Sommelierselects.com

https://dylanamadan.wixsite.com/wine-collect-select

Yonathan Adye
Dylan J. Goode
Djurre Hoogterp
Sjoerd F. Kelder
Réka Szuromi
UvA
Amsterdam

1 OVERVIEW

This report presents the design and implementation of Sommelierselects.com, an information organisation project aimed at serving the diverse needs of wine enthusiasts. The system is designed to catalog a static dataset, providing a snapshot of the wine domain at a particular time, while also being flexible enough to accommodate future updates. The project is motivated by the need for a comprehensive and user-friendly platform that can cater to casual drinkers and professional sommeliers alike.

2 WHAT IS BEING ORGANISED?

Our research into the intricate realm of wine underscored the need for an unambiguous and robust data dictionary (see figure 1). This was primarily because wines, inherently detailed in their descriptions, require clarity and uniformity to facilitate meaningful user interactions, irrespective of the user's depth of knowledge in the domain (Spence & Wang, 2018).

We recognize three primary categories of resources: physical, digital, and informational. The physical resources pertain to the actual wine bottles that users may be interested in. Digital resources encompass the wine reviews, ratings, and user profiles that are stored and accessed electronically. The informational resources provide details about the wines, including aspects like their origins and grape varieties.

Interestingly, the concept of "time as a resource" is also intertwined with our domain. Wines have a distinct relationship with time due to their vintages and the ageing process. This relationship influences their taste, value, and overall desirability (Echave et al., 2021). Though the original concept emphasised the tangible placement of resources, in our domain, this idea is metaphorically associated with wines. Specifically, a wine's vintage year becomes a significant marker, critically shaping its life cycle, taste, and subsequent value (Echave et al., 2021).

The organising system we designed primarily catalogues an existing and closed resource collection. We made a conscious decision to utilise a static dataset, which, while not capturing the ongoing evolution of wine trends or the latest reviews, offers a consistent and stable reference point. This choice allows users to explore the dataset as a snapshot of a particular time. However, the platform has been structured to accommodate the systematic updating of wine details, the addition of new wines, or the removal of outdated entries. Although the dataset remains static for now, we recognize the dynamic nature of wine reviews and ratings. Therefore, we have incorporated the understanding of the Resource Life Cycle to ensure the potential for periodic updates and reviews in the future.

In the process of organising and structuring our dataset, it became evident that the diverse range of wine reviews contained both unique and overlapping characteristics. Each wine review, influenced by factors such as region and grape type, possesses distinct attributes that set it apart. We ensured that every wine, regardless of similarities it might share with others, was acknowledged as a distinct and non-duplicable entity in our database. Furthermore, the wine domain, replete with its specialised terminology, made it important to establish consistency in language and descriptors. To address this, we adopted a controlled vocabulary approach, which not only ensured that terms were standardised across the platform but also augmented its usability by minimising potential ambiguities.

However, to streamline navigation and enhance the user experience, it was equally important to group similar wines. Thus, adhering to the Principle of Categorisation, wines were systematically grouped into broader categories based on shared attributes like type (e.g., red, white, rosé), region (e.g., Bordeaux, Napa Valley), or grape variety (e.g., Chardonnay, Merlot). This approach facilitated user navigation, allowing them to easily traverse through broader categories while still preserving the unique identity of each wine.

We have noted the significance of the life cycle of the reviews within our repository. Reviews, like many other forms of data, can undergo transformations or even become obsolete over a period of time. Therefore, it is essential to regularly assess this life cycle to maintain the relevancy and vibrancy of the platform. This ongoing evaluation is fundamental in ensuring that the platform consistently delivers current and valuable insights to its users.

