AI-driven big web redesign: two case studies in Italian universities

Andrea Vian[1,†], Daniele Pretolesi[2,†], Lucia Rampino[3] and Annalisa Barla[4,5,*,†]

[1]Dipartimento Architettura e Design, Università di Genova, Genoa, Italy
[2]AIT - Austrian Institute of Technology, Vienna, Austria
[3]Dipartimento di Design, Politecnico di Milano, Milan, Italy
[4]Dipartimento di Informatica, Bioingegneria, Robotica e Ingegneria dei Sistemi, Università di Genova, Genoa, Italy
[5]Machine Learning Genoa Center, Università di Genova, Genoa, Italy

Abstract

This paper explores the challenges of web redesign in Public Administration (PA), particularly within universities. Universities often struggle with fragmented online presences due to distributed editorial models and diverse communication needs across research, education, and dissemination activities. Limited resources further restrict investment in upskilling staff and adopting modern technologies. Open source solutions, though cost-effective, are often chosen without considering user experience. We present a methodology that combines user-centered design, Artificial Intelligence (AI), and "radical collaboration" to achieve a future-proof and scalable redesign. Starting from a case study of a major web redesign project at an Italian university (2014-2020) involving hundreds of websites and over 200 content editors, the paper details the process, including a large-scale content audit using AI, single sourcing with AI-powered content transformation, and user experience (UX) testing with data visualization. This approach resulted in a unified, user-centric online presence and garnered recognition, including the ForumPA award for best innovator. The paper concludes by discussing the applicability of this methodology to other PA institutions facing similar challenges.

Keywords

Large-scale Web Redesign, Artificial Intelligence, Topic Modeling, Academic Communication

1. Introduction

This work addresses the topic of web redesign for Public Administration (PA), which typically constitutes a large-scale institution often tasked with managing a multitude of websites and touchpoints aimed at diverse audiences, and governed by a distributed editorial model, frequently lacking central coordination. This scenario is particularly pronounced within the context of universities, whose communication needs usually respond to three centrifugal driving forces: educational, research, and dissemination activities [1]. Devising a unified communication strategy is therefore a very challenging problem that is compounded by the limited specialized internal resources typically available within universities. This usually leads such institutions to create a disaggregated online presence, supported by technologies with limited scalability. The result is usually a proliferation of touchpoints designed exclusively to respond to the bureaucratic necessities of those who would maintain them, disregarding users' needs. Moreover, due to budget constraints and resource limitations, institutions often lack the capacity to invest in the upskilling and reskilling of their staff. This perpetuates a cycle where existing technological resources are relied upon, even if they are outdated or insufficient for evolving communication needs. Consequently, open source solutions become appealing not because they are a model of knowledge dissemination and data sharing, but as they offer a seemingly cost-effective solution without the need for significant investment in training or professional development. From the viewpoint of human-centered design, it is relevant to acknowledge how this vicious circle can be broken and how the institution can start a maturation process toward the experience of its users [2]. In this paper, we move from our experience in the University of Genova (UniGe) from 2014 to 2020 where some of the authors faced the wicked problem of redesigning a plethora of hundreds of websites that were left abandoned and unmanaged for years and where more than 200 content editors were contributing without any coordination [3]. We illustrate our methodology based on user-centered design, artificial intelligence (AI), and radical collaboration and explain how it allowed us to approach the problem of large-scale web redesign in a principled way that is future-proof and scalable. We show how this methodology allowed us to successfully

Ital-IA 2024: 4th National Conference on Artificial Intelligence, organized by CINI, May 29-30, 2024, Naples, Italy
*Corresponding author.
†These authors contributed equally.
andrea.vian@unige.it (A. Vian); daniele.pretolesi@ait.ac.at (D. Pretolesi); lucia.rampino@polimi.it (L. Rampino); annalisa.barla@unige.it (A. Barla)
0000-0003-0629-0427 (A. Vian); 0000-0001-9075-0187 (D. Pretolesi); 0000-0002-7591-9324 (L. Rampino); 0000-0002-3436-035X (A. Barla)

redesign all touchpoints and how this approach can be applied to other existing realities.

