for those whose information was not available when consolidating;
- A new CENELEC standard on declaration of CRM content in EEE is under development (CENELEC 2017), and it can be expected that eventually, reporting will become mandatory for

---

[6] Refers to a number of product-specific studies conducted in preparation for the EU Ecodesign Directive. EuP = "Energy-using products". European Parliament (2009). "Directive 2009/125/EC of the European Parliament and of the Council of 21 October 2009 establishing a framework for the setting of ecodesign requirements for energy-related products (recast)." Official Journal of the European Union.

selected substances and products. This can potentially lead to a large amount of very valuable data for describing product compositions. The processes leading to such decisions are generally slow, and no new data is expected in the short-term. Moreover, there is no guarantee that such data will become available for the EU-UMKDP. Future work with EEE compositions should nevertheless monitor this development closely.

Despite the data availability and quality limitations, the consolidation exercise and new information sources added provide the most up-to-date EEE composition datasets internationally, covering the most common and CRM-relevant products and secondary raw materials well.

3.3. Vehicles composition data

Availability and type of data

The availability of comprehensive data (i.e. covering a large set of elements, materials and/or components) on the composition of vehicles with respect to CRMs is limited to a handful of studies (Ministry of Environment Japan 2009, Ministry of Environment Japan 2010, Alonso et al. 2012, Cullbrand and Magnusson 2012, Widmer et al. 2015, Andersson et al. 2016, Restrepo et al. 2017). Three of these studies (Alonso et al. 2012, Cullbrand and Magnusson 2012, Andersson et al. 2016) retrieve their data from the vehicle manufacturers' own databases (containing detailed data on the composition of every component in the vehicle) and present these in an aggregated form (i.e. content per vehicle or vehicle subsystem). The remaining studies retrieve their data from direct measurements of components from end-of-life vehicles or output fractions from automobile shredders. Additional data on specific components (e.g. catalytic converters) or materials (e.g. steel and aluminium), originating from manufacturers' data, are also available and have been used within ProSUM.

On the one hand, considering the complexity of the product, available composition data for vehicles is far from satisfactory to obtain good estimates of average vehicle composition. On the other hand, variability within the product group is lower in comparison to the product group of EEE, so that the required sample size to obtain a good estimate of the average composition of new vehicles is lower than for EEE.

Quality of data

The two types of data available (manufacturers' data and independent measurements), are obtained using entirely different methods, and their quality is therefore discussed separately in the following.

The three studies with a comprehensive assessment of CRM content from manufacturers' data all utilize the International Material Data System (IMDS), a database that has been created by a consortium of automobile manufacturers (Audi, BMW, Daimler Chrysler, Ford Motor Company, Opel, Porsche, Volvo and Volkswagen) to facilitate compliance with European recycling targets and other regulations such as REACH. The system lets each manufacturer access compositional data on components relevant for their vehicle models, with detailed information about components, sub-components and substances contained in these. This information is provided through data sheets by the component suppliers. IMDS contains a vast amount of detailed data on individual components. This data is normally only accessible to the manufacturers. In the available studies, access was granted for research purposes, but results were only allowed to be published on a highly aggregated level of components/subsystems. A number of quality issues with using IMDS to quantify CRM content have been pointed out in the study by Cullbrand and Magnusson (Cullbrand and Magnusson 2012):

- For newer vehicle models, all datasheets may not be present in the database, i.e. there are components with no information on the *e-c* level. In Cullbrand and Magnusson

(2012), the missing information was either obtained by direct contact with the suppliers, or excluded from the study;
- Substances may be reported as “confidential substances”, “misc. not to declare” etc.;
- There are few controls of the data reported by suppliers, so it is likely that some data are reported incorrectly.

The remaining studies are based on direct measurements of the chemical composition of vehicle components or shredder output fractions. Measurement of shredder output fractions has the advantage that it in principle can achieve a close to full mass balance of the vehicle composition by measuring from all output fractions. In practice, this is not possible, as many components are dismantled before shredding, and there are losses in the shredder process, e.g. magnetic materials may remain attached to the shredder equipment. The outputs are highly heterogeneous, which means that in a first step, very large samples must be taken, homogenized and subsampled to obtain a representative sample for chemical analysis. In practice, such large samples sizes are normally not feasible. Moreover, analysis of shredder output fractions cannot provide any information on the location of CRMs, i.e. component-level data.

