Appendix: Methodological Approach

A1. Quantitative Approach

Sample of Analysis

The 30 Italian cities were selected on an equal basis within the territory to guarantee the high heterogeneity of the sample. To set the final sample, we based it on several constructs. First, we based on the homogeneity of the variables included in our model that has been considered uniform to assure the model approach. Second, we construct the sample of analysis based on the reference context as it is characterized by a different city size with a population ranging from 67,200 to 285,200. Finally, to guarantee the robustness of our sample, we apply the probability-proportional-to-size sampling in each stratum of the medium and large cities. This distinction between large and medium size has also been used as a control variable in our model.

Using this approach, the city with the largest population in each stratum emerges as a representative sample to be considered (Levy & Lemeshow, 2011). As a result, the final sample of analysis considered 30 of the 46 most populated Italian cities to guarantee the heterogeneity of the analysis sample. Thus, we use geo-software to draw the Italian map with our sample of analysis emerging from the sampling based on average from 2010 to 2021 (See Fig. A1), which shows the distribution and size of the cities under consideration. This empirical investigation focus on the Italian context as it currently is at the center of the academic and political debate on firms' green transformation, social policies and the current trends in sustainable advance, constituting a prominent empirical focus for the debates on sustainable development in Europe compared to other countries (i.e., Asia, US and Latin America).

Estimation Model

Due to the intrinsic characteristics of the panel dataset, which consists of a wide cross-section and a relatively short time period of 11 years, the dynamic Generalized Methods of Moments (GMM) estimation technique was employed in this article. By including time and city-fixed effects, the time-varying nature of the data was accounted for, and unobserved
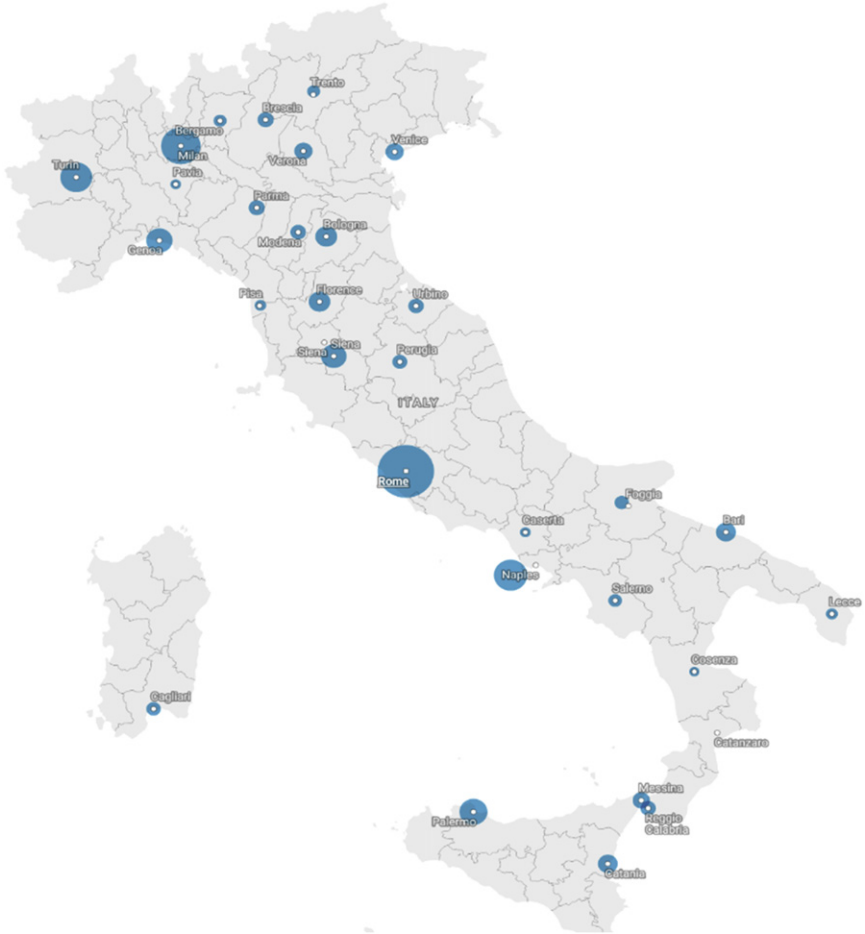
Fig. A1.   Sample of Analysis. *Source*: Created by author.