3 WHY IS IT BEING ORGANISED?

The primary motivation behind organising the wine data on our website is to create a comprehensive and user-friendly platform that caters to the diverse needs of wine enthusiasts, ranging from casual drinkers like Emily (Figure 3) to professional sommeliers like David (Figure 2). The project stems from the recognition of the vast variety of wines available globally, with more than 10,000 wine grape varieties, albeit only a few dozen have achieved widespread popularity and acclaim (Masterclass, 2021). This diversity, while enriching, can potentially overwhelm casual drinkers, necessitating a platform that facilitates easy discovery and understanding of wines from various regions, varieties, and vintages.

The exploration of new wines is a notable trend among consumers, as suggested by the resurgence in white winemaking in Washington, indicative of an interest in exploring new varieties (Washington Tasting Room, 2021). This aligns with the persona of Emily, who, as a casual drinker, might find the platform a conducive space to expand her palate by exploring new wines. On the other hand, the persona of David, a seasoned sommelier, resonates with the segment of consumers who exhibit a more pronounced variety-seeking behaviour, as he might be keen on diving deep into the nuanced distinctions between various wines, regions, and vintages (Caracciolo et al., 2022).

The platform is designed to support interactions such as personalised wine recommendations and dataset exploration. These features are likely to cater to the distinct needs of Emily and David. While Emily might find personalised recommendations exceedingly helpful in navigating the extensive wine variety, David might appreciate the dataset exploration feature, enabling him to delve deeper into different wine categories.

Moreover, the platform's potential utility extends beyond personal use to institutional applications. It is positioned to be a valuable tool for restaurants, wine sellers, or educational institutions offering courses in hospitality or wine tasting. This dual utility underscores the platform's scalability and its capability to serve a broad spectrum of wine enthusiasts, irrespective of their level of expertise or professional affiliation.

Continual user feedback and behavior analysis are integral to the adaptive nature of our platform, aiding in the periodic update of user personas and, by extension, the refinement of our information architecture. This iterative approach underscores our commitment to not only meeting the current needs of our diverse user base but also anticipating and adapting to their evolving preferences and the ever-expanding world of wines.

4 HOW MUCH IS IT BEING ORGANISED?

The prioritisation and arrangement of data are crucial considerations for websites that seek to provide information to consumers. The arrangement of information on a wine recommendation website is of significant importance in improving user experience, facilitating search functionality, and enabling users to make well-informed judgements on their wine choices. This section examines the extent to which the data in this dataset is organised. It mainly addresses the level of description, depth, and organisation that has been implemented for each resource. As well as, the classifications, the impact of resource description and arrangement on user interactions.

The wine dataset demonstrates a significant variation in the level of detail and structure used for each resource. The dataset exhibits a strong and effective arrangement of resources into distinct groups or categories, which is determined by essential criteria in order to optimise the organisation of information. In the following paragraph, each important category for user interaction will be clearly explained with regard to the organisation of the data.

(1) Classification by Country: The wines are classified based on their country of origin, serving as a primary organising concept.
(2) Province: Further granularity is achieved through the categorization of wines by province, contributing to the dataset's organisation.
(3) Regional Classification: By going a step further from the province, the wines are also classified by their specific region of origin.
(4) Winery: The organisation also encompasses the winery as a datapoint, allowing users to delve into wines produced by particular producers.
(5) Variety and Colour: The significance of variety and colour in the classification of wines cannot be overstated, as they allow consumers to effectively sort and choose wines based on their preferred grape type and its associated wine colour.
(6) Price: Every wine has a price per bottle associated with it which is included in the database.
(7) Points: Reviewers award points to the wines they reviewed, these points are included in our data collection.