2. A big web redesign for UniGe homepage and education websites

Established in 1481, UniGe has a long history of academic excellence and innovation. The university enrolls about 3000 faculty and staff and it offers more than 200 undergraduate and graduate programs across various disciplines, including arts, sciences, engineering, law, economics, medicine, and more, attracting tens of thousands of students every year. UniGe's major web redesign project launched in 2014 in sync with and in support of the university's strategic goals at the time, summarized by the five keywords of the governance vision: simplification, participation, welcoming, integration and growth. *Simplification* guided the redesign work from the very beginning. The starting point is to let the communication system deal with inherent content complexity and present the users with selected information that is defined by their unique profile. *Participation* materialized in the system of content single-sourcing, which made possible the coexistence of central editing and distributed editing. *Welcoming* consisted in adopting the user-centered design approach that guided the entire project: from user research to user profiling. *Integration* proved to be one of the biggest challenges. It first originated from the adoption of the principle of decoupling of the backend and frontend [4] and the principle of headless development(Koenig, 2018). It then guided us to develop a middleware layer, which allowed for the coexistence of new and legacy systems and enabled data interconnection [5]. Lastly, *growth* was the result of design activities geared toward substantial change in the services offered to students. The aim here was to build an experiential reputation based on improving the UX rather than forcibly pursuing a transformational facade.

The first interventions required by the new governance were a feasibility study and an assessment of the effort needed to redesign the UniGe's website in light of the keywords of the new strategic vision. An in-depth analysis revealed a common situation among Italian PAs: the website structure mirrored the institution's ambiguous internal processes. It was, in fact, a multi-site. The homepage was the cohabitation space for an inordinate number of pages that make up a multitude of independently managed websites supported by different technological solutions. Overall, this panorama of websites lacked a unified design and even more so a cohesive Information Architecture (IA). Webpages were linked in a maze of cross-references, with entire sections forgotten and completely disconnected, reachable only and not always, by search engines. This condition is clearly observable by looking at Figure 1, where we show the structure of one of the major components of the UniGe website panorama, *studenti.unige.it*, before the redesign process. It is easily noted that the graph has a highly entropic structure, displaying subgraphs composed only of PDF files, clique-like components of pages all connected to one another, and disconnected subgraphs which are only reachable through search engines.

Content was in even worse condition: incomplete or redundant, outdated, self-contradictory, systematically drafted in an involuted form and with bureaucratic and sectorial language, incomprehensible to the users to whom they are addressed. In the context of such a transformation project, the evolutionary web redesign mode works as long as it is free to scale up. But when the complexity of the domain or intervention reaches significantly increased complexity, a technological and process paradigm shift is required. Without this, the resources and procedures in use up to that point prove wholly inadequate to handle the complexity involved. The risk is the indefinite absorption of all available resources, resulting in the intervention not being successful [6]. In the context of a complex socio-technical system, such as a large PAs, this leads to a waste of public resources and the realization of inadequate outcomes. Paradoxically, it also fosters internal resistance to change, manifesting as a rejection of a goal—user experience—that is now seen as entirely alien to the culture of PAs.

3. AI-driven big web redesign methodology

In this section we describe the methodologies employed to tackle the redesign issue according to the pipeline illustrated in Figure 2. The redesign process starts from a large-scale website whose structure is highly entropic. Borrowing methods from design and computer science, we combine quantitative and qualitative approaches to achieve a multidisciplinary future-proof solution capable of adapting to different users' needs whose result is a newly designed website that maximizes findability and maintainability over time.

