Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Design and development of the first electronic multicentric data registry of obstructive sleep apnea in Italy: clinical workflow integrated with digital data management and artificial intelligence

Chiara Maria Palo
Istituti Clinici Scientifici Maugeri
Valentina Tibollo
Istituti Clinici Scientifici Maugeri
Francesco Fanfulla
francesco.fanfulla@icsmaugeri.it

Istituti Clinici Scientifici Maugeri
Vanessa Tang
Istituti Clinici Scientifici Maugeri
Riccardo Bellazzi
University of Pavia
Andrea Romigi
Istituto Neurologico Mediterraneo
Maria Rosaria Bonsignore
Ospedale Vincenzo Cervello
Fabio Colombo
Istituto Auxologico Italiano
Carolina Lombardi
Istituto Auxologico Italiano

Research Article

Keywords: Sleep medicine, Obstructive sleep apnea, Multicentric study, Electronic data collection, REDCap platform, Artificial intelligence, Patient reported outcomes, Data governance, Data management, e-infrastructure.

Posted Date: March 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8968393/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Obstructive Sleep Apnea (OSA) is a high prevalent breathing condition, well characterized in Europe. However, data on the Italian population are still lacking in clinical presentation, polysomnographic characteristics, and comorbidities. The Italian PheNotypes Obstructive Sleep Apnea Study (IPNOS) aims to build the first electronic national registry for OSA research, investigating a population of more than 5,000 patients with suspected OSA.

Methods

The electronic platform REDCap (Research Electronic Data Capture) has been chosen to develop the registry. Analyzing key user, digital and ethical requirements, we designed a tool that complies with current regulations on health data management and able to host a multicenter data collection. Customized computer routines ensure automatic data import from different hospital information systems and real-time quality checks.

Results

The IPNOS registry currently includes 27 electronic forms and 803 variables. To date, data from more than 5,300 patients have been collected across 16 Italian sleep centers. Patients can complete digitally questionnaires with automatic saving of responses within the REDCap platform, thereby reducing the burden associated with manual data transcription. An artificial intelligence tool enables the automated extraction of structured data from text polysomnography reports for direct integration into the registry.

Conclusions

The IPNOS registry is a new, interoperable platform for OSA research in Italy. It ensures multicenter collaboration, guaranteeing an easy expansion of the centers' network. This registry lays the groundwork for future studies and strengthens Italy's contribution in global sleep medicine.

Trial registration

ClinicalTrials.gov, NCT06872242. Registered 12 March 2025. Retrospectively registered.

1. Background

Obstructive sleep apnea (OSA) is a high prevalent disorder with substantial clinical impact [1]. In Italy, approximately 12 million people experience sleep disorders, and OSA may affect up to 15% of individuals over the age of 40 [2–3].

The European Sleep Apnea Database (ESADA) study provided the European profile of OSA in recent years [4–5]. Marked differences have been described across European macro-regions in terms of symptoms and associated complications [6].

However, the Italian population remains largely unexplored, with limited evidence on the prevalence of OSA, its clinical presentation, and associated comorbidities. Moreover, substantial variability exists across the country in terms of diagnostic availability, technological resources, and scientific approaches to the management of this condition. This heterogeneity may become even more pronounced in the near future, given the rapid expansion of centers involved in sleep disorders.

1.1 The project

The Italian Phenotypes Obstructive Sleep Apnea Study (IPNOS) consortium – an Italian network of sleep medicine centers endorsed by the Italian Academy of Sleep Medicine (AIMS) [7] – was established to investigate OSA in Italy. The consortium is structured according to a *hub-and-spoke* model, with four centers hubs (including the coordinating center and three principal units) and twelve spokes centers that support the recruitment process, as shown in Fig. 1.

The main coordinating center is Istituto Auxologico Italiano in Milan, which collaborates with three leading centers: Istituti Clinici Scientifici Maugeri in Pavia (Northern Italy), IRCCS Neuromed Pozzilli in Isernia (Central Italy), and Ospedali Riuniti Villa Sofia-Cervello in Palermo (Southern Italy). This organizational core (the *hubs*) is connected to the AIMS network, which includes other twelve centers. The full list of sleep medicine centers involved in the IPNOS consortium is provided in *Supplementary Materials* – section 1.

The IPNOS study [8] is supported by a grant of Italian Ministry of Health (PNRR-MAD-2022-12375812) [9] and firstly approved by the internal Ethics Committee (EC) of Istituto Auxologico Italiano of Milan on December 21th 2022 (protocol number 2022_12_21_01). This study aims to analyze patients with suspected OSA through a national electronic registry designed to collect standardized clinical data across multiple centers. The objectives of the registry are as follows: i) to establish the largest registry of Italian patients with OSA; ii) to conduct cross-sectional, prospective, or long-term follow-up studies aimed at improving diagnosis and treatment; iii) to promote education, and standardization of procedures among participating Italian centers; iv) to foster collaborative links with other databases, such as ESADA.

2. Methods

The methodological approach used during the creation of the IPNOS registry is structured as shown in Fig. 2.

An inventory of user needs, information technology (IT) and ethical requirements has been conducted to identify the most suitable IT platform available to develop the IPNOS registry, also considering the restrictions of current regulations [10–11]. A dedicated multidisciplinary working team has been established, including: i) clinicians responsible for the clinical management of the disorder; ii) engineers and computer scientists providing the IT competencies required during the registry design; iii) statisticians supporting the analysis of the collected data [12]. The methodology followed during registry's development is shown in Fig. 3.

2.1 Requirements evaluation and variables definition

The development of an electronic registry in healthcare requires protection of data ownership and controlled use and sharing of patient information, as defined in the General Data Protection Regulation (GDPR) [10, 13]. The Data Protection Officer (DPO) guarantees compliance with regulations by each center and data security is strongly ensured through the following measures: i) avoid storing highly sensitive data from patients; ii) keep track of every type of action performed on data (logging); iii) limit visibility on data; iv) guarantee the Findable, Accessible, Interoperable and Reusable (FAIR) principles and data minimization [14, 15]; v) ensure pseudonymization of data [16].

The respective ECs have approved a data management framework whereby each center has access exclusively to its own data. Full access to the complete dataset is restricted to the study's coordinating investigators. An external company [17], contracted within the IPNOS study to manage the platform hosting the registry, has been formally designated as the Data Processor, due to its comprehensive visibility over all centers' data for maintenance purposes. Each participating center acts as an independent Data Controller for its own data and has executed specific agreements to formally appoint the company in this role.

Registry users can be categorized as data providers and data collectors, each with distinct requirements. Data providers – including patients, research laboratories, clinicians, and hospital information systems – require a system capable of accommodating multiple data formats, integrating heterogeneous data sources, and capturing patient-reported outcomes. Data collectors – such as clinicians, laboratory technicians, patients, and IT staff – require rapid, structured, and intuitive data entry processes. IT operators additionally require a highly customizable digital platform with access to database queries and application programming interfaces (APIs) [18].

