Umanistica Digitale – ISSN: 2532-8816 – n. 20, 2025: Special Issue
G. Renda, M. Daquino – *Ragù. Preserving Italian Cookbooks in a Sustainable Crowdsourced Digital Collection*
DOI: http://doi.org/10.6092/issn.2532-8816/21113

Ragù. Preserving Italian Cookbooks in a Sustainable Crowdsourced Digital Collection

Giulia Renda

University of Bologna
giulia.renda3@unibo.it

Marilena Daquino

University of Bologna
marilena.daquino2@unibo.it

Abstract

Recipes of popular origin and handwritten cookbooks have often been overlooked by scholars, due to the difficulties derived from 1) collecting and organising a representative – let alone complete – set of documents, 2) contextualising scattered sources in a broader historical context and communicate their value to a broad audience, and 3) costs such a collection campaign would require. Nonetheless, having such collections available for qualitative and quantitative analysis could highlight peculiarities of the culinary history of a place, as well as considerations on its material history. Ragù is a pilot project that tries to fill in this gap by crowdsourcing and digitising a collection of cookbooks belonging to the Italian traditional cuisine, particularly from Emilia-Romagna, and making it accessible via a digital platform for exploration and analysis. The digital project here aims at contributing to two research lines: a) to identify agile methods for publishing data in a low-cost crowdsourcing project, and b) to devise an effective storytelling journey to present the Ragù project.

Keywords: cultural heritage preservation, crowdsourcing, data management, digital storytelling

Le ricette di origine popolare e i ricettari manoscritti sono stati spesso trascurati dagli studiosi, a causa delle difficoltà derivanti da 1) la raccolta e l'organizzazione di un insieme rappresentativo – per non dire completo – di documenti, 2) la contestualizzazione delle fonti sparse in un contesto storico più ampio e la comunicazione del loro valore a un vasto pubblico, 3) e i costi che una tale campagna di raccolta richiederebbe. Tuttavia, avere a disposizione tali collezioni per un'analisi qualitativa e quantitativa potrebbe mettere in luce le peculiarità della storia culinaria di un luogo, oltre a considerazioni sulla sua storia materiale. Ragù è un progetto pilota che cerca di colmare questa lacuna attraverso il crowdsourcing e la digitalizzazione di una collezione di ricettari appartenenti alla cucina tradizionale italiana, in particolare dell'Emilia-Romagna,

rendendola accessibile attraverso una piattaforma digitale per l'esplorazione e l'analisi. Il progetto digitale si propone di contribuire a due linee di ricerca: a) identificare metodi agili per la pubblicazione di dati in un progetto di crowdsourcing a basso costo e b) ideare un percorso narrativo efficace per presentare il progetto Ragù.

Parole chiave: conservazione del patrimonio culturale, crowdsourcing, gestione dei dati, narrazione digitale

1. Introduction[1]

Recipes of popular origin and handwritten cookbooks are often overlooked by scholars, despite representing a precious primary source of cultural insight. In fact, cookbooks carry information on the influence of diverse food cultures and their relationship with local traditions, as well as witnessing the evolution of the material history of a place or community [6]. However, such collections pose several issues to scholars, who must organise a representative collection of documents which often lack the fundamental paratextual information (who, when, where, and why they were written) [1], hence hampering a correct contextualisation of scattered sources in a broader historical context and therefore undermining a correct communication of their value. Yet, a few projects were born to collect, transcribe, and preserve such a precious intangible cultural heritage [21]. However, a long-term plan for maintenance and continuity of most projects seems to be missing, since most of the notable collections we found in the literature are currently not updated or have different scope and goals than collecting and serving historical recipes [7]. Moreover, while a few of such projects provide means for crowdsourcing information and documents, they lack an interface for dissemination, exploration and analysis, therefore hampering the intended (correct) use of the collection itself. Such issues apply to a larger landscape of small-medium size Humanities projects, which present similar workflows and, unfortunately, results.

Ragù is a recent crowdsourcing project that tries to fill in this gap by collecting and digitising a collection of cookbooks belonging to the Italian traditional cuisine, and making it accessible via a digital platform. The digital project and the whole digitisation process have been designed in a way that effectively reduces costs of maintenance, therefore fostering its long term run and ensuring its preservation, while making it accessible and user-friendly to volunteers that contribute to the collection. Like similar projects, Ragù is a low-budget project managed by a few volunteers, and it presents several criticalities. Firstly, the primary sources need to be physically collected, transcribed and digitised and a mid-term data management plan for preservation and exploitation is needed. Additionally, mechanisms for effective metadata extraction and query must be designed to populate user-friendly interfaces. Given the resource constraints of the project, such methods must ensure automatic updates (of both metadata and user interfaces) so as to require little or no intervention of technical personnel when new sources are added. Lastly, since the web platform aims to engage with both lay users open to serendipitous discovery and scholars interested in exploring historical data with more sophisticated queries (although with no or little technical skills), appropriate strategies for dissemination and search interfaces design must be in place to accommodate various information seeking behaviours.

