STATISTICAL INTEROPERABILITY NODE

In Spain, a **working group (InterSTAT) was created within the CITE** (Interterritorial Committee of Statistics) to establish the legal and technical foundations for the Statistical Interoperability Node, in accordance with Royal Decree 4/2010, of January 8, which regulates the **National Interoperability Framework** (ENI) in Electronic Administration.

The InterSTAT group of the CITE highlighted that **bureaucracy and obsolete exchanges** complicate access to statistical data and proposed a **national strategy** to improve interoperability and **simplify data exchange** among various stakeholders.

As a result, this project has been developed with the objective of **defining a Statistical Interoperability Node**.

This Node must establish **data management policies** that enable data exchange across all **levels of Spanish management** (regional and national).

Furthermore, it must **facilitate data exchange with European Union (EU) bodies** and **improve decision-making** in public policy matters.

All of this is achieved through **standardized common tools and procedures**.

Unión Europea

Instituto Nacional de Estadística y Órganos de Gobierno
Institutos de Estadísticas de las CCAA
Entidades locales, universidades y otros entes regionales

The project has been undertaken through the **European Union Technical Support Instrument** (TSI), designed to assist Member States in **implementing reforms that enhance their administrative and technical capacities**. This support is essential for designing, developing, and implementing effective and sustainable public policies that address the structural and situational challenges faced by European Union (EU) countries.

It has been **led** by the Andalusian Institute of Statistics and Cartography (IECA), the National Institute of Statistics (**INE**) and the Canary Islands Institute of Statistics (**ISTAC**), with the support of **DG-Reform**.

Objectives

The purpose of this project is to **define a Statistical Interoperability Node** that establishes statistical data management policies.

Definition of **standards, architectures, and common methodologies** for statistical data to enable data exchange between Spanish and EU agencies.

Increase the value of statistical data sets to improve government decision-making regarding public policies in Spain by providing high-quality data sets.

Test the technical and functional architecture of the Statistical Interoperability Node, defined by the beneficiaries, in order to refine it with best practices and lessons learned for future adoption by interested agencies.

Assessment

After **assessing the current state** of statistical data exchange between the different levels of administrative management in Spain, the following was identified:

1.- Processes: Statistical data exchanges were established reactively, creating manual and human dependencies, without standardization or formal agreements in many cases, resulting in disparate data structures.

2.- Technology: The exchanges rely on rudimentary and obsolete systems that hinder interoperability. These systems require intermediaries and multiple points of access, which increase maintenance and evolution costs, in addition to obstructing data traceability and standardization.

3.- Legal framework: The governmental decentralization in Spain results in regulatory dispersion concerning statistical regulation, making the homogenization and unification of criteria difficult. Furthermore, the agreements governing data exchange between administrations are bureaucratic and costly, slowing down the process and increasing administrative costs.

Moreover, a **significant diversity of flows and exchange processes** involving the beneficiaries was detected.

For the same entity, there were **various procedures and types of files**, highlighting the need to homogenize and standardize the different exchanges, as well as to create a common node.
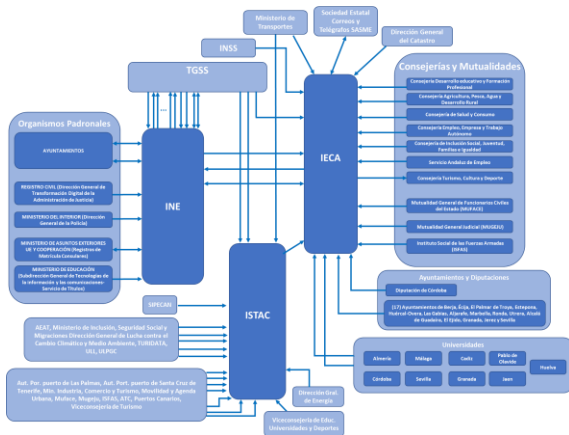
Functional requirements

Following the analysis of the current situation, the **requirements** that the Statistical Interoperability Node must address **were identified**, which will therefore serve as the basis for its design

REQ1 – Management of agreements and supply contracts

REQ2 – Data catalogue

REQ3 – Coordinated governance model

REQ4 – Process automation

REQ5 – Rules, validations and transformations

REQ6 – Secure Exchange channels

REQ7 - Synchronous and asynchronous integrations

REQ8 – Flexibility to future needs

REQ9 – External user access

Selected tool

Once the current situation was analyzed, as well as the requirements that the future Node must contain, a **market analysis was conducted to identify the available solutions** that meet the needs of the Spanish statistical system.

