CARMA 2026

8th Int. Conf. on Advanced Research Methods and Analytics
1-3 July 2026, Valencia, Spain
DOI: https://doi.org/10.4995/CARMA2026.2026.21959

Industry 5.0 technologies for automated data reporting in official statistics

Gabriella Fazzi, Paola Bosso, Giovanni Gualberto Di Paolo, Diego Distefano, Barbara Iaconelli, Alessandra Nurra, Pasquale Papa
Istat, Italian National Institute of Statistics, Italy.

How to cite: Fazzi, G.; Bosso, P.; Di Paolo, G. G.; Distefano, D.; Iaconelli, B.; Nurra, A.; Papa, P. (2026). Industry 5.0 Technologies for Automated Data Reporting in Official Statistics. In: 8th International Conference on Advanced Research Methods and Analytics (CARMA 2026). Valencia, 1-3 July 2026. https://doi.org/10.4995/CARMA2026.2026.21959

Abstract

This paper explores the potential of automated data reporting in the context of Industry 5.0 and organisational AI adoption. Through a multi-source approach to official business statistics, the study focuses on industrial production volume while aiming to extend findings to other business variables. The digitalisation of business processes presents opportunities for reducing data collection costs and respondent burden, enhancing timeliness and accuracy. As statistical institutes integrate new technologies and alternative data sources, there remains a lack of comprehensive literature on the topic. The experimental study unfolds in two phases: the first assesses data generated by advanced ERP systems through semi-structured interviews with providers; the second evaluates the acceptability of this reporting mode among businesses in the monthly industrial production survey. Sampling draws on a record linkage of the IIP survey (5,243 firms) and the ICT survey, yielding a target population of 299 firms. Preliminary outcomes highlight significant benefits in efficiency and data quality, though successful adoption hinges on overcoming technical, organisational, and trust-related challenges.

Keywords: *Industry 5.0; AI adoption; ERP systems; automated data reporting; M2M; official statistics.*

Article in Press

1. Background and Objectives

The digital transformation driven by Industry 5.0 significantly enhances the availability, granularity, and timeliness of real-time data generated by enterprises. Industry 5.0 represents the natural evolution of Industry 4.0, that is focused on digitalisation and automation to

All authors contributed to and approved the paper. Specifically, P. Bosso wrote §, 2.1, G.G. Di Paolo wrote § 3, D. Distefano wrote § 1, G. Fazzi wrote § 2.4, B. Iaconelli wrote § 2.3 and 2.3.1, A. Nurra wrote § 2.2, P. Papa wrote § 4.

This work is licensed under a Creative Commons License CC BY-NC-SA 4.0
Editorial Universitat Politècnica de València

maximise process integration and efficiency through cyber-physical systems, IoT, and machine-to-machine communication, streamlining the extraction, processing, and validation of information from enterprise systems, sensors, and digital platforms. Industry 5.0 includes all of these technologies and has an explicit focus on environmental sustainability, energy efficiency, system resilience, and worker well-being (European Commission, 2021).

Organizations are increasingly adopting a holistic perspective on their operations, viewing the value chain as an interconnected system. As supply chains become more complex and customer-focused, data emerges as a strategic asset. Growing technologies are transforming these supply chains into adaptive, intelligent networks, marking a crucial shift towards integrated, data-driven ecosystems that enhance business performance and yield modern, responsive statistical outputs.

The application of AI models and new technologies within the Industry 5.0 framework offers significant opportunities for official business surveys (Snijkers et al., 2023; Snijkers et al., 2025). Comparable pilots in Portugal and the Netherlands (Snijkers et al., 2025) have demonstrated the feasibility of automated collection, while also surfacing challenges around ERP heterogeneity and data quality assurance. This paper presents an experimental study designed to improve data integration and statistical production through a multi-source approach that harnesses AI and innovative digital solutions. Automating business reporting processes can lower costs and respondent burden while enhancing timeliness and accuracy (UNECE, 2024).

