Extracting Structured Information About Food Loss and Waste Measurement Practices Using Large Language Models: A Feasibility Study

Junoš Lukan
junos.lukan@ijs.si
Jožef Stefan Institute
Department of Intelligent Systems
Jožef Stefan International
Postgraduate School
Ljubljana, Slovenia

Maori Inagawa
maoriinagawa@keio.jp
Jožef Stefan Institute
Department of Intelligent Systems
Ljubljana, Slovenia

Mitja Luštrek
mitja.lustrek@ijs.si
Jožef Stefan Institute
Department of Intelligent Systems
Jožef Stefan International
Postgraduate School
Ljubljana, Slovenia

Abstract

Waste Quantification Solutions to Limit Environmental Stress (WASTELESS) project aims to develop and test innovative tools and methodologies for measuring and monitoring food loss and waste (FLW). A key objective is to create a decision support toolbox that helps food actors across the entire supply chain, including consumers, select the most suitable method for measuring and monitoring FLW. To help with this decision, existing, already tested FLW measurement practices can be consulted, which are currently published as short documents. In this work, we show how the data about them can be extracted using large language models (LLMs). Additionally, we propose how this data can be structured and represented as an ontology. With this process, we can help users find relevant data without needing to browse through many documents.

Keywords

food loss and waste, large language models, data extraction, ontology

1 Introduction

The project *Waste Quantification Solutions to Limit Environmental Stress* (WASTELESS; https://wastelesseu.com/) is designed to develop and test a mix of innovative tools and methodologies for food loss and waste (FLW) measurement and monitoring. One of the tasks is also to create a decision support toolbox [10]. It should help all profiles of food actors, i.e. across the whole food supply chain (FSC), including consumers, who want to measure and monitor their FLW, to select the most appropriate method.

There have been several attempts to harmonise FLW measurement methods. The *Food loss and waste accounting and reporting standard* (FLW Standard; [7]) stands out as a good structured attempt. It was produced by the Food Loss & Waste Protocol, a multi-stakeholder partnership with involvement by Food and Agriculture Organization of the United Nations (FAO) and World Resources Institute among others.

The FLW standard establishes the scope of an FLW inventory. Furthermore, it provides definitions of boundary elements and recommendations for classifications that should be used to describe them. For classifying food into categories, it suggests the FAO's and World Health Organization's Codex General Standard for Food Additives [5]. We might add that alternatively, Annex II of "Regulation (EC) No 1333/2008 of the European Parliament and of the Council" can also be used. For lifecycle stage, the International Standard Industrial Classifications of All Economic Activities (ISIC) or the Statistical Classification of Economic Activities in the European Community (NACE) [4] should be used. Finally, for geographical boundary classification UN region or country codes should be used or Nomenclature of Territorial Units for Statistics (NUTS) [2] in the European context.

The FLW standard also provides guidelines on how to decide which quantification method to use for FLW measurement or monitoring. The *FLW Quantification method ranking tool* was prepared by the Waste and Resources Action Programme (WRAP) and includes eleven questions. Most of the questions serve as exclusion criteria. For example, a negative response to either "Do you have existing records that could be used for quantifying FLW?" (Q9) or "Do you have access to those records?" (Q10) excludes the method of records. As another example, a negative response to "Can you get direct access to the FLW being quantified" (Q3) immediately excludes direct weighing, counting, assessing volume, and waste composition analysis, since these all need such access to be feasible. These questions encapsulate the most important characteristics by which these methods distinguish from one another and lend themselves to particular needs of users.

In this paper, we build upon this work by proposing a unified structure through which to describe various practices of FLW measurement and reduction. This is a step towards systematic representation of these data that can enable further analysis of the practices thus described and their comparison and validation.

2 Methods

We first outline the structure of desired shortened descriptions, report on the process of using large language models (LLMs) to automatically extract them and finally evaluate the results by comparing them to human annotations.

