microscopIA SAS
Chamber of Commerce of Bogotá
NIT 901263621-0

AI training for automated data extraction from clinical records

Jorge Racedo

Director

jorge.racedo@microscopiatech.com

Department of Clinical Research

Division of Translational Research and Data Science

microscopIA

Disclaimer

Intellectual property belongs wholly to Jorge Racedo on behalf of microscopIA, including but not limited to concepts, implementation, and any other kind of information. microscopIA as well has funded completely ongoing projects under this macro. No external institution has participated financially or in development and cannot claim any copyright, trademark, or intellectual property under any circumstance. Infringements will be lawfully processed under Colombian and international jurisdiction by a microscopIA attorney. microscopIA has been legally registered as a technological organization against the Chamber of Commerce of Bogotá, and its research group reported against Colciencias to conduct clinical research and agroindustrial projects. As this is an introductory document, much confidential information has been omitted deliberately.

EXECUTIVE SUMMARY

Most hospitals and government offices in Colombia have no centralized systems to efficiently collect data for evidence generation for academic purposes and policymaking. In addition, physicians in small-to-middle-size cities do not participate in conducting clinical research due to significant gaps in resources and personnel skilled in data science. It results in a lack of information about the prevalence and impact of diseases, limiting municipalities' power to decide when developing prevention, screening-diagnosis, and treatment programs. microscopIA, through the current approach, has grown a platform to allow physicians to collect data matching a research project while collecting massive data for AI training and policymaking.

NEED IDENTIFICATION AND SCREENING

microscopIA as a technological organization has different operative focuses, including developing automated devices for healthcare and agroindustry, training AIs for detecting human and animal and plant diseases, as well as democratizing data generation as a first approach to centralize healthcare data in Colombia for further implementation in Latin America, Asia, and Africa. This document briefly explores the microscopIA process to centralize healthcare and human parameter data relevant to governments, pharmaceutical companies, and other industries demanding data on population health for their business models or community interventions.

Unmet Need Exploration

Improving healthcare research capacity is a global priority, especially for low- and middle-income countries. Evidence is needed to enhance decisions in

patient management. There is a call for government and healthcare institutions to tackle such disparities evident in a suboptimal output, especially in regions like Latin America, primarily underrepresented in major journals due to a lack of robust infrastructure and resources[1],[2]. Gaps in clinical research result from significant barriers such as a lack of a mentorship culture and financial support, overloaded physicians' patient care, nonhomogeneous distribution of resources, scare infrastructure, and nonproficiency in technical skills and non-English speaking[3].

Besides hospitals and academic institutions, governments also demand data for addressing policymaking, as evident in major infectious disease outbreaks like the COVID-19 pandemic. Generating data is the basis for effective healthcare management delivery, especially in the most vulnerable communities[4]. In Colombia, few hospitals have well-established facilities to generate evidence (clinical guidelines and academic and professional output), and there is little information concerning how many providers and institutions conduct clinical and translational research nationwide to provide such data. Few universities have a monopoly on evidence generation, and microscopIA will target them as a new actor in that market by reducing the need for hospitals and doctors to partner with universities through expensive agreements with limited real-life applications.

Colombian governments have neither no resources to centralize real-time data from healthcare providers. As a result, hospitals provide information suboptimally, with general statistics without valuable information on patient's profiles for community intervention. Each hospital may utilize a different software to manage clinical records, and healthcare government offices have not standardized the process efficiently to access it. The lack of information

makes it nonfeasible to implement more efficient programs for healthcare delivery. It affects mainly rural areas in which healthcare providers have scarce resources. Figure 1 presents the unmet need addressed by microscopIA.

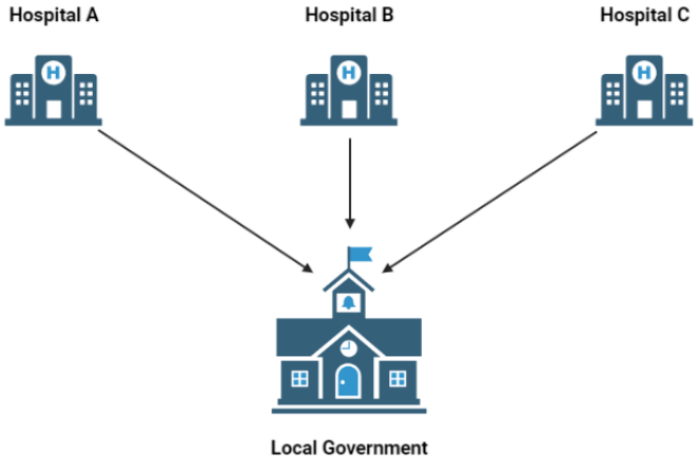


Figure 1. Each hospital collects clinical data relevant to governments through their software. However, the process is not standardized, and the resources utilized may differ widely from paper form to digital ones.