The study unfolds in two main phases: (i) semi-structured interviews with advanced ERP system developers to evaluate data structures, compliance with statistical requirements, and AI potential for automated transmission; (ii) semi-structured interviews with enterprises to assess feasibility and acceptance of M2M data collection. The methodology and preliminary findings aim to facilitate the future scaling of similar approaches.

2. Methodology and Experimental Steps

2.1. Experimental Domain and ERP Providers Assessment

The variable chosen for experimentation is industrial production expressed in volume (Bosso et al., 2025). This choice is motivated by both practical and methodological considerations: the information is collected exclusively through direct statistical surveys (Monthly Survey on Industrial Production - IIP; Annual Survey on Industrial Production - PRODCOM) and is not available from administrative sources, while Industry 5.0 digital platforms offer strong capabilities for capturing these data.

The experimental study engages major ERP solution providers to evaluate operational feasibility. Advanced ERP environments integrated with Manufacturing Execution System

Article in Press

(MES) modules represent the most promising infrastructure. Preliminary assessments focus on potential gains in time and costs for respondents. A dedicated interview protocol integrates two dimensions: knowledge-related aspects (user profile, AI-based services, data management) and operational issues crucial for structured data flows (security, privacy, data retention). The focus on AI integration into modern ERP solutions supports functionalities such as sales forecasting, predictive inventory management, and automated document classification.

2.2. Characteristics of the Digital Nature of Enterprises

In this context the evaluation of the digital maturity of enterprises is aimed at identifying firms potentially capable of supporting automated Machine-to-Machine (M2M) data exchange with Istat in the context of IIP and Prodcom; IIP survey plays a crucial due to its repetitive nature and strict timeliness requirements. The methodological framework assumes that M2M capability depends on three components: digital integration, data management, and analytical sophistication. On this basis, two composite indexes practices are constructed from observable ICT practices, providing a consistent basis for ranking enterprises by digital readiness, guiding the selection and preparation of interviews, and assessing each firm's propensity to support the transition towards data-driven official statistics. The empirical analysis integrates three sources: the IIP survey for 2025 (5,243 enterprises) and the EU ICT Usage and E-Commerce Survey for 2023 and 2025 (21,232 and 21,576 respondents respectively). Deterministic record linkage by firm identifier produces a matched dataset of 2,379 enterprises (45.4% of the IIP population): 1,356 observed in both ICT waves, 606 only in 2025, and 417 only in 2023. The matched sample has a higher prevalence of large enterprises, reflecting the ICT survey design.

The analytical population is restricted to firms adopting ERP systems (erp_use=1), representing a minimum requirement for structured data management and interoperability. Applying the most recent available ICT information reduces the population from 2,379 to 2,191 enterprises (excluding 188 firms without ERP). A more restrictive subset of 401 firms is obtained by requiring cloud-based ERP (clouderp_use=1) and sensor-based data analysis (an_sensor=1).

Ten binary variables from the ICT survey are grouped into four domains: (1) data exchange and integration (EDI, supply chain sharing); (2) digital management (CRM, BI); (3) data analysis capabilities (transactional data, sensor/M2M data); (4) AI adoption (production and financial applications).

Index1 is the simple sum of nine binary variables (excluding the supply chain variable, only available in 2023), ranging from 0 to 9, with a normalised version Index1n $\in [0,1]$. It captures the breadth of digital adoption and is conceptually related to Eurostat's Digital Intensity Index (Eurostat, 2026), with a specific focus on production-related and data-driven capabilities.

Index2 adopts a multidimensional structure, averaged across the four domains above, analogous to the DESI methodology. The first dimension takes the maximum of EDI and supply chain

Article in Press

usage; the second averages CRM and BI; the third averages transactional and sensor data analysis; the fourth averages AI in production and finance. General-purpose variables (overall AI use, internal data analysis) are excluded to avoid redundancy and focus on operational capabilities.