Through close collaboration between healthcare professionals and technical staff, the full set of variables to be collected has been defined. Clinical data, validated and customized questionnaires were structured for longitudinal collection. These steps constituted the foundation for the selection and development of the registry platform.

2.2 Platform selection

Among different open-source electronic data collection (EDC) platforms, Research Electronic Data Capture (REDCap) has emerged as one of the most robust and established solutions, combining high reliability with a stable governance model and comprehensive documentation [19]. More details about REDCap platform are provided in *Supplementary Materials* – sections 2.

2.2.1 REDCap in multicenter studies

In case of multicenter studies, the Data Access Groups (DAGs) functionality of REDCap allows us to manage the different centers involved. Each center is identified by a representative name and a unique code and it is associated with its respective users. All centers' data are stored on a common REDCap-based server, as shown in Fig. 4.

2.3 Platform design

The REDCap registry design phase began in January 2024 and took approximately six months to complete. Before proceeding with eCRFs implementation, the virtual environment to host REDCap has been set with firewalls to protect data access and the activation of periodic backup procedures. More details about the virtual environment set are provided in *Supplementary Materials* – section 3.

eCRFs has been designed using the REDCap's Online Designer functionality. Through various REDCap specific features, it has been possible to limit missing data and apply customized compilation logic to data collection.

Before starting the collection of real data, the registry has been tested by healthcare personnel to highlight any issues, shortcomings, or improvements to be made. After validation, the test platform was deactivated, and the production environment has been updated with all IPNOS users (Fig. 5).

3. Results

The IPNOS electronic data registry is active since October 8, 2024. It is the first IT platform in Italy to host such a large collection of data on OSA patients. Data are entered in tabular format, with rows representing pseudonymized patients and columns corresponding to eCRF variables.

The data registry structure consists of 27 eCRFs and 803 variables, which can be grouped into the main topics shown in the tables below (Table 1 and Table 2).

Table 1
eCRFs of the IPNOS data registry.

Patient History	eCRF	Variables' number
	Personal Data	**26**
	Driving History	24
	Symptoms and Signs	37
Clinical Assessment	Comorbidities	74
	Drugs	6
	Anthropometric Measurements	13
Clinical Exams	Cardiological Assessment	46
	Blood Chemistry	69
	Sleep Study	38
	Sleep History	74
	Pulmonary Function Tests	32
	Screening ABPM	12
Therapies	Therapy Prescription	77
Questionnaires & Tests	Anamnestic Questionnaire	44
	Psychomotor Vigilance Task	13

Table 2
Validated questionnaires of the IPNOS data registry.

eCRF	Variables' number
Epworth Sleepness Scale	12
Insomnia Severity Index (ISI)	11
Morningness-Eveningness Questionnaire (MEQr)	9
Pittsburgh Sleep Quality Index (PSQI)	35
Hospital Anxiety Depression Scale (HADS)	20
Fatigue Severity Scale (FSS)	14
CGI/PGI Severity Scale	4
CGI/PGI Improvement Scale	4
Restless Legs Syndrome Rating Scale (IRLS)	12
Difficulties in Emotion Regulation Scale (DERS)	30
Perth Emotional Reactivity Scale (PERS)	28
Montreal Cognitive Assessment (MOCA)	12

Several eCRFs have been configured as "repeatable," allowing them to be completed multiple times for the same participant. The repeatable Drugs eCRF is designed to collect data for one medication at a time, including its Anatomical Therapeutic Chemical (ATC) code. REDCap supports the linkage of eCRF variables to ontologies available through the National Center for Biomedical Ontology (NCBO) BioPortal, an online repository hosting more than 450 biomedical ontologies [20]. Biomedical ontologies provide standardized, machine-readable vocabularies that enhance semantic interoperability. Within REDCap, integration with BioPortal enables the automatic adoption of standardized terminology and has been implemented for the ATC code variable in the Drugs eCRF.

Patients undergo multiple clinical evaluations. Polysomnography, considered the gold standard for diagnosing sleep disorders such as OSA [21], records parameters including brain activity, heart rate, respiratory patterns, oxygen saturation, and snoring during sleep. Information about prescribed therapy for the treatment of OSA (Positive Airway Pressure - PAP, weight loss, etc.) are also collected.

Various validated and customized sleep questionnaires are completed directly by patients. Validated questionnaires consist of standardized items that generate a final score, which is automatically calculated by the registry using the Calculated Field functionality. Finally, the Psychomotor Vigilance Task (PVT) is administered to assess patients' psychomotor reactivity [22].

IPNOS data collection is organized across multiple time points, specifically including a baseline assessment and a follow-up evaluation conducted at least 6 months after the initial visit. Additional

follow-up time points can be readily incorporated as needed.

At present, 16 centers and 65 authorized users contribute data to the registry, each operating under predefined role-based access privileges.

3.1 Electronic patient reported outcomes

Traditionally, sleep questionnaires are administered to patients in paper form during their initial visit. However, paper-based patient-reported outcomes are associated with higher rates of missing data, transcription errors, and limited standardization

To address these limitations, digital questionnaires have been implemented to standardize patient responses and enable direct data capture within the registry, thereby reducing the burden of manual transcription from paper forms. By leveraging the Survey functionality within REDCap, it has been possible to configure all the questionnaires for online administration. The majority of items have been set as mandatory fields significantly decreasing the proportion of missing data. The digital questionnaires have been administered on-site during hospital visits, optimizing patient waiting times through the use of tablets provided by the facilities. Patients received support from dedicated healthcare personnel, who were available to assist both content-related questions and technical issues.

This approach has been implemented in more than 100 patients at the Maugeri Institute in Pavia and Auxologico in Milan, the centers with the highest enrollment rates. Among the 150 critical issues identified, 30% were technical in nature (primarily related to tablet usage), whereas 60% concerned difficulties in interpreting questionnaire content. These findings led to targeted revisions and partial refinement of several questionnaire sections.

3.2 Quality checks of data collection

The availability of high-quality data is essential for conducting robust analyses and generating reliable results supported by sound scientific evidence [23–24]. REDCap provides predefined validation rules to identify missing data, out-of-range values, and errors in automated calculations. In addition, customized rules have been implemented using the Data Quality module to detect more specific inconsistencies and omissions. To date, 47 ad hoc data quality rules and 58 dedicated reports have been implemented to investigate specific characteristics of the study population. A data manager has been appointed to supervise ongoing data quality activities, and a designated monitor figure is responsible for conducting periodic quality checks and resolving identified issues.

3.3 Overall system architecture

The developed electronic registry is part of a more complex IT architecture that ensures interoperability among different systems, as shown in Fig. 6.

REDCap also includes a Data Import Tool that supports the upload of Comma-Separated Values (CSV) files formatted according to platform specifications. This functionality has been used to integrate the

registry with hospital information systems, enabling the automated import of pre-existing data through customized procedures. In particular, a dedicated routine was developed to import blood chemistry test results already available in digital format from external systems.

For administration of the PVT test, each center has been equipped with a dedicated workstation on which the specific test software is installed. This software generates a CSV output file, that is subsequently formatted according to REDCap requirements and uploaded via the Data Import Tool.