Variable Name	Measurement Units	Allowed Units	Definitions of the Variable	Description (optional)
Country	Text	Any valid country name	Country is a distinct territorial body or political entity recognized as an independent nation.	Represents the nation from which the wine originates, providing a macro-level understanding of the wine's geographical origin.
Province	Text	Any valid province name	Province is a principal administrative division of certain countries or empires.	Narrows down the wine's origin to a specific province or state within the country.
Region	Text	Any valid region name within a province or state	Region_1 is a part of a province or state, having definable boundaries or characteristics.	A further specification of the wine's origin, indicating the wine-growing area within a province or state.
Winery	Text	Any valid winery name	Winery is the name of the winery that is responsible for the creation of the wine.	A qualitative assessment of the winery that made the wine.
Description	Text	Any valid textual description	Description is a spoken or written representation or account of a person, object, or event.	A qualitative assessment of the wine, often penned by a sommelier, encompassing the wine's taste, aroma, appearance, and mouthfeel.
Points	Numeric	1-100	Points are a numerical representation that evaluates the quality or value of something.	A quantitative measure of the wine's quality, as rated by WineEnthusiast. Only wines scoring 80 or above are reviewed.
Price	Numeric	Any positive value	Price is the amount of money expected, required, or given in payment for something.	Denotes the cost of a single bottle of the wine in question.
Title	Text	Any valid name of the wine	Title is a name or title signifying the specific wine, also including the year of harvest of the grapes.	Offers a more granular insight into the wine's origin, pinpointing the year of the grape harvest and the name of the wine.
Variety	Text	Any valid grape variety	Variety is a specific kind or sort of something that belongs to a larger group or class.	Refers to the type of grapes used in the wine's production, such as Pinot Noir.
Wine Colour	Text	Red, White, Rosé, etc.	Wine Colour is a characteristic of visual perception that enables one to differentiate otherwise identical objects.	Describes the visual hue of the wine, for instance, red, white, or rosé.

Figure 1: Data Dictionary

The dataset exhibits a notable level of organisation via its multi-tiered categorization system, thereby greatly augmenting the user experience. Users have the ability to engage with the website by choosing wines according to many qualities, including country of origin, province, region, grape variety, and colour. The categorization of the information into five distinct categories has a profound impact on user interactions, enabling users to efficiently explore and engage with the dataset. This technique enables consumers to make well-informed decisions when selecting wines.

4.1 Our Ontology

Our ontology (see Figure 4) serves as a robust framework for organising wine-related data, tailored for enthusiasts expanding their collections. As you can see in Figure 4, comprising five key entities—Color/Type, Grape Variety, Taste, Origin/Region, and Price—the ontology employs a controlled vocabulary to ensure consistency and clarity. These entities are linked through relationships like "Has Grape Variety" and "Is Located In," offering a comprehensive view of each wine's characteristics.

The controlled vocabulary adheres to industry standards (SKOS), ensuring that the terms used are universally understood. This enhances data management and facilitates effective retrieval of information. However, one of the challenges we faced was in balancing the granularity of the vocabulary. Too many sub-categories could overwhelm users, while too few could limit the utility of the ontology. Another challenge was in ensuring that the new terminology introduced was both intuitive and aligned with existing industry terms.

Despite these challenges, the ontology is designed to be both user-friendly and rigorous, aiding in making informed selections based on taste, origin, and budget constraints.

5 WHEN IS IT BEING ORGANISED?

The organisation of resources on this platform is primarily a snapshot organisation, occurring at the time of resource creation and collection assembly. This is a basic, static form of organisation that does not dynamically adapt in real-time to user interactions. In this model, resources such as wine data, reviews, and vendor information are organised at the point they are added to the collection, and this organisational structure remains largely unchanged thereafter. This approach aligns with "just-in-case" organisational strategies, where resources are organised in a manner that anticipates the most common types of user interactions.

Given the platform's snapshot nature and basic website architecture, there is limited scope for complex regulatory compliance mechanisms. However, basic legal mandates, such as age verification for alcohol-related content, are adhered to. The organisation of resources is not significantly shaped by industry practices or cultural traditions, as the primary focus is on basic user accessibility and information retrieval.