Measurement of dismantled components has the advantage that one can obtain very detailed information on the composition of individual components. The main disadvantage is that, due to the resources required, it is only feasible to perform such a study for a small number of components and vehicles. Therefore, such studies can only contribute with “hot-spot” data, focusing on components where it is suspected that most of the metal in question resides. All studies based on measurements from end-of-life vehicles will necessarily provide information that is mainly relevant for vehicle models of a certain age, typically reflecting the average lifetime of vehicles, and will not capture the latest technological developments.

Recommendations for working with vehicles data

Apart from the small amount of data available, the main challenge with vehicle composition data concerns how to represent such a complex product in a way that allows for harmonization of data from different sources. Unfortunately, there is no agreed-upon or standardized way of describing and grouping different vehicle components, and the groups and subsystems used are not clearly defined. Different studies present data with entirely different levels of detail, referring to a different set of component groups, materials or subsystems[7]. As a result of these differences in vehicle descriptions, the consolidation of data within ProSUM was performed at a highly aggregated level, for example by grouping all electrical and electronic components into one subsystem. This naturally affects the usefulness of data in cases where detailed information about where to locate critical or valuable materials in ELVs are sought. It would improve the dataset on vehicles substantially if (in the continued work on the UMKDP) this component group (electrical and electronic system) would be disaggregated into its constituent devices or control systems. However, this requires that more detailed data becomes available. To consolidate new and more detailed data with existing aggregate data, it would be necessary to allocate components from the new and expanded list to the subsystems and component groups used in existing studies and perform statistical data reconciliation.

In future studies, efforts should be made to produce data that are compatible with existing datasets, and to develop a standardized list of components and subsystems for vehicles (Du et

[7] For example, the study by Alonso et al. refers to the following 13 subsystems: info and controls; interior; steering and brakes; suspensions; tires and wheels; transmission; closures; electrical; exterior; engine; fuel and exhaust; HVAC; while the study by Cullbrand and Magnusson refers to the following 14 subsystems: powertrain mounts; wheels and wheel suspension; climate system; doors, boot and hood; infotainment; engine system; safety electronics; restraints and air bags; body structure; standard components; transmission; seating; exterior lighting; security and body electronics.

al. 2015). Some work has been done in this direction by Restrepo et al. (Restrepo et al. 2017), who classify 300 automotive electrical and electronic devices as controllers, actuators or sensors, and group them into 71 different control systems. This approach is useful, since it takes the physical (or wireless) connections between components as the starting point for grouping, rather than less clearly defined categories such as “safety electronics”. IMDS, although confidential, is a potentially important future data source, and its structure should be taken into consideration.

The following set of recommendations for future work on composition data for vehicles is recommened:

- Ensure that new data are compatible with existing datasets;
- Provide clear definitions of subsystems or component groups used, by listing all components included in the given subsystem;
- Make data gaps explicit: when a hot-spot approach is used, point out which components and materials were *not* included in addition to those that were included;
- Provide extensive description of vehicles investigated, including brand, model, model year, fuel type, mass and an indication of equipment level (class) if possible.

3.4. General recommendations

The work with product composition data in ProSUM has involved an extensive review of available data, evaluation of data quality, harmonization of data from different sources and calculation of representative product compositions. Throughout this work, the general working process was to first define (as well as possible) a procedure for the work (e.g. templates and code lists), and then implement it in practice. The first step is crucial, but is usually difficult to do without the practical experience of the second step. In practice, revisions to the procedure were always made. This iterative process of testing and revising the procedures is however time-consuming, and due to the goal-oriented approach of ProSUM, it was often necessary to move on with the practical implementation although further revisions to the procedure would have been welcome. In this chapter, the most important experiences from the work with composition data are summarised, and recommendations provided for how to continue the work in the future.

