frontiers in Artificial Intelligence

PERSPECTIVE
published: 23 February 2021
doi: 10.3389/frai.2020.621577

Using Natural Language Processing and Artificial Intelligence to Explore the Nutrition and Sustainability of Recipes and Food

Marieke van Erp[1]†, Christian Reynolds[2†], Diana Maynard[3], Alain Starke[4], Rebeca Ibáñez Martín[5], Frederic Andres[6], Maria C. A. Leite[7], Damien Alvarez de Toledo[6], Ximena Schmidt Rivera[8], Christoph Trattner[4], Steven Brewer[9], Carla Adriano Martins[10], Alana Kluczkovski[10], Angelina Frankowska[10], Sarah Bridle[10], Renata Bertazzi Levy[11], Fernanda Rauber[11], Jacqueline Tereza da Silva[10] and Ulbe Bosma[12]*

[1]KNAW Humanities Cluster, Amsterdam, Netherlands, [2]Centre for Food Policy, City, University of London, London, United Kingdom, [3]Natural Language Processing Group, Department of Computer Science, The University of Sheffield, Sheffield, United Kingdom, [4]Department of Information Science and Media Studies, University of Bergen, Bergen, Norway, [5]Meertens Institute (KNAW), Amsterdam, Netherlands, [6]National Institute of Informatics, Chiyoda-ku, Japan, [7]Department of Mathematics and Statistics, College of Arts and Sciences, University of South Florida, St. Petersburg, FL, United States, [8]Equitable Development and Resilience Research Group, Institute of Energy Futures, College of Engineering, Design and Physical Science, Brunel University London, Uxbridge, United Kingdom, [9]Text Mining Solutions Ltd., York, United Kingdom, [10]Department of Physics & Astronomy, Faculty of Science and Engineering, The University of Manchester, Manchester, United Kingdom, [11]University of São Paulo, São Paulo, Brazil, [12]International Institute of Social History (KNAW), Amsterdam, Netherlands

OPEN ACCESS

Edited by:
Tome Eftimov,
Institut Jožef Stefan (IJS), Slovenia

Reviewed by:
Stefano Campese,
Omnys SRL, Italy
Eftim Zdravevski,
Saints Cyril and Methodius University of Skopje, North Macedonia

***Correspondence:**
Marieke van Erp
marieke.van.erp@dh.huc.knaw.nl

†These authors have contributed equally to this work

Specialty section:
This article was submitted to AI in Food, Agriculture and Water, a section of the journal Frontiers in Artificial Intelligence.

Received: *26 October 2020*
Accepted: *10 December 2020*
Published: *23 February 2021*

Citation:
van Erp M, Reynolds C, Maynard D, Starke A, Ibáñez Martín R, Andres F, Leite MCA, Alvarez de Toledo D, Schmidt Rivera X, Trattner C, Brewer S, Adriano Martins C, Kluczkovski A, Frankowska A, Bridle S, Levy RB, Rauber F, Tereza da Silva J and Bosma U (2021) Using Natural Language Processing and Artificial Intelligence to Explore the Nutrition and Sustainability of Recipes and Food. Front. Artif. Intell. 3:621577. doi: 10.3389/frai.2020.621577

In this paper, we discuss the use of natural language processing and artificial intelligence to analyze nutritional and sustainability aspects of recipes and food. We present the state-of-the-art and some use cases, followed by a discussion of challenges. Our perspective on addressing these is that while they typically have a technical nature, they nevertheless require an interdisciplinary approach combining natural language processing and artificial intelligence with expert domain knowledge to create practical tools and comprehensive analysis for the food domain.

Keywords: natural language processing, semantic web, computational recipe analysis, food history, interdisciplinary, recommender systems, food science, food computing

INTRODUCTION

Today's big societal challenges are increasingly analyzed from a data-driven perspective (van Veenstra and Kotterink, 2017), while the universal pervasiveness of food and its inherent multidisciplinary nature (Deutsch and Miller, 2007) enable it as an accessible window into every culture and time period. Many global challenges are directly related to food, nutrition, and sustainability.[1] At least 6 of the UN's Sustainable Development Goals involve food (UN, 2015). The food system is linked to 30% of total greenhouse gas emissions (Mbow et al., 2019), and healthcare costs are increasing due to diet-related issues (Schulze et al., 2018; Branca et al., 2019); 60%+ of adults in the United Kingdom and

[1]Food analysis was an essential entry point to explore cultures in early anthropological works, for example, as a proxy to understand material cultures, material practices, taboos, and social relations and rituals. Later, food has become an important focus of analysis to theorize colonialism and the machinery of empire, globalization, and more recently, urbanization and political ecology.

United States are now obese or overweight (WHO, 2020). Food is also central to many countries' economies (11% of total employment in the US and the Netherlands (Hausmann et al., 2014; FAO, 2019)) and cultural heritage (Richard and Coste, 2020). The ability to research food and recipes can help us address the challenges of sustainable and healthy eating in diverse cultural contexts, particularly given the current need to move to a more plant-based diet (Willett et al., 2019).

However, making the information on food accessible is far from trivial. Analysis of digitized or digital recipes is a new and upcoming field of research, with publications linked to nutritional and health studies (Reinivuo et al., 2009; Trattner et al., 2017), computational linguistics (Jurafsky, 2015), computational gastronomy (Jain et al., 2015), shopping (Aiello et al., 2019), allergen detection (Alemany-Bordera et al. 2016; Amato and Cozzalino, 2020), and the Semantic Web (Haussmann et al., 2019). Several research challenges are at the crossroads of data engineering, intelligent food, and cooking recipes, as discussed at the recent IEEE DECOR@ICDE workshop series (Andres et al., 2020). Furthermore, contemporary recipe analysis is underdeveloped in terms of links to sustainability—beyond publications by Reynolds and collaborators (Reynolds, 2017, Reynolds, 2019; Quadros et al., 2019; Reynolds et al., 2019), Andres and collaborators (Andres et al., 2018; Fritzen et al., 2018; Andres et al., 2019; de Toledo et al., 2020; Toledo et al., 2020), Asano and Biermann (2019), and Herrera (2020).