By contrast, all questionnaires and sleep assessment scales completed digitally by patients using Survey functionality required neither manual nor automated CSV-based data entry. Patients responses are captured and archived in real time within the platform in a structured and standardized format.

4. Discussion

The present paper describes the methodology and decision-making process underlying the development of the first electronic data registry in Italy dedicated to sleep medicine, with a specific focus on OSA disorder.

The registry has demonstrated its capacity to centralize heterogeneous data while ensuring consistency, completeness, and accessibility across diverse clinical workflows. It standardizes data collection among all participating centers, thereby minimizing challenges related to data harmonization and integration. The electronic data capture framework reduces transcription errors, facilitates longitudinal follow-up, and enables rapid aggregation and analysis of results.

4.1 Limitations

Despite these advances, the initial phase of data collection still frequently relies on paper-based methods. The use of EDC tools requires basic digital literacy, which may represent a barrier, particularly among elderly patients, in the absence of caregivers or dedicated support staff. Furthermore, paper-based data collection does not require dedicated IT infrastructure, which may be limited in hospital settings due to financial and organizational constraints.

Consequently, the time and operational burden associated with manual data entry procedures by healthcare professionals remain substantial. Although electronic tools such as REDCap streamline and accelerate data entry processes, dedicated personnel for paper-to-electronic transcription are often still required.

The hybrid strategy employed in the IPNOS study, which combines both paper-based and digital data collection, represents a pragmatic solution to mitigate these challenges [25].

4.2 Future perspectives

To further reduce manual data entry, the current architecture is being enhanced through the implementation of a system capable of autonomously processing Portable Document Format (PDF) documents containing unstructured information. This process generates a structured file containing the data extracted from the original PDF, which can then be automatically adapted and input into REDCap.

Within the IPNOS study, this solution has been developed to analyze the PDF text reports produced by the polysomnography examinations, which typically produce unstructured PDF document for each patient. This system is currently under development by Auxologico and Maugeri research teams in collaboration with an external company specialized in medical systems and devices [26]. The system implements the Reliable Extractive Question Answering (REQA) algorithm, which streamlines the creation of objects representing structured report templates [27], thus fostering a certifiable interpretation of the information contained within the relevant reports. It employs generative Artificial Intelligence (AI) and advanced Natural Language Processing (NLP) preprocessing pipelines and language models to effectively parse documents and respond to user-defined inquiries. The REQA system produces a JavaScript Object Notation (JSON) or CSV output, associating a confidence score to the extracted information. The flow described is presented in Fig. 7.

Figure 8 shows the workflow followed for each patient: at each center, every polysomnography produces a PDF report, which is provided as input to REQA algorithm to extract the relevant sleep data in a structured format. A developed IT routine processes the output file and formats it as required by REDCap. The incorporation of AI tools into Aux Reader introduces a significant innovation in IPNOS data collection.

We would like to further enhance the ability of patients to digitally complete all IPNOS questionnaires. The tested tablet-based questionnaires can only be completed while patients are in the hospital. Furthermore, it is always necessary to identify dedicated personnel responsible for managing this activity, taking away time from other outpatient activities.

For this reason, further steps will allow patients to fill out questionnaires digitally directly from home, via email, or on their smartphones. To do this, the registry can be integrated with MyCap, a free mobile application for patients that can be installed on iOS and Android devices to capture patient-reported results for any REDCap project [28]. Patients will be able to receive the list of questionnaires directly on the app installed on their devices, and their responses will be immediately synchronized and stored into the registry via a Secure Sockets Layer (SSL) connection. Patients create a 6-digit Personal Identification Number (PIN) that is used to open the app, ensuring maximum protection for their collected data. The use of MyCap offers significant advantages, but requires further ethical considerations as it involves managing patient healthcare data outside the hospital. MyCap is currently being reviewed by the DPO in order to reformulate the EC procedures necessary for its use.

To enhance the registry's applicability in future follow-up and multicenter studies and facilitate its use in international research, it will be standardized with globally recognized frameworks such as HL7 Fast Healthcare Interoperability Resources (FHIR) and Observational Medical Outcomes Partnership (OMOP).

5. Conclusions

Sleep medicine is a rapidly evolving field, and in recent years Italian research centers have increasingly contributed to its international advancement. In this context, access to a large, structured, and interoperable data source is a strategic asset: it enhances research quality, promotes multicenter collaboration, and facilitates the comparison of findings with those from other international settings. The IPNOS registry is an innovative and valuable resource that can help bridge the current gap in detailed national data on OSA phenotypes, thus strengthening Italy's role within the international research community.

The creation of a shared and interoperable database therefore responds to an immediate need for standardization and quality, but at the same time, the methodology presented here lays the foundation for a sustainable research model that can be replicated in other research fields.

Abbreviations

OSAs: Obstructive Sleep Apnea Syndrome. ESADA: European Sleep Apnea Database. IPNOS: Italian PheNotypes Obstructive Sleep Apnea Study. AIMS: Italian Academy of Sleep Medicine. IT: Information Technology. GDPR: General Data Protection Regulation. IRCCS: Scientific Hospital and Care Institutes. DPO: Data Protection Officer. EC: Ethics Committees. FAIR: Findable, Accessible, Interoperable, Reusable. API: Application Programming Interface. PVT: Psychomotor Vigilance Task. EDC: Electronic Data Collection. eCRF: Electronic Case Report Form. REDCap: Research Electronic Data Collection. DAGs: Data Access Groups. NCBO: National Center for Biomedical Ontology. PAP: Positive Airway Pressure. CSV: Comma-Separated Values. REQA: Reliable Extractive Question Answering. AI: Artificial Intelligence. NLP: Natural Language Processing. JSON: JavaScript Object Notation. SSL: Secure Sockets Layer. PIN: Personal Identification Number. ATC: Anatomical Therapeutic Chemical Classification. FHIR: Fast Healthcare Interoperability Resources. OMOP: Observational Medical Outcomes Partnership.

Declarations

Ethics approval and consent to participate

The study protocol was approved by the internal Ethics Committee of Istituto Auxologico Italiano of Milan on December 21th 2022 (protocol number 2022_12_21_01). Informed consent was obtained from all subjects in accordance with the Declaration of Helsinki.

Consent for publication

Informed consent for publication was obtained from all subjects in accordance with the Declaration of Helsinki.

Availability of data and materials

The data supporting the results of this study are used under license for the ongoing study and not publicly available. The data are available from the authors upon reasonable request and with the permission of Carolina Lombardi, Istituto Auxologico Italiano IRCCS, Milan, Italy.

Competing interests

The authors declare that they have no competing interests.

Funding

This publication was funded by the Italian Ministry of Health under the National Recovery and Resilience Plan (PNRR), Project “Italian PheNotypes Obstructive sleep Apnea Study (IPNOS)”, Grant No. PNRR-MAD-2022-12375812.

Authors’ contributions

The authors confirm contribution to the paper as follows: study conception and design: All Authors. Analysis and interpretation of results: All Authors. Draft manuscript preparation: CMP, VT. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgements

Not applicable.