Code lists

Consolidation of data from different sources into a harmonized system rests on the use of code lists to designate the meaning of the data. A large number of code lists were used for the work with product composition data, the most important[8] of which are the code lists for products (*UNU key*, *UNU sub-key, UNU device type, BATT key, BATT sub-key and vehicle key*), components, (*component, component group*) and material (*material, material type*), which are used to define what object the data refer to. The definition of the UNU keys (for EEE) had been done prior to ProSUM. They are (for the most part) clearly defined, allowing little room for interpretation. The BATT keys follow logically from the different chemistries of the most common battery types. The vehicle keys were defined within ProSUM, but were defined to match the description of vehicle types used in EU statistics. They are also clearly defined.

The main challenge therefore lies in the code lists for materials and components. These code lists should firstly, to conform with the objectives of ProSUM, allow for describing products in a way that facilitates recycling though identification of CRM “hotspots” (Baldé et al. 2015). Secondly, the code lists should facilitate the harmonization and consolidation of data from a wide variety of data sources. While the first point could be reasonably addressed after an initial review

[8] Important in the sense that the definition of these code lists largely determines the further work with recording and consolidating data.

of main data sources, the second point requires detailed knowledge of most of the data. However, such detailed knowledge can only be obtained by systematically bringing data together. This leads to a recursive problem: a good definition of code lists is difficult before recording the data, but recording the data cannot be done well without the definition of code lists. In practice, the code lists of ProSUM were first defined according to the objectives, and later revised to accommodate data that did not fit in the existing lists. Proper definition of code lists is a very time consuming process, as it requires clear definitions of every code. At the time of writing, the ProSUM component code list contains 295 components, and the material code list contains 444 materials. Due to the complexity of the products of interest in ProSUM[9] and the number of different representations found in data sources, it was not possible to provide detailed definitions and correspondence tables for all components. This led to some difficulties in interpretation and consolidation of different data sources later. However, even perfect definition of code lists would not have resolved all such problems, as primary sources rarely provide unambiguous descriptions in the first place. Based on these experiences, it is recommended to invest more time in developing good component and material code lists.

Aggregation of materials and components

The main difficulty with consolidating data from different sources lies in the fact that they use different sets of components and materials to describe products. One data source may provide the following composition of a laptop: 1% cables, 14.5% polymers, 6.5% printed circuit boards, 18.5% liquid crystal display, 1% LED backlighting, 35% metal alloys, 4% others. Another source may provide this composition: 109g keyboard, 177g CD drive, 8g RAM, 206g motherboard, 4g CPU, 800g display, 369g top casing, 292g bottom casing. It is clearly difficult to consolidate the data from these two sources, as they refer to entirely different sets of materials and components. In ProSUM, two different approaches have been used for consolidating data that refer to different sets of components and materials. The first, most common approach has been to aggregate data to a level that works for all data sources. This approach is illustrated for two different data situations in Figure 4 and Figure 5. In the data example above, which is conceptually illustrated in Figure 4, aggregation is difficult, since there is no common set of components and materials that works for both sources. Data source A and B use component/material sets that are incompatible. To take both data sources into account, one may aggregate all the way to the product level, working only with the calculated e-p data, but this leads to a loss of most of the information.

[9] A passenger vehicle consists of about 30'000 individual components.

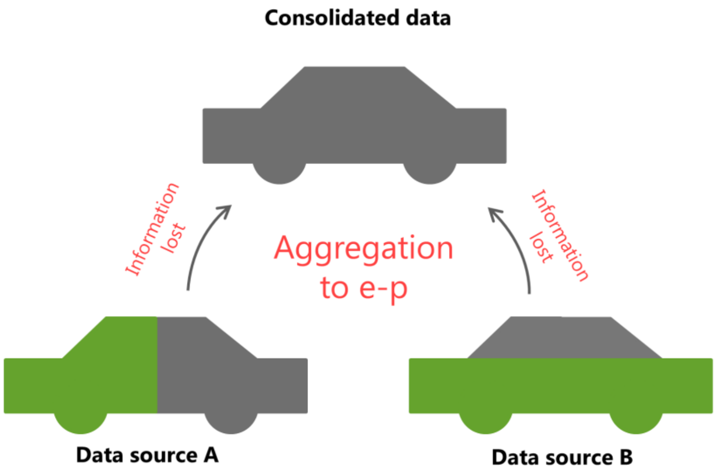


Figure 4 Consolidation of data from two incompatible sources. Data sources A and B have similar levels of detail, but the sets of components/materials used are incompatible. To take into account data from both sources, they are both aggregated to the product level, leading to a loss of information. In this case the component/material level disappears altogether. Alternatively, one of the sources could be discarded.