[1] Giulia Renda is responsible for section 2,3, and 4; Marilena Daquino is responsible for section 1,5 and 6.

In this article, we present the strategies we adopted to realise the Ragù project. In particular, we describe existing approaches to crowdsourcing and storytelling applied to cookbooks in section *2. State of the* art, we propose an agile model for data management in low-cost projects in section *3. An open source* approach to publish low-cost crowdsourcing campaigns, and we present the digital storytelling journey automatically populated from crowdsourced data in section *4. A storytelling journey on* cookbooks. We discuss our results in the light of the state of the art in section *Discussion* and we conclude with future works.

2. State of the art

In the last decades, the massive digitisation process of large libraries, museums and archives has been successfully carried out by specialists in galleries, libraries, archives and museums (GLAM) [14]. However, the business model applied by national and international institutions is not always applicable to small and medium-sized institutions, cultural associations or individuals, who may struggle to find sustainable and cost-effective strategies.

Crowdsourcing [11] is the process of outsourcing tasks or problems to a large group of people, typically using the web as a platform for information exchange. Crowdsourcing methods have been adopted in GLAM and heritage organisations to carry out several tasks and cope with the lack of human resources. Typical tasks delegated to volunteers include metadata creation, analysis of cultural heritage objects, as well as contributions of private objects or experiences into a platform [18], [3], [15]. The latter task is particularly challenging although pivotal to the development of material history enquiry, since public (digital) collections rarely contain materials belonging to private collections or individuals, e.g. household documents. Therefore crowdsourcing such primary sources may be the only way to support research enquiries based on witnesses of our material history [19].

An exemplar case of crowdsourced campaign is *Europeana 1914-1918*,[2] where users could upload pictures of memorabilia to a web platform to be digitised and to populate a digital archive. However, projects on a smaller scale in terms of both finances and workforce encounter problems related to continuity and sustainability [5], including socio-technical, technological, and dissemination challenges. As a matter of fact, most crowdsourcing projects in GLAM rely on expensive and not easy-to-customise platforms for collecting citizens' contributions [5]. More importantly, existing projects do not often provide dissemination interfaces to communicate results of the crowdsourcing campaign (whose development and maintenance adds an additional cost to the project). When it exists, the resulting interface may not be able to engage a wide variety of visitors with curated – and again, expensive – research paths, hence completely lacking interpretative frameworks on top of the collections [5].

Scholars and practitioners in the Humanities have highlighted two types of interactive behaviour in web platforms for information seeking purposes. The first strategy focuses on reducing the cognitive load of users by breaking down information, displaying it in small chunks [12], however losing the overall picture. The second approach advocates instead for more *generous interfaces*, arguing that hiding the overall picture of content from the user can disorient them and therefore become a source of frustration [22]. Considering that visitors on cultural heritage websites don't always have a specific task in mind, methods for casual browsing can effectively help them to

[2] https://www.europeana.eu/en/collections/topic/83-world-war-i

discover and find their way to a specific task or interest. Such an approach has been summarised as "overview first, zoom and filter, then details on demand" [17]. To overcome the overwhelming feeling caused by the large amount of materials that a digital collection can contain, a common practice in the cultural heritage domain is to design "paths" to help the user navigate content, and to use data visualisation techniques to provide graphical aids that effectively improve their experience [8], [10]. However, data visualisations can be hard to understand and interpret for everyday users, especially in the Humanities landscape of digital projects, where practitioners often borrow graphical solutions (e.g. charts) developed in other domains with different purposes and assumptions and rarely repurpose them critically [2]. Moreover, visualisations often prevent users from seeing the interpretive framework that scholars used to design the chart themselves, which often simplify complexity of the Humanities research instead of highlighting its added value. To mitigate the error-prone interpretation of data visualisations, storytelling techniques can be used to effectively convey complex information, complementing unavoidable simplifications made by visualisations [9], [12], [16] and acting as an interpretative framework [4].

To the best of our knowledge, very few examples of platforms presenting digital collections of cookbooks rely on generous interfaces, storytelling and data visualisation techniques. *The Early American Cookbooks* [21] collects cookbooks from 1800 to 1920, and serves them on a website organised in thematic categories, which feature blog-like articles using data visualisations to offer insights into the collection. Developed by a digital humanist at the New York University Libraries, the workflow for data management does not seem to be replicable, and the storytelling strategies are limited to few reiterated charts that do not allow further exploration of the catalogue. Other examples of collections of private cookbooks include *The South African Jewish Cookbook Project*,[3] although the project website does not present the collection using any form of data visualisations or storytelling strategies. Rather it offers traditional interfaces for exploring items of the collection thematically. Other ongoing research projects, such as ArDIRe,[4] suggest that more interpretative approaches to recipe collections are of growing interest, although the implementation is in the early stages.

