Del dato a la inteligencia fiscalizadora compartida: un Espacio Nacional de Datos de Control Externo (ENDICEX).

Leonardo Visconti Cox
Auditor del Tribunal de Cuentas

Samuel Gómez Vidal
Juan Millán Santana
Auditores de la Cámara de Cuentas de Andalucía

Revista Auditoría Pública nº 88. Diciembre 2026. Páginas: 155-164

Resumen: Este artículo propone el ENDICEX como modelo de cooperación entre las instituciones de control externo (ICEX) para compartir datos, indicadores, metodologías y conocimiento fiscalizador. Aprovechando las infraestructuras existentes y preservando la autonomía institucional, plantea una nueva etapa de transformación digital orientada a generar inteligencia fiscalizadora compartida y facilitar la incorporación progresiva de la inteligencia artificial (IA) a la fiscalización pública.

Como aplicación exploratoria, se analizan 111 entidades locales mediante indicadores socioeconómicos y financieros. El análisis de componentes principales (PCA) y el algoritmo de agrupamiento K-means permiten identificar 5 perfiles diferenciados entre 101 entidades, mientras que las 10 restantes se examinan separadamente por incumplimientos de rendición de cuentas. Los perfiles obtenidos permiten proponer actuaciones fiscalizadoras adaptadas a las características de cada grupo.

Los resultados ilustran el potencial del ENDICEX para orientar la planificación fiscalizadora basada en riesgos y favorecer la reutilización de metodologías analíticas reproducibles entre instituciones. La propuesta constituye un marco de cooperación para avanzar desde la explotación individual de datos hacia la generación de conocimiento fiscalizador compartido.

Palabras Clave: inteligencia fiscalizadora; espacios de datos; inteligencia artificial; aprendizaje automático; control externo.

Abstract: This article proposes ENDICEX as a cooperation model among external audit institutions (ICEX) to share data, indicators, methodologies, and audit knowledge. Building on existing infrastructures while preserving institutional autonomy, it outlines a new stage of digital transformation aimed at generating shared audit intelligence and facilitating the progressive integration of artificial intelligence into public sector auditing.

As an exploratory application, 111 local government entities are analyzed using socioeconomic and financial indicators. Principal Component Analysis (PCA) and the K-means clustering algorithm identifies 5 distinct profiles among 101 entities, while the remaining 10 are examined separately due to non-compliance with financial reporting obligations. The resulting profiles support the proposal of audit procedures tailored to the characteristics of each group.

The results illustrate ENDICEX's potential to guide risk-based audit planning and promote the reuse of reproducible analytical methodologies across institutions. The proposal provides a cooperation framework for advancing from individual data analysis towards the generation of shared audit knowledge.

Keywords: audit intelligence; data spaces; artificial intelligence; machine learning; external audit.

1. Introducción.

Las ICEX disponen de fuentes de datos y herramientas, pero su desarrollo aislado dificulta su reutilización. Además, tienen el reto de transformar los datos disponibles en **inteligencia fiscalizadora**, para generar valor público conforme a los principios de la INTOSAI-P 12, *El valor y beneficio de las Entidades Fiscalizadoras Superiores: marcando la diferencia en la vida de los ciudadanos* (INTOSAI, 2019a).

La **inteligencia fiscalizadora** es la capacidad institucional de combinar datos, metodologías analíticas y conocimiento profesional para generar información útil, verificable y reutilizable, bajo el juicio de los auditores. Cuando esta capacidad se comparte entre distintas instituciones, se genera la **inteligencia fiscalizadora compartida**.

Este trabajo plantea 3 objetivos:

1. Analizar la evolución de la forma de remitir información para generar inteligencia fiscalizadora.

2. Proponer crear el ENDICEX como modelo de cooperación basado en la interoperabilidad e integración de fuentes de datos (*data pooling*).

3. Explorar su potencial para compartir metodologías reproducibles, incorporar la IA y favorecer la transparencia y la participación ciudadana mediante el uso de datos abiertos.

Para ilustrar esta propuesta, se analizan datos financieros y socioeconómicos de entidades locales mediante técnicas estadísticas y de aprendizaje automático. El objetivo es identificar patrones que permitan orientar las fiscalizaciones según los riesgos detectados y las características socioeconómicas de cada municipio.