In more fortunate cases, sets of components and materials from different sources are compatible, so that one source may be aggregated to the level of the other. This data situation is illustrated in Figure 5. Here, information is lost only from one source. The vehicles data set was created using the approach illustrated in Figure 4 and Figure 5. For example, the component *electrical and electronic system* does not appear in any of the primary sources, but was introduced in ProSUM to accommodate data from several sources with data on a more detailed level.

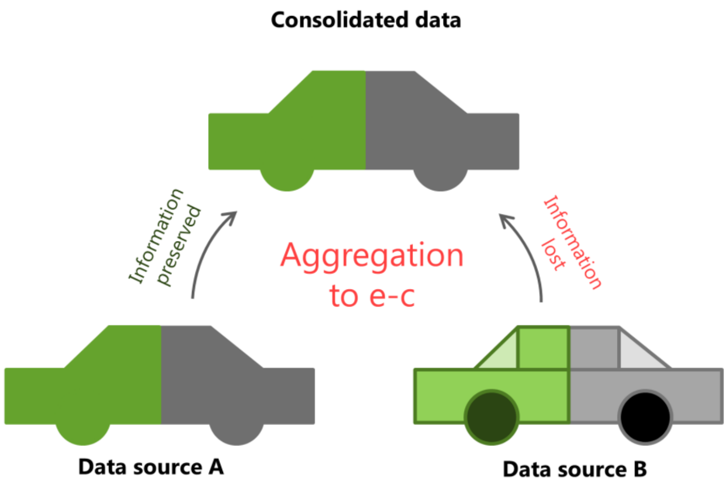


Figure 5 Consolidation of data from two compatible sources. Sources A and C use different sets of components/materials. Source C provides a higher level of detail than source A, using a set of components/materials that is compatible with the smaller set from source A (i.e. the components/materials of source A can be obtained by grouping the components/materials of source C). To utilize both sources, data from source C are aggregated, leading to a loss of information from this source only.

The second approach, which was used for the EEE data works in the following way: instead of aggregating, the details from all sources are kept in the calculation of the representative composition. The approach is illustrated in Figure 6. All components and materials from all sources are kept, but their masses are adjusted so that they still add up to 100% of the product mass. As a consequence, the components and materials of the consolidated data are not mutually exclusive, rather there is some overlap (two of the car wheels in the example are covered by the aluminium material instead). The advantages of this approach are that: 1) it allows for a quicker consolidation of data from many different sources without going through the time-consuming step of manually selecting a reasonable aggregation level; and 2) it allows for retaining all information on the e-c and e-m level in the final consolidated data. The obvious drawback is that the data on the m-p and c-p levels are meaningless in isolation. For EEE data, the approach works well because the components that are most relevant for CRM content (e.g. printed circuit boards) are nearly always included in the component lists of the primary sources, and hence there is no bias introduced in their c-p data.