3. An open source approach to publish low-cost crowdsourcing campaigns

Ragù began as a serendipitous discovery of handwritten cookbooks in a cellar to be cleared out. While the creation of the collection posed a number of challenges that could have hampered the final result and its reusability, it soon became an opportunity to systematically tackle and solve many of those issues that are shared across Humanities projects. Most important aspects that we account for concern approaches to the design and management of 1) methods for data entry, 2) web interfaces population, and 3) mid and long-term data management. The overall goals of the approach here proposed are 1) to define a reusable technical workflow for collecting and publishing a collection of curated data and images, wherein editorial intervention is possible at any point without the intervention of technical staff, and 2) to design a solution for

[3] https://sajewishcookbooks.org.za/

[4] https://dium.uniud.it/en/ricerca/progetti-corso/archivi-biblioteche-edizioni/ardire-archivio-e-dizionario-digitali-dei-primi-ricettari-artistici/

dissemination, exploration, and analysis that responds to information seeking needs of experts and curious visitors.

Methods for data entry. The historian leading the project made a press release with a call for contributions on various channels (newspapers, magazines, academy dissemination channels), and started building the collection thanks to the contribution of numerous volunteers, who offered their private cookbooks to be digitised and shared publicly with an open license.

Whenever a new cookbook is proposed and evaluated, it is photographed with a smartphone and then returned to the owner. The sensible reason to use low-cost equipment for the digitalisation is due to the nature of the sources themselves, that once their content has been transcribed and annotated, do not present noteworthy peculiarities related to the handwriting styles and/or the material support – which would otherwise require more sophisticated scanning and digitising techniques. Medium-quality images are archived in an online drive folder (free of charge) and are later reused to populate the web interfaces. Image files are named so as to recall the identifier of the cookbook, the date of digitisation, and the sequential number of the page depicted.

Cookbooks contents are annotated via a Google spreadsheet, available online to facilitate access and collaboration.[5] Figure 1 shows an excerpt of the main table, including its first headings and rows. The table was devised by data curators that volunteered in the project, and it includes all fundamental bibliographic metadata needed to identify cookbooks and recipes – e.g. recipe title, year, city of origin, author –, information on the image file, and annotations relevant to historical research enquiries, e.g. type of recipe, ingredients, quantities, and linguistic variations of names of ingredients. Controlled vocabularies and folksonomies have been defined in a separate table (the tab "vocabolari" of the spreadsheet) that mainly address ingredients names, recipes categories, geographical information (city, region, and country), scope, procedure, and units of measurement. It's worth noting that while the main unit of analysis can be considered the recipe, each row of the table represents the occurrence of a single ingredient that appears in a recipe. This is due to the many variable aspects that can characterise the usage of the ingredient in the context of a recipe, such as qualifiers (e.g. "green" olives), linguistic variant forms of the name (e.g. "Bazzotto" cheese), and alternative ingredients (e.g. oil instead of butter).

	B	C	D	E	F	G	H	I	J	K	L	M
1	Notebook	From	To	Time qualifier	Place	Region	Country	Author	Img	Pag.	Title chapter	Title Recipe
2	title	year	year		town			surname name	nome	numero		
8	Le ricette di zia Dina	1960	1970	ca.	Rimini	Emilia R...	Italy	Dina	Quaderno 1_Rimini_29ago2019_2.jpg	2	Minestre	Maccheroni con besciamella
9	Le ricette di zia Dina	1960	1970	ca.	Rimini	Emilia R...	Italy	Dina	Quaderno 1_Rimini_29ago2019_2.jpg	2	Minestre	Maccheroni con besciamella
10	Le ricette di zia Dina	1960	1970	ca.	Rimini	Emilia R...	Italy	Dina	Quaderno 1_Rimini_29ago2019_2.jpg	2	Minestre	Maccheroni con besciamella
11	Le ricette di zia Dina	1960	1970	ca.	Rimini	Emilia R...	Italy	Dina	Quaderno 1_Rimini_29ago2019_2.jpg	2	Minestre	Maccheroni con besciamella
12	Le ricette di zia Dina	1960	1970	ca.	Rimini	Emilia R...	Italy	Dina	Quaderno 1_Rimini_29ago2019_2.jpg	2	Minestre	Maccheroni con besciamella
13	Le ricette di zia Dina	1960	1970	ca.	Rimini	Emilia R...	Italy	Dina	Quaderno 1_Rimini_29ago2019_2.jpg; Quaderno 1_Rimini_29ago2019_3.jpg	2; 3	Minestre	Polenta alla lombarda
14	Le ricette di zia Dina	1960	1970	ca.	Rimini	Emilia R...	Italy	Dina	Quaderno 1_Rimini_29ago2019_2.jpg; Quaderno 1_Rimini_29ago2019_3.jpg	2; 3	Minestre	Polenta alla lombarda

Figure 1An excerpt of the table for annotating recipes and ingredients, in Google Spreadsheet.