References

1. Gell LK, Mehta K, Esmaeili N, Taranto-Montemurro L, Sands SA, Pittman SD, et al. Performance evaluation of a ring-worn pulse oximeter for the identification and monitoring of obstructive sleep apnea. Front Sleep. 2025;4:1549272. 10.3389/frsle.2025.1549272.
2. Benjafield AV, Ayas NT, Eastwood PR, Heinzer R, Ip MSM, Morrell MJ, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. Lancet Respir Med. 2019;7(8):687–98. 10.1016/S2213-2600(19)30198-5.
3. American Academy of Sleep Medicine. In: Sateia M, editor. The international classification of sleep disorders: diagnostic and coding manual. 2 ed. Westchester (IL): American Academy of Sleep Medicine; 2005. pp. 1–297.

4. Bonsignore MR, Pepin JL, Anttalainen U, Schiza SE, Basoglu OK, Pataka A, et al. Clinical presentation of patients with suspected obstructive sleep apnea and self-reported physician-diagnosed asthma in the ESADA cohort. J Sleep Res. 2018;27(6):e12729. 10.1111/jsr.12729.
5. Bonsignore MR, Hedner J, ESADA study group. The European Sleep Apnoea Database (ESADA) ERS Clinical Research Collaboration: past, present and future. Eur Respir J. 2018;52(4):1801666. 10.1183/13993003.01666-2018.
6. Bailly S, Grote L, Hedner J, Schiza S, McNicholas WT, Basoglu OK, et al. Clusters of sleep apnoea phenotypes: a large pan-European study from the European Sleep Apnoea Database (ESADA). Respirology. 2020;25(12):1232–40. 10.1111/resp.13969.
7. Accademia Italiana di Medicina del Sonno. AIMS. 2025. https://sonnomed.it/. Accessed 19 Dec 2025.
8. Italian PheNotypes Obstructive Sleep Apnea Study. ClinicalTrials.gov. 2025. https://clinicaltrials.gov/study/NCT06872242. Accessed 23 Dec 2025.
9. Istituto Auxologico Italiano. PNRR-MAD-2022-12375812: Studio italiano sui fenotipi delle apnee ostruttive del sonno (IPNOS). OpenPNRR. n.d. https://openpnrr.it/progetti/184870/.. Accessed 23 Dec 2025.
10. European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union. 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679. Accessed 11 Dec 2025.
11. Timotijevic L, Astley S, Bogaardt MJ, Bucher T, Carr I, Copani G, et al. Designing a research infrastructure (RI) on food behaviour and health: Balancing user needs, business model, governance mechanisms and technology. Trends Food Sci Technol. 2021;116:405–14. 10.1016/j.tifs.2021.07.022.
12. O'Cathain A, Murphy E, Nicholl J. Multidisciplinary, interdisciplinary, or dysfunctional? Team working in mixed-methods research. Qual Health Res. 2008;18(11):1574–85. 10.1177/1049732308325535.
13. Repubblica Italiana. Decreto legislativo 10 agosto 2018, n. 101, recante disposizioni per l'adeguamento della normativa nazionale al Regolamento (UE) 2016/679, art. 110-bis. Gazz Ufficiale della Repubblica Italiana. 2018;(205).
14. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. Comment: the FAIR guiding principles for scientific data management and stewardship. Sci Data. 2016;3:1–9. 10.1038/sdata.2016.18.
15. Biega AJ, Potash P, Daumé H, Diaz F, Finck M. Operationalizing the legal principle of data minimization for personalization. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York (NY): Association for Computing Machinery; 2020. pp. 399–408. 10.1145/3397271.3401034

16. Shin SY, Kim HS. Data Pseudonymization in a Range That Does Not Affect Data Quality: Correlation with the Degree of Participation of Clinicians. J Korean Med Sci. 2021;36(44):e299. 10.3346/jkms.2021.36.e299.
17. Biomeris. Informatica per la ricerca clinica. 2012. https://www.biomeris.it/.. Accessed 22 Dec 2025.
18. Daniel C, Tannier X, Kalra D. Clinical research informatics. Yearb Med Inf. 2022;31(1):161–4. 10.1055/s-0042-1742530.
19. Franklin JD, Guidry A, Brinkley JF. A partnership approach for electronic data capture in small-scale clinical trials. J Biomed Inf. 2011;44(Suppl 1):S103–8. 10.1016/j.jbi.2011.05.008.
20. Ochs C, He Z, Zheng L, Geller J, Perl Y, Hripcsak G, Musen MA. Utilizing a structural meta-ontology for family-based quality assurance of the BioPortal ontologies. J Biomed Inf. 2016;61:63–76. 10.1016/j.jbi.2016.03.007.
21. Pisani MA, Friese RS, Gehlbach BK, Schwab RJ, Weinhouse GL, Jones SF. Sleep in the intensive care unit. Am J Respir Crit Care Med. 2015;191(7):731–8. 10.1164/rccm.201411-2099CI.
22. Evangelista E, Rassu AL, Lopez R, Biagioli N, Chenini S, Barateau L, et al. Sleep inertia measurement with the psychomotor vigilance task in idiopathic hypersomnia. Sleep. 2022;45(1):zsab220. 10.1093/sleep/zsab220.
23. Bian J, Lyu T, Loiacono A, Viramontes TM, Lipori G, Guo Y, et al. Assessing the practice of data quality evaluation in a national clinical data research network through a systematic scoping review in the era of real-world data. J Am Med Inf Assoc. 2020;27(12):1999–2010. 10.1093/jamia/ocaa245.
24. Diniz-Silva F, Ferreira JC. Enhancing research integrity and data quality through standardized electronic case report forms. J Bras Pneumol. 2024;50(4):e20240292. 10.36416/1806-3756/e20240292.
25. Chaudhuri S, Bagepally BS, Bhar D, Reddy Singam UK. Electronic versus paper-based data collection for conducting health-care research: a cost-comparison analysis. Indian J Public Health 2022 Oct–Dec;66(4):443–7. 10.4103/ijph.ijph_1271_21
26. Vivisol V. 2025. https://www.vivisol.it/. Accessed 17 Dec 2025.
27. Zhao B, Bai J, Li C, Zhang J, Rong W, Ouyang Y, Xiong Z. Enhancing biomedical ReQA with adversarial hard in-batch negative samples. IEEE/ACM Trans Comput Biol Bioinform. 2023 Sep–Oct;20(5):2933–44. 10.1109/TCBB.2023.3261315.
28. MyCap MC. 2025. https://projectmycap.org/. Accessed 17 Dec 2025.