Opportunities

Managing clinical data is sensitive and protected lawfully. However, it results in making nonefficient attempts to centralize data as individuals would access information of patients. The human factor makes ethically nonfeasible centralizing data with current methods. The microscopIA approach is standardizing data collection despite hospitals having different software for clinical records by training AIs to collect massive data while reducing as much as possible the participation of human readers to ensure confidentiality in such documents. Figure 2 presents the microscopIA approach to centralize data from healthcare providers for then providing it to governments, non-government organizations (NGOs), and international agencies.

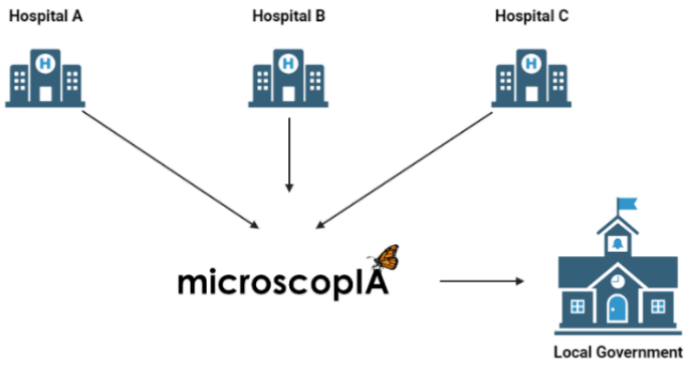


Figure 2. microscopIA's AIs will connect to hospital software to extract real-time data relevant to government and decision-makers without a human reader. They will then become customers from microscopIA data generation for their needs.

Core Problem and background

Managing clinical records for public health issues demands hospitals to have a team, including epidemiologists and statisticians. Unfortunately, considering limitations in funding and resources in the Colombian healthcare system, it is not feasible for all hospitals to have such personnel. As a result, most physicians and providers are marginalized from evidence generation and cannot provide a source for healthcare policymaking.

In August 2017, there were 1534 healthcare providers just in the metropolitan area of Bogotá. Most of them small-to-middle size institutions[5]. Nationwide, there were 1800 institutions categorized as hospitals and clinics to the same date. About 1 in 5 institutions from the public sector faced financial deficits[6]. In general, hospitals and clinics, like any other healthcare provider, would be targeted as potential sources of data, being the local government responsible for funding microscopIA services. The lack of resources represents the opportunity for microscopIA to provide cost-

efficient services in organizing data for decision-makers and formulate the demanded documents. For most hospitals, contracting personnel or having a research unit is not convenient considering internal financial deficits.

Desired Outcome

microscopIA during the short term will be collecting data from clinical records from different hospitals in Colombia. Currently, six hospital departments are providing data. Physicians will provide documents against manuscripts to be published while our team trains AIs. Writing academic manuscripts to be published will justify the interaction with local physicians and hospital departments, as shown in Figure 3. In the long term, AIs are expected to be integrated with the software used by hospitals, clinics, and healthcare providers to extract information in real-time and centralize it in microscopIA databases to them provide it to governments or any other organization in need of them.

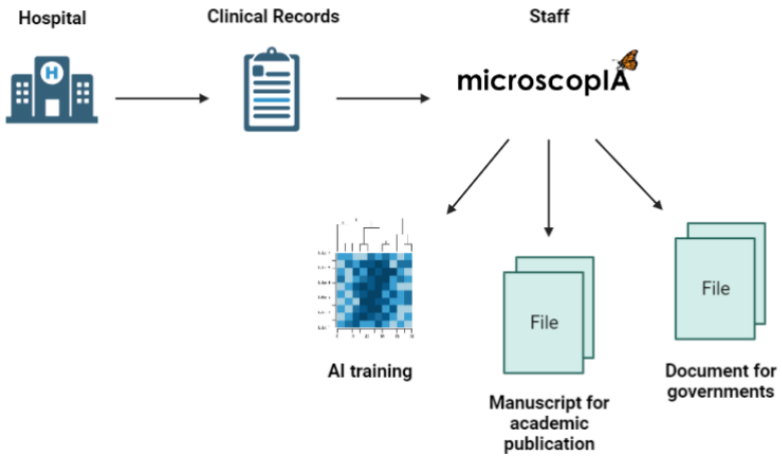


Figure 3. In the short term, the workflow implemented by microscopIA to train AIs and generate output for academic and government purposes.