3.1. Design and interdisciplinarity

The first condition for a successful redesign is an evolution of the radical collaboration approach [7]: in fact, when dealing with the UniGe use case, the first author conducted the design and prototyping phases in a multidisciplinary group, involving faculty with diverse expertise and other stakeholders within the university. This led to the subsequent structuring of a small, permanent

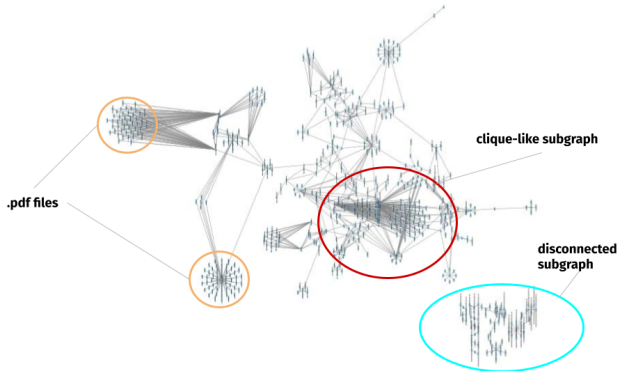

Figure 1: studenti.unige.it website structure (mid-2017). The website structure displays a set of idiosyncrasies, such as clique-like subgraphs, many clusters of PDF files linked from a few node pages, and an independent connected component that is a disconnected subgraph.

multidisciplinary group consisting of designers, developers, process analysts, copywriters and data scientists. As the complexity and project burdens grew, the prototyping group evolved into the permanent radical collaboration group, eventually becoming a design group as well. At the same time, the group's traction also expanded: the design research driving the original initiative increasingly needed data and the ability to analyze it to interpret the complexity of the socio-technical context and to guide design action. This, in turn, continued to exert a profound influence on technological choices and the choice of questions to be answered in the data. The second condition necessary to achieve change in complex settings, designed and built around the ascertained needs of users, is to have the deep and ongoing support of governance. In a PA or complex entity, services and processes are interconnected and depend - to name just a few factors - on the people sustaining them, the organizational models, the technological infrastructure, the operational practices, the available skills, the investments in reskill and upskill, human resource recruitment and management policies, the regulatory environment, and the financial situation. These are all mutually conditioning factors, and as the size of the institution grows, they grow more than proportionally, until they generate a system of constraints and cascading repercussions involving the entire institution. The human-centered design or redesign of a service impacts this system of interconnected

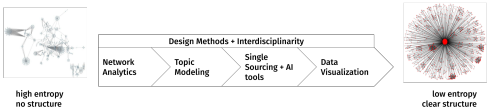

Figure 2: Pipeline representing the AI-driven big web redesign methodology

constraints, eventually involving compartments of the institution even quite distant from those directly involved in service delivery [8].

3.2. Network Analytics and AI for large-scale web redesign

The process of an effective web redesign that is more profound and radical than a "Face Lift" usually starts with a thorough content audit of the entire existing set of pages. This process is time-consuming and cannot be done manually if the number of pages ranges in the order of tens of thousands - a typical figure for large-scale websites. This was the main reason that guided us into leveraging AI approaches to understand the structure of the existing website, and define the optimal IA by categorizing the existing content.

Statistical testing to assess large-scale website structure The statistical analysis of the existing website starts by considering the website as a directed network of documents (nodes) connected through hyperlinks (edges). One example is represented in Figure 1, where the main structure of the *studenti.unige.it* (mid-2017) is depicted. By exploiting network analytics [9], we explored the distributions of the in-degree and out-degree, that are the number of in-coming links, and the number of out-going links, respectively. We argue that the scale-free property, that characterizes the *World-Wide-Web* as a whole [10], is no longer evident for specific sub-categories of websites. Particularly, we defined a method that could be used to better characterize topological properties deriving from different generative principles: central or peripheral. The first category is typically characterized by a strong central control in the design and evolution of the IA and content generation. Conversely, the last

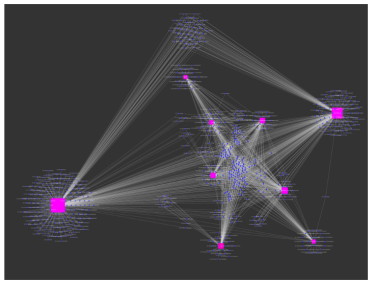

Figure 3: Topic modeling on studenti.unige.it website structure (mid-2017). The analysis identified 8 main topics, with many pages linked to multiple topics.