2. Espacio de datos: hacia un nuevo ecosistema de conocimiento fiscalizador.

Un **espacio de datos** es un ecosistema que facilita el intercambio y la reutilización de información entre organizaciones mediante un marco común de gobernanza, interoperabilidad y confianza (Centro de Referencia de Espacios de Datos [CRED], 2025, apartado 1.1, p. 4). La Guía de Gobernanza de Espacios de Datos contempla, entre otras formas organizativas, un modelo en el que los participantes conservan su autonomía operativa y soberanía sobre los datos, mientras cooperan mediante mecanismos de interoperabilidad y principios compartidos de gobernanza (CRED, 2025, apartado 3.3, p. 15).

Un referente para ENDICEX es el **Espacio Nacional de Datos de Salud**, concebido como una plataforma distribuida, compuesta por un nodo central y una red de nodos autonómicos (Ministerio de Sanidad, s.f.).

En el ámbito del control externo, existen experiencias que muestran el potencial de este enfoque:

- La **Cour des comptes** de Francia publica, a través del portal de datos abiertos del Gobierno francés, información sobre finanzas públicas, informes y actividad fiscalizadora, facilitando su acceso y reutilización (Cour des comptes, s. f.).

- El **Tribunal de Cuentas Europeo**, mediante su Decisión n.º 6-2019, de 16 de abril, sobre la política de datos abiertos y la reutilización de documentos, establece un marco para facilitar la reutilización de sus documentos públicos y contempla su disponibilidad en formatos abiertos y legibles por máquina, cuando resulte posible y adecuado (Tribunal de Cuentas Europeo, 2019).

- Por su parte, la **Sindicatura de Comptes de la Comunitat Valenciana** ha desarrollado el **Sistema Integral de Gestión de Datos de la Sindicatura**, en el marco de su Plan Estratégico del Gobierno del Dato, con el fin de centralizar y gobernar la información institucional, así como facilitar la interoperabilidad con las entidades fiscalizadas y otras fuentes de datos.

Se propone ampliar la cooperación entre las ICEX para compartir datos, indicadores, metodologías y productos analíticos reproducibles. El ENDICEX aprovecha y complementa las infraestructuras existentes mediante mecanismos comunes de interoperabilidad, análisis y reutilización de resultados, al tiempo que preserva la autonomía institucional.

3. De la rendición telemática a la inteligencia fiscalizadora compartida.

La transformación digital de las ICEX puede estructurarse en tres niveles de evolución.

1. **La recepción electrónica** constituye la etapa consolidada en las ICEX. Comprende la remisión telemática de cuentas anuales e información sobre contratos, convenios y control interno, mediante plataformas que facilitan su disponibilidad y posterior tratamiento.

2. **La explotación analítica** representa la transición que actualmente están desarrollando las ICEX mediante la creación de modelos de datos propios y la integración de fuentes externas. El uso de indicadores, técnicas estadísticas e IA, junto con la estructuración de datos, permite identificar patrones, anomalías y señales de riesgo. Sin embargo, el desarrollo aislado de estas capacidades limita su comparabilidad y reutilización. En esta evolución destaca el proyecto **VIGÍA-A** de la Cámara de Cuentas de Andalucía, orientado a impulsar una fiscalización más inteligente, predictiva y basada en riesgos mediante la incorporación de la IA y de la analítica avanzada de datos (Cardenete Flores, 2026).

3. **La inteligencia fiscalizadora compartida** supone avanzar del análisis individual hacia la generación y reutilización de conocimiento entre las ICEX. El ENDICEX facilitaría esta evolución mediante el intercambio de datos, indicadores, metodologías y resultados analíticos reproducibles, incorporando el juicio profesional de los auditores. Este modelo sentaría las bases para integrar progresivamente la IA en el ciclo fiscalizador.

4. El ENDICEX: concepto y aportación diferencial.

El ENDICEX se propone como espacio compartido que conecta fuentes de datos, herramientas y conocimiento fiscalizador mediante reglas comunes de interoperabilidad, aprovechando infraestructuras existentes, como la Plataforma de Rendición Telemática de las Instituciones de Control Externo (PICEX), sin necesidad de centralizar la información.

La propuesta se alinea con la INTOSAI-P 12 y la ISSAI 100, sobre valor público y principios de fiscalización (INTOSAI, 2019a, 2019b), y con las GPF-OCEX 1315, 1503 y 5371, relativas a la identificación de riesgos, la evidencia electrónica y el uso de la IA generativa, respectivamente (ASOCEX, 2022a, 2022b, 2026). Este marco favorece la aplicación de nuevas tecnologías, preservando la autonomía institucional y el juicio profesional del auditor.