While AIs are trained and tested, microscopIA involves doctors and medical students as associate researchers to extract information from clinical records. To keep documents safe, all parts involved sign confidentiality agreements with microscopIA and the microscopIA Research Groups has been built against Colciencias, with an inner Committee for Research, to accomplish regulation and prepare for offering services in data generation (publishable papers and other academic output) to physicians, doctors in training (resident doctors), students, and hospital departments. All records have to be de-identified. Besides the scholarly production, results are used to elaborate recommendation documents for the government. For example, microscopIA has conducted five projects during the spring term 2021, one of them titled "A lack of access to genetic counseling among women with breast cancer- why are we not testing enough patients?". This study aimed to evaluate the sociodemographic and clinicopathological features associated with testing and counseling decisions in Hispanic-Latino individuals with breast cancer in a low-to-middle income setting. Results showcased a suboptimal referral and testing under scarce resources, allowing authors to identify needs to be addressed to tackle such disparities. The findings from this study enable microscopIA to currently write a document for the Colombian Ministry of Health and Social Protection's National Cancer

Control Program to tackle the suboptimal access to genetic counseling among patients with breast cancer to provide proper treatment to risk-associated variant carriers and implement programs for screening and early diagnosis of their unaffected relatives. Results suggest conducting specific programs for underrepresented groups, such as ethnic minorities, patients with government-subsidized providers or not affiliated to the healthcare system, and Venezuelan immigrants. All this is a strategy to gain visibility as a promising figure in generating solid evidence nationwide on healthcare delivery. All the ongoing projects have been proposed based on priorities in healthcare nationwide and receiving attention from domestic and foreign organizations, including the access of women to sexual and reproductive services, predictors for fetal-maternal morbidity and mortality, cost-effectiveness in accessing specialized services, and the overall prevalence of diseases relevant to providers. Operatively, microscopIA will follow a federated learning framework (FLF) justified by extracting data for research and policymaking, as in Figure 4.

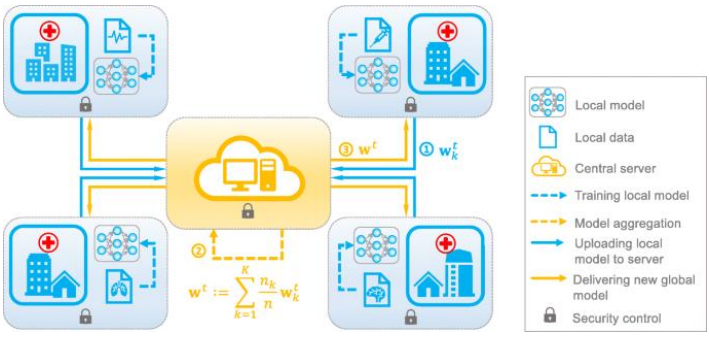


Figure 4. Federated learning framework. Extracted from [7].

In the next four months, the team will be deciding on the most feasible method for FLF (Figure 5), considering the highly variable clinical records and regulations in Colombia, both legal as institutional, for the enrolled hospitals and clinics.

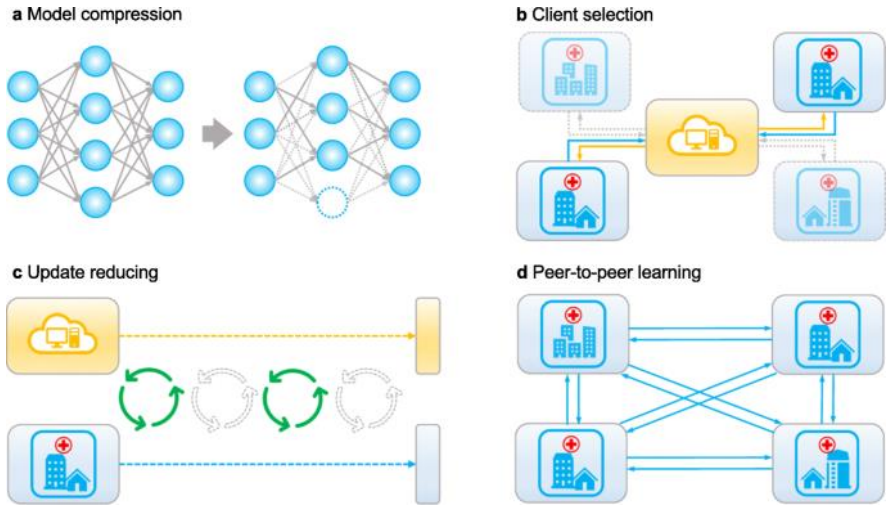


Figure 5. Methods for federated learning framework. Extracted from [7].