Figures

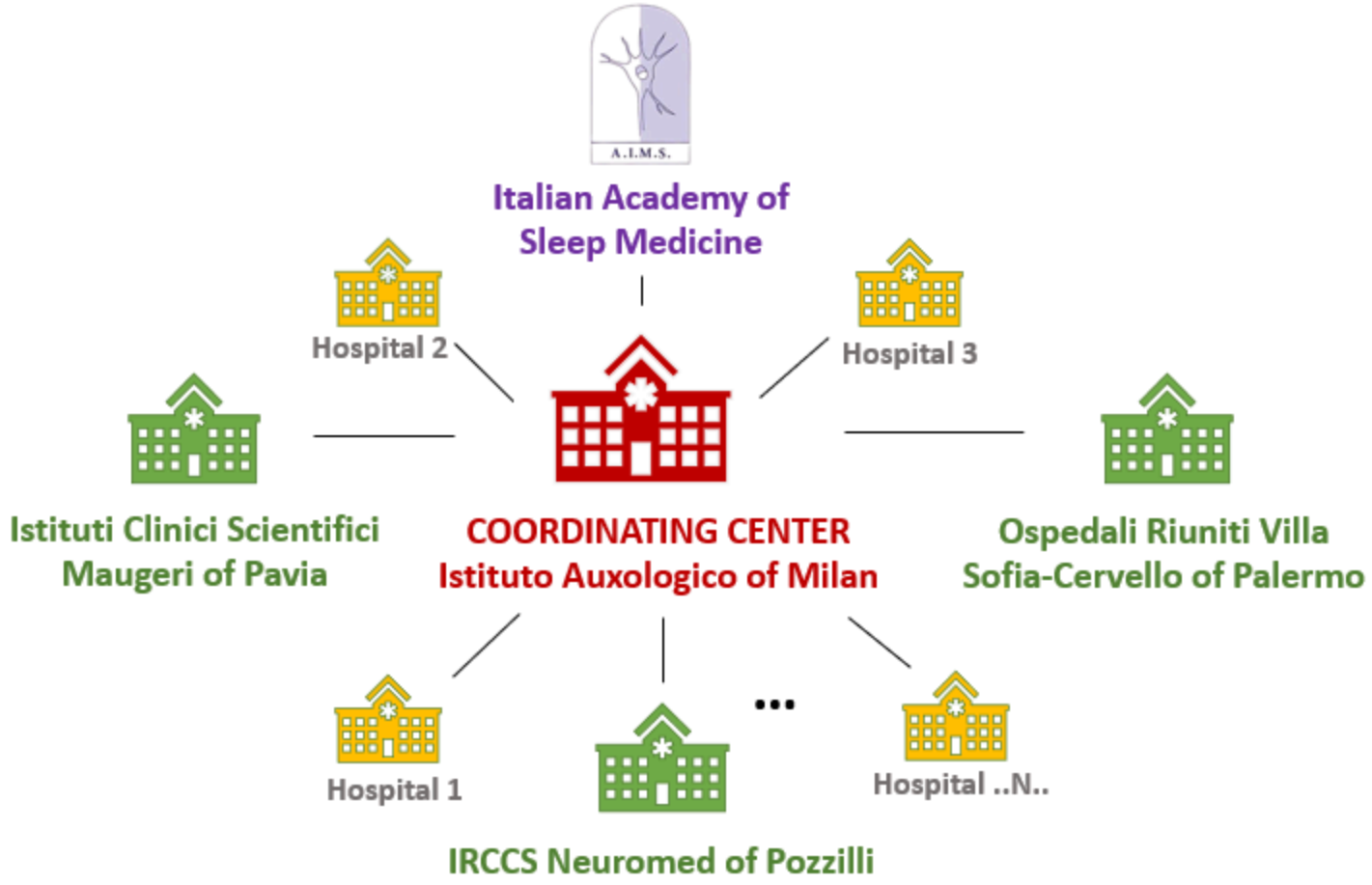


Figure 1

IPNOS hub & spokemodel

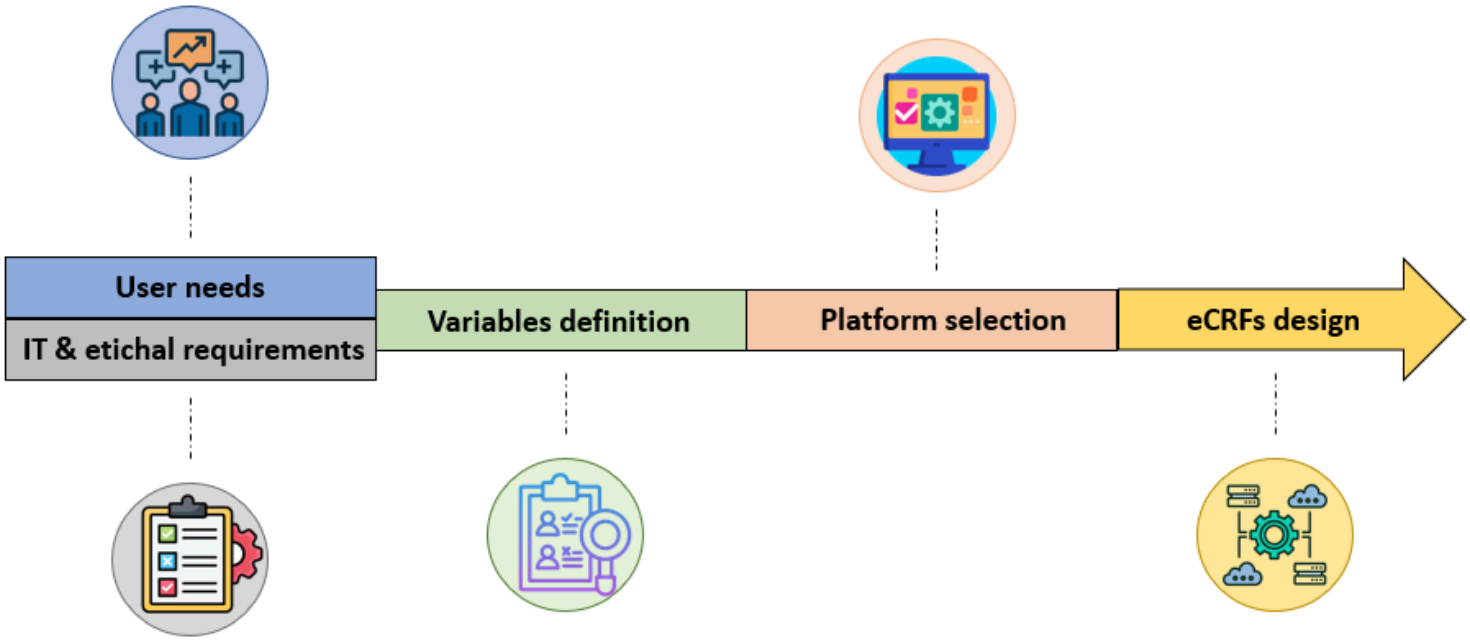


Figure 2

Overall methodological approach followed during IPNOS registry's development