La **figura 1** sintetiza 4 capacidades: integrar fuentes de datos; compartir métodos analíticos y herramientas de IA; aplicar resultados al ciclo fiscalizador; e incorporar el juicio profesional para interpretar y mejorar los procedimientos.

Figura 1. Capacidades del ENDICEX.



Fuente: elaboración propia con IA generativa (diseño gráfico).

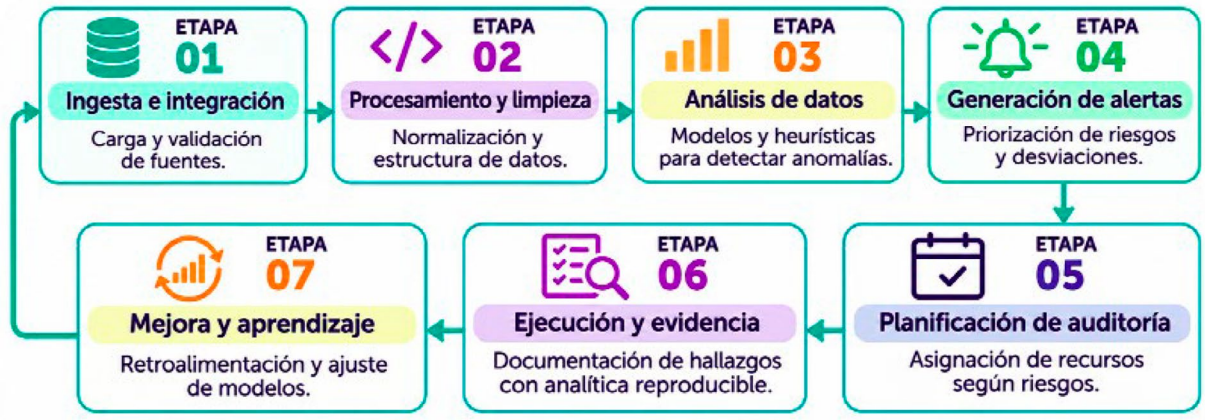
La aportación diferencial del ENDICEX es transformar análisis individuales en conocimiento fiscalizador compartido, impulsando su reutilización, comparabilidad y mejora continua. Por ejemplo, si una ICEX crea un modelo de endeudamiento atípico, puede compartir código y metodología. Así, la inversión individual genera capacidades colectivas, preservando la autonomía y el juicio profesional auditor. Su eficacia requiere validación entre ICEX.

5. Arquitectura funcional del ENDICEX.

El ENDICEX se propone como arquitectura funcional que conecta fuentes de datos, mecanismos de interoperabilidad técnica y semántica y servicios analíticos, bajo principios comunes de gobernanza, calidad, seguridad y trazabilidad.

En la **figura 2** se detallan las etapas interrelacionadas del ENDICEX.

Figura 2. Arquitectura funcional del ENDICEX.



Fuente: elaboración propia con IA generativa (diseño gráfico).

La arquitectura se proyecta en el ciclo fiscalizador: integra y analiza datos, identifica riesgos, genera alertas y facilita la obtención de evidencias.

Su desarrollo requeriría gobernanza común sobre calidad, actualización y disponibilidad de datos, así como acceso, validación y reutilización de modelos analíticos.

El ENDICEX podría implantarse progresivamente, comenzando por compartir indicadores, metodologías y código; continuando con su validación conjunta entre distintas ICEX; y avanzando hacia servicios analíticos interoperables y soluciones de IA compartidas.

6. Caso de uso: integración de datos y segmentación multidimensional de entidades locales.

Para ilustrar las posibilidades del ENDICEX, se integraron indicadores socioeconómicos y económico–financieros de 111 entidades locales, obtenidos de fuentes públicas. El objetivo es identificar perfiles de entidades que contribuyan a orientar la planificación fiscalizadora mediante la combinación de información financiera y socioeconómica de la población de sus municipios.

La muestra y el procedimiento de análisis se describen en el **anexo metodológico**.

6.1. Entidades sin cuentas rendidas: incumplimiento y vulnerabilidad socioeconómica.

A 31 de octubre de 2026, de las 111 entidades seleccionadas, 10 no habían rendido las cuentas de 2024; de ellas, 6 tampoco habían rendido las de 2023.