We suggest that the reason this problem has not been addressed in an integrated manner is partly due to the complexity of linking environmental impact databases to food terminology, which is time-consuming without artificial intelligence (AI) and natural language processing (NLP) tools. It is only in the last few years that these methods have been applied to combining recipes, food texts, and other environmental, nutritional, and economic databases, but this work is still incipient.

In this article, we encourage an interdisciplinary approach to the exploration of nutrition and sustainability. We highlight challenges and opportunities of using AI to analyze the food domain through recipes and present use cases that form the basis of a collaborative movement to provide a multifaceted and data-driven analysis of nutrition and sustainability. First, we explore issues around collecting and integrating food, nutrition, and sustainability data. Second, we review the NLP and other AI methods currently employed in linking and analyzing these data sources. We conclude by discussing how such techniques can be used to engage and translate food challenges to stakeholders and forecast possible future applications such as novel kinds of recommender systems that encourage positive behavioral change.

Challenges With Food, Recipe, Nutrition, and Sustainability Data

A major challenge with recipe datasets is that no standard online collections exist yet due to restrictions in terms of use, language, etc. Previous research often uses proprietary materials through APIs including crawls from online recipe collections and databases such as epicurious.com, allrecipes.com, RecipeDB, CulinaryDB, Hawa World, TarlaDalal.com, and chefkoch.de (Shidochi et al., 2009; Ahn et al., 2011; Teng et al., 2012; Ahnert, 2013; Jain et al., 2015; Kusmierczyk et al., 2015; Bogojeska et al., 2016; Tallab and Alrazgan, 2016; Kikuchi et al., 2017; Sajadmanesh et al., 2017; Bagler and Singh, 2018; Chang et al., 2018; Min et al., 2018; Asano and Biermann, 2019; Batra et al., 2019; Trattner and Elsweiler, 2019; Herrera, 2020; Sharma et al., 2020). To be useful in practical applications, these untapped sources require structuring, linking, and analysis via NLP techniques.

High-quality nutrition databases are compiled by multiple global organizations (e.g., Louie et al., 2016; FAO/WHO, 2020; UK, 2020; USDA, 2020; EFSA, 2020; and RIVM, 2020) for their respective geographies. However, each has its own coding standards and hierarchy, making them inflexible, and thus time-consuming and difficult to combine, compare, or integrate. For example, the USDA has a large archive of its national nutritional recommendations organized chronologically, allowing researchers to investigate changes in nutritional recommendations across time. The FAO, on the other hand, organizes its data around global food systems with a strong mission to fight malnutrition and hunger and incorporate global UN programs. AI tools are already being used to link and reconcile databases, fill data gaps, and establish common ontological frameworks (Eftimov et al., 2016; Eftimov et al., 2017; Ispirova et al., 2017; Dooley et al., 2018; Ispirova et al., 2019; Ispirova et al., 2019; Popovski et al., 2019), but the problem is not trivial.

Sustainability data are less coherent and not consistently available. Over the last twenty years, databases of aggregated meta-analysis of life cycle analysis (LCA) studies have emerged, providing sustainability information linked to specific food products (e.g., climate change, water, land use, or biodiversity impacts). Additionally, there are paywalled or consultancy LCA databases. Although most of these follow standards (e.g., ISO14040/44 or BSI-PAS 2050), key aspects that influence the results, such as the scope of the study (e.g., cradle-to-gate and cradle-to-grave), functional units used (e.g. mass, volume, and calories), and assumptions made are not always clear or well-documented, make it difficult for nonexperts to interpret, use, and apply this content to more comprehensive studies like those about healthy and sustainable food. Recently, Ghose et al., (2019) and Ghose (2020) proposed NLP methods for semantic investigation of LCA databases. However, while the sustainability data might be available for individual ingredients, it is still rare for entire recipes. The computation of a recipe's sustainability data, such as its carbon footprint, includes taking into account the combination and volumes of different ingredients.

It is clear that while a number of knowledge sources are available, a major challenge is that there are no (or limited) direct links between and among nutrition, sustainability, and recipe databases, with differing levels of data granularity even in databases of the same type. Furthermore, there are often linguistic, conceptual, and terminological gaps between the different kinds of knowledge sources, and while in principle, the immense amount of data allows for very detailed views of specific knowledge domains; the lack of any interconnecting framework makes this information largely incommensurable across different dimensions.

Furthermore, most analysis is limited to small-scale manual efforts that do not have a temporal aspect, with little connection between quantitative and qualitative methods. Linking approaches in the Semantic Web sphere are, in some cases, more well-developed but mostly do not relate to the sustainability aspect, and they are more targeted at shopping and healthy recipe applications. Current applications typically also focus on digital data that are already at least semistructured and do not require complex NLP.

Challenges for NLP in Computational Recipe Analysis

Contemporary recipe analysis is a well-researched field (Reinivuo et al., 2009; Trattner et al., 2017). However, once recipes do not come from the same source document or are not digitally born, automatic recipe analysis becomes a complex problem for language technology tools (van Erp et al., 2018).

When analyzing older data, artifacts from the digitization process may insert errors in the text and units of measurement and language usage may differ according to the source, region, or time period. This needs to be addressed first to enable comparison between recipes over time and space. In this section, we present a use case on automatically analyzing sugar quantities from historical apple pie recipes to illustrate some of the challenges.

We analyzed apple pie recipes for Dutch, American, French, and German and found that differences in coverage of the sources, data access via the different portals,[2] and classification of recipes (as not all retrieved articles mentioning apple pie are indeed recipes) required tailoring the tools to each resource. Artifacts of the digitization process, such as Optical Character Recognition[3] errors hamper these processes, as not all characters are recognized correctly, rendering parts of a sentence or even entire documents unreadable (e.g., "% Pfund Zucker" for "¼ Pfund Zucker"). This is as yet an unsolved problem (van Strien et al., 2020).