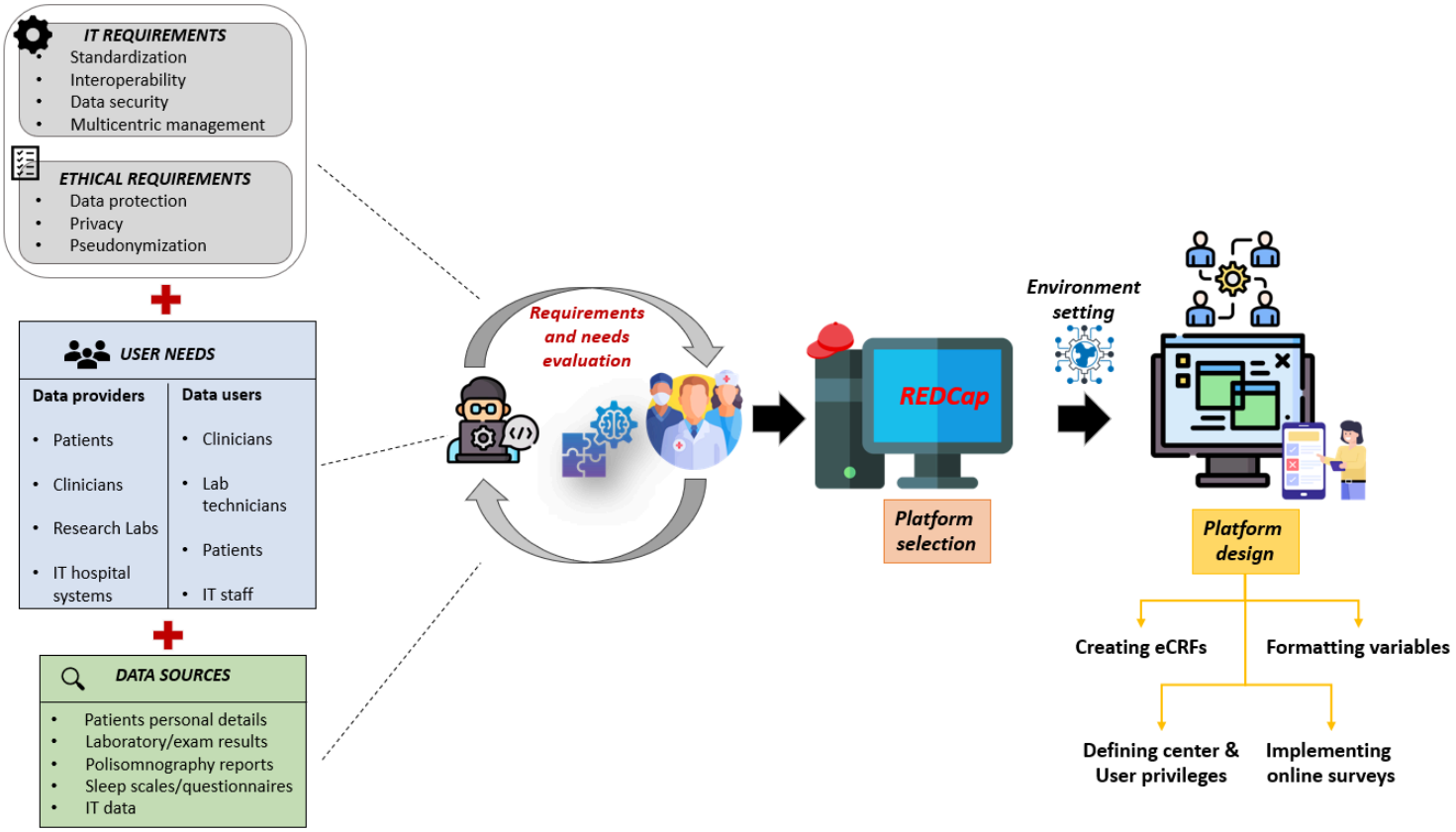


Figure 3

Methodological phases followed during registry development

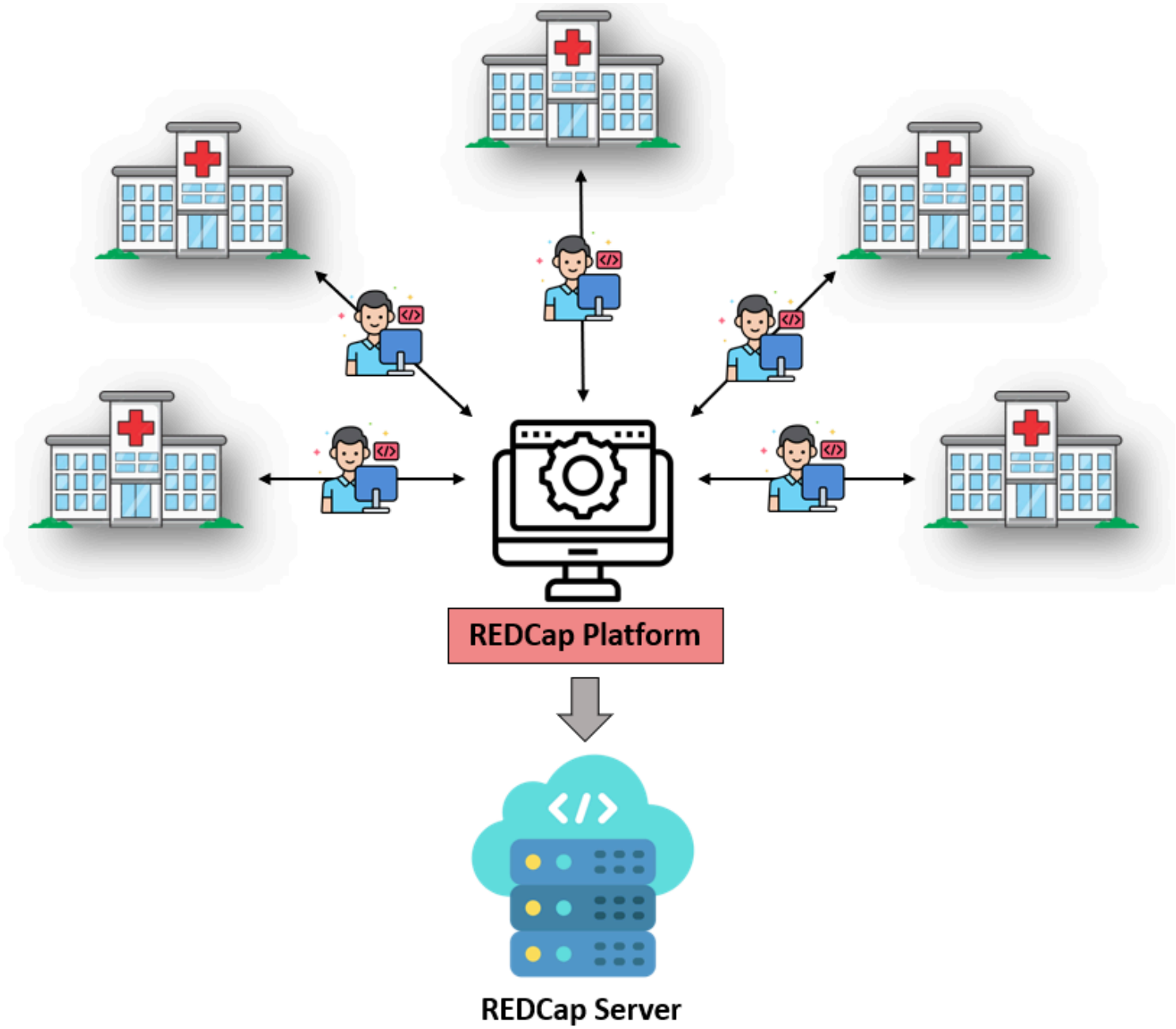


Figure 4

REDCap operating model in multicenter studies

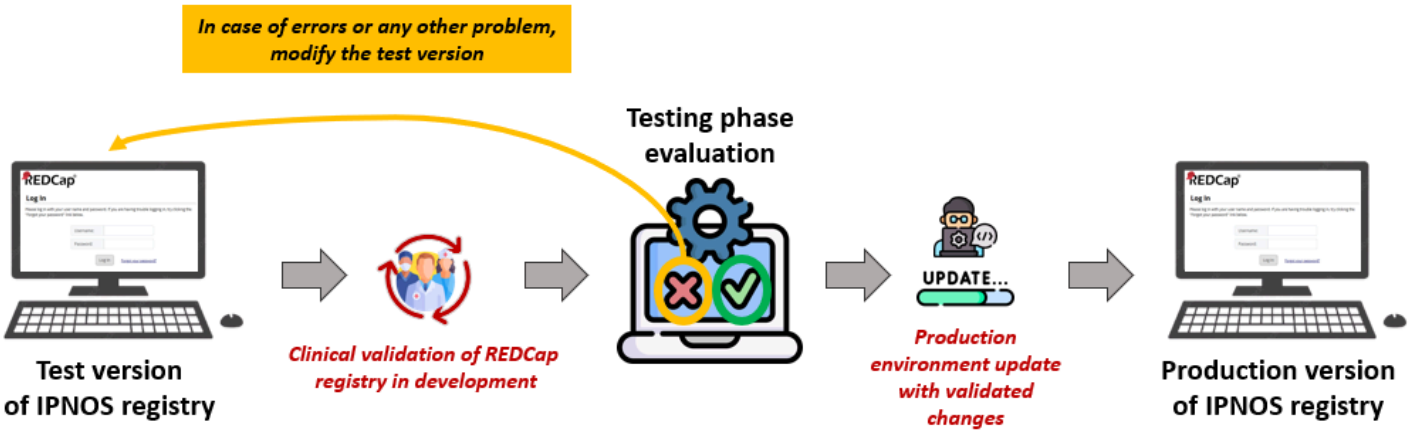