Information Society 2025, Ljubljana, Slovenia

https://doi.org/10.70314/is.2025.skui.8598

2.1 Structure of Extracted Information

Based on the previously mentioned *FLW Quantification method ranking tool* and domain-expert knowledge, we determined the following characteristics of FLW measurement methods and practices to be of the most importance:

(1) FLW method.
FLW measurement and reduction practices might describe very specific technologies and techniques. To make the information more general, we decided to classify as one of ten categories of quantification methods. These are described in detail in the Supplement [8] to the *FLW Accounting and Reporting Standard* [7].

(2) Region of interest.
European Union (EU) member countries have diverse legislation that is of relevance to FLW measurement (see [13] for a review). Some have legislation actions that are legally binding, such as laws and regulations, and as such prescribe methods of monitoring and FLW measurement as well as the ways of reporting the data. On the other hand, some countries only approach the topic through non-binding legislation actions, such as agency orders and policy papers. As such, not every method might be appropriate for every country or region.

(3) Food supply chain (FSC) stage.
Food loss and waste can occur at any stage of the food supply chain, starting from farmers and other producers, through food manufacturers and processors, distributors and shippers, grocery stores and restaurants, all the way to the customers and consumers. Some methods are more appropriate for certain stages in this chain. For example, a household might keep a diary of their FLW, while sellers such as grocery stores, generally manage their stock more systematically and precisely.

(4) Accuracy.
FLW measurement methods need also to be considered from the point of desired accuracy. The highest accuracy can be achieved by directly weighing the waste or separating it into components (waste composition analysis), while diaries or volume assessment produce data of medium accuracy. At the lowest end, proxy data can be used to assess FLW, for example by using data from another region to extrapolate findings to another; keeping in mind that such data will not be very accurate.

(5) Food category.
Depending on the type of food and how it is packed, we might only be able to use some FLW measurement methods, but not others. For example, when dealing with packed food items, wasted products can be simply counted and their weight inferred. Meanwhile, when waste occurs with liquid food, such as milk, volume assessment can be fairly accurate to estimate the weight of FLW.

(6) Direct access to FLW.
Some food waste cannot be measured directly, such as by weighing, counting, or waste composition analysis. For example, when waste is discarded directly into the drain in the process of food processing, it might be mixed with other waste water exiting the processing plant. In cases like this, non-direct methods need to be employed, such as modelling or mass balance.

To be able to suggest specific FLW practices according to the criteria described above, we need to first describe them in terms of these characteristics. For harmonious representation, we used already mentioned NUTS and NACE classifications for region of interest and FSC stage, respectively. We also used a simplified version of FAO's *Global individual food consumption* (GIFT; [6]) classification to describe food category. For accuracy, we opted for three categorical levels of "low", "medium", and "high", while direct access to FLW can be represented with a simple Boolean.

2.2 Extraction of Data

To test the extraction of data, we used 11 FLW measurement and reduction practice descriptions. This included 3 descriptions of practices developed and piloted in the WASTELESS project as well as 8 practices developed in other European projects [16].

To extract data from FLW practice descriptions, we used two LLMs: ChatGPT 5 Auto [12] and Le Chat [11]. The prompt consisted of the following:

(1) Introduction: general summary of the whole extraction process;
(2) Main instructions:
   (a) Information to be extracted: a list of questions, the answers to which represent the data that is to be extracted from the practice description;
   (b) Data types and values: a list of possible values and their types for each of the data field, including lists of NUTS and NACE codes and food categories;
   (c) Missing information: instructions on how to deal with missing, incomplete, or unclear data;
   (d) Format: description of the format of expected output (`.csv` data);
(3) Example:
   (a) Input: a short, synthetic description of a FLW practice;
   (b) Reasoning: values for all data fields and their relationship to original text, indicate missing values;
   (c) Output: the expected line of data output.

We included all reference classifications as `.csv` files as well as the *Guidance on FLW Quantification Methods* as a PDF.

Following this initial prompt, practice descriptions were uploaded one by one and the output saved. The lead author of this paper also extracted the same information from the descriptions manually.