microscopIA has over 700 clinical records in its serves from ongoing projects. In addition, two research papers have been finished to be published for the physicians who collaborate, two manuscripts for public policies in underlying subjects are being written, and coding experiments for extracting data started recently in association with a German data scientist.

Existing Solutions

Healthcare data is fragmented between providers, and hospitals still cannot access data from other institutions due to conflict of commercial interest and sensitive information on records protected by regulations such as the Health Insurance Portability and Accountability Act (HIPAA)[7]. In Colombia, no solutions are attempting centralizing clinical data for decision-

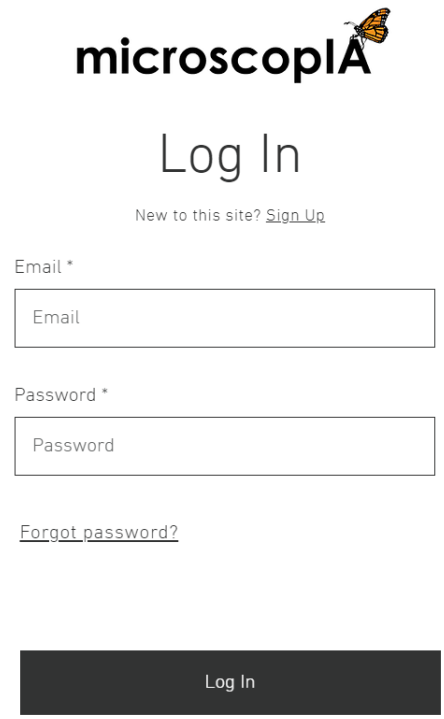
makers based on AI. No significant projects on FLF being conducted in Colombia.

Outside Colombia, approaches for FLF are developed by the projects OpenMined's PySyft[8], Google's TensorFlow Federated[9], Webank's FATE[10], and Duan's Tensor/IO[11]. In the short term, microscopIA will focus on extracting data for public health and research projects while training AIs, but the mid-term goal is to reconstruct individual profiles by tracing any admission, diagnosis, and information in any provider server. This information will then be provided to governments, providers, and any other institution demanding while maintaining the privacy and de-identifying patients to protect their rights. In-hospital FLF has been used for different healthcare, significantly to predict diseases and mortality. Concerning FLF, microscopIA will extract data, with human supervision, for surveys targeting research projects in the first stages to replace human readers for machine learning. Security and privacy will be addressed in the late stages of coding. Most approaches in FLF have been for research purposes[7].

HOW DOES IT WORK?

1. Physicians and hospital representatives go to our website (https://www.microscopiatech.com/)

2. Log In with a personalized user and password.



3. Estudios section on the member area.

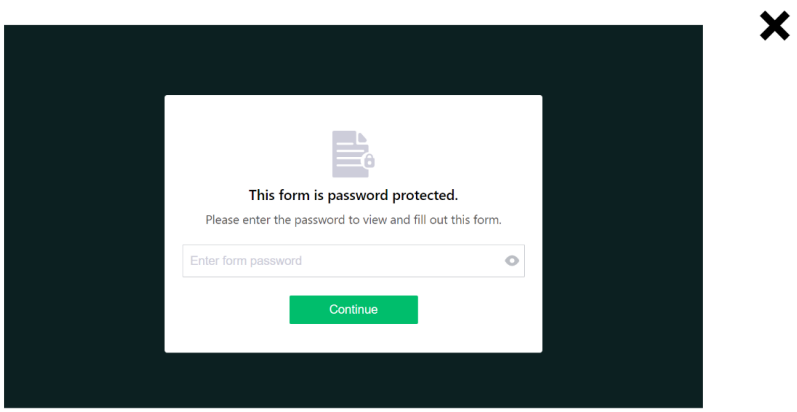
4. All ongoing studies are displayed.