Web interfaces population. The digitisation and transcription is an ongoing process, and cookbooks are often added to the collection. However, to showcase (partial) results of the data collection as soon as possible, we decided to create a website that could be easily and dynamically updated every time new recipes are added to the spreadsheet. To achieve this task, we leverage

5 https://docs.google.com/spreadsheets/d/1terFx_mYVspOjvDUJfcBqHR7j_OlvXVwUesHNHAl_wU/edit?usp=sharing

one of the most well-known platforms for depositing and versioning open source code, i.e. GitHub, which allows us to publish a static website and to run scripts that manipulate data into appropriate formats for being reused in the web interfaces. The GitHub repository of Ragù project is available online.[6] It includes the versioned copy of the aforementioned table ("ragu_disci - metadati.tsv", "ragu_disci - vocabolari.tsv"), the GitHub actions and the code used for populating the website ("script.py", "github/workflows/actions.yml"), and the source code of the website.

In Figure 2 we illustrate the procedure that an editor needs to perform in order to update the website content and the technical workflow that we implemented to automatically populate the interfaces. The population and subsequent update of the web interface require an editor to first download the working table from Google spreadsheets as .tsv files (including the metadata and vocabulary tables) and the images from the drive folder, and upload them to the GitHub repository. Commits of the uploaded files are labelled as "data update". Non-expert users can upload the data using a user-friendly drag-and-drop interface of GitHub. Similarly, images are uploaded in the *recipe_photos* folder.

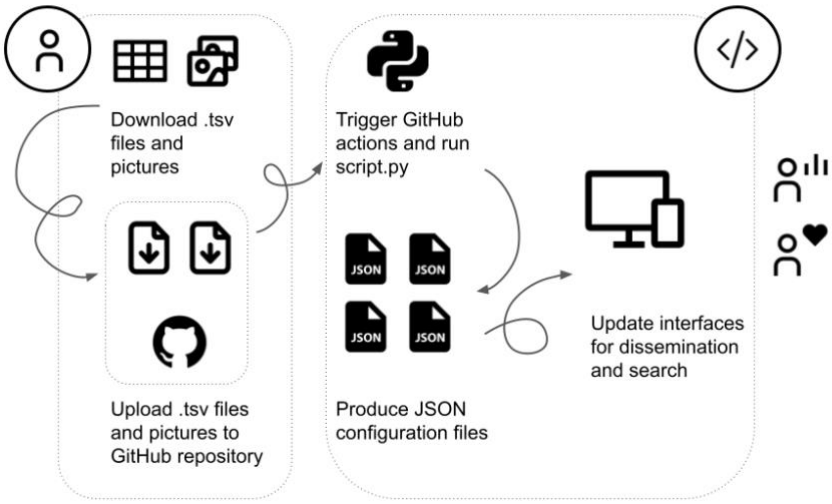

Figure 2 The editorial and technical workflow to populate web interfaces of Ragù.

The submission to GitHub triggers a GitHub action that runs the *script.py* file, which includes methods to extract information from the latest version of tsv files and write configuration files that will be used by the web platform. The JSON format was chosen to represent the configuration files, since it is lightweight, it is easily manipulated by both Python and JavaScript languages (later used in the website), it is highly readable also by humans, and in the future it can be easily transformed into JSON-LD to add semantics to the data. The configuration files automatically created by GitHub actions are the following:

[6] https://github.com/raguproject/raguproject.github.io

- *general.json* contains statistics to be used in the homepage – e.g. the number of recipe books, recipes, and ingredients – and paths to other files (recipe books and filters).
- Other files contain recipes grouped by letter (*alphabet.json*), type of course (*categories.json*), ingredients (*ingredients.json*) and provenance (*provenance.json*), which are used to populate the advanced search of the website.
- Bespoke files are created to build visualisations divided by type, namely: *map.json*, *matrix.json*, *network.json* and *piechart.json*, each organised according to the requirements of Charts.js,[7] the data visualisation library chosen to create the charts.
- Finally, each cookbook has its own file, in which we find general information such as title, year, origin, and a dictionary including all the recipes, which in turn contains the annotations extracted from the working table.

Every time a GitHub commit is performed by the editors, the above described automatic workflow is run again and configuration files are replaced. So doing, the web interfaces that read content from the JSON files are automatically updated (e.g. charts, catalogue search, and so on) and the website offers a seamless experience of updated, correct, information.