heterogeneity across cities was controlled. Following Roodman's approach (Roodman, 2009a, 2009b, 2009c) to the estimation model, internal instruments from lagged variables were also utilized to address the issue of over-identification (Windmeijer, 2005).

Roodman's approach (Roodman, 2009a, 2009b, 2009c) was employed in our estimation model to address the critical issue of identification and ensure the validity of our results. This approach relies on the use of internal instruments derived from lagged variables, and it was chosen for several reasons. Firstly, identification is a crucial aspect in estimating econometric models as it allows for the unique attribution of the effects of independent variables on the dependent variable. However, in the presence of endogenous

or correlated variables, identification can become challenging. The use of internal instruments, such as lagged variables, helps overcome this issue by providing additional information that aids in the identification process. Roodman's approach offers a robust solution for addressing over-identification, a common concern in econometric analysis. Over-identification occurs when there are more instruments than necessary to identify the parameters of interest, leading to inefficient estimates and potential biases. By employing internal instruments from lagged variables, we ensure that the model is neither under-identified nor over-identified, striking a balance that promotes accurate and reliable estimation (Wooldridge, 2002). The inclusion of lagged variables as instruments is particularly valuable in dynamic models, where the dependent variable's past values can serve as relevant predictors. By capturing the persistence and lagged effects of the dependent variable, the lagged variables act as valid instruments, allowing us to account for the dynamic nature of the relationship under investigation. Furthermore, Roodman's approach is well-suited for panel datasets with a wide cross-section and a relatively short time period, as is the case in our study (Roodman, 2009c). By incorporating time and city-fixed effects, we account for the time-varying characteristics of the data and control for unobserved heterogeneity across cities. This helps mitigate potential biases and provides a more accurate estimation of the effects of interest.

To ensure the validity of the dynamic GMM estimations and assess the robustness of the results, various diagnostic tests were conducted. Specifically, the Arellano and Bond test, the Sargan test, and the Wald Chi-Square test were employed (Bond & Arellano, 2012; Magazzini & Calzolari, 2020; Roodman, 2009b; Sargan, 1958).

The Sargan test was used to examine the joint validity of the instruments and detect over-identification in the model (Sargan, 1958). The choice of using the Sargan test instead of the Hausmann test was made, as it is known to be more effective in detecting over-identification and less susceptible to the instrument proliferation problem, as noted by Roodman (Roodman, 2009c). Specifically, unlike the Hausmann test, the Sargan test is known to be more effective in detecting over-identification and less susceptible to the instrument proliferation problem (Bond & Arellano, 2012; Breitung & Salish, 2021; Debarsy, 2012). By conducting the Sargan test, we ensured that our instruments were not correlated with the error term and that the model was properly identified. This test further enhanced the validity and reliability of our estimation results, contributing to the robustness of our analysis.

In addition, the Arellano and Bond test (Bond & Arellano, 2012; Lee & Yu, 2014) was applied to assess any autocorrelation in the idiosyncratic disturbance term. Further investigation of potential autocorrelation issues was conducted through the AR(1) and AR(2) tests. These tests helped evaluate the presence of autocorrelation and address it appropriately.