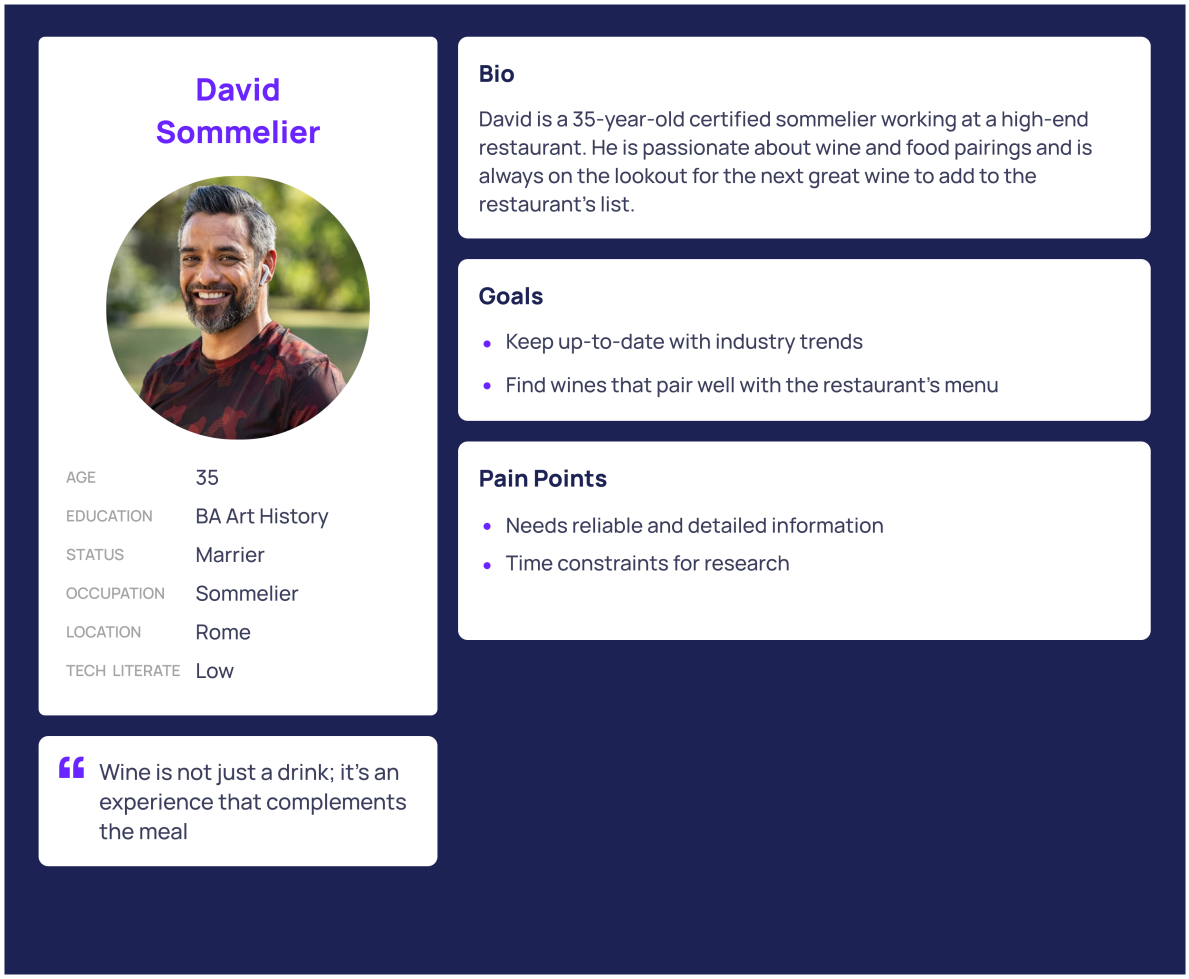

Figure 2: User Persona: David.

6 HOW OR BY WHOM IS IT BEING ORGANISED?

The data is organised by a combination of our team and the automated system. Starting with an empty database, data is collected by the team either by web scraping or questionnaires. This results in an extensive excel sheet containing all this information. Preferably this data would be cleaned for grammar mistakes and other inconsistencies. Different algorithms could be utilised for this however, this is not done with the current data because this is not strictly necessary for the project. The next step is to check if all data entries contain a whine colour, white, red, or rose. If this value is not available in the entry, the grape variety is put through our machine learning algorithm, which will return this information. Once this process is completed, the data is added to the database and from there on available via the dashboard.