Mid and Long-term data management. While for images there is no versioning system in place (i.e. once an image is replaced, the old version is deleted), the working table relies on Google Spreadsheet versioning system to store changes made by contributors and on GitHub versioning system to store versions accepted by the editors, i.e. ready to be published. In summary, GitHub provides us with five (free-of-charge) important features, namely:

1. Hosting of the dataset, i.e. a versioned copy of the aforementioned working table.
2. Hosting and publication of the web platform, developed using sensibly web technologies and frameworks.
3. Runnable methods, i.e. GitHub Actions, to dynamically extract data from the table and populate the website.
4. Sufficient editorial strategies to prevent people from publishing content before a review of the table has been done.
5. Strategies for the long term preservation of data, code, and application via synchronisation with Zenodo.

GitHub has been unanimously recognised as a scholarly venue for code dissemination and discovery, although it does not offer all the features that a long-term preservation repository should offer (e.g. addition of bibliographic metadata, stability and citability of persistent resources on the web). Nonetheless, the possibility to integrate third-party applications, such as Zenodo for code synchronisation, storage, versioning and DOI assignment, allows us to cope with the aforementioned limitations of a day-to-day mid-term working environment, delegating long-term preservation tasks to appropriate agents.

[7] https://www.chartjs.org/

4. A storytelling journey on cookbooks

To accommodate requirements of different target audiences, the digital collection is presented on a website[8] that offers three sections, namely: an introductory storytelling journey (*homepage*), a search page based on filters and facets (*recipes*), and an ebook-like browser of digitised cookbooks (*cookbooks*, currently under construction). Such a division of content is representative of expected information seeking behaviours of identified target users, which can be simplified in "overview first, detail on demand" (Schneiderman 1996). In particular, the ratio used when designing Ragù interfaces falls under the general paradigm of the generous interfaces (Whitelaw 2015), which allow incrementally detailed interfaces to be served on demand and therefore accompanying users with more sophisticated information needs in the discovery and exploration of the collection.

The *homepage* offers a digital storytelling journey that leverages the narrative element to showcase the content of the collection and stimulate curiosity. The overview of the collection contents is organised in a number of views, each expanding on a specific topic of interest, e.g. ingredients, categories of dishes, in a sequential reading path. The concept idea connecting visualisations is the red thread of history and tradition, and uses animation (the animated red line guides the user into looking in the focus areas, directing their attention to the most important content) and scrollytelling to collate several sections. Sections generally include a chart accompanied by a (sequential) introductory text (Figure 3). Such a narrative uses data visualisation elements to present data in an intuitive way and offers stimuli that can be leveraged later in searches. In detail, the red line graphically guides users through five topics and invites them to explore recipes and cookbooks based on such summarisation, which show significant data patterns that cannot be immediately appreciated by performing a search on the catalogue of recipes (i.e. the page *recipes*). The steps of the storytelling journey are the following:

1. **Overview**. Users are introduced to the topic of cookbooks, and the motivation of the project (preservation and dissemination) with a short impactful text description. Some figures, i.e. the number of cookbooks, recipes, and ingredients, give immediately an idea of the extent of the collection, therefore clarifying the level of representativeness of the collection and letting the reader estimate its potential added value. By starting the journey with the motivation underlying the birth of the project, we aim at creating a first moment of empathy, i.e. the user agrees on the importance of preserving such a wealth of information and is willing to explore it to gain some new knowledge about it.
2. **Provenance**. The journey continues with a map displaying the provenance of collected cookbooks, so as to complete the previous figures on the extent of the collection. Again, the map is designed to stimulate empathy. Indeed, a user will likely look for recipes from their region. Currently the web application includes only a selection of digitised cookbooks from the Emilia-Romagna region, and more have to be included, in order to show where the project originated, and how it evolved by engaging with more donors of handwritten notebooks.
3. **Ingredients**. As aforementioned, while the recipe is the natural unit of measure of the items included in the collection, the focus of searches is the usage of ingredients. Three data visualisations give the visitor an overview of which and how ingredients are used in recipes, showing the most representative ones and stimulating interest in potential

[8] https://raguproject.github.io/

combinations. In detail, 1) a graph network shows recipes grouped by ingredients, highlighting which ingredients are the most represented in the Italian diet, 2) the iconic pie chart (represented as a measuring cup) shows how units of measures are described in recipes, highlighting how the experience of the cook, which often measures by eye ("q.b.") has a pivotal role in the Italian intangible heritage, and 3) a bubble chart shows the co-occurrence/correlation of ingredients and lets users grasp the constituents and common combinations of the Italian popular cuisine.

4. **Dishes**. A set table shows all the types of courses that are annotated in the dataset. While not providing any numeric insight at this point of the journey, the visualisation aims at introducing important concepts that will become facets in the search interface and to trigger the memory of the reader, who will be, at this point of the journey, thinking about their private experience and the dishes that are part of their life journey.
5. **Gender**. A final remark is put on the fact that almost all recipes are written by women and that the project aims at saving their memories from loss and forgetfulness. Again, no numeric values are presented here. At the end of the journey, the aim is to move the attention of the reader from mere quantitative aspects relevant to the collection at hand towards more qualitative, affective, features characterising the sources of the collection, which can make a reader sympathise with the project, e.g. by triggering memories about grandmothers, aunties, mothers, neighbours that have once cooked a dish that stuck in their memory, but of which they cannot recall (or they never knew) all the details.