Quantities can be expressed by numerals and fractions or spelled out, and units are expressed as metric, imperial, or other measurements such as teacups. Here, not just conversion tables but contextual knowledge are needed as, for example, teacups vary in size between North American and Europe. Quantities are also not always specified (e.g., "Honig oder Zucker nach Süße der Äpfel und Gusto" or "2 sucre") or are difficult to assess when recipes use preprocessed ingredients such as compote and/or ready-made pastry, with unknown sugar content. These problems often also exist with modern recipes.

Additionally, often recipes do not mention the number of portions produced, and it is unknown how often people eat pie and how big a portion they typically eat. Therefore, we could not automatically normalize these to a "per person" or "portion" quantity. One could use a typical portion size, calculated from similar recipes, or have a portion size based on calories, but this requires further analysis and transformation.

Even for contemporary, digital-born recipes enriched with structured data, quantity extraction is not trivial. In analyzing recipes from the American site Allrecipes.com and its British site Allrecipes.co.uk, we found that while both were ostensibly from the same organization, the webpage structure of the two domains was quite different; thus, different analysis scripts had to be created for each. It is not always easy to retrieve the publication dates of these recipes, making it difficult to correctly assess the recipe's publication date and to use it in recipe trend analysis.

Challenges of Analyzing Contemporary Recipes for Nutrition

In the field of nutrition, the presentation and analysis of recipes is usually done through "technical preparation sheets." Traditionally, these sheets contain a list of ingredients, culinary techniques, preparation times of the dishes, necessary equipment, and portion sizes. They also quantify the calorific value and macro (and micro) nutrients of each recipe. This latter quantification is carried out by manually linking to food composition tables or automatically with specific nutrition software (Tufts University, 2020). This not only enables the standardization of recipe preparation but also acts as a support tool for the composition of nutritionally balanced menus (see Akutsu et al., 2005). Another way of presenting and analyzing recipes in terms of nutrition is through the Nutriscore scale (Chantal et al., 2017), a European index for foods that was developed as part of the French Health Law in 2016. Its goal is to improve the nutritional information shown on food packages to help consumers make healthier purchases. NLP techniques are currently scarcely adopted by the professional nutritionist community, who rely largely on manual techniques. There are also commercial APIs (e.g., provided by Edamam.com or Spoonacular.com) that offer nutrition integration into recipes using NLP. These have found wide customer bases but are not widely used in the nutrition-practitioner community.

Similar critical NLP issues exist when linking recipes to nutritional data. First, many recipes are still only found in printed and handwritten books, and modern recipe books have much irregular formatting. Even modern recipes contain vernacular ingredients, cooking methods and cooking temperatures, and units (e.g. 25 g sprigs of mint, 1 slice of large kohlrabi, a bunch of coriander, 3 baby Brussels sprouts, and 1 pinch salt), and developing ways to normalize and interpret these is time-consuming. Likewise, geographic differences may cause ambiguity (e.g., United States and United Kingdom tablespoon size). Methods to handle noisy recipe data and process it efficiently have been discussed (Trattner et al., 2019), with pipelines used to predict nutrient values of recipes (Trattner and Elsweiler, 2017; Rokicki et al., 2018).

[2]For Dutch recipes, we used the Delpher.nl portal maintained by the Dutch National Library, for American, the Chronicling America portal maintained by the US Library of Congress, for French, Gallica maintained by the Bibliothèque Nationale de France, and for German, the Europeana Newspapers portal.

[3]Optical Character Recognition is the process of automatically converting text from an image (scanned) to machine readable format.

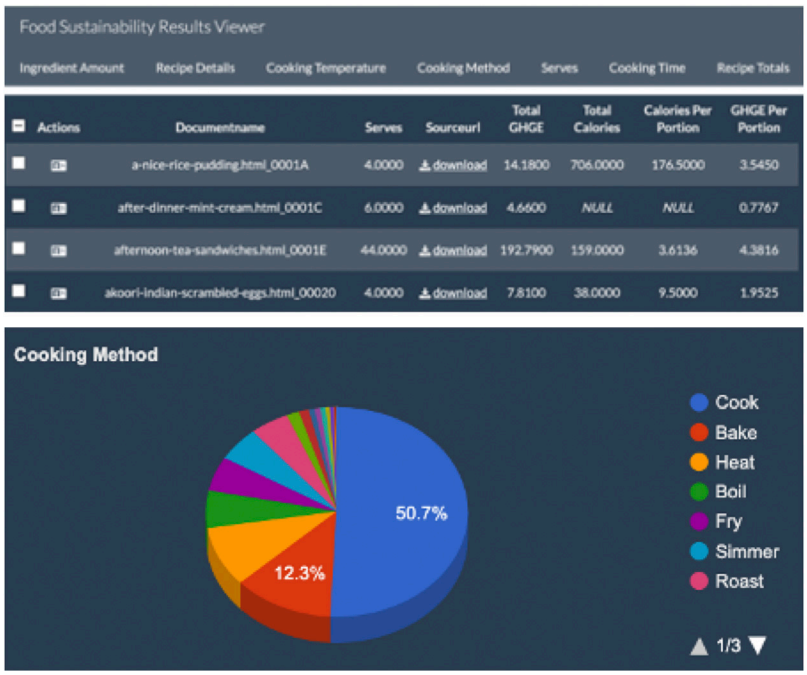

FIGURE 1 | Screenshots from the pilot food sustainability results viewer produced by Text Mining Solutions Ltd.