The team uses a Decision Tree algorithm to decide the colour of the given wines depending on the grape variety. The Decision Tree is trained via a data set of 130.000 items. Before the training run could be started, the colour data had to be added to the entries. To achieve this goal, two additional algorithms were developed. The goal of the first first algorithm was to collect grape varieties including their colour. This was achieved by a web scraper, scraping Wikipedia data. The goal of the second algorithm was to match the entries

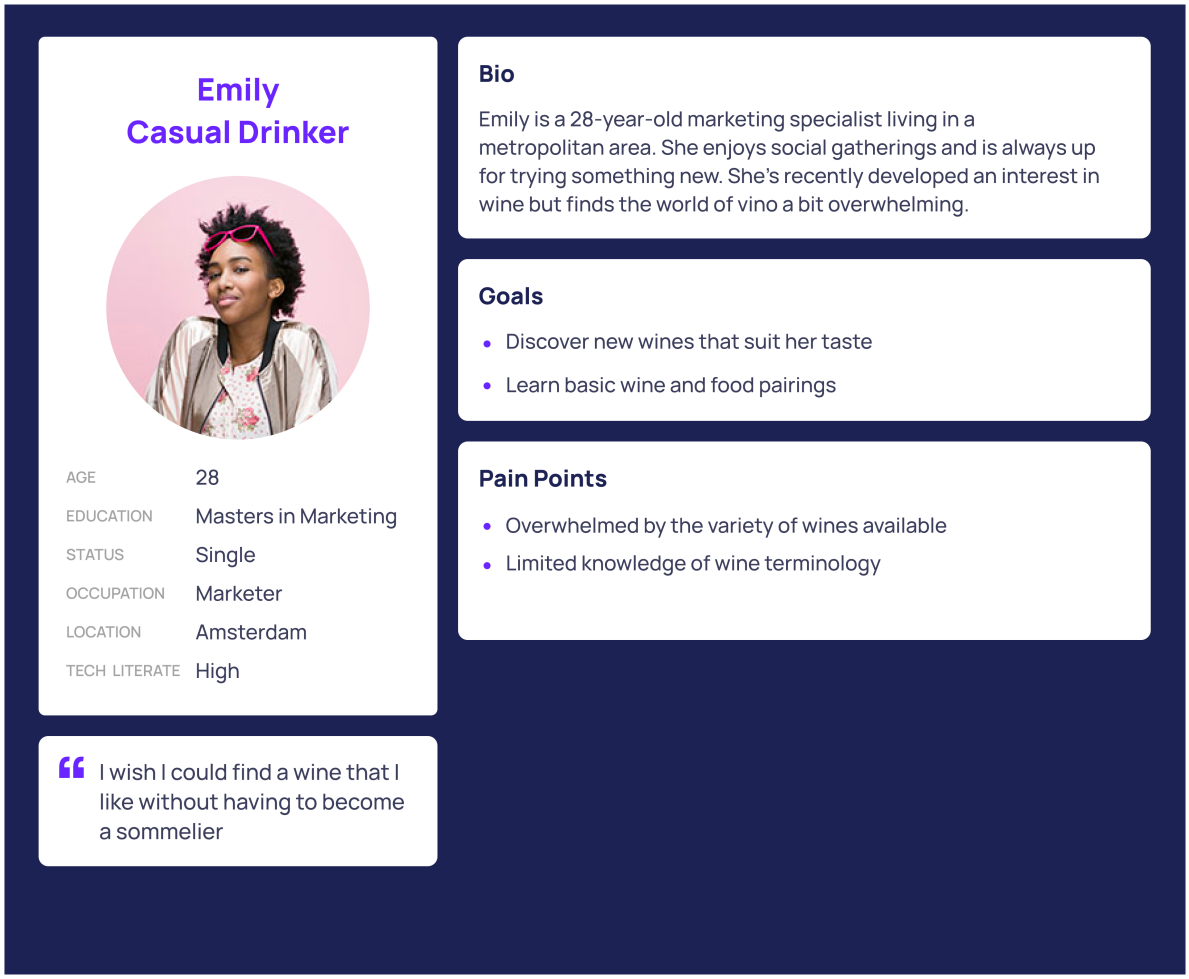

Figure 3: User Persona: Emily.

of the 130.000 data set to the scraped data set and match a wine colour to the entries. The matching algorithm matched items on the variety column of both data sets. There had to be a 90 percent match to get a positive result. This 90 percent match allows for some small grammatical differences, such as capital character versus lowercase character, or characters that are not in the right spot in the word or sentence. When this process was completed, the data was imputed to the Decision Tree algorithm for training.

7 WHERE IS IT BEING ORGANISED?

The arrangement of resources inside the dataset is influenced by a combination of factors including design considerations and regulatory limitations. Design-wise, the dataset is structured to categorize wines based on key attributes, including country of origin, province, region, variety, and color. This organisation enables users to effectively filter and choose wines according to their own tastes, so augmenting the overall user experience. In this particular context, the nature of the resource placement is crucial in relation to its characteristics of changeability, mobility, and interdependency.

Firstly, the information is structured according to a static resource location model inside the Wix Cloud online database. Once a wine is included in the dataset, its position