2.3 Evaluation of Results

To evaluate the extraction of data by LLMs, we compared the output by these models to human annotations. Here, the cases of multiple possible values and missing data need to be considered. First, some characteristics can objectively contain several values. For example, a FLW measurement practice might be applicable to several FSC stages and more than one food category. Secondly, some data cannot be determined from the description of practice.

For a characteristic with more than one possible value, consider two subsets of all possible values (U): human annotations (H) and machine-extracted values (M). The following list gives the scores that were used in the evaluation for all possible relationships of these two sets.

- +2; when the subsets were equal, $H = M$.
- +1; when an LLM extracted more values than a human, but including those, $\emptyset \neq H \subset M \neq U$.
- 0; when the sets were overlapping, but neither contained the other, that is, there was a partial match in values, $H \cap M \neq \emptyset, H \not\subseteq M, M \not\subseteq H$.
- 0; when there was data available, but LLM extracted no information or returned all possible values, $\emptyset \neq H \subset U$, but $M = \emptyset$ or $M = U$.
- -1; when an LLM failed to extract all values that a human did, $U \supseteq H \supset M \neq \emptyset$.
- -2; when the subsets had no values in common, i.e., were disjoint, $H \cap M = \emptyset$.

Note that for simple true or false values, this list simplifies to the extreme cases; thus they were scored as $+2$ and -2, respectively.

The reasoning behind the scoring is that we prefer to describe a practice in broader terms, even if some extracted values are inapplicable, rather than miss a particular value. As an example, it is better to describe a practice as suitable for all food categories than missing the one that it is actually suitable for. Similarly, when no information is extracted, we can conservatively assume all values apply. In such a case, an LLM failed objectively, but it is not punished for it. In the worst case scenario, an LLM "extracted" or hallucinated some values, but they have nothing in common with human annotations; for this two points are deducted.

3 Results

To evaluate the extraction of data by the LLMs, we scored their answers as described in Section 2.3. We summarised these scores for each practice characteristic in Table 1, where shown are the sum of scores and the number of perfect scores, that is the number of times the LLM completely agreed with the human rater. The number of practices tested was 11, which is therefore the maximum number of perfect scores, while the maximum sum is 22.

Table 1: Agreement scores for each characteristic of a FLW practice between a human rater and two different LLMs. The sum of scores and the number of perfect extractions are shown.

Model	ChatGPT		Le Chat	
Metric	Sum	Perfect	Sum	Perfect
FLW method	13	8	3	5
Region	12	7	13	5
FSC stage	8	7	12	6
Accuracy	−2	4	−5	3
Food category	22	11	21	10
Direct access	6	7	14	9
Total	59	44	58	38

Both models achieved similar scores in total across all practice characteristics. ChatGPT did, however, perfectly agree with the human rating more often. Of all the characteristics, food category was the easiest for the LLMs to extract. This is a simple classification and usually, the type of food is mentioned explicitly. The FLW quantification method was inferred with moderate success. On the other hand, accuracy of methods was very poorly described.

4 Discussion

In this work, we have shown how using two LLMs, the data from unstructured FLW measurement and reduction practice descriptions can be extracted into structured data. We achieved satisfying if imperfect results.

The most important data point, which is the class of the FLW measurement method was extracted with moderate success. It needs to be pointed out that extracted information was not wildly inaccurate in most cases, despite of what the scores might suggest. For example, a method of tracking waste on a blockchain was classified as using records, where in fact, the data were collected with surveys before being, indeed, *recorded*. Similarly, one practice described weighing waste as it was collected in the wastebasket, while simultaneously taking photos of the material. Here, the true measurement method was direct weighing, but the LLMs classified it as waste composition analysis. By using photos, such an analysis could in theory be done, but was not in such case. Thus, to improve the relevance of the FLW measurement method, we might instead group them by some other characteristics. For example, we could drop the data field of direct access and instead consider groups of methods separated in terms of needing direct access to waste.