Challenges of Analyzing Recipes for Sustainability

Contemporary recipe analysis is underdeveloped in terms of links to sustainability. One method of assessing sustainability of a recipe is to link it to an existing quantified environmental measure. Multiple groups have now used some NLP to map specific measures such as greenhouse gas emission (GHGE) and nutrient variables to standard food classifications such as FoodEx2 (Eftimov et al., 2017; Mertens et al., 2019; Quadros et al., 2019; Reynolds et al., 2019). Additional web visualization tools examining the GHGE of foods have been developed for manual recipe analysis (US EPA, 2020; The Vegan Society, 2020).[4,5] One of the most advanced web tools is the NAHGAST Online Tool, which can provide economic, health, social, and environmental (material footprint, carbon footprint, water use, and land use) footprints of user submitted recipes (Speck et al., 2020). This has had 1,509 user-submitted recipes in the first research phase, with focus on empowering Out-of-Home Catering Sector users to reduce their impacts. Despite the proliferation of these web-based tools, for recipe exploration, little advantage has been taken of the full range of NLP's capabilities.

Reynolds and collaborators (in unpublished pilot work) collaborated with Text Mining Solutions Ltd. to map sustainability information to recipes. This pilot project used the GATE NLP toolkit (Cunningham et al., 2002), as the framework for extracting information extraction. They then applied environmental impact and calorie data to each ingredient per portion and calculated an overall figure for the recipes' footprints to understand the environmental impact and trade-offs of the recipes. Text Mining Solutions Ltd. also created a web visualization tool (see **Figure 1**) to enable citizen engagement and interactive exploration of trade-offs between recipes, sustainability, and nutrition.

Concurrently, Andres and collaborators (Andres, 2018; Andres et al., 2018; Andres et al., 2019; de Toledo et al., 2020; Toledo et al., 2020) created the CROPPER service (CaRbon fOotprint reciPe oPtimizER). CROPPER improves an input recipe by updating its ingredients and cooking procedures to reduce its carbon footprint while keeping it savoury. CROPPER is part of the CRWB group research[6] created in memory of the cook Nicole Andres (1942–2016) to manage her 60-year cooking recipe collection legacy and enhance it for data science research.

Asano and Biermann (2019) used NLP-driven recipe analysis to examine dietary transitions toward sustainable diets but did not link this to environmental impacts, examining instead the number and composition of vegetarian and vegan recipes submitted. Finally, Herrera (2020) used a recommender system to minimize food waste and recommend recipes using

[4]Take a Bite out of Climate Change (2020). Available at: https://www.takeabitecc.org/calculator.html (accessed October 23, 2020).

[5]Greenhouse Gas Footprint Calculator. Available at: https://www.tuco.ac.uk/ghgcalculator/index.html (accessed October 23, 2020).

[6]https://www.researchgate.net/project/Cooking-Recipes-Without-Border-Research-group.

(organic) locally grown food. Interestingly, this provides a link between recipes, supply chain, and modes of production.

When discussing recipe analysis, the environmental impacts of cooking cannot be ignored, accounting for as much as 61% of total emissions associated with specific foods (Frankowska et al., 2019). However, only de Toledo et al. (2020) provided a published framework for the combination and extraction of cooking data from recipes. This calculation needs additional data including cooking time, the energy consumption of home appliances (per kW h), and the carbon emissions (per kW h) related to the specific energy grid. To date, there has been no real-world application that calculates the environmental impacts of recipes, though various strands of research are underway from Reynolds, Andres, and Trattner, along with collaborators (see the "**Introduction**" section).

Future Directions

Our research opens up an avenue of new possibilities for food personalization and engagement in shifts toward healthy sustainable diets and cooking.

In particular, recommender technology can be integrated into current recipe websites and apps to improve support for users who wish to adopt healthier and/or more sustainable eating habits. A disadvantage of such personalized systems is that they typically reinforce existing eating habits (Starke, 2019), encouraging users to buy more of the same products rather than try healthier alternatives, and even so-called "persuasive" agent-based recommenders may still be based on existing lifestyle choices and social network activity (Palanca et al., 2014). NLP-based methods not only make it easier to compute the healthiness or sustainability of recipes but could also allow the design of personalized interventions that are rapidly explainable, updatable, and deployable, highlighting different categories that cater to different eating goals, such as health or sustainability. Additionally, Fritzen et al. (2018) have begun to focus on the integration of collective intelligence and AI via social networks to evaluate the gap between citizens' dish expectations and tasting experiences.[7]

A further limitation of current NLP recipe analysis is related to geographically contextualizing diets, nutrients, and food footprints, which is critical for global relevance. Current nutrient and environmental impact databases are not detailed enough to provide analysis and recommendations at different geographic levels (e.g. Western Europe and East Asia have very different requirements).

If adopted and implemented correctly, recipes analyzed and contextualized with NLP and linked to recommender systems will be useful to the general public as well as providing an analytical tool for specialists (including nutritionists, historians, chefs, educators, and policymakers). Enhancing recommender systems with multimedia capabilities (taste, texture, and smell) (Ghinea et al., 2011) could enable a better comprehension of recipes and target dishes. The food industry and supermarkets are obvious adopters of this technology through the Internet of Food Things;[8] while archives and libraries can use this technology to engage citizens with their collections. Government and nongovernment organizations can use this technology to monitor gastronomy, food culture, and dietary patterns and form comprehensive and adaptive policies.

