

CDIP/34/INF/2

ORIGINAL:  ENGLISH

DATE:  February 27, 2025

# Committee on Development and Intellectual Property (CDIP)

**Thirty-Fourth Session
Geneva, May 5 to 9, 2025**

SUMMARY OF THE “SYSTEM FOR STANDARDIZATION, ENRICHMENT AND ECONOMIC ANALYSIS OF INTELLECTUAL PROPERTY AND INNOVATION DATA TO SUPPORT POLICY DESIGN (VERSION 2.0)”

*prepared by the Secretariat*

 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 2.0)”.

 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).

 *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 2.0)”**

This document outlines the major improvements made in version 2.0 of the system designed to standardize, enrich, and analyze intellectual property (IP) and innovation data, aimed at providing comprehensive economic analysis to inform policy design. Developed under 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” (CDIP/26/4), version 2.0 significantly builds upon the foundational work laid by version 1.0. Key updates are derived from experience in pilot countries, including El Salvador, Bhutan, and Indonesia, incorporating a more robust, scalable, and insightful system to meet the diverse needs of national governments and IP Offices.

## KEY IMPROVEMENTS IN VERSION 2.0

### Enhanced Data Quality and Rigorous Checks

Applying the version 1.0 of the system to analyze the IP data provided by the participating pilot countries highlighted numerous data quality challenges. In response, version 2.0 has implemented enhanced data quality checks at various stages of data processing. Drawing lessons from El Salvador’s pilot, where inconsistencies in IP data were found, more stringent quality assessments now ensure the accuracy and completeness of input data before analysis begins.

For instance, by cross-referencing data across multiple national sources and international databases, the system is better equipped to identify discrepancies and gaps in the data. These improvements guarantee that the system outputs are based on reliable information, thus enhancing the credibility of policy recommendations drawn from the data. Bhutan and Indonesia’s inclusion has further strengthened these procedures, providing valuable validation across different data environments and institutional setups.

### Collaborative Digital Working Environment

A significant update in version 2.0 is the establishment of a GitHub repository. This repository allows for the centralized management of all system-related code, datasets, and methodologies, fostering an open, collaborative workspace for both WIPO and participating IP Offices. By utilizing GitHub, stakeholders across different regions can access and contribute to the system’s ongoing development, facilitating the exchange of knowledge and best practices in real-time.

Although the repository is currently only available for use at WIPO, it will be made publicly accessible by the end of the project. This would enable other countries, beyond the initial pilot group, to leverage the system for their own IP data analysis, scaling the system’s impact on a global level. This approach will create a transparent, open-source environment for IP data management and analysis, empowering national IP Offices with the tools to independently manage their datasets.

### Streamlined Data Ingestion and Cleaning

In version 2.0, the process for data ingestion has been improved with the introduction of a universal dictionary of variables. This dictionary harmonizes the labels and classifications used across different countries’ IP data, ensuring uniformity and comparability of data inputs. By standardizing raw data at the point of ingestion, the system significantly reduces errors related to inconsistent labeling, misclassification, or incomplete entries.

The new cleaning processes automatically apply these harmonized labels, streamlining the initial preparation phase. This enables the system to quickly and effectively process large datasets, allowing for faster turnaround times in producing actionable insights. Additionally, the system’s flexibility to integrate new country data enables future pilot countries to adopt these processes more quickly.

### Development and Codification of Indicators

One of the standout features of version 2.0 is the codification procedure for the development of indicators. This new process automates the creation of indicators from raw IP and innovation data, providing a direct pathway from data ingestion to policy-relevant insights. These indicators provide concrete information on the results of the data analysis processed delivered by the system modules, providing users with analytical tools and resources to better comprehend the dynamics of the national IP system and innovation ecosystem.

### New Features in the Innovation Complexity Module

The [Innovation Complexity Module](https://dacatalogue.wipo.int/projectfiles/DA_1_4_10_35_37_01/Systematization_Doc/EN/System%20for%20Standardization%2C%20Enrichment%20and%20Economic%20Analysis%20of%20IP%20and%20Innovation%20Data%20to%20Support%20Policy%20Design_V1.pdf) has been significantly expanded in version 2.0, incorporating additional visualizations and indicators that allow the analyst to deliver more actionable insights. Notably:

* The distinction between established and endangered capabilities to help countries identify potential threats of losing valuable knowledge.
* The breakdown of relatedness density into the different dimensions of innovation to understand which sources of a country’s knowledge are bringing them opportunities to diversify their knowledge.
* The distinction between global and local innovation opportunities, allowing to add additional layers to hierarchize diversification paths based on the risks they encompass.
* The identification of the number of innovation outputs that are required to achieve a capability, allowing to add additional layers to hierarchize them.
* An indication of an innovation field that is relatively complex to a unit. This makes the complexity indicators relative to the perspective of the country being analyzed.
* The profiler function that automatically generates a set of figures and indicators for any combination of units (countries).

### Redesign of Project Outputs and Automation

Version 2.0 rethinks the overall structure of project outputs. In response to feedback from the pilot countries, the system has incorporated automated generation of indicators for any country, based on WIPO data. This ensures that reports, graphs, and other outputs are produced efficiently and with minimal manual intervention.

A redesign of the [Innovation Ecosystem Fundamentals Module](https://dacatalogue.wipo.int/projectfiles/DA_1_4_10_35_37_01/Systematization_Doc/EN/System%20for%20Standardization%2C%20Enrichment%20and%20Economic%20Analysis%20of%20IP%20and%20Innovation%20Data%20to%20Support%20Policy%20Design_V1.pdf) has been considered to enable the content of this module to smoothly introduce users to the technical aspects of the subsequent system modules with focus on: gender innovation gaps, innovation hotspots and innovation complexity.

## LESSONS LEARNED AND NEXT STEPS

The development of the version 2.0 revealed the importance of having a flexible, yet standardized approach to managing diverse national datasets. The process of addressing the data quality challenges encountered in version 1.0, especially during the El Salvador pilot, emphasized the critical need for rigorous validation and harmonization processes. Lessons from Bhutan and Indonesia highlighted how varying national data environments require tailored ingestion and cleaning procedures. Additionally, establishing a collaborative, digital working environment on GitHub has been underscoring the value of transparency and real-time collaboration between IP Offices and WIPO. The introduction of a universal dictionary of variables proved to be a key milestone, as it allowed for standardized data classification, irrespective of the country providing the input. This standardization, combined with enhancements to the Innovation Complexity Module, showcased the effectiveness of applying universal tools that also allow for country-specific insights.

Moving forward, version 3.0 will include further refinement based on continued interaction with existing pilot countries and new data expected to be provided by the latter. The focus will be on automating key aspects of indicator generation and analysis, making the system scalable and accessible to a broader range of countries. Furthermore, work is in progress for the development of the training materials to make them available on GitHub, expecting to deliver the trainings for the pilot countries during the second half of 2025. These steps are essential to ensuring the system remains relevant, flexible, and adaptable to the evolving needs of both pilot countries and the broader global innovation landscape.

[End of Annex and of document]