Both indexes show logical coherence (higher scores for more digital firms) and dimensional coherence (larger firms tend to score higher). Redundancy among constituent variables is limited, with high correlations observed only in a small subset. The indexes provide a consistent basis for ranking enterprises, selecting pilot units, guiding the transition towards data-driven statistical production, and supporting the understanding of the enterprises to be interviewed as well as the conduct of the interviews.

2.3. Target Enterprises and Sampling Strategy

The identification of target enterprises follows a multi-step elimination process. The production units most suited to the study are those characterised by traditional technological processes — specifically, mass manufacturing with large batch production, continuity of operations, and measurability through sensors or smart devices. To support automated M2M data transmission, a firm must satisfy at least two minimum technological requirements: (i) availability of technologies enabling the electronic measurement of production volumes; (ii) availability of technologies enabling electronic M2M data transmission.

Productions that are difficult to measure electronically are excluded, including those involving long and complex processes (shipbuilding, aircraft, robotics, machinery) where output is expressed in hours worked or monetary value, and products subject to frequent quality changes. Productions measured in hours or value in the IIP survey are also excluded, along with those belonging to NACE sectors 29.10 (motor vehicles) and 19.20 (refined petroleum products).

Among the available ICT indicators, the two most suitable proxies for the minimum technological requirements are: the presence of sensor-based data analysis (an_sensor) and the adoption of a cloud-based or hybrid ERP system (clouderp_use). The selection condition applied is:

IIP ∩ (ICT_2023 ∪ ICT_2025) | an_sensor=1 | clouderp_use=1 | uom=p.q[2]. | Code 29.10=0 | Code 19.20=0

After applying these criteria, the final target population consists of 299 legal entities and 568 productions.

[2] The expression stands for unit of measurement = phisical quantity (tonnes, cubic metres, litres, etc.)

Article in Press

2.3.1. Company Selection and Sectoral Strategy

For unit selection, a broad set of variables was considered: the monthly response profile for 2025, all products per local unit, units of measurement, NACE codes, value-added weights, annual production flow (2025), inclusion in the PRODCOM survey (2024 edition), turnover of the legal unit, and number of employees (2024), the level of digital maturity.

The sectoral strategy prioritises depth over breadth: selecting multiple units within the same sector rather than few units across many sectors, in order to assess the degree of technology diffusion within each sector (Patton, 2002). The four sectors selected are: food and beverages, glass and tiles, plastic products, and chemicals. The first three were chosen for their reliance on mass production methods; the chemical sector was included to introduce analytical complexity.

Within each sector, units were ranked in descending order by product-class weight and then by firm weight within the class. Preference was given to firms with a higher number of products and multiple local units, allowing assessment of electronic measurement across heterogeneous product types and plant configurations, reducing the required sample size while maximising analytical coverage (Eisenhardt, 1989). No additional preferences were introduced for digitalisation level beyond the initial selection criteria, in order to avoid over-representation of technologically advanced firms (Cochran, 1977).

The exercise identified a core sample of 23 firms plus a reserve list of 20, all of which have 100 or more employees and a medium–high level of digital maturity, based on ICT composite indexes, with 11 in food and beverages, 5 in chemicals, 5 in glass and tiles, and 2 in plastic packaging. The reserve list follows a similar sectoral distribution.

2.4. Design and Administration of Semi-structured Interviews

The experimental phase adopts semi-structured, in-depth qualitative interviews as its primary data collection strategy (Kvale & Brinkmann, 2009). This design is based on the well-established strengths of qualitative interviewing: depth, contextual sensitivity, access to participants' situated practices, and the ability to surface unanticipated issues (Corbetta, 2003; Patton, 2002). Two distinct but conceptually aligned interview guides were developed, one for enterprises, one for ERP providers, sharing a common analytical framework while reflecting each actor's different position in the data ecosystem.