Figure 5

REDCap data registry validation phase

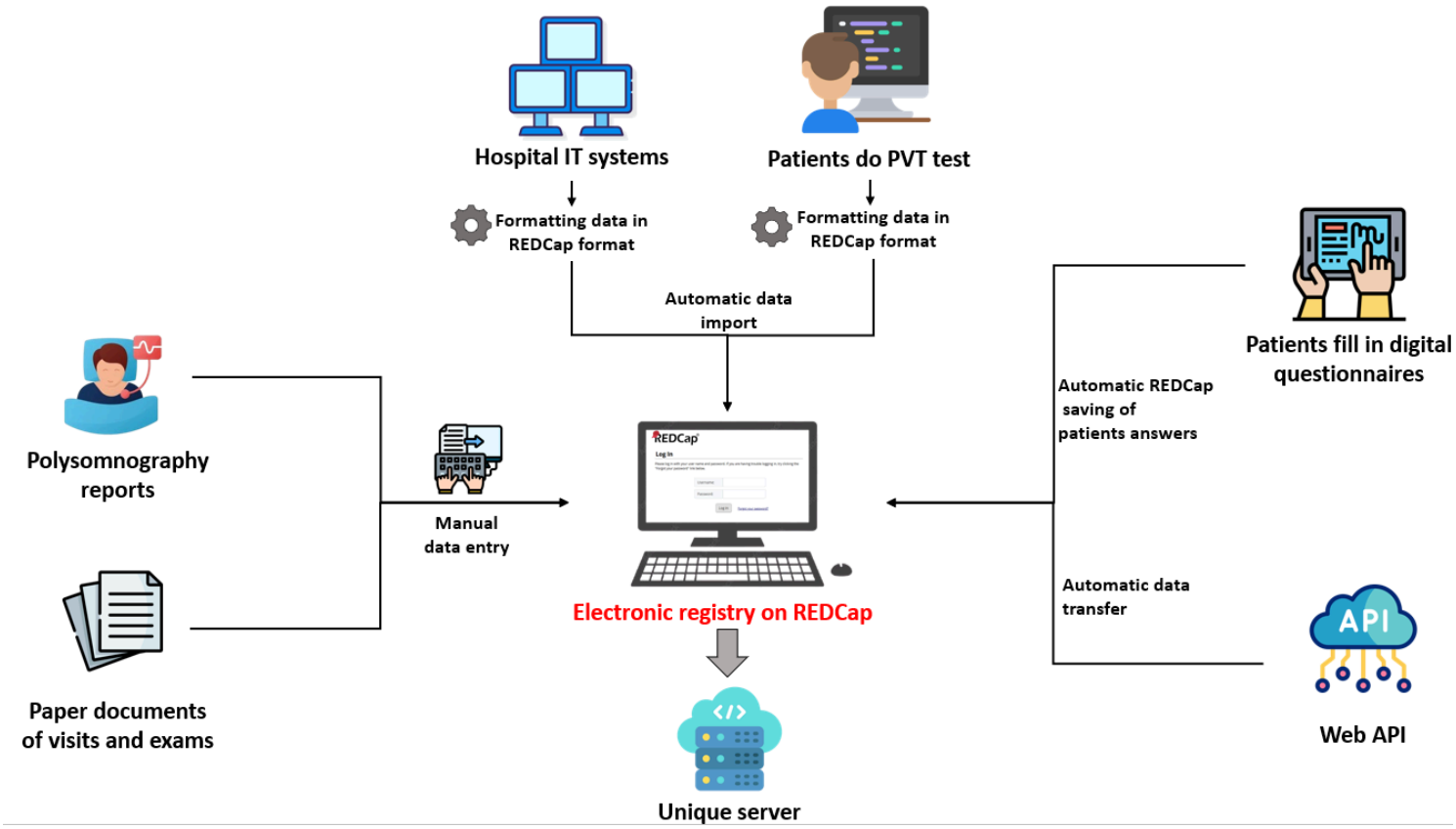


Figure 6

IPNOS infrastructure

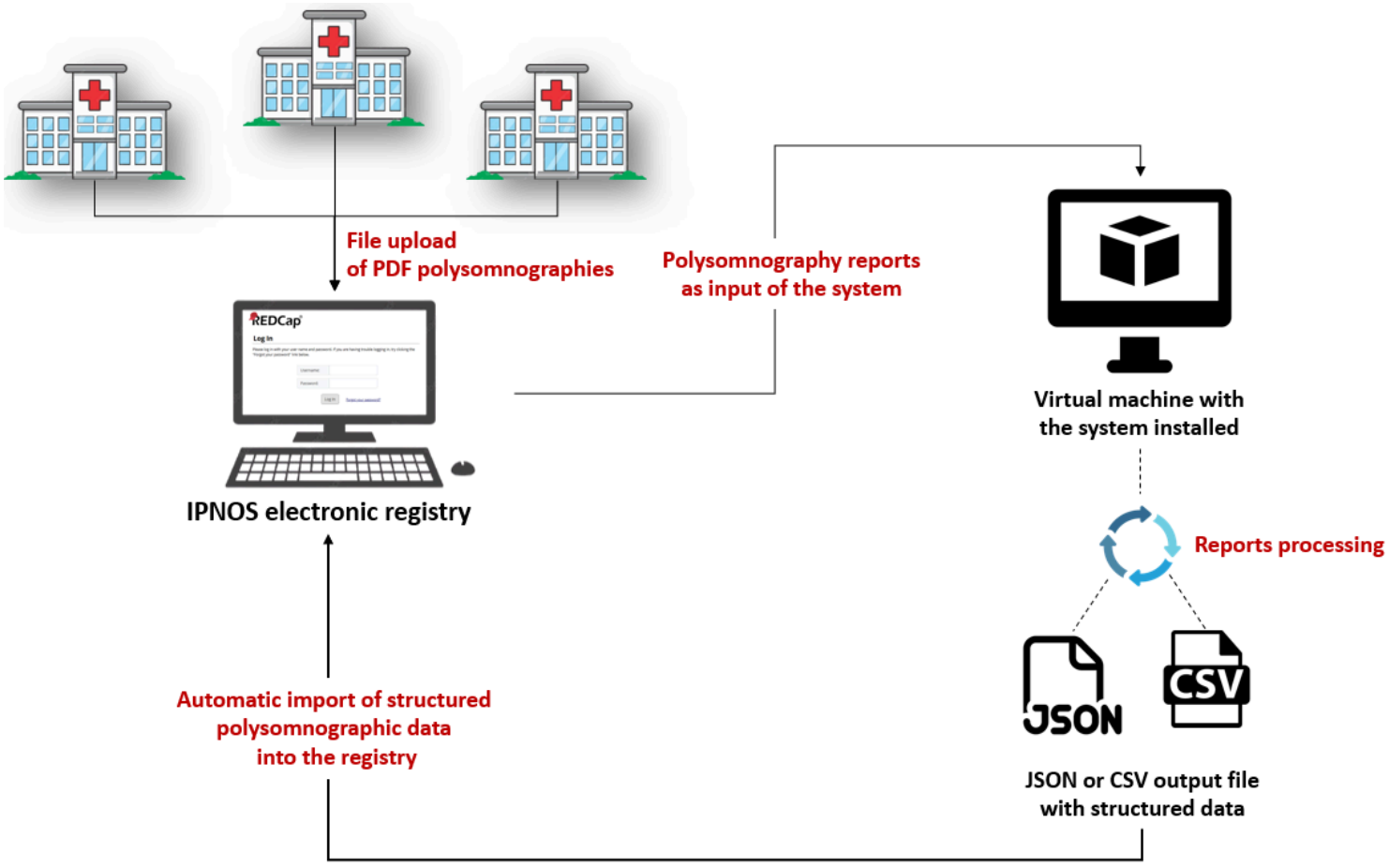


Figure 7