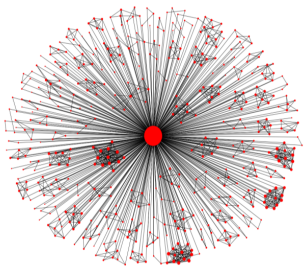

Figure 4: corsi.unige.it website structure.

category is completely user-guided and its evolution is likely to be random [11]. This method may be used to trace and monitor the evolution over time of the website content also according to the editorial model in use: a few people that are allowed to write anywhere in the website and tightly control the structure or a multitude of contributors that are allowed to write based on collaborative editing and community moderation. The result on the UniGe website denoted a combination of central and peripheral editorial strategies that, over time, led to a chaotic arrangement, confirming the need for a profound redesign.

Topic Modeling to infer an optimal Information Architecture We then proceeded by defining an AI-driven approach to define the optimal IA. Specifically, we exploited topic modeling methods [12, 13] to identify how many topics were discussed on the website and visualize how the pages often presented multiple unrelated topics, leading to a confusing UX. The result of such an experiment on *studenti.unige.it* is visualized in Figure 3, where the fuschia squares represent the topics and the blue dots represent the pages. Even at first glance, it is clear that many pages are linked by two or more topics, with a central cluster of pages connected to up to 5 topics.

3.3. AI-based component management system

To implement a new strategy for content creation and management, we devised a system that exploits single sourcing and AI methods to transform information into structured and reusable data, making it easier to create, maintain, and update content across multiple channels and user profiles. For example, a single piece of information can be automatically tailored to the needs of different audiences, such as students, faculty, and staff. The system also integrates with machine learning methods for image recognition and automatic translation, further reducing the burden on human editors. This overcomes the struggle typical of traditional content management systems to handle the ever-changing regulations, information overload, and new digital technologies that large organizations face [14]. The system highlights the importance of human-centered design throughout the content management process. It is designed to be user-friendly and easy to learn, even for non-technical users. This ensures that content creators can focus on creating high-quality content, rather than struggling with the technology. This process resulted in the publication of a new set of federated websites whose structure and content were the result of the entire pipeline just described. Figure 4 illustrates its hierarchical structure.

3.4. Assessing results through data visualization

To assess the overall usability of the newly designed website and the effectiveness of the entire procedure, we resorted to a set of UX tests conducted on a sample of 60 students divided into two groups ($M_{age} = 22.7yrs$), with the purpose of comparing the usability of the *corsi.unige.it* website, designed with the methods above, and of *studenti.unige.it*, which historically UniGe uses to communicate with its students. Each user must answer three questions related to a course of study, navigating exclusively on one of the two sites. The questions were the following. *Question A*: Is there an exam to enroll in the Business Administration BSc? What does it consist of? *Question B*: When is the deadline for Erasmus+ application? *Question C*: Find where to ask for an internship in car design as a MSc student in Design.

The exercise was deemed successful if the subject correctly answered the question at the end of the search. Researchers evaluated the effectiveness of the site by measuring the number of clicks and the time taken to complete the exercise, with each question allotted a maximum response time of 600 seconds. There were three potential outcomes: correct, partially correct, or incorrect answers. The vertical bubble chart in Figure 5 shows the result of the UX test. Each circle corresponds to a

	corsi.unige.it	studenti.unige.it
Time (s)	150.5	253.5
# Clicks	5	16
Success Score	1	-1
Success Rate	75%	7%

Table 1
Median time, median number of clicks, median success score, and global success rate between the two websites over the three questions.

user and is arranged along the vertical axis according to the number of clicks. The diameter is proportional to the time required to perform the exercise while the color represents its success (correct answer, partially correct, incorrect).