Food category, however, was very reliably extracted. This indicates that in the further process of the extracted data, we could make the best use of the food type. Accuracy of the method described was not extracted well, but this is most likely due to the subjectivity of this characteristic. The authors of FLW practice descriptions never explicitly addressed the question of accuracy, so it needed to be estimated roughly by other characteristics, such as the general accuracy of the FLW method class. This also suggests that a three-level accuracy is probably too fine grained and it should be described only as "low" and "high".

We should note that our evaluation only compares the performance of LLMs to manual extraction of data performed by a single person. It is expected that people would also differ in their extractions, i.e., would not achieve perfect inter-rater agreement. Thus, the evaluation should not be interpreted as how well the LLMs captured the "objective" truth.

With this process, LLMs enabled us to transform the descriptions from simple PDF files into structured CSV files in a semi-automatic way. In terms of the five-star rating of open data [9] which describes how to get from data in proprietary formats to linked open data, we thus increase their level from one star to three stars. We can extend this further and increase the rating of this data to five stars: publish truly linked data.

The first step that can follow directly the results of this work is to transform the structure described in Section 2.1

Listing 1: A snippet of the ontology in Turtle language [1]

```
@prefix : <http://purl.archive.org/fwo/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dbpedia: <http://dbpedia.org/resource/> .
@base <http://purl.archive.org/fwo/> .

<http://purl.archive.org/fwo> rdf:type owl:Ontology .

#####################  Classes  #####################
:FoodLossWasteMeasurementPractice rdf:type owl:Class ;
     rdfs:label "Food Loss and Waste Measurement
          Practice"@en .

:Region rdf:type owl:Class ;
        rdfs:label "A NUTS code of the region" ;
        owl:equivalentClass
             dbpedia:Nomenclature_of_Territorial_Units-
        _for_Statistics .

:FoodCategory rdf:type owl:Class ;
              rdfs:label "Food Category" .

:DairyAndEggs rdf:type owl:Class ;
              rdfs:subClassOf :FoodCategory ;
              rdfs:label "Dairy & Eggs" .

:Milk rdf:type owl:Class ;
      rdfs:subClassOf :DairyAndEggs ;
      rdfs:label "Milk" .
# ... more classes defined ...
################  Object Properties  ################
:hasTitle rdf:type owl:DatatypeProperty ;
    rdfs:domain :FoodLossWasteMeasurementPractice ;
    rdfs:range rdfs:Literal ;
    rdfs:label "with the title" .

:hasRegion rdf:type owl:ObjectProperty ;
    rdfs:domain :FoodLossWasteMeasurementPractice ;
    rdfs:range :Region ;
    rdfs:label "applied in regions" .

:hasFoodCategory rdf:type owl:ObjectProperty ;
        rdfs:domain :FoodLossWasteMeasurementPractice ;
        rdfs:range :FoodCategory ;
        rdfs:label "applicable to food categories" .

:hasAccuracy rdf:type owl:DatatypeProperty ;
    rdfs:domain :FoodLossWasteMeasurementPractice ;
    rdfs:range "low"^^xsd:string,
         "medium"^^xsd:string, "high"^^xsd:string .
```

into an ontology. We illustrate this idea in Listing 1 which encodes the characteristics as classes and how to connect these to an individual practice using object and datatype properties. Once we represent the structure like this, we can encode a specific instance of FLW measurement practice as:

```
:MyDairyWastePractice a
     :FoodLossWasteMeasurementPractice ;
    :hasTitle "Tracking Waste of Dairy in Slovenia" ;
    :hasFoodCategory :WholeMilk ;
    :hasAccuracy "high"^^xsd:string ;
    :hasRegion :SI0.
```

The data on FLW measurement practices can then be easily linked to other published data and the closest candidate ontology is the *Food Waste Ontology* by Stojanov et al. [15]. The dataset described by this ontology is already vast and is being extended through *FoodWasteEXplorer* [14]. By leveraging it, we plan to publish the practice descriptions as five-star data in future work.

Acknowledgments

This work was carried out as a part of the WASTELESS project which is funded by the European Union's Horizon Europe Research and Innovation programme under Grant Agreement No. 101084222.