CONCLUSION

Our perspective is that food and recipe research to help solve health and sustainability issues needs to be addressed in an interdisciplinary fashion, integrating NLP and other AI techniques with historical food research, food science, nutrition, and sustainability expertize. As outlined in this paper, multiple technical challenges still need to be solved. However, a purely technical approach is not sufficient: despite numerous advances in NLP, technology needs to be tightly interwoven with expert knowledge, highlighting the need for engagement with a wider interdisciplinary community. The collaborative work demonstrated in this paper shows that the combined viewpoints and expertize can make extremely encouraging steps toward addressing and resolving these critical issues.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article, and further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

ME was responsible for general setup, to abstract, introduction, and challenges for NLP in recipe analysis, article polishing, and submission. CR was responsible for conceptualisation and funding capture and revised and edited the manuscript. DM helped to shape both the idea and focus of this paper and consolidated the various components. She contributed to all sections, especially the introduction, discussion, and NP and AI aspects of the paper, as well as general editing. AS involved in conceptualisation and writing future directions, recommender systems, and NP challenges and review of the manuscript. RI was responsible for general setup, contextualisation of food as a relevant proxy for research, analysis of the different coding strategies of nutritional databases, and Section 2. FA 60-year CRWB cooking recipe collection and Flavorlens social network. ML DECOR and carbon footprint optimisation. DA carbon footprint optimisation and Nutriscore evaluation in recipes. XR involved in conceptualisation and writing of climate challenges, LCA database review, and review of the manuscript. CT was responsible for data, challenges, and

[7]https://www.researchgate.net/project/Big-Data-LOD-benchmark-meets-Intelligent-Food-and-Cooking-Recipe.

[8]https://www.foodchain.ac.uk/

analysis and commented on successive drafts of the manuscript. CM contributed to challenges and future directions and has critically revised the manuscript. AK, AF, SB, RB, FR, and JS contributed to conceptualisation and writing of nutrition and climate challenges and review of the manuscript. UB contributed to conceptualisation and apple pie use case.

FUNDING

This research activity was funded through multiple research grants from Research Councils UK, the University of Manchester, the University of Sheffield, the STFC Food Network+ and the HEFCE Catalyst-funded N8 AgriFood Resilience Programme with matched funding from the N8 group of Universities. Specific named projects that funded this research include the the AHRC funded AHRC US-UK Food Digital Scholarship Network (Grant Reference: AH/S012591/1), STFC GCRF funded project "Trends in greenhouse gas emissions from Brazilian foods using GGDOT" (ST/S003320/1), the STFC funded project "Piloting Zooniverse for food, health and sustainability citizen science" (ST/T001410/1), and the STFC Food Network+ Awarded Scoping Project "Piloting Zooniverse to help us understand citizen food perceptions". Funding was also supplied from the ESRC via the University of Sheffield Social Sciences Partnerships, Impact and Knowledge Exchange fund for "Recipe environmental impact calculator"; and through Research England via the University of Sheffield QR Strategic Priorities Fund projects "Cooking as part of a Sustainable Food System – creating an wider evidence base for policy makers", and "Food based citizen science in the UK as a policy tool". This research project arose from the N8 AgriFood-funded project "Greenhouse Gas and Dietary choices Open-source Toolkit (GGDOT) hacknights.' Ximena Schmidt Rivera was supported through Brunel University internal Research England GCRF QR Fund. Alana Kluczkovski and Carla Adriano Martins were supported through The University of Manchester GCRF QR Visiting Researcher Fellowship. Andres and colleagues would like to express their deepest appreciation to the National Institute of Informatics, Japan for the ongoing research support.

REFERENCES

Ahn, Y.-Y., Ahnert, S. E., Bagrow, J. P., and Barabási, A.-L. (2011). Flavor network and the principles of food pairing. *Sci. Rep.* 1, 196. doi:10.1038/srep00196

Ahnert, S. E. (2013). Network analysis and data mining in food science: the emergence of computational gastronomy. *Flavour* 2, 4. doi:10.1186/2044-7248-2-4

Aiello, L. M., Schifanella, R., Quercia, D., and Del Prete, L. (2019). Large-scale and high-resolution analysis of food purchases and health outcomes. *EPJ Data Sci.* 8, 14. doi:10.1140/epjds/s13688-019-0191-y

Akutsu, R. de C., Botelho, R. A., Camargo, E. B., Sávio, K. E. O., and Araújo, W. C. (2005). A ficha técnica de preparação como instrumento de qualidade na produção de refeições (The technical cards as quality instrument for good manufacturing process). *Rev. Nutr.* 18, 277–279. doi:10.1590/S1415-52732005000200012

Alemany-Bordera, J., Heras Barberá, S. M., Palanca Cámara, J., and Julian Inglada, V. J. (2016). Bargaining agents based system for automatic classification of potential allergens in recipes. *ADCAIJ: Adv. Distr. Comput. Artif. Intell. J.* 5 (2), 43–51.

Amato, A., and Cozzolino, G. (2021). "SafeEat: extraction of information about the presence of food allergens in recipes," in *Advances in intelligent networking and collaborative systems. INCoS 2020. Advances in intelligent systems and computing*. Editors L. Barolli, K. Li, and H. Miwa (Cham: Springer). doi:10.1007/978-3-030-57796-4_19

Andres, F., d Orazio, L., Papin, J., and Irvin Grosky, W. (2018). "The cooking recipes without border data set: FAIR challenges," in International workshop on data science—present and future of open data and open science, Mishima, Shizuoka, Japan.

Andres, F., Ghinea, G., Grosky W., W., and Leite, M.C.A. (2020). "Overview of the 3rd DECOR workshop," in 2020 IEEE 36th international conference on data engineering workshops. Dallas, TX: ICDEW, 162–164. doi:10.1109/ICDEW49219.2020.9123828

Andres, F., Jouseau, R., and Ly, T. (2019). *Cooking recipes without border ecosystem project*. Poster, NII Open House. doi:10.13140/rg.2.2.27794.50886

Andres, F. (2018). "The CRWB rsbench: toward a cooking recipe benchmark initiative," in IEEE 34th international conference on data engineering workshops (ICDEW), 154–156. doi:10.1109/ICDEW.2018.00032

Asano, Y. M., and Biermann, G. (2019). Rising adoption and retention of meat-free diets in online recipe data. *Nat. Sustain* 2, 621–627. doi:10.1038/s41893-019-0316-0