The contact strategy is gradual and non-coercive. Initial engagement is via email invitation framing participation as exploratory and voluntary, clearly separated from compulsory survey obligations. A telephone follow-up is used only for scheduling. Interviews are conducted remotely via Computer Assisted Video Interviewing (CAVI) — a flexible mode positioned between face-to-face and telephone, particularly suited to busy professionals familiar with digital tools. Each interview is expected to last 45–60 minutes. Interviewers are thematic

Article in Press

specialists trained in active listening, probing strategies, and minimising interviewer bias (Knott et al., 2022).

The enterprise guide investigates: current organisational workflows for producing and transmitting production-in-volume data; roles involved and coordination across sites; internal data transformations (unit conversions, recoding); enabling technologies (ERP, MES, sensors, analytics); and conditions for accepting automated transmission, with explicit attention to data control, security, privacy, and allocation of responsibilities.

The ERP-provider guide focuses on: availability and generation processes of key variables and metadata (including MES and sensor integration); feasible M2M channels and formats (APIs, SFTP, CSV, JSON, XML); security and auditability requirements; continuity parameters; and the feasibility of a standard "Istat module". A complementary variable checklist may be left with respondents for post-interview completion.

Both guides are flexible frameworks rather than fixed scripts, relying on open-ended prompts and probing to encourage respondent-led elaboration while maintaining stable thematic anchors for cross-case analysis. The number of interviews is guided by saturation principles: data collection continues until additional interviews yield diminishing analytical returns, accounting for both code saturation and meaning saturation (Guest et al., 2006; Hennink et al., 2017).

3. Early outcomes and discussion

Preliminary outcomes from Istat's Business Statistical Portal indicate substantial potential benefits in terms of time savings. For short-term surveys, reporting time decreases by 68% for enterprises with more than 20 employees and by 86% for those with over 250 employees. Across the entire population of enterprises with 20 employees or more, this corresponds to an estimated annual saving of about 165,000 working hours. Large enterprises would save on average 17 hours per year, equivalent to approximately €4,943,000.

To identify strengths, weaknesses, opportunities, and threats associated with this approach, the project team conducted a SWOT analysis (Table 1). The SWOT analysis constitutes the conceptual framework for the analysis of the interviews and digitalisation indicators, making it more directly actionable. In particular, the ST and WT strategies reflect concrete barriers identified in comparable pilots (Snijkers et al., 2025): ERP heterogeneity, legal responsibility allocation, and cultural resistance to shifting from web-based to M2M reporting. It is important to note that benefits may not be fully realised in the short term, as much will depend on the extent to which solutions can be scaled and standardised.

The experiment connects to broader research initiatives by leading statistical institutions. Many organisations are progressively moving towards multi-source production models that integrate alternative data inputs, enabled by generative AI, machine learning, and web scraping. This

Article in Press

approach is fully consistent with that strategic direction. Enterprises are not merely data sources but active participants whose trust and consent are integral to the model's sustainability. Practical challenges include initial engineering and integration costs, uneven digitalisation across Italian SMEs, and the need for complex statistical and organisational coordination to ensure coherence between traditional survey data, administrative records, and M2M flows.

Further steps involve alignment with statistical information requirements to evaluate whether data extracted from enterprise platforms comply with official statistical standards. A subsequent step concerns activation of a pilot M2M reporting flow with volunteer companies to refine procedures and assess timeliness and data quality. Finally, comparative analysis will compare traditional collection methods with M2M-based transmission to evaluate data reliability and process efficiency.