Datos- only for microscopIA staff to extract data on personalized surveys

Read me- information on inclusion criteria to be checked by collaborators

Subir- allows physicians to submit records

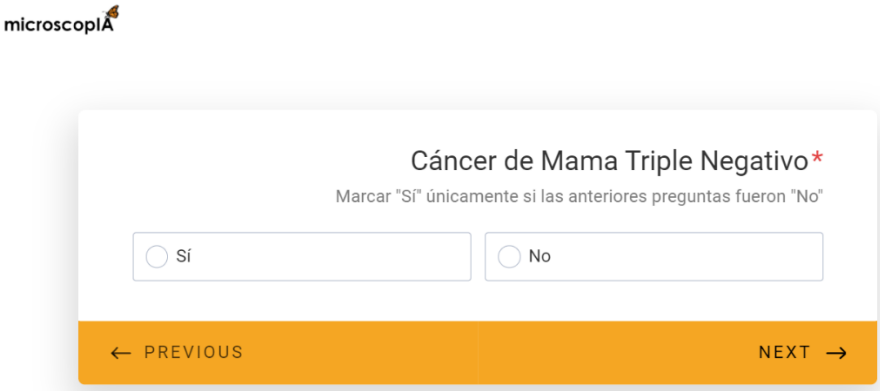
5. Only doctors enrolled physicians can access specific Read Me and Subir areas, sure they are password protected. It means that information is not accessible to anyone.



6. Once inside, the physician submits clinical records. All submissions are tracked as the collaborator is logged in.

7. All submissions synchronize with a specific folder for microscopIA staff to access them. Only microscopIA staff enrolled in a project can access it after signing the contract and confidentiality agreements. microscopIA Research Director is responsible for de-identifying patient information from records before the team starts collecting data.

8. microscopIA staff fills multiple-choice surveys to standardize data collection for easy manipulation.



REIMBURSEMENT AND FINANCIAL STRATEGY

MicroscopIA intends to tackle such gaps in generating bio and healthcare data for hospitals, academic institutions, and governments. It will be performed in a market generating phase and three operative stages, as presented in Figure 6.

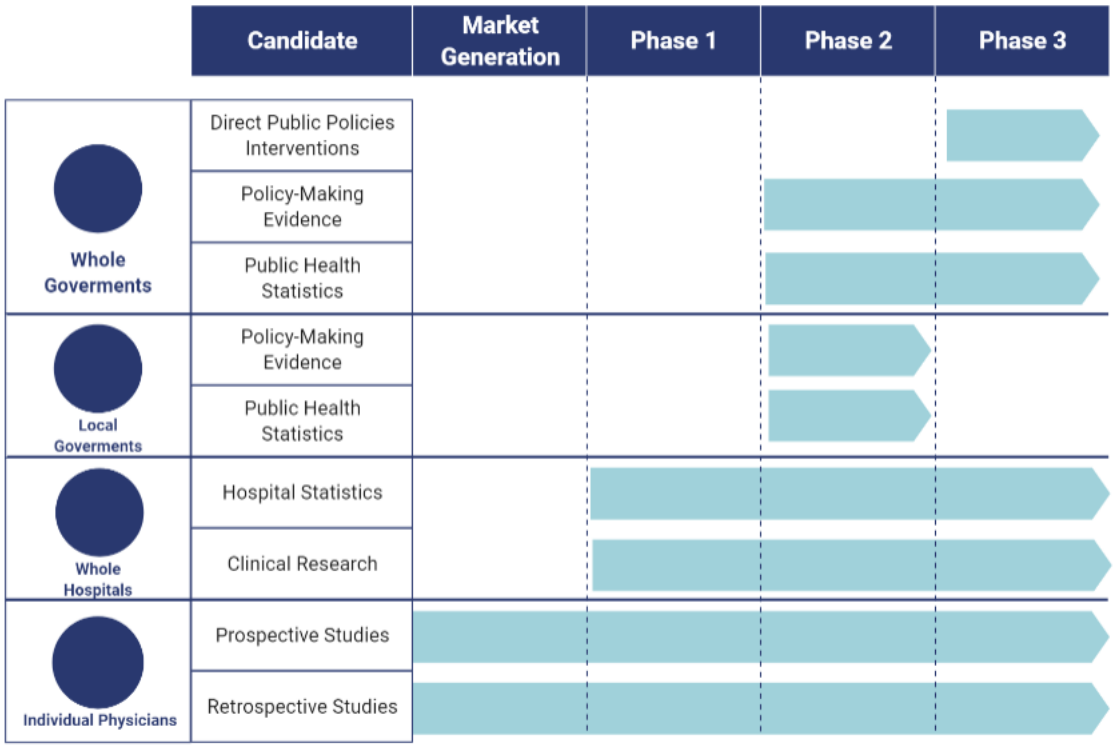


Figure 6. Stages for business development.