The **solution that best fits** this project is the **General State Administration Data Platf**orm, based on the IBM Cloud Pak for Data product. This platform integrates a data sharing module with the following functionalities:

| Functions of the General State Administration Data Platform	
Creation of the public sector data space	Generation of data catalogs (data products) for each organization
Uniform access to data sets	Interoperability between organizations and with other sectoral data spaces

Future Node

The future Statistical Interoperability Node will serve as the **national space for statistical data**, ensuring the **standardization of exchange processes** and facilitating access to data by the various actors involved in these processes
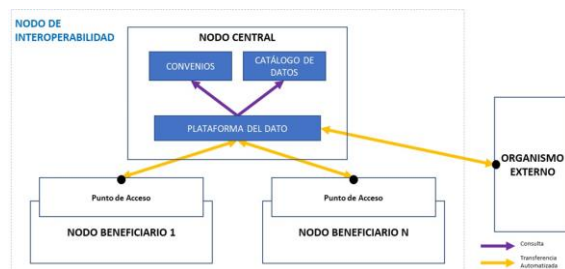
It has been defined as a **node of nodes**, with the approach of a **federated data space**, where the **digital sovereignty of each participating organization is respected**. In this context, a Central Node has been established to implement, **centralize, and govern the exchanges** of statistical information among the various participating organizations, whether regional or national.

## Use Case

CU1 PUSH - INE → IECA

Exchange in file format and according to the supply agreements, with automated sending to the destination node.

CU2 PULL - ISTAC → INE

Exchange with registration and publication of the file in the Catalog, partial consultation of the information on the platform.

CU2 (bis) PULL - ISTAC → INE

The same exchange as CU2, but the file reception is automatic, using an sFTP as the destination node.
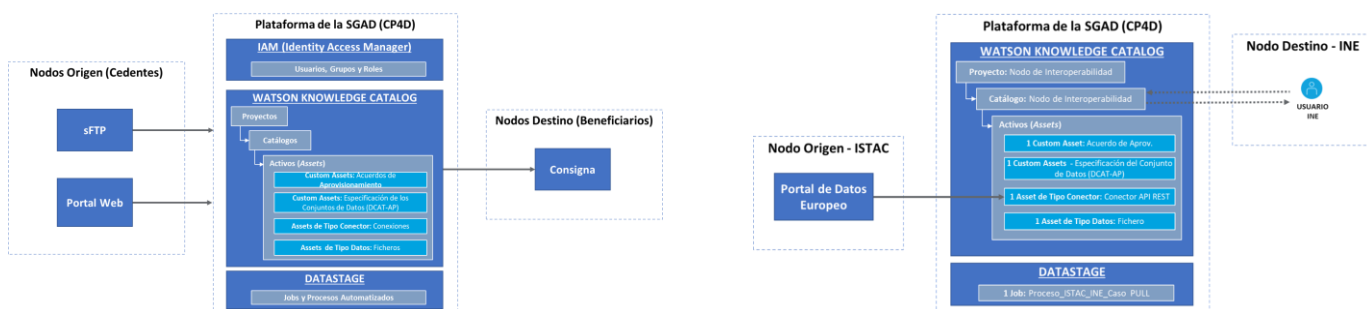
Conclusions after the PoC development

The platform **meets the main requirements** identified and evaluated during the analysis and diagnosis phase.

There are **some key constraints that limit, but do not block**, the use of the SGAD platform as a technical solution for the implementation of the Statistical Interoperability Node:

User **connectivity** to the catalog must **always occur within the Sara Network**; there cannot be a public catalog on the Internet. However, access to data sources or data destinations outside the Sara Network will be allowed, provided that such connectivity is approved and managed by the SGAD team.

There are **limitations in configuring the granularity of security** for the catalog elements and the actions users can perform on them. A significant evolution in this aspect is expected in the next version 5 of the product (currently 4.8.

For the **files published in the catalog, there is no possibility of direct download without being persisted on the platform**. For the execution of scheduled exchanges or those carried out using the Datastage product, temporary persistence of the information will be necessary during the execution of the process. No one, except the process itself, will have access to this information..

Lessons Learned

In order to ease the future implementation of the Statistical Interoperability Node, the following **lessons or considerations must be taken into account**, which are derived from the present project::

- There is no **market product that fully meets the technical requirements** of the Node. Therefore, to completely cover these requirements, **additional custom developments would be necessary**.

- The **platform may have specific technical requirements that must be met** for its proper implementation and operation, limiting the available options for deployment.

- There is a **need for adjustments or updates in regulatory** matters within the platform to maintain compliance.

- It is recommended to **standardize** and ensure that all access points use the same technologies and protocols.

- **Data governance will always remain in the hands of each member**, and access to information will be conducted solely through local access points.

- The **platform must define the definitive standards** for information processing to be followed. The use of DCAT-AP for the definition of information metadata and DDI for the definition of field metadata has been proposed.

- **Identity management** in the IBM Cloud Pak for Data product **has limitations** in the evaluated version (4.8). This management should be analysed in future versions of the product.