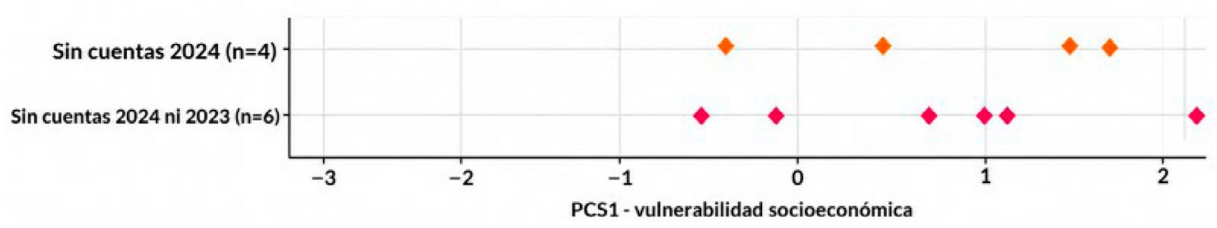
Estas entidades se analizaron separadamente, combinando la persistencia del incumplimiento con la vulnerabilidad socioeconómica de su población, incorporando esta última como criterio adicional para orientar la fiscalización. Para el análisis se aplicó la técnica de análisis de componentes principales (*principal component analysis*, PCA), que permite sintetizar los indicadores socioeconómicos en un componente (PCS1), como se detalla en el **anexo metodológico**[1].

1 Los datos, métodos y resultados están disponibles hasta el 30/04/2027 en GitHub (https://github.com/endicex/Endicex.git) para su consulta y reutilización.

La **figura 3** representa ambas dimensiones. El eje vertical distingue las situaciones de rendición y el horizontal muestra el componente socioeconómico (PCS1), cuyos valores más elevados indican mayor vulnerabilidad socioeconómica relativa.

Esta información permite complementar la priorización de fiscalizaciones de cumplimiento con criterios socioeconómicos, sin que estos impliquen por sí mismos un mayor riesgo de irregularidades.

Figura 3. Incidencias de rendición y vulnerabilidad socioeconómica.



Fuente: elaboración propia.

6.2. Análisis multivariante y construcción de conglomerados (clusters).

Para comparar las entidades se utilizaron 2 técnicas complementarias. Primero, el PCA permitió resumir los 5 indicadores originales en 3 componentes que conservan gran parte de su información, uno socioeconómico (PCS1) y dos financieros (PCF1 y PCF2). Después, el algoritmo de aprendizaje no supervisado **K-means** agrupó las 101 entidades según sus semejanzas en estos componentes, equilibrando el peso de los bloques socioeconómico y financiero. Así se identificaron 5 perfiles que permiten comparar situaciones diferentes y orientar posibles actuaciones fiscalizadoras, recogidos en el **cuadro 1**.

Cuadro 1. Caracterización socioeconómica y económico-financiera de los clusters de entidades.

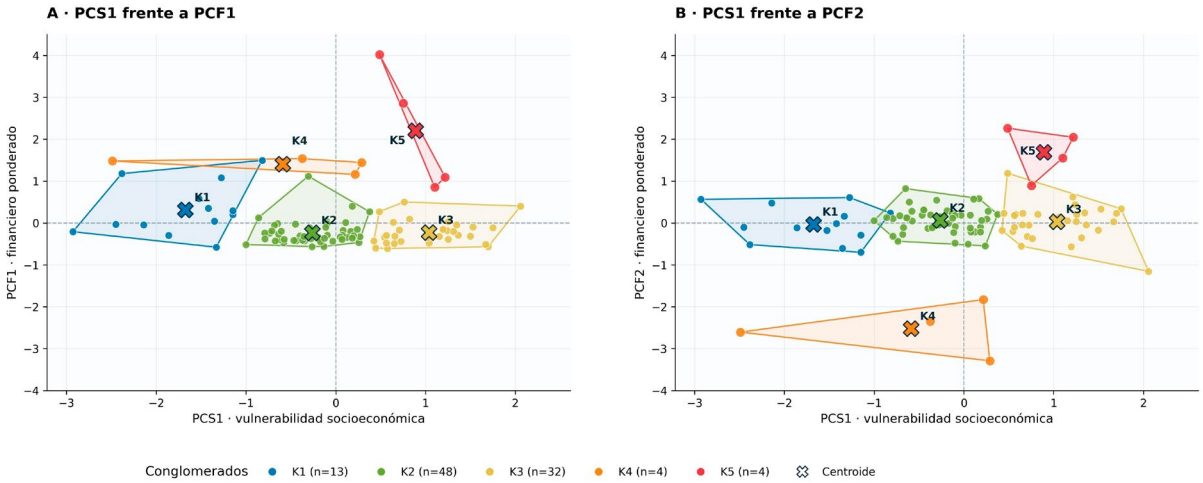
Perfil	Características identificadas
K1 (n=13)	**Baja VSE relativa** y heterogeneidad económico-financiera.
K2 (n=48)	**VSE intermedia** y baja singularidad económico-financiera relativa.
K3 (n=32)	**Elevada VSE** y singularidad económico-financiera relativamente moderada.
K4 (n=4)	**Baja VSE** relativa y elevada singularidad económico-financiera, especialmente inversora.
K5 (n=4)	**Elevada VSE** y acusada singularidad económico-financiera, especialmente vinculada al endeudamiento.