Figure 3 An example of a chart with introductory text in the storytelling journey of Ragù.

The second section of the website is the search interface, called *recipes*, where the user can view recipes in alphabetical order (like in an encyclopaedia of recipes) or select them via three facets that can filter recipes by type of meal, ingredients, city of origin (Figure 4a). When clicking on a recipe, the user can access the digitised picture of the original cookbook and the transcribed metadata (Figure 4b). For example, the collection includes two recipes for the "Pasticcio di maccheroni", one written by Mrs Dina and the other by Mrs Anna Maria (Figure 4a). Opening both recipes allows one to appreciate the differences in the making, discover the variants of the names that these women used to call the ingredients, which ingredients were the most commonly

used and where. Being free in the exploration allows the user to find complex recipes such as "Tortellini" that come in at least 5 different variations, or navigate into poorer recipes, perhaps dating back to war times.

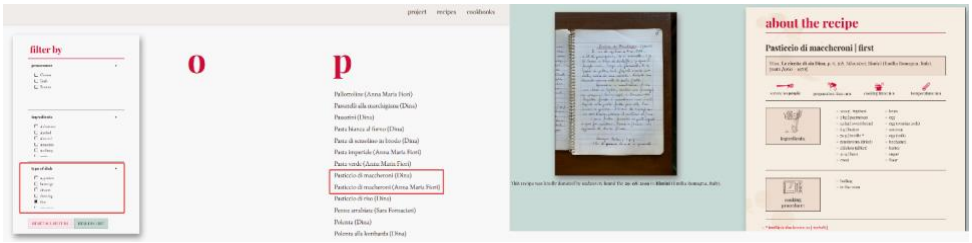

Figure 4a) Recipes search interface (left) and b) an example of recipe (right).

In future works, we aim at building the third section of the website, i.e. "cookbooks", the e-book reader of cookbooks. While the journey and the search respectively introduce and provide access to single recipes, scholars and visitors may be interested in appreciating the co-occurrence of recipes (instead of ingredients) in the same cookbook, therefore seeing combinations of recipes as part of a family tradition. Moreover, only by reading the digitised cookbook users can appreciate the full cooking process, which is only briefly annotated in the metadata and where the sequence of cooking steps is not recorded. No scholarly annotations are foreseen at this stage, as the project primarily focuses on the material restitution and open access to the sources. However, each digitised cookbook will include a basic record of the recipes it contains, including their titles and metadata already extracted during transcription.

In summary, the exploratory journey provides an overall view of the collection, informing users of what type of information they could expect later in the search of recipes in the *recipes* section, and establishing a trust relation with users, which are informed ahead on the boundaries (and therefore the limitations) of the collection. So doing, we aim at preventing users' frustration when discovering later in the search journey that the collection is not (yet) complete enough to perform a sophisticated search that responds to certain historical enquiries. Moreover, the journey is designed to create an empathic relation with the user, presenting those aspects that appear in most cookbooks and that can trigger memories in potential contributors, therefore attracting new donors to participate and enrich the collection with new private materials.

5. Discussion

We presented an approach for data entry and web interfaces update that has several socio-technical advantages.

Agile and sustainable software development. We leverage the power of the GitHub hosting platform, hence preventing us from installing and configuring existing software in a dedicated server and developing sophisticated custom solutions for ensuring the continuous update. On the one hand, it can be argued that we had to develop from scratch the static website instead of relying of content management systems (CMS) that offer catalogue browsing and searching interfaces as part of a out-of-the-box software solution, e.g. OmekaS, ResearchSpace. On the other hand, the design of the storytelling journey would have required significant customisation

of such CMSs, therefore narrowing down the time/coding related benefits. Moreover, static web resources do not require extensive maintenance or software update and can continue running several years without interventions. Server and software updates are delegated to the hosting server (GitHub), which does not charge users for these operations and usually performs such changes without disrupting existing code running on their servers. **Sustainable budget for everybody.** GitHub is a free-of-charge solution for hosting and versioning code, which prevents commissioners and developers from reserving resources (economic for the former and software/hardware for the latter) for maintenance of a dedicated server and it is therefore an appealing sustainable solution in the mid-term. So doing, commissioners do not need to make agreements with a web agency to maintain the code, nor to pay them a fee for the hosting. The initial costs for web development are in this way the only ones to sustain.

Overall control on non-tech-savvy people. The download of spreadsheets and drag-and-drop of tables into GitHub is a rather common operation that non-technical people can perform without supervision or instructions. Such a solution prevents people involved in the project from understanding complex concepts related to data management and web development. In particular, it does not require collaborators to learn sophisticated software solutions for 1) crowdsourcing (tables are relatively easy to understand compared to CMS and give the impression to the editor to be in control on the whole dataset), 2) update (drag-and-drop interfaces hide complexities of version-control-systems), and 3) publication (the web publication is completely automated).