Bagler, G., and Singh, N. (2018). "Data-driven investigations of culinary patterns in traditional recipes across the world," in *IEEE 34th international conference on data engineering workshops*, 157–162. doi:10.1109/ICDEW.2018.00033

Batra, D., Diwan, N., Upadhyay, U., Kalra, J. S., Sharma, T., Sharma, A. K., et al. (2019). Recipedb: a resource for exploring recipes. *SSRN J.* 33, 14–19. doi:10.2139/ssrn.3482237

Bogojeska, A., Kalajdziski, S., and Kocarev, L. (2016). "Processing and analysis of Macedonian cuisine and its flavors by using online recipes," in *ICT Innovations 2015 advances in intelligent systems and computing*. Editors S. Loshkovska and S. Koceski (Cham: Springer International Publishing), 143–152. doi:10.1007/978-3-319-25733-4_15

Branca, F., Lartey, A., Oenema, S., Aguayo, V., Stordalen, G. A., Richardson, R., et al. (2019). Transforming the food system to fight non-communicable diseases. *BMJ.* 364, l296. doi:10.1136/bmj.l296

Chang, M., Guillain, L. V., Jung, H., Hare, V. M., Kim, J., and Agrawala, M. (2018). "Recipescape: an interactive tool for analyzing cooking instructions at scale," in Proceedings of the 2018 CHI conference on Human Factors in computing systems—CHI'18. New York, (New York, NY: ACM Press), 1–12. doi:10.1145/3173574.3174025

Chantal, J., and Hercberg, S.World Health Organization (2017). Development of a new front-of-pack nutrition label in France: the five-color Nutri-Score. *Public Health Panor.* 3 (04), 712–725. doi:10.1016/s2468-2667(18)30009-4

Cunningham, H., Maynard, D., Bontcheva, K., and Tablan, V. (2002). "GATE: a framework and graphical development environment for robust NLP tools and applications, GATE: a framework and graphical development environment for robust NLP tools and applications," in Proceedings of the 40th anniversary meeting of the association for computational linguistics (ACL'02). Philadelphia.

de Toledo, D. A., d Orazio, L., Andres, F., and Leite, M. C. A. (2020). "Cooking related carbon footprint evaluation and optimisation," in *ADBIS, TPDL and EDA 2020 common workshops and doctoral consortium: international workshops: DOING, MADEISD, SKG, BBIGAP, SIMPDA, aiminscience 2020 and doctoral consortium, lyon, France, august 25–27, 2020, proceedings Communications in computer and information science*. Editors L. Bellatreche, M. Bieliková, O. Boussaïd, B. Catania, J. Darmont, E. Demidova, et al. (Cham: Springer International Publishing), 122–128. doi:10.1007/978-3-030-55814-7_10

Deutsch, J., and Miller, J. (2007). Food studies: a multidisciplinary guide to the literature. *Choice Rev. Online.* 45, 393–401. doi:10.5860/choice.45.03.393

Dooley, D. M., Griffiths, E. J., Gosal, G. S., Buttigieg, P. L., Hoehndorf, R., Lange, M. C., et al. (2018). FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. *NPJ Sci. Food* 2, 23. doi:10.1038/s41538-018-0032-6

EFSA (2020). The EFSA comprehensive European food consumption database. Available at: https://www.efsa.europa.eu/en/food-consumption/comprehensive-database (Accessed October 22, 2020).

Eftimov, T., Korošec, P., and Koroušić Seljak, B. (2017). StandFood: standardization of foods using a semi-automatic system for classifying and describing foods according to FoodEx2. *Nutrients* 9, 121–148. doi:10.3390/nu9060542

Eftimov, T., Koroušić Seljak, B., and Korošec, P. (2016). "Grammar and dictionary based named-entity linking for knowledge extraction of evidence-based dietary recommendations," Proceedings of the 8th international Joint conference on knowledge discovery, knowledge engineering and knowledge management. London: SCITEPRESS - Science and Technology Publications, 150–157. doi:10.5220/0006032401500157

FAO (2019). *World food and agriculture—Statistical pocketbook 2019*. Rome.

FAO/WHO (2020). FAO/WHO GIFT|Food and agriculture organization of the united nations. FAO/WHO GIFT|global individual Food consumption data tool. Available at: http://www.fao.org/gift-individual-food-consumption/en/ (Accessed October 22, 2020).

Frankowska, A., Reynolds, C., Bridle, S., Rauber, F., da Silva, J., Kluczkovski, A., et al. (2019). How do UK cooking methods contribute to climate change? *Clim. Change Law Collect* 7, 111–178. doi:10.1163/9789004322714_cclc_2016-0148-022

Fritzen, A., Andres, F., and Leite, M. (2018). "Introducing flavorlens: a social media platform for sharing dish observations,"in Proceedings of the 3rd international Workshop on Multisensory Approaches to Human-food interaction (MHFI'18). New York, NY, USA: Association for Computing Machinery. doi:10.1145/3279954.3279961

Ghinea, G., Andres, F., and Gulliver (2011). *Multiple sensorial media advances and applications: new developments in MulSeMedia*. London: IGI publisher. doi:10.4018/978-1-60960-821-7 344p

Ghose, A., Lissandrini, M., Weidema, B. P., and Hose, K. (2019). *An open source dataset and Ontology for product footprinting (Awarded best poster—ESWC 2019)*. Unpublished. doi:10.13140/rg.2.2.21495.16800

Ghose, A. (2020). *Modeling LCA data on the semantic web*. (LCA Food).

Hausmann, R., Hidalgo, C. A, Bustos, S., Coscia, M., Chung, S., Jimenez, J., et al. (2014). *The Atlas of economic complexity—Mapping paths to prosperity*. Berlin: Springer.