The Arellano and Bond test is a widely used method to detect autocorrelation in dynamic panel data models. It is based on the first-differenced transformation of the model equation and compares the coefficient estimates of lagged dependent variables with their predicted values. The test examines whether the lagged dependent variables are correlated with the first-differenced residuals, indicating the presence of autocorrelation. To further investigate potential autocorrelation issues, we conducted the AR(1) and AR(2) tests. The AR(1) test examines first-order autocorrelation by regressing the residuals on lagged residuals from the previous time period. Similarly, the AR(2) test examines second-order autocorrelation by regressing the residuals on two lags of residuals. These tests help us evaluate the presence and magnitude of autocorrelation in the error terms. If significant autocorrelation is detected, it suggests that the error terms are correlated over time, violating the assumption of independence. To address autocorrelation, we can employ various techniques such as adding lagged dependent variables, including additional lagged independent variables, or using alternative estimation methods such as Generalized Method of Moments (GMM) or Fixed Effects (FE) models (Allison, 1994; Elhorst, 2003; Wooldridge, 2002). By applying the Arellano and Bond test, along with the AR(1) and AR(2) tests, we can identify and address autocorrelation in our model. This ensures that our estimations account for the time-dependent patterns and improves the reliability of our results. By taking these rigorous steps to detect and mitigate autocorrelation, we enhance the validity and robustness of our analysis, providing more accurate insights into the relationship between variables in our dynamic panel data model.

Finally, to ensure the validity of the model and address heteroskedasticity, the Wald Chi-square test was conducted along with additional checks. Heteroskedasticity refers to the presence of unequal variances in the error terms, which can affect the efficiency and accuracy of the estimations (Lee & Yu, 2014; Roodman, 2009c). The Wald Chi-square test is a statistical test that examines whether the coefficients in the model are jointly significant. It assesses the overall validity of the model and helps determine if the estimated relationships between variables are statistically meaningful. To mitigate the effects of unobserved heterogeneity and endogeneity issues, the Roodman GMM dynamic approach (Roodman, 2009c) was adopted. This approach is a robust estimation method that addresses endogeneity and unobserved heterogeneity by using instrumental variables. In this case, the lagged-dependent variable was utilized as an instrument to capture the persistence of the dependent variable. By including the lagged-dependent variable as an instrument, potential biases caused by the endogeneity of the dependent variable and the presence of unobserved factors that may affect the relationship between variables are controlled. By employing the Wald Chi-square test, implementing additional checks, and adopting the Roodman GMM dynamic approach, the validity and reliability of the model are

enhanced. These methodological approach enable the control of hetero-skedasticity, address endogeneity and unobserved heterogeneity, and provide more accurate and robust estimations (Lee & Yu, 2014; Magazzini & Calzolari, 2020; Roodman, 2009a). As a result, meaningful conclusions can be drawn, and reliable inferences about the relationships and dynamics within the model can be made.

As a last check, the presence of multicollinearity, which refers to high correlation among independent variables in a regression model, was assessed using the Variance Inflation Factor (VIF). The VIF measures the extent to which the variance of an estimated regression coefficient is inflated due to multicollinearity (Akinwande, Dikko, & Samson, 2015). In our analysis, a VIF value of less than four is generally considered acceptable, indicating that the independent variables in the model are not highly correlated with each other. This suggests that multicollinearity is not a significant issue in our model. By ensuring low levels of multicollinearity, we enhance the stability and reliability of the results. High levels of multicollinearity can lead to unreliable coefficient estimates, making it difficult to interpret the impact of individual variables on the dependent variable. However, in our case, the VIF values below four indicate that the variables in our model are relatively independent and do not suffer from multicollinearity issues. Having a stable and reliable model allows us to make accurate inferences and draw meaningful conclusions from our analysis. It ensures that the estimated coefficients are more robust and the relationships between variables are accurately represented. Therefore, by assessing and addressing multicollinearity, we can have greater confidence in the validity and accuracy of our results.

By employing these rigorous statistical techniques and diagnostic tests, the robustness of the analysis was enhanced, and potential biases were minimized. These methodological choices strengthen the validity of the findings and provide a solid foundation for drawing meaningful conclusions from the research.