VSE: vulnerabilidad socioeconómica del municipio.
n= número de entidades/observaciones

Fuente: elaboración propia.

La **figura 4** representa los 5 conglomerados mediante 2 gráficos. En ambos, el eje horizontal (PCS1) indica la vulnerabilidad socioeconómica, mientras que los ejes verticales representan, respectivamente, los componentes financieros PCF1 y PCF2. Cada punto corresponde a una entidad, los colores identifican los grupos y las cruces sus centroides, que representan la posición media dentro de cada grupo.

Figura 4. Conglomerados K-means en 2 dimensiones (PCA).

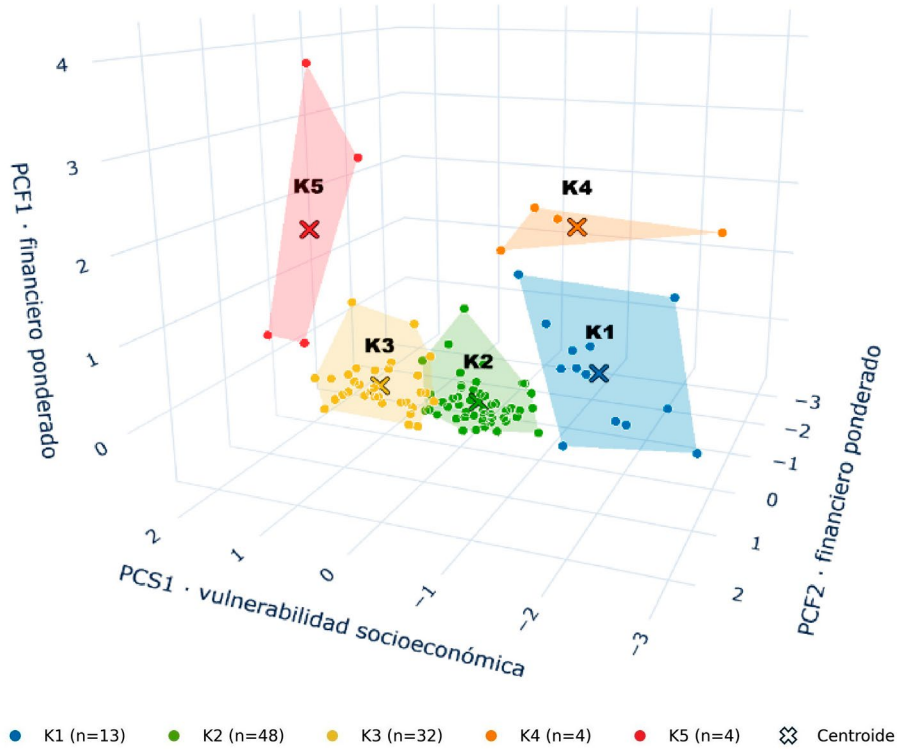


Fuente: elaboración propia.

Los gráficos sugieren que los municipios con distintos niveles de vulnerabilidad socioeconómica semejante presentan perfiles económico–financieros diferenciados. Así, K3 y K5 comparten una elevada VSE, pero K5 destaca por su singularidad financiera vinculada al endeudamiento. Por su parte, K4 presenta una elevada singularidad inversora pese a su menor VSE.