Profits and expansion are expected since Phase 1. In the late stages, coverage, coding, and payment will be entirely afforded by governments and individuals demanding services. Concerning reimbursement, direct pricing will work in Phase 1, while the bid for contracts with governments and hospital networks will be the expected model after it. As physicians and researchers, individual clients might pay directly for the product demanded or come from governments or hospital networks through contracts, including

specific requirements. In the early phases, products and services will result in published papers and assistance in developing, conducting, planning, and posting clinical guidelines. The last stages will demand developing new services and product lines associated with public health instead of just publishing academic output. Clients and governments will be the payers and reimburse directly to microscopIA against all the products and services they consume directly or indirectly by them or their beneficiaries. microscopIA, due to regulatory issues, will develop its Committee for Research and Transference (mCRT) to become independent of external institutions when deciding on ongoing and further studies and projects.

mCRT will serve as a strategy for validation to maintain microscopIA branding and operative standards among the general public, customers, and governments. mCRT will maintain services such as the registry of randomized controlled trials under national and international regulatory measures. mCRT is an initiative for general activities inside microscopIA. mCRT will provide additional services as a decision-maker and evaluator to universities and academic institutions to avoid them conforming their own and make the process more efficient after Phase 2. At Phase 3, it is also expected that international pharmaceutical and biotechnological companies will become users to supervise clinical trials and market studies for their products and services in Colombia and Latin America.

Due to the noncentralized model inside microscopIA, we will have to develop and implement its market strategy and have an administrative staff responsible for profit as an interdependent microscopIA division. Market research and planning are currently conducted during the Market Generation

stage based on experience and introducing our products and services to local physicians in Northwestern Colombia.

CANVAS – BUSINESS MODEL

This project is currently in the early prototype and service development stage, and there is little data. Figure 7 presents the operative model during the next 12 months. Figure 8 summarizes the elements of the business canvas model during the Market Generation stage.

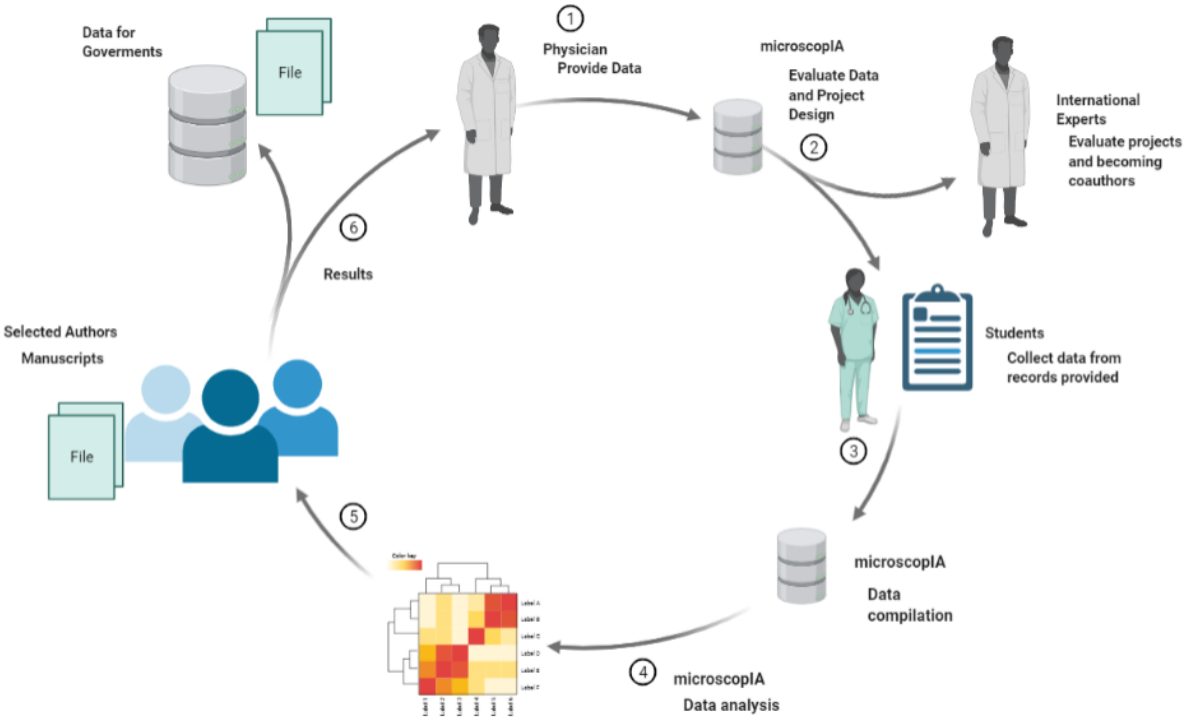


Figure 7. Operative model in 12 months.