Haussmann, S., Seneviratne, O., Chen, Y., Ne'eman, Y., Codella, J., Chen, C.-H., et al. (2019). "FoodKG: a semantics-driven knowledge graph for food recommendation," in *The semantic web – ISWC 2019: 18th international semantic web conference, auckland, New Zealand, october 26–30, 2019, proceedings, part II Lecture notes in computer science*. Editors C. Ghidini, O. Hartig, M. Maleshkova, V. Svátek, I. Cruz, A. Hogan, et al. (Cham: Springer International Publishing), 146–162. doi:10.1007/978-3-030-30796-7_10

Herrera, J. C. S. (2020). *Sustainable recipes. A food recipe sourcing and recommendation system to minimize food Miles*.arXiv

Ispirova, G., Eftimov, T., Korošec, P., and Koroušić Seljak, B. (2019). MIGHT: statistical methodology for missing-data imputation in food composition databases. *Appl. Sci* 9, 4111. doi:10.3390/app9194111

Ispirova, G., Eftimov, T., Koroušić Seljak, B., and Korošec, P. (2017). "Mapping food composition data from various data sources to a domain-specific ontology," in *Proceedings of the 9th international Joint conference on knowledge discovery, knowledge engineering and knowledge management*. London: SCITEPRESS - Science and Technology Publications, 203–210. doi:10.5220/0006504302030210

Jain, A., N KR.and Bagler, G. (2015). Analysis of food pairing in regional cuisines of India. *PLoS One* 10, e0139539. doi:10.1371/journal.pone.0139539

Jurafsky, D. (2015). *The language of food: a linguist reads the menu*. Berlin: Springer.

Kikuchi, Y., Kumano, M., and Kimura, M. (2017). "Analyzing dynamical activities of Co-occurrence patterns for cooking ingredients," in IEEE international conference on data mining workshops, 17. doi:10.1109/ICDMW.2017.10

Kusmierczyk, T., Trattner, C., and Nørvåg, K. (2015). "Temporality in online food recipe consumption and production," in Proceedings of the 24th international conference on world wide web - WWW' '15 companion. New York, New York, USA: ACM Press, 55–56. doi:10.1145/2740908.2742752

Louie, J. C. Y., Barclay, A. W., and Brand-Miller, J. C. (2016). Assigning glycemic index to foods in a recent Australian food composition database. *Eur. J. Clin. Nutr*. 70, 280–281. doi:10.1038/ejcn.2015.186

Mbow, C., Rosenzweig, C., Barioni, L. G., Benton, T. G., Herrero, M., Krishnapillai, M., et al. (2019). "Chapter 5: Food Security," in Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. IPCC).

Mertens, E., Kaptijn, G., Kuijsten, A., van Zanten, H., Geleijnse, J. M., and van 't Veer, P. (2019). SHARP-Indicators Database toward a public database for environmental sustainability. *Data Brief*. 27, 104617. doi:10.1016/j.dib.2019.104617

Min, W., Bao, B.-K., Mei, S., Zhu, Y., Rui, Y., and Jiang, S. (2018). You are what You eat: exploring rich recipe information for cross-region food analysis. *IEEE Trans. Multimed* 20, 950–964. doi:10.1109/TMM.2017.2759499

Palanca, J., Heras, S., Botti, V., and Julián, V. (2014). "receteame.com: a persuasive social recommendation system," in *Advances in practical applications of heterogeneous multi-agent systems. The PAAMS collection. PAAMS 2014. Lecture notes in computer science* Editors Y. Demazeau, F. Zambonelli, J.M. Corchado, and J. Bajo (Cham: Springer). doi:10.1007/978-3-319-07551-8_40

Popovski, G., Seljak, B., and Eftimov, T. (2019). "FoodOntoMap: linking food concepts across different food ontologies," in Proceedings of the 11th international Joint conference on knowledge discovery, knowledge engineering and knowledge management. London: SCITEPRESS - Science and Technology Publications, 195–202. doi:10.5220/0008353201950202

Quadros, V. P., Balcerzak, A., Sousa, R. F., Ferrari, M., Schmidt Rivera, X., Reynolds, C. J., et al. (2019). *Using individual food consumption data to estimate the environmental impact of diets: the potentiality of the FAO/WHO GIFT platform*. Rome: The American University of Rome Graduate School.

Reinivuo, H., Bell, S., and Ovaskainen, M.-L. (2009). Harmonization of recipe calculation procedures in European food composition databases. *J. Food Compos. Anal*. 22, 410–413. doi:10.1016/j.jfca.2009.04.003

Reynolds, C. J. (2017). Energy embodied in household cookery: the missing part of a sustainable food system? Part 1: a method to survey and calculate representative recipes. *Energy Proc* 123, 220–227. doi:10.1016/j.egypro.2017.07.245

Reynolds, C., Rivera, X. C. S., Frankowska, A., Kluczkovski, A., Da Silva, J. T., Bridle, S., et al. (2019). *A pilot method linking greenhouse gas emission databases to the FoodEx2 classification*. Unpublished. doi:10.13140/rg.2.2.15990.34889

Reynolds, C. (2019). "Sustainable Gastronomy; power and energy use in food—Is it possible to fight climate change though cookery?," in *Proceedings of the Oxford food Symposium*. Editor M. McWilliams (London: Prospect Books).

Richard, M., and Coste, M. (2020). Food is Culture—EU policy brief on food and cultural heritage. *Eur. Nostra Slow Food*. 16, 324–388. doi:10.5040/9781474296250.0031

RIVM (2020). *Dutch food composition database|RIVM*. Available at: https://www.rivm.nl/en/dutch-food-composition-database. (Accessed October 22, 2020).

Rokicki, M., Trattner, C., and Herder, E. (2018). *The impact of recipe features, social cues and demographics on estimating the healthiness of online recipes*. New York, NY: ICWSM.