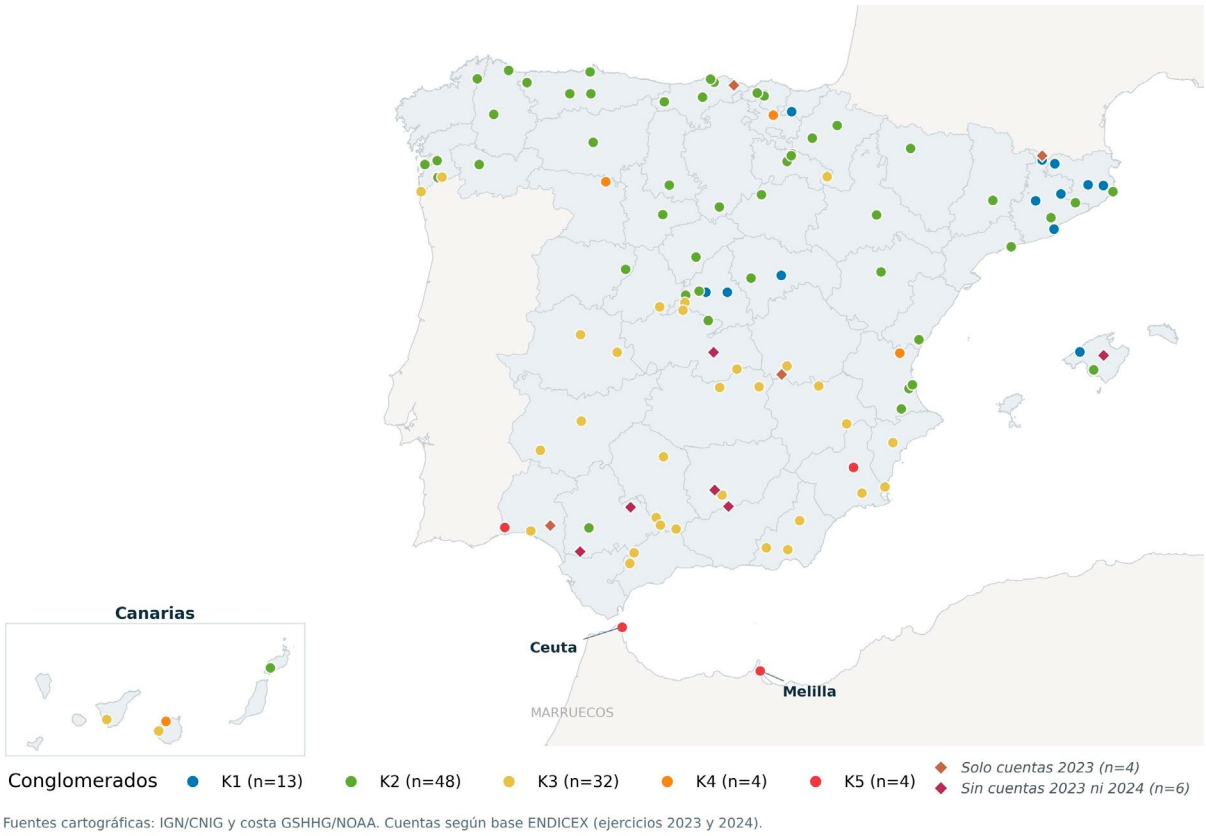
La **figura 5** representa conjuntamente los 3 componentes del análisis y la **figura 6** muestra su distribución territorial.

Figura 5. Conglomerados K-means en 3 dimensiones (PCA).



Fuente: elaboración propia.

Figura 6. Distribución territorial de las agrupaciones K-means y rendición de cuentas.



Fuente: elaboración propia.

6.3. Resultados y aplicación a la planificación fiscalizadora.

Los perfiles identificados permiten orientar decisiones diferenciadas de planificación fiscalizadora. Así, las entidades sin cuentas rendidas (NR) podrían priorizarse para fiscalizaciones de cumplimiento; K5, para examinar el endeudamiento y la sostenibilidad financiera; K4, para analizar la contratación y ejecución de inversiones públicas; y K3, para evaluar la cobertura y los resultados de los servicios públicos en contextos de mayor vulnerabilidad socioeconómica. Por su parte, K1 y K2 podrían mantenerse inicialmente como grupos de comparación o seguimiento, sin excluir posibles actuaciones sobre entidades concretas. El **cuadro 2** presenta una propuesta exploratoria de priorización.



Cuadro 2. Propuestas de actuación.

