E

# Committee on Development and Intellectual Property (CDIP)

**Thirty-Second Session**
**Geneva, April 29 to May 3, 2024**

SUMMARY OF THE "SYSTEM FOR STANDARDIZATION, ENRICHMENT AND ECONOMIC ANALYSIS OF INTELLECTUAL PROPERTY AND INNOVATION DATA TO SUPPORT POLICY DESIGN (VERSION 1.0)"

*prepared by the Secretariat*

1.      The Annex to this document contains the Summary of the "System for Standardization, Enrichment and Economic Analysis of Intellectual Property and Innovation Data to Support Policy Design (version 1.0)."

2.      This system was developed in the context of the Development Agenda (DA) Project on "Systematization of Statistical Data and the Design and Implementation of a Methodology for Developing Impact Assessments on the Use of the Intellectual Property System" (document CDIP/26/4).

> *3.      The Committee is invited to take note of the information contained in the Annex to the present document.*

[Annex follows]

**Summary of the "System for Standardization, Enrichment and Economic Analysis of Intellectual Property and Innovation Data to Support Policy Design (version 1.0)"**

This document presents the first version of the system aiming at standardizing and enriching intellectual property (IP) and innovation data with the purpose of providing economic analysis to support policy design.  It was developed in the framework of the DA project on "Systematization of Statistical Data and the Design and Implementation of a Methodology for Developing Impact Assessments on the Use of the Intellectual Property System" (document CDIP/26/4).

Following the project's objective, the system intends to structure the methodologies for the economic analyses of IP and related policies at the national level.  The system's outcomes should support the design of public policies that could be adopted in accordance with national development plans and help to strengthen the interplay between social capital, productive undertakings and IP.  It also aims at building the human and technical capacity of the relevant entities of the beneficiary countries and other Member States.

The present document is organized as follows:  section I provides the context for the main rationale behind choosing El Salvador as the inaugural pilot country, detailing the anticipated insights from this selection;  section II outlines the strategic approach for developing various methodologies, offering a detailed guide through the data processing utilized by the system;  section III discusses the project outputs and their dissemination to stakeholders;  and section IV addresses the challenges encountered in the initial phase of the project, highlighting the improvements slated for implementation in the subsequent phase.

I.    CONTEXT

This document focuses on the system as designed for the first pilot country, El Salvador.  During the initial assessment, design and implementation of the system's first version, El Salvador was deemed a valuable pilot country due to several country conditions, such as size, socio-economic development, demographics, and geography.  Additionally, El Salvador met the project's data needs, specifically as it had:

a)    sizable amounts of national IP data that are unaccounted in international IP databases;

b)    a noticeable IP use by national stakeholders outside the country (*i.e.* not collected by the national IP Office);  and

c)    an active focal points in the national IP Office and other government entities.

II.    METHODOLOGY DESIGN

Designing the system's methodology involved a review of technical reports and scientific publications from the World Intellectual Property Organization (WIPO), IP Offices having economic units, and the academia, making use of IP and innovation data with the purpose of providing economic analysis to inform policy design.

The first version of the system:  a) draws on international best practices and methodologies in order to design a novel methodology that assesses, standardizes, and enriches IP and innovation data;  and b) creates new IP and innovation indicators and datasets ready to use for economic analyses.  Within this process, methodological design refinements were also undertaken by incorporating user-oriented insights obtained during technical meetings held with IP office focal points of the participating pilot country and other experts.

The resulting methodology and its first implementation lay the ground for a version of the system for each pilot-country's iteration.  The system has three main sub-elements:  (a) data assessment and collection;  (b) data enrichment and database consolidation;  and (c) inputs for economic analysis.  Each sub-element relies on independent modules, performing a function necessary for the system's overall performance.  These sub-elements are as follows:

a.  Data Assessment and Collection

The first sub-element of the system includes the data preparation and diagnostic modules.  Those modules are designed to gather and aggregate data from both the partner country and international data collections.  The initial step taken in developing the modules is gathering the required IP data inputs and verifying the highest possible data availability by coordinating several technical meetings with focal points of the IP Office.  In El Salvador's case, collaboration with the National Centre of Registries and other government entities enabled to collect substantial amounts of IP data, inputs, and information for conducting further analyses.  Subsequently, a variety of data engineering techniques are applied to standardize the data.  Ultimately, the process involves meticulous quality evaluations of the IP data provided by the IP Office, including a comparison with existing national and international IP data sources, to ensure the integrity and reliability of the information utilized for analysis and policy development.