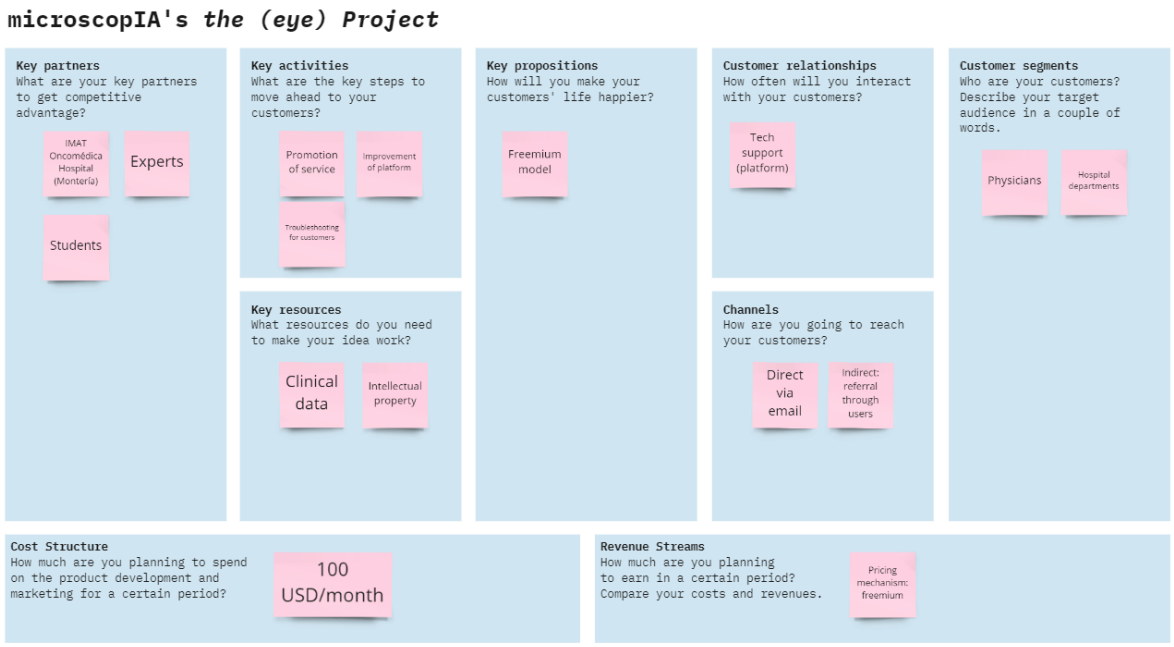


Figure 8. Business Canvas in 12 months.

MARKETING, STAKEHOLDERS, SALES, AND DISTRIBUTION STRATEGY

The first studies are being conducted under a freemium modality. From August 2021, microscopIA will offer physicians and hospital departments the same service to ensure financial operations and resources. A proposal presented to physicians is attached to this document. To engage physicians, microscopIA provides complementary services such as managing their CvLAC curriculum (Colciencias), organizing academic meetings, and supporting them in having abstracts for national and international conferences. Data is microscopIA main goal.

- Owners and investors: microscopIA owns *the (eye) project* entirely and is responsible for funding it. No outside investors are needed during the current stage.
- Creditors: no involved nor expected at the current phase.
- Trade unions: no involved nor expected at the current phase.
- Employees: microscopIA has a team of students who participate in the current stage against coauthoring the resulting papers. No money-type payment actually but intended in the future. No employees or staff from microscopIA involved in *the (eye) project* besides Jorge Racedo.
- Collaborators: due to confidentiality, a current and further international experts list cannot be presented in this document.
- Governments: no involvement nor expected at the current phase.