Data Collection and Variables

This research uses city-level variables and collects data from databases at the national (ANPR, AGCOM, ISTAT, GreenItaly, GSE, OECD, MIUR CoC and IBS) and international (EUROSTAT and OECD) levels. Specifically, in Tables A1 and A2, describe the variables adopted in the empirical analysis showing the source, type of variables, operationalization, and short description. Table A3 displays the results of the GMM model.

Table A1. Variables Description and Construction.

Variable	Description	Operationalization	Type	Source
Employment	Percentage rate of employment in the city (over 30 years)	Natural logarithm of the variable over the population yearly	Constant	ISTAT
Business creation	Number of companies registered in the "Chamber of Commerce" of each city per year	Natural logarithm of the variable over the total number of companies yearly	Constant	ISTAT, CoC
Start-ups	Number of companies registered in the "Start-Up" section in the "Chamber of Commerce" of each city per year	Natural logarithm of the variable over the total number of companies yearly	Constant	ISTAT, CoC
Innovative companies	Number of innovative companies registered in the "Chamber of Commerce" of each city per year	Natural logarithm of the variable over the total number of companies yearly	Constant	ISTAT, CoC
Female entrepreneurship	The total number of companies founded by female entrepreneurs in the city per year	Natural logarithm of the variable over the total number of companies yearly	Constant	ISTAT, CoC
Green companies	Companies with ISO 9001, ISO 14001, and EMAS (Eco-Management and Audit Scheme) certifications founded yearly in the city	Natural logarithm of the variable over the total companies in the city yearly	Constant	ISTAT, GreenItaly
SC digital implementation	Variable constructed based on nine indicators relevant to the technological development of the city (refer to Table A2)	Index [0-1] based on the indicators, populations, and distribution in the sample	Index	Multiple sources
SC amenities and facilities	Variable constructed based on seven indicators relevant to the liveability of the smart city (refer to Table A2)	Index [0-1] based on the indicators, populations, and distribution in the sample	Index	Multiple sources
City development	Economic development of the city according to the division proposed by the European community	Dummy variable that considers cities in the most developed urban areas (1) and cities in transition areas (0)	Dummy	ISTAT

Table A1. *(Continued)*

Variable	Description	Operationalization	Type	Source
City size	Size of the city considering that 300,000 is the threshold between medium and large cities	Dummy variable considers (1) cities that have a population greater than 300,000 and (0) cities that have a lower population	Dummy	ISTAT
Employment	Percentage rate of employment in the city (over 30 years)	Natural logarithm of the variable over the population yearly	Constant	ISTAT
City GDP	Gross domestic product produced in each city in the year *n* considered	Natural logarithm of the variable over the population in the city	Constant	EUROSTAT – OECD
Population	Descriptive number of the total population in the city in the year "*n*"	Natural logarithm of the variable	Constant	ISTAT-ANPR
Airports	Total number of airport (public and private) in the city and within 50 km	Natural logarithm of the variable over the population in the city	Constant	ISTAT
Private R&D	Total amount of private sector investment in research & development in the city per year	Natural logarithm of the variable over the population in the city	Constant	ISTAT
Public R&D	Total amount of public sector investment in research & development in the city per year	Natural logarithm of the variable over the population in the city	Constant	ISTAT
Total companies	Total number of firms active in the city, based on the registration on the Chamber of Commerce	Natural logarithm of the variable over the population in the city	Constant	IBS

Source: Table created by author.

ANPR (Anagrafe Nazionale della Popolazione Residente); ISTAT (Italian National Institute of Statistics); GSE (Energetic Services Management); GreenItaly; Terna (Italian Energetic Network); OECD (Organisation for Economic Co-operation and Development); IBS (Italian Business Register); EUROSTAT (Statistical office of the European Union).

Table A2. Independent Variables Construction.