Grupo	Señal identificada	Fiscalizaciones propuestas	Prioridad indicativa
NR	Ausencia de rendición de cuentas de 2024, persistente en 6 entidades.	Cumplimiento de la rendición y transparencia	Muy alta
K5	Elevada vulnerabilidad socioeconómica (VSE) y singularidad financiera vinculada al endeudamiento.	Endeudamiento, sostenibilidad y servicios públicos	Alta
K4	Baja VSE y elevada singularidad financiera, especialmente inversora.	Planificación, ejecución y evaluación de la inversión	Alta
K3	Elevada VSE y singularidad financiera moderada.	Cobertura, accesibilidad y equidad de servicios	Media
K2	VSE intermedia y baja singularidad financiera.	Grupo de comparación.	Baja
K1	Baja VSE y heterogeneidad financiera.	Análisis de desviaciones financieras específicas	Baja

Fuente: elaboración propia.

El análisis permite orientar la selección de entidades y materias fiscalizables, sin evaluar la calidad de la gestión ni acreditar irregularidades. Las propuestas requieren información adicional y aplicación de juicio profesional. Su alcance podría ampliarse incorporando datos sobre control interno, contratos y convenios disponibles en la PICEX.

La metodología utilizada constituye una aplicación de IA mediante el algoritmo K-means, una técnica de aprendizaje automático no supervisado que permite identificar grupos de entidades con características semejantes (Jain, 2010), sin definir previamente las categorías. Además, la elaboración conjunta del caso por auditores de distintas ICEX y la publicación del código Python empleado ilustran cómo compartir metodologías reproducibles permite reutilizarlas, adaptarlas y mejorarlas entre instituciones, materializando el principio de inteligencia fiscalizadora compartida.

El caso permite constatar la viabilidad técnica de integrar fuentes heterogéneas y generar análisis reproducibles. No obstante, la eficacia del ENDICEX como modelo de cooperación institucional debería contrastarse mediante experiencias de aplicación conjunta entre ICEX que permitan evaluar la interoperabilidad, reutilización y mejora compartida de metodologías analíticas. Su desarrollo deberá garantizar la transparencia metodológica y trazabilidad de los resultados, preservando la autonomía institucional y el juicio profesional de los auditores.

7. Conclusiones.

- El ENDICEX propone transformar soluciones individuales en capacidades analíticas compartidas y reutilizables, favoreciendo la cooperación entre las ICEX.
- Para ello, propone conectar datos, indicadores, metodologías y resultados reproducibles mediante interoperabilidad y gobernanza comunes. Aprovechando las infraestructuras existentes y respetando la autonomía institucional, permitiría publicar, validar, versionar y reutilizar modelos, mejorar su comparabilidad y reducir duplicidades.
- El caso exploratorio integra indicadores socioeconómicos y económico-financieros de 111 entidades locales. Mediante PCA y aprendizaje automático no supervisado, se identifican 5 perfiles entre 101 entidades, mientras que las 10 restantes se analizan por incumplimientos de rendición de cuentas. Los resultados orientan fiscalizaciones diferenciadas, sin acreditar irregularidades ni sustituir el juicio profesional. El código Python y la metodología compartidos facilitan su replicabilidad.
- Esta experiencia ilustra el potencial del ENDICEX para impulsar una etapa de transformación digital de las ICEX, basada en la generación de inteligencia fiscalizadora compartida. Su desarrollo permitiría avanzar hacia modelos cooperativos de aprendizaje automático e inteligencia artificial en el ciclo fiscalizador. Bajo principios de transparencia, trazabilidad y supervisión profesional, este enfoque contribuiría a fortalecer la fiscalización basada en evidencias, facilitar el acceso y la reutilización de datos por la ciudadanía y generar mayor valor público, en consonancia con la INTOSAI-P 12 (INTOSAI, 2019a).

Bibliografía.

Asociación de Órganos de Control Externo Autonómicos (ASOCEX). (2022a). GPF-OCEX 1315. Identificación y valoración del riesgo de incorrección material (revisada).

Asociación de Órganos de Control Externo Autonómicos (ASOCEX). (2022b). GPF-OCEX 1503. La evidencia electrónica de auditoría.

Asociación de Órganos de Control Externo Autonómicos (ASOCEX). (2026). GPF-OCEX 5371. Utilización de la IA generativa en las fiscalizaciones.