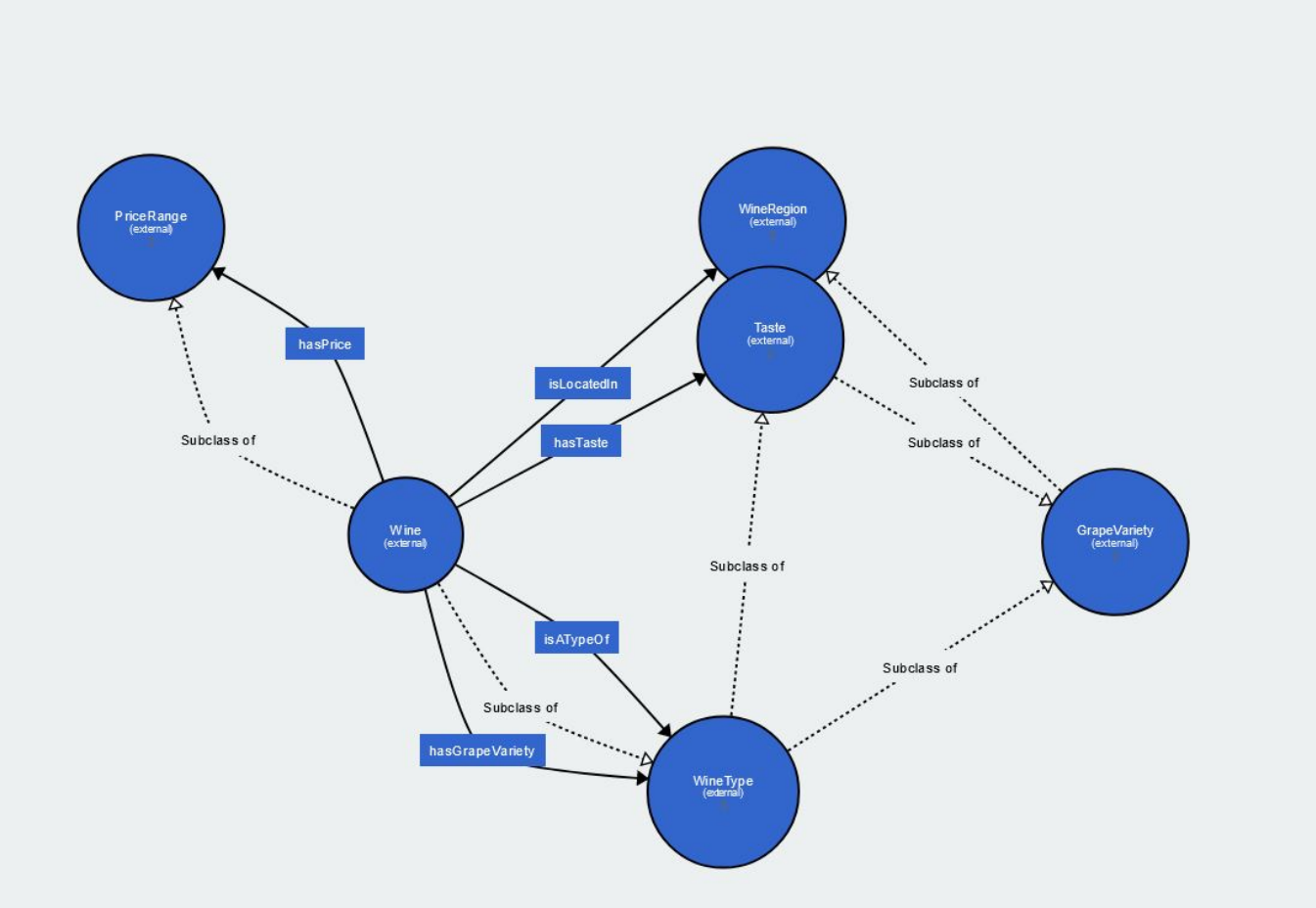

Figure 4: Wine Ontology.

stays unchanged throughout time. Consistent access to information on wines facilitates predictability and enhances user-friendliness in the dataset's functionality.

Secondly, the resources do not experience any kind of transitions. The items in question maintain an established position and do not undergo any changes inside the assortment. The organisation of the information does not facilitate dynamic resource mobility; instead, it offers a reliable and consistent environment for users to traverse.

Thirdly, in addition to the fundamental static organisational structure, the dataset includes dynamic aspects that allow for potential changes in wine suggestions depending on user input. For instance, comprehensive details on the year of production are included for each wine. The ability for users to apply filters to wines based on certain years or periods adds a dynamic component to the process of locating resources. This functionality allows users to investigate wines that have been made throughout specified time periods, so adapting the changing tastes of individuals over time.

In brief, the organisational structure of the wine website dataset integrates both design-driven and regulatory-informed approaches, providing consumers with a diverse collection of criteria to facilitate wine selection. The system adheres to static resource placement rules in order to ensure consistency. However, the inclusion of dynamic factors enables the system to react to changing preferences. The manner in which the dataset is structured significantly impacts user interactions and their capacity to efficiently retrieve information, hence enhancing the overall user experience.

8 OTHER CONSIDERATIONS

Over the course of designing and implementing Sommeliers-elects.com, several considerations have come to the fore that are not explicitly covered in the main sections of this report. One such aspect is the tension between standardisation and

customisation. While the use of a controlled vocabulary and a robust data dictionary aids in creating a standardized user experience, it also poses challenges in accommodating the rich, diverse terminologies often used by wine connoisseurs.

The choice of machine learning algorithms for tasks such as grape variety classification also warrants discussion. While the Decision Tree algorithm has shown promise, the quality of its output is heavily dependent on the quality and comprehensiveness of the training data, which was sourced through web scraping. This opens up a dialogue about the ethical and quality considerations in using publicly available data for machine learning.

Lastly, the static nature of the organisational structure, while beneficial for consistency, poses limitations on real-time adaptability to user behavior. This is particularly relevant in the context of emerging technologies like dynamic ontologies that could offer more fluid, user-responsive organizational systems.