Variable	Description	Operationalization	Type	Source
Smart City – Digital Technology Implementation				
Online services	Total number of online services proposed by the city on the municipal website per year	Natural logarithm of the variable	Constant	ISTAT
Broadband access	Percentage of families who have access to ADSL	Natural logarithm of the variable operationalized over the population in the city per year	Constant	AGCOM
Municipal app	Municipal app download number per year	Natural logarithm of the variable over the population in the city	Constant	ISTAT
Home-banking diffusion	Number of users who utilize home banking in the city per year	Natural logarithm of the variable over the population in the city	Percentage	ISTAT
Digital transparency	Number of public data concerning the investments of the city per year	Natural logarithm of the variable over the population in the city	Constant	ANAC
Digital openness	Total number of public access databases in the city per year	Natural logarithm of the variable over the population in the city	Constant	FPA
Social public administration	Variable based on engagement, productivity, and use of public online services in the city per year	Natural logarithm of the variable over the population in the city	Construct	FPA
Public Wi-Fi	Variable based on the number of access points, quality of service, and communication in the city per year	Natural logarithm of the variable over the population in the city	Construct	ISTAT
IoT development	The total amount of investments in IoT and ICTs in the city per year	Natural logarithm of the variable over the population in the city	Percentage	ISTAT
Smart City – Amenities and Facilities				
Health care	Total number of public and private hospitalization services in the city per year	Natural logarithm of the variable over the population in the city	Constant	ISTAT
Health personnel	Total number of employees in public and private hospitals in the city per year	Natural logarithm of the variable over the population in the city	Constant	ISTAT-IPS

Table A2. *(Continued)*

Variable	Description	Operationalization	Type	Source
Elderly assistance	Total number of public and private elderly reception services in the city per year	Natural logarithm of the variable over the population in the city	Constant	ISTAT-IPS
Hospital emigration	Total number of hospital emigrations to other regions for ordinary hospitalizations in the city per year	Natural logarithm of the variable over the population in the city	Constant	ISTAT-IPS
Childcare	Total number of childcare assistance based on private and public services in the city per year	Natural logarithm of the variable over the population in the city	Constant	ISTAT
Job integration	Total number of new residents who work permanently in the city and new residents who found a job in the first six months in the city per year	Natural logarithm of the variable over the population in the city	Percentage	ISTAT
Coworking area	Total number of coworking services in the city per year	Natural logarithm of the variable over the total number of companies in the city	Percentage	FPA-OECD
Education development	Number of people in the city who interacted with services offered in the educational field (public schools, courses, master's, or services in the educational field) in the city per year	Natural logarithm of the variable over the population in the city	Percentage	MIUR

Source: Table created by author.

AGCOM (Autorità per le Garanzie nelle Comunicazioni); ANAC (Autorità Nazionale Anticorruzione); FPA (Innovation in Public Administration); ISTAT (Italian National Institute of Statistics); IPS (Intrusion Prevention System); OECD (Organization for Economic Co-operation and Development); MIUR (Ministero dell'Istruzione e del Merito).

Table A3. GMM Model Estimation Results – The Effect of SC Implementation on Economic Outcomes.