During the next 12 months, customers are intended to be physicians and hospital departments in the North area of Colombia and the Antioquia region. They are contacted directly through email or referred by other customers. A freemium plan is offered in which they receive the first manuscript to be published (in which microscopIA has to be included as an author) and then negotiate pricing for further projects. Price is set based on complexity and resources demanded. To date, a hospital and seven individuals physicians started collaboration in the free-pricing period. More physicians providing data are expected from a referral from current collaborators. A period of 4 months is scheduled for having a manuscript.

COMPETITIVE ADVANTAGE AND BUSINESS STRATEGY

No companies provide a service like the one described in this document nationwide. microscopIA, through this approach, intends to reduce the need for companies such as governments, universities, and hospitals to contract

personnel to conduct clinical and biomedical research and supervise more sophisticated projects such as clinical trials and evidence for policymaking. For generating evidence, before scaling our services, hospitals and governments still make a direct contract of statistics professionals and epidemiologists, and support staff. microscopIA value consists in reducing their participation in evidence generation by only demanding them data. microscopIA will be responsible for all the personnel and resources needed to end the desired product for the customer (policymaking proposal, public health statistics, and academic papers).

OPERATING PLAN AND FINANCIAL MODEL

By the end of the next 12 months, microscopIA expects to finish at least 12 manuscripts to publish in the freemium plan for physicians. It is expected about 600-800 USD monthly as an investment in resources to operate. All the expenses will be funded entirely by microscopIA. These concluded studies will support starting Phase 1 after showing potential customers, creating a niche, and starting a profitable model.

TEAM

microscopIA currently has three part-time employees and three associate researchers working in development (devices), machine learning, and designing clinical research. Besides, two other persons are working in the microscopIA Foundation, which is the humanitarian brand of the organization working in impacting vulnerable communities while advancing our products and data collection while supporting vulnerable communities. microscopIA is also supported by five researchers at international universities and nine Colombian physicians who actively participate in ongoing projects.

Information on team members and collaborators will not be revealed in this document.

REFERENCES

[1] F. Zegers-Hochschild, “Barriers to conducting clinical research in reproductive medicine: Latin America,” *Fertil. Steril.*, vol. 96, no. 4, pp. 802–804, 2011, doi: 10.1016/j.fertnstert.2011.08.043.

[2] K. Chomsky-Higgins *et al.*, “Barriers to clinical research in Latin America,” *Front. Public Heal.*, vol. 5, no. APR, p. 57, Apr. 2017, DOI: 10.3389/FPUBH.2017.00057.

[3] M. O. Angel *et al.*, “Mentoring as an opportunity to improve research and cancer care in Latin America (AAZPIRE project),” *ESMO Open*, vol. 5, no. 6. BMJ Publishing Group, Nov. 24, 2020, DOI: 10.1136/esmoopen-2020-000988.

[4] L. Sigfrid *et al.*, “Addressing challenges for clinical research responses to emerging epidemics and pandemics: a scoping review,” *BMC Med.*, vol. 18, no. 1, p. 190, Jun. 2020, DOI: 10.1186/s12916-020-01624-8.

[5] Datos Abiertos Colombia, “Listado de Instituciones de Salud.” https://www.datos.gov.co/Salud-y-Protecci-n-Social/Listado-de-Instituciones-de-Salud/fgk8-hnys (accessed Jun. 06, 2021).

[6] EL TIEMPO, “La verdadera historia de hospitales y clínicas al borde de la quiebra.” https://www.eltiempo.com/salud/la-verdadera-historia-de-hospitales-y-clinicas-al-borde-de-la-quiebra-247988 (accessed Jun. 06, 2021).

[7] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, “Federated Learning for Healthcare Informatics,” *J. Healthc. Informatics Res.*, vol. 5, no. 1, pp. 1–19, Mar. 2021, DOI: 10.1007/s41666-020-00082-4.

[8] T. Ryffel *et al.*, "A generic framework for privacy-preserving deep learning," Nov. 2018, Accessed: Jun. 14, 2021. [Online]. Available: http://arxiv.org/abs/1811.04017.

[9] Google, “TensorFlow Federated.” https://www.tensorflow.org/federated?hl=es-419 (accessed Jun. 14, 2021).

[10] We bank, "FedAI.org – Federated AI Ecosystem." https://www.fedai.org/cn/ (accessed Jun. 14, 2021).

[11] R. Duan *et al.*, “Learning from electronic health records across multiple sites: A communication-efficient and privacy-preserving distributed algorithm,” *J. Am. Med. Informatics Assoc.*, vol. 27, no. 3, pp. 376–385, Mar. 2020, DOI: 10.1093/jamia/ocz199.