Polysomnographic reports' processing with Aux Reader solution

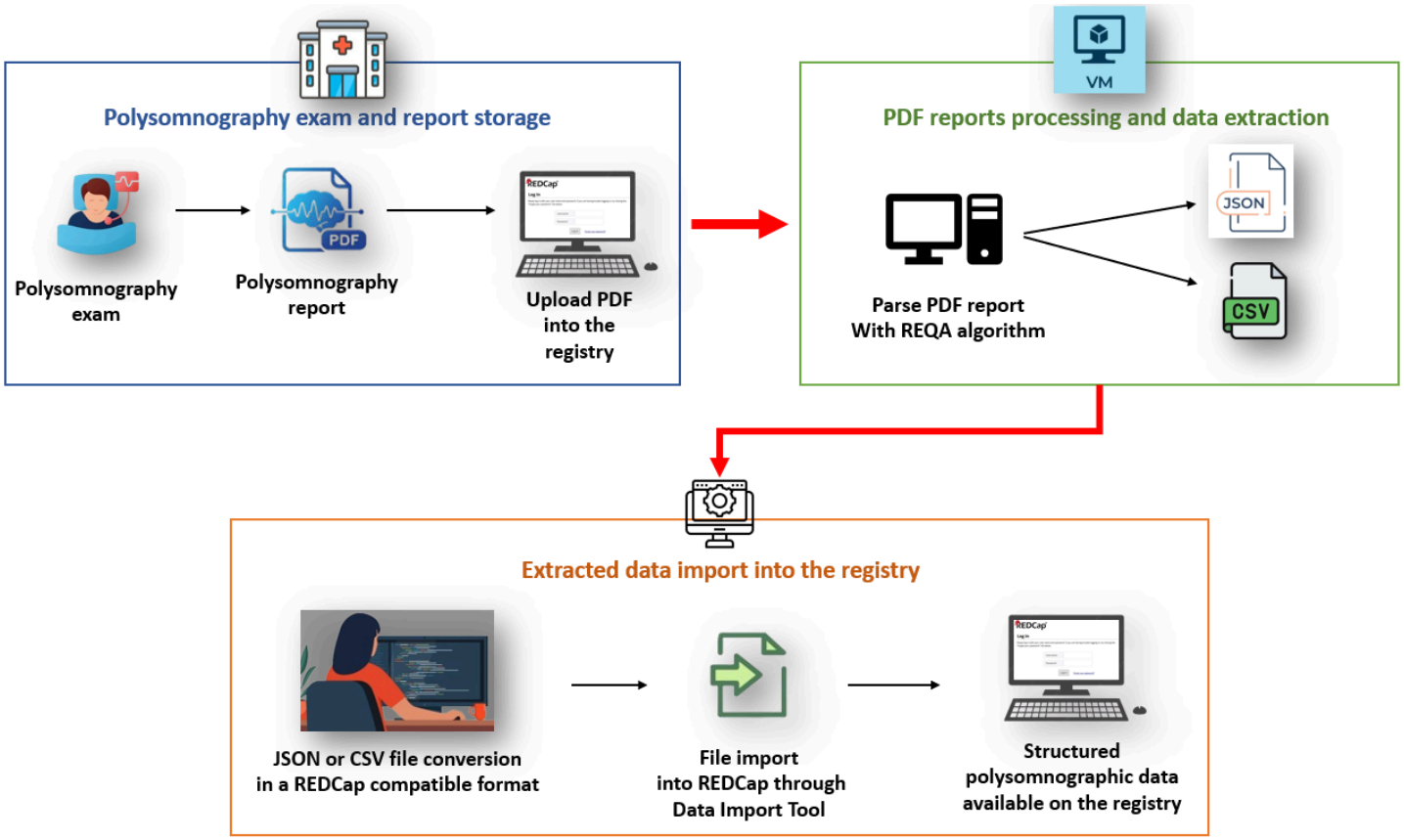


Figure 8

Data processing flow from patient polysomnography to structured data entry

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterials.docx