Table 1 shows how on average users performance on *studenti.unige.it* was significantly worse than on *corsi.unige.it*. In particular, in the case of *corsi.unige.it*, three out of four users answer correctly with just a few clicks and in a short time, with a success of 75 percent, while less than one in ten users responds correctly using *studenti.unige.it*. Although our experience was successful in achieving the objective we set in the beginning, this was not without hiccups and obstacles. In particular, a strong resistance to change was observed at the moment

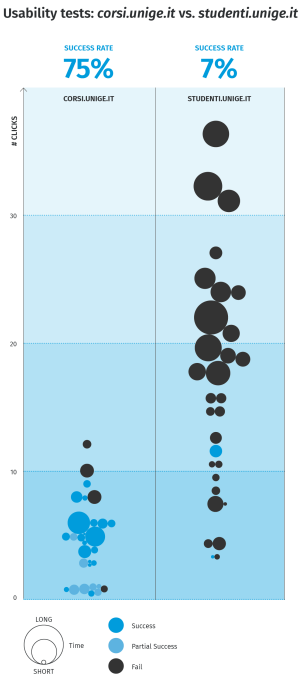

Figure 5: Usability tests: corsi.unige.it vs. studenti.unige.it

of implementing the newly redesigned system in the university procedures. Thus, human-centered design approaches and codesign and shared prototyping activities have proved invaluable allies in gaining consensus, incorporating the knowledge and skills of all stakeholders into the design process, and disseminating the culture of UX. However, splitting the design-driven digital innovation project into autonomous modules was necessary

This behavior, as predicted by [15], is to be expected when working within complex organizations to step up in the UX maturity model, particularly if the starting point is the bottom of the ladder, i.e. absent or limited UX. In this case, besides working on technological improvements, the institution must focus on a profound cultural change that supports UX knowledge and data awareness.

4. A new challenge: PhD Programme at PoliMi

Our successful experience in UniGe resonated across several national outlets and was presented in several news channels and conferences [14, 16] and culminated in winning the ForumPA award for best innovator [17]. The exposure we received from these initiatives allowed us to get in touch with other realities that were facing similar issues as the ones we faced in UniGe. In this line, we were able to begin a tight-knit collaboration with the Department of Design at Politecnico di Milano (PoliMi), where the Design PhD Programme at PoliMi was facing the problem of redesigning their website. The PhD Programme in Design at PoliMi is the largest PhD design course in Italy, with almost 90 PhD students enrolled.

Their goal was to devise a system that could allow for an optimal presentation of the research activity carried on by the PhD alumni and candidates as well as a website that could describe procedures to potential new students (admission procedure, courses requirements, etc.) and stakeholders. Our involvement since the very beginning of the redesign process, allowed us to adopt the methodology defined for UniGe. Together, we identified the major pain points from their old websites, mainly consisting of a lack of automation, an inactive homepage, and a content strategy that did not leverage single-sourcing, user profiling, or structured information. The result was a set of websites that penalized the research efforts. Applying our methodology allowed us to identify the root of these inefficiencies. In particular, PhD procedures lack a connection between the legacy data, processes and the website content. Also, in research dissemination pages, the existing data structure does not account for the interactions among researchers and therefore it is difficult to keep up-to-date and coherent. Currently, the project is focusing on designing a data-driven system which may enhance the usability of the websites through refining the

quality of the underlying information structure. A primary objective is to establish an architecture capable of robust automation. This architecture aims to convert natural language and simple information about researchers' activity in structured data. By exploiting legacy data in a seamless flux, the system reduces maintenance effort at the bare minimum and keeps data in sync and up to date. Also, the system will behave by considering research as the result of heterogeneous networks of people and topics. In doing so, the data produced by the research efforts become structured and easily reusable across different touchpoints.