9 INDIVIDUAL CONTRIBUTIONS

Throughout the course of this project, we maintained a weekly meeting schedule, where every team member actively participated in brainstorming sessions, provided status updates, and fostered a culture of collaboration. Our collective approach enabled us to jointly address challenges and offer assistance when needed. Additionally, we created a WhatsApp group for seamless communication, ensuring that we remained interconnected for more effective work. Importantly, significant directional decisions were made collectively.

Contributions by Réka: I contributed to the project by crafting presentations and generating a mockup of the website. Additionally, I conducted web scraping activities to compile a comprehensive and tidy list of grape varieties and their respective colours. This dataset served as a foundational resource for our machine learning. Finally, I also assisted Sjoerd in executing machine learning scripts.

Contributions by Djurre: I transformed the mockup created by Réka into a functional website using the wix.com platform. During the translation process, I prioritised the implementation of a user-friendly navigation system and aimed to maintain a predominantly one-page structure for the website. Subsequently, I worked on the table interaction on the website, ensuring that users can filter through our dataset, using a number of dropdown menus as well as input fields. Additionaly, I uploaded/cleaned the dataset to be accepted by and used within the context of wix.com. Lastly, I contributed to writing sections "How much is it being organised?" and "Where is it being organised?".

Contributions by Sjoerd: I contributed to the project by writing several python scripts. The first script aimed to match all data entries of the main data sheet with a wine colour, red, white, or rosé. The script took the main data sheet, and the scraped data which Réka provide, as input. Then it tried to add a wine colour to each entry of the main data set by matching on the variety columns of both data sheets. The second script was focus in the machine learning model. For this project a Decision tree algorithm is used. The script took the main data set, including the matched colour data, as input. This input was split in a 0.3:0.7 ratio, testing data and training data respectively. Additionally I wrote some code that created a visual representation of the generated decision tree as a JEPG file. Another contribution of mine is the "How or by whom is it being organised" part of the report.

Contributions by Yonathan: I contributed to the project by working on the front-end and back-end of the website. I also assisted by migrating the cleaned data to MySQL Oracle cloud database that I created to be used to for matching users to the best rated wine. I assisted Djurre on fixing a bug in the search functionality of the table in the website. I created the ontology/vocabulary in a standard format and I also assisted with the defining of the ontology and controlled vocabulary.

Contributions by Dylan: I contributed by developing the User Experience (UX) Design by conducting user research to inform the design and functionality of the website. This involved creating user personas and validating them through research. I also took on project coordination tasks such as establishing the project timeline, roles, and setting up the Google Drive for team collaboration. I also researched the best approaches for our project, such as setting up the WIX website. I also took on writing and research tasks for the report, ensuring we were reflective of the concepts covered in class and wrote the “Overview”, “What”, “Why”, “When”, and “Other Considerations” sections, ensuring they were aligned with the project’s objectives and findings.

REFERENCES

Spence, C., & Wang, Q. J. (2018). What does the term ‘complexity’ mean in the world of wine?. *International Journal of Gastronomy and Food Science, 14*, 45-54.

Masterclass. (2021). What Are the Different Types of Wine Grapes? A Guide to the Various. Retrieved from https://www.masterclass.co
are-the-different-types-of-wine-grapes-a-guide-to-the-various-types-of-red-and-white-wine-grapes-in-the-world

Washington Tasting Room. (2021). The Year in Wine: 2021 Consumer Trends. Retrieved from https://www.washingtontastingroom.co
year-in-wine-2021-consumer-trends

Schaefer, R., Olsen, J., & Thach, L. (2018). Exploratory wine consumer behavior in a transitional market: The case of Poland. *Wine Economics and Policy, 7*(1), 54-64.

Caracciolo, F., Furno, M., D'Amico, M., Califano, G., & Di Vita, G. (2022). Variety seeking behavior in the wine domain: A consumers segmentation using big data. *Food Quality and Preference, 97*, 104481.

Echave, J., Barral, M., Fraga-Corral, M., Prieto, M. A., & Simal-Gandara, J. (2021). Bottle aging and storage of wines: A review. *Molecules, 26*(3), 713.