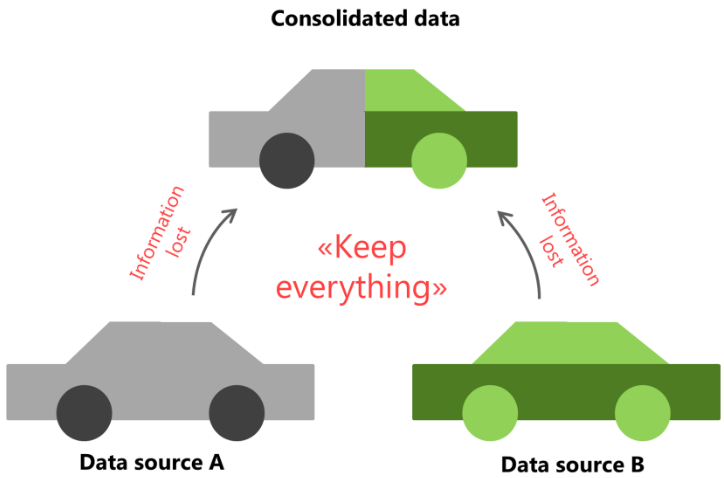


Figure 6 Consolidation of data from two sources with incompatible component/material sets. Think of the bottom left car as two components (wheels and body) and the bottom right car as two materials (steel and aluminium). After consolidation, we describe the top car as consisting of 2 wheels, half a body, some steel and some aluminium. Obviously, this makes the consolidated m-p and c-p meaningless in isolation as they grossly underestimate the mass of each component/material. However, on the e-c and e-m level a lot more detail is retained and the overall estimate of e-p should be more accurate due to full utilization of the data.

Both of the approaches used in ProSUM have the significant drawback that they lead to a substantial loss of information. In the case of incompatible component/material lists, it is not possible to retain the information from all sources in the calculation of an average composition. In the case of component/material lists that are compatible, a more desirable outcome would be as illustrated in Figure 7. Here, the consolidated data retain the level of detail from the most detailed data source, while still including the more aggregate information from source A. This is possible by performing statistical data reconciliation (Cencic and Frühwirth 2015). The method is used for consolidating mass flow systems with redundant information (e.g. detailed inputs and total output of a process are known), but could be applied in exactly the same way to composition data (which also have to respect mass balance). Data reconciliation requires that the correspondence tables between all components and materials are properly defined, and that a probability distribution is specified for each input parameter. Due to the large amount of work required, statistical data reconciliation was not implemented in ProSUM, but would be an interesting alternative to improve the consolidated data sets in the future. It would still be necessary to choose a level of detail for the consolidated data, and it would require a substantial amount of work to define the correspondence tables between all components and materials. There would still be some loss of information from sources that use component/material lists that are incompatible with the list chosen for the consolidated data (as illustrated in Figure 4).

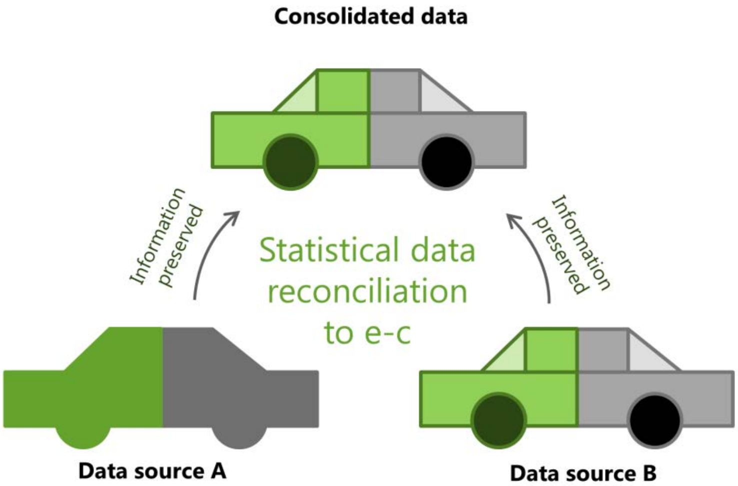


Figure 7 Data reconciliation in the consolidation process. Data from two sources are consolidated. The data situation is the same as in Figure 5, but instead of aggregating the data from source B, statistical data reconciliation is performed. In effect, the data from source B are adjusted up or down based on the observations from data source A. The higher the uncertainty of data source B, and the larger the difference between source A and B, the larger the adjustment to the data. This procedure was not used within ProSUM.

4. References

Alonso, E., T. Wallington, A. Sherman, M. Everson, F. Field, R. Roth and R. Kirchain (2012). "An Assessment of the Rare Earth Element Content of Conventional and Electric Vehicles." SAE International Journal of Materials & Manufacturing, SAE International Journal of Materials & Manufacturing.