5. Discussion and Conclusion

Our experience with two large scale universities, suggests that Italian academia still lacks concrete strategies and methods to tackle the ever increasing amount of information a PA faces today. Particularly in the context of web redesign, we defined a methodology that goes beyond a shallow visual redesign in favor of a new methodology where existing data is first used to guide the definition of an optimal IA and then structured to feed the touchpoints based on single-sourcing, interoperability, and user profiled approach. Our methodology exploits design thinking principles, User-Centred Design methods, agile programming, prototyping and innovators (young fearless developers and designers) to define a new way of thinking about information and how to distribute it across complex organizations. The result is a system that is able to govern a multitude of different and interconnected websites thanks to AI, Natural Language Processing and data management principles.

Academia represents a unique type of PA, tasked with the crucial responsibility of not only accumulating vast knowledge but also effectively sharing it with the broader public. However, like many PAs, academia tends to exhibit a conservative stance toward digital advancements, often resisting change and clinging to outdated methods. Rather than embracing true innovation, there's a tendency to equate progress with simply digitizing existing processes. This approach typically prioritizes documents over recognizing the central importance of data, which should serve as the foundation for all operations, particularly within sprawling institutions characterized by numerous interconnected yet autonomous departments.

Acknowledgments

A. Barla is part of the RAISE Innovation Ecosystem funded by the Recovery and Resilience Plan (NRRP) (ECS00000035)

References

[1] C. Pinho, M. Franco, L. Mendes, Web portals as tools to support information management in higher education institutions: A systematic literature review, International Journal of Information Management 41 (2018) 80–92.
[2] J. Nielsen, H. Loranger, Prioritizing web usability, Pearson Education, 2006.
[3] A. Vian, A big web redesign: data driven design research through practice and implementation, Fuori Collana, Genoa University Press, 2020.
[4] S. Grünwald, H. Bergius, Decoupling content management, in: Proc. of WWW2012, 2012.
[5] C. Bizer, T. Heath, T. Berners-Lee, Linked data: The story so far, in: Semantic services, interoperability and web applications: emerging concepts, IGI global, 2011, pp. 205–227.
[6] L. Rosenfeld, P. Morville, Information architecture for the world wide web, O'Reilly Media, Inc., 2002.
[7] B. Burnett, D. Evans, Designing your work life: how to thrive and change and find happiness at work, Knopf, 2020.
[8] D. A. Norman, P. J. Stappers, DesignX: Complex Sociotechnical Systems, She Ji: The Journal of Design, Economics, and Innovation 1 (2015).
[9] A.-L. Barabási, Network science, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371 (2013).
[10] A.-L. Barabási, Scale-free networks: a decade and beyond, Science 325 (2009) 412–413.
[11] D. Garbarino, V. Tozzo, A. Vian, A. Barla, A robust method for statistical testing of empirical power-law distributions, in: Proc. of WAW 2020, Springer, 2020.
[12] J. Chang, D. Blei, Relational topic models for document networks, in: Artificial intelligence and statistics, PMLR, 2009, pp. 81–88.
[13] J. Chang, D. M. Blei, Hierarchical relational models for document networks (2010).
[14] A. Barla, M. Cuneo, S. R. Nunzi, G. Paniati, A. Vian, AI-based component management system for structured content creation, annotation, and publication, in: Proc. of IHSI 2022., volume 22, AHFE International, 2022.
[15] K. Pernice, K. Moran, K. Whitenton, S. Gibbons, The 6 levels of ux maturity, 2024. URL: https://www.nngroup.com/articles/ux-maturity-model/.
[16] E. Capone, Intelligenza artificiale e big data, il futuro dell'università di genova passa da internet, 2020. URL: https://bit.ly/3TPJIed.
[17] ForumPA, Andrea vian: "la trasformazione digitale dei servizi informativi", 2023. URL: https://bit.ly/3J6eM4b.