Editorial control. The simple workflow we designed allows one to perform editorial control without requiring yet another interface to learn. Only collaborators of the repository can upload the versioned tables on the GitHub repository to trigger the update of the website. While it would have been possible to link changes in the original spreadsheet to trigger the update of the website, the choice of an intermediate step of downloading from Google spreadsheet and manually uploading to GitHub was made to prevent accidental publication and to empower editors.

Citability and long-term preservation. The synchronisation between GitHub and Zenodo provides scholars with those features that should characterise a scholarly product, such as bibliographic description in a catalogue for data discovery and the assignment of a DOI for citability purposes. Zenodo synchronisation also accounts for changes in the data sources and assigns a versioned DOI to the latter.

Scalability and limitations. While the process described here does not present any innovative technical aspect – especially if compared to existing workflows that leverage state-of-the-art content management systems – the noteworthy aspects that make the whole process potentially interesting to scholars are its reasonable scalability and sustainability. The python script we developed runs on GitHub servers, hence preventing scholars from buying or subscribing to a server wherein to run and store code and website. GitHub actions can perform nicely over great amounts of data, although we cannot estimate its behaviour on very large datasets. In case of several TB of data to be ingested, we can assume different data manipulation strategies should be found (e.g. chunking data and parallelising scripts) and a premium account on GitHub would be needed, since the "freemium" version of GitHub only offers a limited number of GitHub action runnings (although we can confidently claim these are sufficient for small and mid-size projects). Moreover, the website resulting from the process is a static website (i.e. it does not rely on a database management system of any kind), therefore preventing maintenance of the server as well as of applications running on it. Again, we can assume that search interfaces for large datasets may have to rely on more powerful software solutions that the ones developed

only client-side for Ragù. The limitations of using GitHub as a pivotal element in a scholarly project have already been addressed [20]. However, users with different levels of technical skills are getting more and more acquainted with its interfaces and it seems a reasonable technical environment wherein to (partially) handle editorial workflows for small- mid-size projects.

Storytelling for the History of Cookbooks. The journey we presented is a rather unique endeavour, since competing projects on the same topic are hard to find in the literature or on the web. Most sources of recipes are online databases designed for different purposes [7], and with few exceptions these do not account for historical recipes, therefore aiming at a different audience and different objectives (i.e. provide instructional information on cooking procedures only) than those peculiar of historical enquiry. Nonetheless, we believe the adopted approach to present the collection better reflects needs of Humanistic research, which requires bespoke interfaces to support awareness in exploratory analysis and foster interpretation [22].

6. Conclusion

Despite limited resources, the approach applied to *Ragù*, combining crowdsourcing, good practices in data management, and automated updates through a popular platform such as GitHub, showcases a pragmatic and scalable solution that can be implemented in similar projects. The digital collection's website, structured as an exploratory journey with digital storytelling elements, successfully balances overview and detailed exploration. By leveraging data visualisation techniques, it provides users with insights into the history, provenance, ingredients, dishes, and gender aspects of the cookbooks. The emphasis on an exploratory journey serves not only to enrich understanding but also to invite new contributions, ensuring a dynamic and continually expanding digital collection. Future works will address current work in progress operations, such as the inclusion of new recipes, the finalisation of the ebook-like browser of recipes, and the generalisation of the source code used to build the repository, so that similar projects can leverage out-of-the-box solutions to publish their crowdsourced collections. Moreover, we are currently exploring the integration of a lightweight crowdsourced peer review mechanism to support collaborative quality control of transcriptions and annotations. This would allow contributors and users to help validate content and, potentially, assign confidence scores to the resources. Finally, considering that the primary audience is predominantly Italian, we plan to provide an Italian translation of at least the project's informational pages to ensure accessibility and engagement.

Acknowledgements

The *Ragù* project is directed by Dr. Mila Fumini and coordinated by the Digital Humanities Advanced Research Centre (/DH.arc, Unibo): Marilena Daquino (supervision), Giulia Manganelli (web design, web development, copywriting), and Giulia Renda (data management). Data editing: Roberta Balduzzi.

This work has been partially funded by Project PE 0000020 CHANGES - CUP J33C22002850006, NRP Mission 4 Component 2 Investment 1.3, Funded by the European Union - NextGenerationEU.

References

[1] Albala, Ken. 2012. 'Cookbooks as Historical Documents'. In *The Oxford Handbook of Food History*, edited by Jeffrey M. Pilcher, 227–40. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199729937.013.0013.

[2] Allington, Daniel, Sarah Brouillette, and David Golumbia. 2016. 'Neoliberal Tools (and Archives): A Political History of Digital Humanities'. Los Angeles Review of Books. 2016. https://lareviewofbooks.org/article/neoliberal-tools-archives-political-history-digital-humanities.