Table 1. Use of ERP platforms for automated reporting: SWOT analysis

	Opportunities (O)	**Threats (T)**
Strengths (S)	**SO Strategies (Maxi–Maxi)** *Leverage strengths to exploit opportunities* • Capitalise on structured collaboration with ERP providers to align with EU digital data policies. • Use strong technical expertise to integrate AI into ERP-based data extraction. • Exploit automation potential to scale the model across multiple statistical domains. • Position the initiative within Industry 5.0 ecosystems to attract institutional support.	**ST Strategies (Maxi–Mini)** *Use strengths to mitigate threats* • Develop common data standards to reduce ERP heterogeneity. • Strengthen governance and cybersecurity frameworks leveraging internal expertise. • Formalise partnerships with ERP providers to reduce fragmentation and risks. • Promote transparency and methodological clarity to reduce resistance and trust issues.
Weaknesses (W)	**WO Strategies (Mini–Maxi)** *Use opportunities to overcome weaknesses* • Use EU digital transition policies to support SME ERP adoption. • Promote standardisation initiatives to address heterogeneous data structures. • Leverage funding under digitalisation programmes. • Use AI-based tools to compensate for structural ERP differences.	**WT Strategies (Mini–Mini)** *Minimise weaknesses and avoid threats* • Develop phased implementation to manage resource constraints. • Prioritise sectors with higher ERP penetration to ensure coverage. • Establish clear legal agreements to reduce regulatory risks. • Implement change management strategies to address cultural resistance.

4. Conclusions

The automated data reporting model offers significant potential benefits in terms of process efficiency, data quality, and reduction of respondent burden. This study proposes a strategic framework to assess the feasibility of a generalised M2M data reporting model, leveraging specialised ERP modules designed for Industry 5.0 environments. The potential advantages

Article in Press

include a substantial reduction in response burden, lower long-term costs, and improvements in the quality and timeliness of statistical outputs, addressing several dimensions of the Total Survey Error (TSE) framework. On this regard, the analysis confirms the need to evolve the traditional Total Survey Error (TSE) paradigm (Groves, R.M. et al., 2009) towards an Extended TSE (ETSE) framework, capable of incorporating new forms of error. The methodological contribution of this paper lies in the combination of a record-linkage-based sampling framework, two composite digitalisation indicators, and a structured qualitative interviewing protocol, a design that can be replicated in other statistical domains or national contexts. A central contribution of the study lies in the recognition that digital maturity has become a key stratification variable for official statistics. A more comprehensive evaluation will become possible as the experimentation progresses. Policy implications involve a robust governance framework capable of ensuring interoperability, transparency, and effective collaboration among statistical institutions, ERP providers, and the business community associated to gradual and experimental implementation strategy. By aligning technological innovation with appropriate governance mechanisms, this approach has the potential to contribute to the evolution of official business statistics towards a more efficient, timely, and sustainable statistical production system.

References

Bosso, P., Di Paolo, G. G., Distefano, D., & Papa, P. (2025). *Acquisition of new sources for official statistics through Industry 5.0: Preliminary results*. Expert Meeting on Statistical Data Collection, UNECE.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). Wiley.

Corbetta, P. (2003). *Social research: Theory, methods and techniques*. Sage.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532–550.

European Commission. (2021). *Industry 5.0: Towards a sustainable, human-centric and resilient European industry*. Publications Office of the EU.

Eurostat. (2026). *Reference metadata on Digital Intensity Index (DII)*. https://ec.europa.eu/eurostat/cache/metadata/en/isoc_e_dii_esmsip2.htm

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2nd ed.). John Wiley & Sons

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1), 59–82.

Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017). Code saturation versus meaning saturation: how many interviews are enough? *Qualitative Health Research*, 27(4), 591–608.

Knott, E., Rao, A. H., Summers, K., & Teeger, C. (2022). Interviews in the social sciences. *Nature Reviews Methods Primers*, 2, Article 73.

Article in Press

Kvale, S., & Brinkmann, S. (2009). *InterViews: Learning the craft of qualitative research interviewing* (2nd ed.). Sage.

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Sage.

Snijkers, G. (Ed.). (2023). *Advances in business statistics, methods and data collection*. Wiley.

Snijkers, G., Bavdaž, M., Biffignandi, S., Jones, J., Petersen, T., Schouten, B., & Van Delden, A. (2025). Automated business data collection: Experiences in Portugal and the Netherlands. *Statistical Journal of the IAOS*, 41(2), 216–233.

UNECE. (2024). *Adopting artificial intelligence in the production and dissemination of official statistics by Norway* (ECE/CES/2024/21). 72nd Plenary Session of the Conference of European Statisticians.

Article in Press