Andersson, M., M. Ljunggren Söderman and B. A. Sandén (2016). "Are scarce metals in cars functionally recycled?" Waste Management.

Avicenne (2016). "Battery Market Development for Consumer Electronics, Automotive, and Industrial." Batteries congresses, Nice, France.

Baldé, K., V. van Straalen, J. Huisman, H. Habib, P. Chancerel, P. Maehlitz, S. Rotter, P. Wäger, F. Schjøth, A. Hallberg, A. Scheepens and D. Cassard (2015). "Review and harmonization of data - Deliverable 5.3 (ProSUM)."

Cassard, D., F. Tertre, F. Schjøth, T. Heijboer, D. Čápová, A. Hallberg and J. Šinigoj (2016). "ProSUM EU-Urban Mining Knowledge Data Platform Specifications."

Cencic, O. and O. Frühwirth (2015). "A general framework for data reconciliation - Part I: Linear constraints." Computers and Chemical Engineering.

CENELEC (2017). "General method to declare the use of critical raw materials in energy related products." https://www.cenelec.eu/dyn/www/f?p=104:110:1546627903880101::::FSP_ORG_ID,FSP_PROJECT,FSP_LANG_ID:2240017,65687,25

Cullbrand, K. and O. Magnusson (2012). "The Use of Potentially Critical Materials in Passenger Cars." Chalmers University of Technology.

Du, X., E. Restrepo, R. Widmer and P. Wäger (2015). "Quantifying the distribution of critical metals in conventional passenger vehicles using input-driven and output-driven approaches: a comparative study." Journal of Material Cycles and Waste Management, Journal of Material Cycles and Waste Management.

European Parliament (2009). "Directive 2009/125/EC of the European Parliament and of the Council of 21 October 2009 establishing a framework for the setting of ecodesign requirements for energy-related products (recast)." Official Journal of the European Union.

Løvik, A. N., P. Wäger, C. Chanson, J. Huisman, M. Ljunggren Söderman, M. Guzman Brechu, A. Haarman, H. Habib, P. Mählitz, M. Rösslein and M. Ueberschaar (2017). "Report on consolidation of data into CRM database: Deliverable 2.5 (ProSUM)."

Ministry of Environment Japan (2009). "Report on FY2009 Survey to Promote Streamlining of End-of-Life Vehicle Recycling (使用済み自動車再資源化の効率化及び合理化推進調査)." Ministry of the Environment, Japan.

Ministry of Environment Japan (2010). "Report on 2010 Survey to Identify the Characteristics of Automotive Shredder Residue (自動車破砕残さにおける性状把握調査業務)." Ministry of Environment Japan, Ministry of the Environment, Japan.

Restrepo, E., A. N. Løvik, P. Wäger, R. Widmer, R. Lonka and D. B. Müller (2017). "Stocks, flows and distribution of critical metals in embedded electronics in passenger vehicles." Environmental Science & Technology, American Chemical Society.

Rotter, V. S., P. Wäger, N. Korf, P. M. Mählitz, R. Figi, M. Ueberschaar, O. Münch, C. Schreiner, M. Rösslein and A. N. Løvik (2017). "Material characterization methods for waste fractions and tailings - Deliverable D4.3 (ProSUM)."

Wäger, P., V. S. Rotter, N. Korf, P. M. Mählitz, R. Figi, M. Ueberschaar, O. Münch, C. Schreiner, M. Rösslein and A. N. Løvik (2017). "CRM Sample Preparation and Chemical Analysis - Deliverable 2.6 (ProSUM)."

Widmer, R., X. Du, O. Haag, E. Restrepo and P. A. Wager (2015). "Scarce Metals in Conventional Passenger Vehicles and End-of-Life Vehicle Shredder Output." Environmental science & technology, American Chemical Society.

Annex 1 – CRM parameter template

The CRM parameter template used for recording composition data from primary sources is supplied as Annex 1 to this deliverable report.

Annex 2 – Data portrayal template

The data portrayal template used to provide consolidated data sets for harvesting to the EU-UMKDP is provided as Annex 2 to this deliverable report.