[3] Bonacchi, Chiara, Andrew Bevan, Adi Keinan-Schoonbaert, Daniel Pett, and Jennifer Wexler. 2019. 'Participation in Heritage Crowdsourcing'. *Museum Management and Curatorship* 34 (2): 166–82. https://doi.org/10.1080/09647775.2018.1559080.

[4] Bruner, Jerome. 1991. 'The Narrative Construction of Reality'. *Critical Inquiry* 18 (1): 1–21.

[5] Daquino, Marilena. 2021. 'Linked Open Data Native Cataloguing and Archival Description'. *JLIS.It* 12 (3): 91–104. https://doi.org/10.4403/jlis.it-12703.

[6] Driver, Elizabeth. 2009. 'Cookbooks as Primary Sources for Writing History: A Bibliographer's View'. *Food, Culture & Society* 12 (3): 257–74. https://doi.org/10.2752/175174409X431987.

[7] Erp, Marieke van, Melvin Wevers, and Hugo Huurdeman. 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*, 217–32. Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-030-00671-6_13.

[8] Fernie, Kate, Jillian Griffiths, Mark Stevenson, Paul Clough, Paula Goodale, Mark Hall, Phil Archer, et al. 2012. 'PATHS: Personalising Access to Cultural Heritage Spaces'. In *2012 18th International Conference on Virtual Systems and Multimedia*, 469–74. https://doi.org/10.1109/VSMM.2012.6365960.

[9] Gershon, Nahum, and Ward Page. 2001. 'What Storytelling Can Do for Information Visualization'. *Communications of the ACM* 44 (8): 31–37. https://doi.org/10.1145/381641.381653.

[10] Grasso, Marco, Marilena Daquino, and Giulia Renda. 2024. 'From Ontology Design to User Experience. A Methodology to Design Interfaces for Information Seeking Purposes'. *Umanistica Digitale*, no. 18, 53–85. https://doi.org/10.6092/issn.2532-8816/19038.

[11] Howe, Jeff. 2006. 'The Rise of Crowdsourcing'. *Wired*, 2006. https://www.wired.com/2006/06/crowds/.

[12] Kiruthika, Jay, Souheil Khaddaj, Darrel Greenhill, and Jarek Francik. 2016. 'User Experience Design in Web Applications'. In *2016 IEEE Intl Conference on*

Computational Science and Engineering (CSE) and IEEE Intl Conference on Embedded and Ubiquitous Computing (EUC) and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering (DCABES), 642–46. IEEE. https://doi.org/10.1109/CSE-EUC-DCABES.2016.253.

[13] Lombardo, Vincenzo, and Rossana Damiano. 2012. 'Storytelling on Mobile Devices for Cultural Heritage'. *New Review of Hypermedia and Multimedia* 18 (1–2): 11–35. https://doi.org/10.1080/13614568.2012.617846.

[14] Monson, Jane D. 2017. *Getting Started with Digital Collections: Scaling to Fit Your Organization*. ALA Editions.

[15] Navarrete, Trilce. 2020. 'Crowdsourcing the Digital Transformation of Heritage'. In *Digital Transformation in the Cultural and Creative Industries*. Routledge.

[16] Renda, Giulia, Marilena Daquino, and Valentina Presutti. 2023. 'Melody: A Platform for Linked Open Data Visualisation and Curated Storytelling'. In *Proceedings of the 34th ACM Conference on Hypertext and Social Media*, 1–8. HT '23. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3603163.3609035.

[17] Shneiderman, Ben. 1996. 'The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations'. In *Proceedings 1996 IEEE Symposium on Visual Languages*, 336–43. IEEE.

[18] Simon, Nina. 2010. *The Participatory Museum*. Santa Cruz: Museum 2.0.

[19] Spennemann, Dirk H. R. 2023. 'Datasets for Material Culture Studies: A Protocol for the Systematic Compilation of Items Held in Private Hands'. *Heritage* 6 (2): 1977–85. https://doi.org/10.3390/heritage6020106.

[20] Spiro, Lisa. 2016. 'Evaluating GitHub as a Platform of Knowledge for the Humanities'. In *Digital Humanities 2016: Conference Abstracts*, 688–90. Kraków: Jagiellonian University & Pedagogical University. https://dh2016.adho.org/abstracts/225.

[21] Stevens, Gioia. 2017. 'New Metadata Recipes for Old Cookbooks: Creating and Analyzing a Digital Collection Using the HathiTrust Research Center Portal'. *The Code4Lib Journal*, no. 37. https://journal.code4lib.org/articles/12548.

[22] Whitelaw, Mitchell. 2015. 'Generous Interfaces for Digital Cultural Collections'. *Digital Humanities Quarterly* 9 (1). https://www.digitalhumanities.org/dhq/vol/9/1/000205/000205.html.