References

[1] David Beckett, Tim Berners-Lee, Eric Prud'hommeaux, and Gavin Carothers. 2014. *RDF 1.1 Turtle. Terse RDF Triple Language*. Ed. by Eric Prud'hommeaux and Gavin Carothers. World Wide Web Consortium (W3C), (Feb. 25, 2014). Retrieved Aug. 29, 2025 from https://www.w3.org/TR/turtle/.

[2] European Parliament and Council of the European Union. 2003. Regulation (EC) no 1059/2003 of the European Parliament and of the Council. On the establishment of a common classification of territorial units for statistics (NUTS). *Official Journal of the European Union*, 154, 1, (June 21, 2003), 1–41. http://data.europa.eu/eli/reg/2003/1059/oj.

[3] European Parliament and Council of the European Union. 2008. Regulation (EC) no 1333/2008 of the European Parliament and of the Council. On food additives. Version 02008R1333-20240423. *Official Journal of the European Union*, 354, 16, (Dec. 16, 2008), 16–33. http://data.europa.eu/eli/reg/2008/1333/oj.

[4] European Parliament and Council of the European Union. 2006. Regulation (EC) no 1893/2006 of the European Parliament and of the Council. Establishing the statistical classification of economic activities NACE revision 2 and amending Council Regulation (EEC) no 3037/90 as well as certain EC regulations on specific statistical domains. *Official Journal of the European Union*, 393, (Dec. 20, 2006), 1–39, 1, (Dec. 20, 2006). http://data.europa.eu/eli/reg/2006/1893/oj.

[5] Food and Agriculture Organization of the United Nations and World Health Organization. 2019. General standard for food additives. Codex STAN 192-1995. (2019).

[6] Food and Agriculture Organization of the United Nations (FAO). 2022. *FAO/WHO GIFT. Global Individual Food Consumption Data Tool*. Retrieved Aug. 30, 2025 from https://www.fao.org/gift-individual-food-consumption/about/en.

[7] Craig Hanson et al. 2016. *Food Loss and Waste Accountingand Reporting Standard*. Version 1.0. World Resources Institute. ISBN: 978-1-56973-892-4.

[8] Craig Hanson et al. 2016. *Guidance on FLW Quantification Methods. Supplement to the Food Loss and Waste (FLW) Accounting and Reporting Standard*. World Resources Institute. ISBN: 978-1-56973-893-1.

[9] Tim Berners Lee. 2006. Linked data. Design issues. Version 2009-06-18. (July 27, 2006). https://www.w3.org/DesignIssues/LinkedData.html.

[10] Mitja Luštrek and Junoš Lukan. 2024. Practice Abstracts – batch 1 – early phase. Deliverable 6.2. Research rep. Jožef Stefan Institute. doi:10.5281/ZENODO.13503261.

[11] [SW] Mistral AI, Le Chat version November 2024, 2024. url: https://chat.mistral.ai/.

[12] [SW] OpenAI, ChatGPT version GPT-5, 2025. url: https://chatgpt.com/.

[13] Zhuang Qian, Wu Chen, and Giorgia Sabbatini. 2023. White book for FLW reduction, measurement, and monitoring practices. Deliverable 1.1. Research rep. Version 1.0. University of Southern Denmark, (Aug. 30, 2023). 116 pp. doi:10.5281/ZENODO.11065358.

[14] REFRESH. FoodWasteEXplorer. https://www.foodwasteexplorer.eu/about.

[15] Riste Stojanov, Tome Eftimov, Hannah Pinchen, Maria Traka, Paul Finglas, Drago Torkar, and Barbara Korousic Seljak. 2019. Food waste ontology. A formal description of knowledge from the domain of food waste. In *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, (Dec. 2019). doi:10.1109/bigdata47090.2019.9006254.

[16] Sustainable Food System Innovation Platform. Practice abstract inventory. Retrieved Sept. 1, 2025 from https://www.smartchain-platform.eu/en/practice-abstract-inventory.