	Model I		Model II		Model III		Model IV		Model V		Model VI	
	Coeff.	**s.e.**	**Coeff.**	**s.e.**	**Coeff.**	**s.e.**	**Coeff.**	**s.e.**	**Coeff.**	**s.e.**	**Coeff.**	**s.e.**
Employment $(t-1)$	0.492***	[0.112]										
Business creation $(t-1)$			0.573***	[0.083]								
Start-ups $(t-1)$					0.668***	[0.096]						
Innovative companies $(t-1)$							0.701***	[0.106]				
Green companies $(t-1)$									0.554**	[0.168]		
Female entrepreneurship $(t-1)$											0.648***	[0.120]
SC digital implementation	0.332**	[0.243]	0.492***	[0.112]	0.613***	[0.105]	0.698***	[0.123]	0.597***	[0.096]	0.736**	[0.108]
SC amenities and facilities	0.468***	[0.612]	0.265***	[0.232]	0.258**	[0.216]	0.227**	[0.183]	0.558***	[0.072]	0.363***	[0.334]
City development	0.106**	[0.212]	0.188***	[0.212]	0.132**	[0.268]	0.144***	[0.283]	0.141*	[0.261]	0.176**	[0.294]
City size	0.271*	[0.442]	0.371**	[0.442]	0.187**	[0.431]	0.199*	[0.466]	0.196**	[0.382]	0.231*	[0.354]
Employment	–	–	0.451*	[0.112]	0.543	[0.331]	0.332	[0.293]	0.198*	[0.301]	0.267*	[0.288]
City GDP	0.216**	[0.185]	0.199**	[0.185]	0.214**	[0.198]	0.226**	[0.213]	0.223*	[0.188]	0.258*	[0.258]
Population	0.324*	[0.188]	0.406**	[0.188]	0.319*	[0.204]	0.331**	[0.243]	0.328*	[0.218]	0.363**	[0.233]
Airports	1.366*	[0.236]	1.107*	[0.236]	1.064	[0.245]	1.076*	[0.230]	1.073***	[0.216]	1.108***	[0.281]
Private R&D	−0.153**	[0.271]	−0.199*	[0.271]	−0.113***	[0.255]	−0.101**	[0.271]	−0.104*	[0.234]	−0.069**	[0.260]
Public R&D	0.309**	[0.138]	0.178***	[0.138]	0.267*	[0.144]	0.255*	[0.152]	0.258**	[0.128]	0.223*	[0.150]
Total companies	1.072*	[0.402]	1.352***	[0.402]	0.853**	[0.578]	0.841***	[0.618]	0.844**	[0.539]	0.809**	[0.701]
City effect	*Included*		*Included*		*Included*		*Included*		*Included*		*Included*	
Year effect	*Included*		*Included*		*Included*		*Included*		*Included*		*Included*	

Wald chi²	365.32	401.23	386.43	379.22	392.67	361.81
AR(2)	1.87	1.34	2.26	2.08	1.95	2.13
p Value	0.113	0.127	0.168	0.142	0.117	0.144
Sargan test	46.35	48.24	59.16	53.18	51.93	64.85
p Value	0.669	0.745	0.692	0.633	0.659	0.704
No observation	330	330	330	330	330	330
No* city	30	30	30	30	30	30

Source: Table created by author.

Employment, business creation, start-ups, innovative companies, green companies and female entrepreneurship as a DV. Significance level: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. Robust standard errors in brackets. VIF (variance inflation factor) < 4. Arellano–Bond AR (2) is used to look for possible autocorrelation issues. Sargan test is used to look for possible over-identification restrictions in the model. The variable "Employment" as a control is not considered in Model 1 as it is used as the DV. The number of observations is reduced (360–330) for the leg effect (−1).

References

Akinwande, M. O., Dikko, H. G., & Samson, A. (2015). Variance inflation factor: As a condition for the inclusion of suppressor variable(s) in regression analysis. *Open Journal of Statistics*, *05*(07). doi:10.4236/ojs.2015.57075

Allison, P. D. (1994). Using panel data to estimate the effects of events. *Sociological Methods & Research*. doi:10.1177/0049124194023002002

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, *58*.

Bond, S., & Arellano, M. (2012). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, *58*(2).