The resulting data preparation and diagnostic modules allow to:  i) apply basic cleaning and standardization techniques on all collected IP data;  and ii) conduct coverage and quality assessments on the available pilot country data.

b.  Data Enrichment and Database Consolidation

The second sub-element of the designed system includes the data enrichment modules.  The second set of modules focuses on adding information to the cleaned and shaped IP and innovation data delivered by the initial system modules.  Advanced techniques, are applied, such as geocoding entities' addresses (*e.g.* applicants, inventors or creators), disambiguating and classifying entities, attributing gender, classifying industries and technologies, which are typically used in IP and innovation data analysis.

The enrichment of IP and innovation data enables to increase the analytical potential of the underlying data.  In addition, the modules in this sub-element consolidate the data in a standardized, enriched database ready for analytical use.  As a result, the consolidated database provides data analysts, economists, statisticians and other IP Office users with a complete data source for their reporting and research activities.

c.  Inputs for Economic Analysis

The third sub-element of the designed system includes modules for producing inputs for economic analysis.  These modules produce thematic indicators, tables, and figures reporting on several dimensions of the innovation ecosystem and its IP use.  The themes cover standard dimensions, such as use of IP in the economic environment of El Salvador (national and global),  and more advanced ones, like innovation capabilities, gender gap, geography of innovation, or innovation complexity.  Indicators are presented in different types of visualizations designed to convey key insight to analysts, economists, or policy makers.

The outcomes of each themed dimension will be shared with the pilot country counterparts, along with an associated training.  Some of these outcomes will be used in an economic study

to be presented during the training sessions.  The training strategy and study are further described in the next section.

### III.    DISSEMINATION STRATEGY

The methodology and project outcomes will be disseminated through several channels:

a)  all codes produced for the three sub-elements constituting the modules will be published in reproducible formats through scripts in open-source software (*e.g.* SQL and Python) and collaborative platforms, such as GitHub;

b)  interactive examples on how to use the different modules and how to produce thematic indicators, tables, and figures will be provided in different forms (*e.g.* Jupyter notebooks and interactive dashboards);  and

c)  all modules and outputs will be further disseminated through technical training workshops tailored to diverse groups of users across the participating IP Office and other government agencies.  These trainings aim to build human capacity, foster collaboration, and maximize the impact of the project across diverse stakeholder groups.

A set of training materials will be produced to provide targeted users with the skills and knowledge needed to take ownership of the system's deliverables.  Three target groups of users will be considered:

- Policymakers:  to be trained on how to interpret IP and innovation indicators produced by the analytical modules, empowering them to make informed decisions and formulate effective policies based on the system's outcomes.

- Data analysts, statisticians, and economists:  to be trained on how to use the analytical modules in order to derive insights from the data that is meaningful for policymaking.

- Data administrators:  to be trained on how to use the data preparation and enrichment modules, equipping them with the necessary skills to effectively manage and enrich their national data sources.

### IV.    LESSONS LEARNED AND NEXT STEPS

Developing the first version of the system encountered the following challenges:

a)  Establishing the right balance of internal and external skills required for conceiving of and implementing the system was not straightforward.  The implementation of the system required an unusual mix of advanced knowledge of IP data, IP and innovation indicators, and programming.  Similarly, the development and implementation of the dissemination strategy requires additional technical skills.

b)  Access to the data required coordinating with different institutional settings within both the participating pilot country and the Secretariat.

c)  The effort needed to research and explore which components to include in the first version was substantial.

However, these challenges served as valuable learning opportunities, highlighting the importance of adaptability and resilience in the project implementation.  Going forward, the lessons learned will provide an invaluable perspective for subsequent pilot countries.  The next steps will be WIPO's delivery of the trainings in pilot country 1 and delivery of the system's improved versions in pilot countries 2 and 3.  By proactively implementing the lessons learned and planning the next steps collaboratively, the project remains poised to overcome challenges and drive sustainable progress throughout developing the targeted set of deliverables.

[End of Annex and of document]