Sajadmanesh, S., Jafarzadeh, S., Ossia, S. A., Rabiee, H. R., Haddadi, H., Mejova, Y., et al. (2017). "Kissing cuisines: exploring worldwide culinary habits on the web," in Proceedings of the 26th international conference on world wide web companion—WWW'17 companion. New York, New York, USA: ACM Press, 1013–1021. doi:10.1145/3041021.3055137

Schulze, M. B., Martínez-González, M. A., Fung, T. T., Lichtenstein, A. H., and Forouhi, N. G. (2018). Food based dietary patterns and chronic disease prevention. *BMJ* 361, k2396. doi:10.1136/bmj.k2396

Sharma, T., Upadhyay, U., Kalra, J., Arora, S., Ahmad, S., Aggarwal, B., et al. (2020). "Hierarchical clustering of world cuisines (ICDEW) (IEEE)-104" in 2020 IEEE 36th international conference on data engineering workshops. doi:10.1109/ICDEW49219.2020.00007

Shidochi, Y., Takahashi, T., Ide, I., and Murase, H. (2009). "Finding replaceable materials in cooking recipe texts considering characteristic cooking actions." in Proceedings of the ACM multimedia 2009 workshop on Multimedia for cooking and eating activities—CEA'09. New York, NY, USA: ACM Press. doi:10.1145/1630995.1630998

Speck, M., Bienge, K., Wagner, L., Engelmann, T., Schuster, S., Teitscheid, P., et al. (2020). Creating sustainable meals supported by the NAHGAST online tool—approach and effects on GHG emissions and use of natural resources. *Sustainability* 12 (3), 1136. doi:10.3390/su12031136

Starke, A. (2019). RecSys Challenges in achieving sustainable eating habits. in (Copenhagen: HealthRecSys'19).

Tallab, S. T., and Alrazgan, M. S. (2016). Exploring the food pairing hypothesis in arab cuisine: a study in computational gastronomy. *Proc. Comput. Sci.* 82, 135–137. doi:10.1016/j.procs.2016.04.020

Teng, C.-Y., Lin, Y.-R., and Adamic, L. A. (2012). "Recipe recommendation using ingredient networks," in Proceedings of the 3rd Annual ACM web science conference on—WebSci'12. New York, New York, USA: ACM Press, 298–307. doi:10.1145/2380718.2380757

The Vegan Society. (2020). Carbon food calculator|the vegan society. Available at: https://www.vegansociety.com/take-action/campaigns/plate-planet/carbon-calculator. (Accessed October 24, 2020).

Toledo, D. A. D., Truffat, I., Andres, F., D'Orazio, L., and Leite, M. C. A. (2020). *Cooking recipes: how to improve carbon footprint? Poster*. NII open house June 2020. doi:10.13140/rg.2.2.18015.89761/1

Trattner, C., and Elsweiler, D. (2017). *Food recommender systems: important contributions, challenges and future research directions*. arXiv preprint arXiv: 1711.02760.

Trattner, C., Elsweiler, D., and Howard, S. (2017). Estimating the healthiness of Internet recipes: a cross-sectional study. *Front Public Health* 5, 16. doi:10.3389/fpubh.2017.00016

Trattner, C., and Elsweiler, D. (2019). What online data say about eating habits. *Nat. Sustain* 2, 545–546. doi:10.1038/s41893-019-0329-8

Trattner, C., Kusmierczyk, T., and Nørvåg, K. (2019). Investigating and predicting online food recipe upload behavior. *Inf. Process. Manag* 56, 654–673. doi:10.1016/j.ipm.2018.10.016

Tufts University (2020). Calculating calories and nutrients in meals - Jean mayer USDA Human nutrition research center on aging. Available at: https://hnrca.tufts.edu/flipbook/resources/restaurant-meal-calculator/. (Accessed: October 23, 2020).

UK, GOV. (2020). Composition of foods integrated dataset (CoFID) - GOV.UK. Available at: https://www.gov.uk/government/publications/composition-of-foods-integrated-dataset-cofid (Accessed October 22, 2020).

UN (2015). *Transforming our world: the 2030 Agenda for Sustainable Development. Division for Sustainable Development Goals*. New York: Springer.

US EPA (2020). Carbon footprint calculator | climate change | US EPA. Available at: https://www3.epa.gov/carbon-footprint-calculator/. (Accessed October 24, 2020).

USDA (2020). FoodData central. Available at: https://fdc.nal.usda.gov/api-guide.html. (Accessed October 22, 2020).

van Erp, M., Wevers, M., and Huurdeman, H. (2018). "Constructing a recipe web from historical Newspapers," in *The semantic web—ISWC 2018: 17th international semantic web conference, monterey, CA, USA, october 8–12, 2018, proceedings, part I Lecture notes in computer science*. Editors D. Vrandečić, K. Bontcheva, M. C. Suárez-Figueroa, V. Presutti, I. Celino, M. Sabou, et al. (Cham: Springer International Publishing), 217–232. doi:10.1007/978-3-030-00671-6_13

van Strien, D., Beelen, K., Ardanuy, M., Hosseini, K., McGillivray, B., and Colavizza, G. (2020). "Assessing the impact of OCR quality on downstream NLP tasks," in Proceedings of the 12th international conference on agents and artificial intelligence. London: SCITEPRESS - Science and Technology Publications, 484–496. doi:10.5220/0009169004840496

van Veenstra, A. F., and Kotterink, B. (2017). "Data-driven policy making: the policy Lab approach," in *Electronic participation lecture notes in computer science*. Editors P. Parycek, Y. Charalabidis, A. V. Chugunov, P. Panagiotopoulos, T. A. Pardo, Ø. Sæbø, et al. (Cham: Springer International Publishing), 100–111. doi:10.1007/978-3-319-64322-9_9

WHO (2020). WHO|global database on body mass index (BMI). Available at: https://www.who.int/nutrition/databases/bmi/en/. (Accessed: October 14, 2020).

Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., et al. (2019). Food in the Anthropocene: the EAT-Lancet Commission on healthy diets from sustainable food systems. *Lancet* 393, 447–492. doi:10.1016/S0140-6736(18)31788-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.