Breitung, J., & Salish, N. (2021). Estimation of heterogeneous panels with systematic slope variations. *Journal of Econometrics*, *220*(2). doi:10.1016/j.jeconom.2020.04.007

Debarsy, N. (2012). The Mundlak approach in the spatial durbin panel data model. *Spatial Economic Analysis*, *7*(1). doi:10.1080/17421772.2011.647059

Elhorst, J. P. (2003). Specification and estimation of spatial panel data models. *International Regional Science Review*, *26*(3). doi:10.1177/0160017603253791

Lee, L. F., & Yu, J. (2014). Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics*, *180*(2). doi:10.1016/j.jeconom.2014.03.003

Magazzini, L., & Calzolari, G. (2020). Testing initial conditions in dynamic panel data models. *Econometric Reviews*, *39*(2). doi:10.1080/07474938.2019.1690194

Roodman, D. (2009a). A Note on the theme of too many instruments. *Oxford Bulletin of Economics & Statistics*, *71*(1). doi:10.1111/j.1468-0084.2008.00542.x

Roodman, D. (2009b). How to do xtabond2: An introduction to difference and system GMM in Stata. *STATA Journal*, *9*(1). doi:10.1177/1536867x0900900106

Roodman, D. (2009c). Practitioners' corner: A note on the theme of too many instruments. *Oxford Bulletin of Economics & Statistics*, *71*(1). doi:10.1111/j.1468-0084.2008.00542.x

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, *26*(3). doi:10.2307/1907619

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, *126*(1), 25–51. doi:10.1016/j.jeconom.2004.02.005

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Booksgooglecom (Vol. 58(2)). doi:10.1515/humr.2003.021

A2. Interview Appendix

To collect data on the internal implementation of the smart city ecosystem (part one of the book), a series of interviews were conducted with public managers and policymakers responsible for various functions and areas within the municipalities. These interviews aimed to explore the implementation of smart city projects, the capabilities involved, key actors, and the connections between the internal environment and institutions in the development of these projects.

The interviews took place between January and May 2023 and involved nine participants from different municipal departments, including the Digitalization office, tourism, mobility, control room, policies, urban investments, and front offices. The respondents held diverse job titles and had responsibilities and competencies relevant to their respective roles. During the interviews, participants were asked a range of questions to delve into the specific aspects of the digitalization projects in their cities. The questions covered topics such as the functioning and phases of the selected project, the activities carried out within each phase, the benefits observed from its implementation, and the challenges encountered. Additionally, the interviews explored the technological aspects of the projects, including the specific technologies used in different phases, how the participants became acquainted with these technologies, and whether they were sourced internally or through collaborations with external actors. The types of data involved in the projects and their collection and processing methods were also discussed. Furthermore, the interviews delved into the soft skills and competencies of the actors involved in the projects. Participants were asked about the actors' roles, their specific competencies, how contact was established with them, and whether the required professional skills were available within the organization or obtained through external collaborations.

The final set of questions focused on the project's evaluation, control, and monitoring. This included inquiries about how the benefits generated from the implementation were identified (ex-ante), how the results were evaluated (ex-post), who was responsible for the project's functioning, and whether the project had a positive impact on the local ecosystem.

Overall, these interviews provided valuable insights into the implementation of smart city projects, the coordination of different dimensions, and the role of various stakeholders. The information collected through these interviews contributes to a deeper understanding of the factors influencing the success and challenges of smart city initiatives.

The complete list of interviewees can be found in Table B1.

Table B1. List of Interviews.

Interview No.	Date	Time	Length (minutes)	Role within the Municipality
1	17-Gen	15:00	46:00	Digital transition sector manager
2	09-Feb	09:30	65:00	Manager of digital service and telematics implementation
3	09-Mar	10:00	38:00	Corruption prevention office manager
4	15-Mar	10:00	45:00	Registry and demographic services manager
5	15-Mar	11:00	50:00	Civil status office manager
6	04-Apr	10:30	52:00	Technical project management office for european projects
7	04-Apr	11:15	30:00	Monitoring and reporting office
8	22-Mag	15:30	64:00	Manager smart mobility
9	23-Mag	09:00	22:00	Manager control room and security

Source: Table created by author.

Exploratory interviews: Municipalities β.