Cardenete Flores, M. A. (2026, 21 de julio). *Vigía-A: una nueva generación de fiscalización al servicio de Andalucía*. La Voz del Sur.

Centro de Referencia de Espacios de Datos (CRED). (2025). Guía de gobernanza de espacios de datos (versión 1.0). Ministerio para la Transformación Digital y de la Función Pública, Secretaría de Estado de Digitalización e IA.

Comisión Europea. (2020). Una estrategia europea de datos [Comunicación COM(2020) 66 final].

Comisión Europea. (2025). Estrategia para una Unión de Datos: desbloquear datos para la IA [Comunicación COM(2025) 835 final].

Cour des comptes. (s. f.). *Jeux de données*. data.gouv.fr.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651–666.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), artículo 20150202. https://doi.org/10.1098/rsta.2015.0202

Ministerio de Sanidad. (s. f.). Cómo funciona el ENDS. https://ends.gob.es/explora-el-ends/como-funciona-el-ends

Organización Internacional de las Entidades Fiscalizadoras Superiores (INTOSAI). (2019a). INTOSAI-P 12. El valor y beneficio de las Entidades Fiscalizadoras Superiores: marcando la diferencia en la vida de los ciudadanos.

Organización Internacional de las Entidades Fiscalizadoras Superiores (INTOSAI). (2019b). ISSAI 100. Principios fundamentales de auditoría del sector público.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65.

Tribunal de Cuentas Europeo. (2019). Decisión n.º 6-2019, de 16 de abril de 2019, sobre la política de datos abiertos y la reutilización de documentos.

Anexo metodológico.

Alcance y análisis.

El tratamiento de los datos, la aplicación de técnicas de análisis multivariante y la identificación de clusters y posibles perfiles de fiscalización tienen una finalidad exclusivamente investigadora y exploratoria. Los resultados no constituyen hallazgos de fiscalización ni implican valoraciones sobre la gestión de las entidades analizadas. Cualquier conclusión fiscalizadora requeriría la aplicación de los correspondientes procedimientos formales de fiscalización.

Ámbito temporal y subjetivo.

El análisis comprende el ejercicio 2024 y se amplía a 2023 para examinar la persistencia de incumplimientos en la rendición de cuentas. Incluye ayuntamientos y las ciudades autónomas de Ceuta y Melilla.

Selección de muestra.

Se seleccionaron 111 entidades (109 ayuntamientos, Ceuta y Melilla) mediante 2 fases reproducibles. La muestra, exploratoria y no representativa, ilustra la integración de fuentes heterogéneas y la segmentación multidimensional.

En la primera fase se seleccionaron 98 ayuntamientos mediante muestreo estratificado en 4 intervalos poblacionales, con asignación proporcional a la raíz cuadrada de los municipios elegibles por estrato y selección pseudoaleatoria[2] sin reemplazamiento.

En la segunda fase se incorporaron Ceuta, Melilla y otros 11 ayuntamientos, seleccionados pseudoaleatoriamente entre provincias no representadas en la primera fase, para ampliar la cobertura territorial.

2 Las semillas usadas para garantizar la reproducibilidad son 20261027 y 20261138, en cada fase respectivamente.

Variables y fuentes de información.

Se utilizaron 5 indicadores agrupados en 2 dimensiones:

- Socioeconómica: renta media por persona y pobreza relativa[3].
- Económico-financiera: gasto, endeudamiento e inversión por habitante.

Los indicadores socioeconómicos se obtuvieron mediante las interfaces de programación de aplicaciones (API) del Instituto Nacional de Estadística (INE), a partir del Atlas de Distribución de Renta de los Hogares de 2024. La información económico-financiera se obtuvo principalmente de la PICEX. Para los municipios de Navarra y el País Vasco se consultaron, respectivamente, el Sistema de Intercambio de Datos de los Expedientes Contables y el portal de datos abiertos del Instituto Vasco de Estadística.

La cartografía oficial de los límites administrativos se obtuvo del Instituto Geográfico Nacional, complementada con información de la base de datos geográfica Global Self-consistent, Hierarchical, High-resolution Geography Database, de la National Oceanic and Atmospheric Administration.

Tratamiento de los datos.

Se normalizaron identificadores, se comprobaron inconsistencias y se trataron los valores ausentes.

Los 3 indicadores económico–financieros se transformaron mediante desviaciones respecto de la mediana, normalizadas por el rango intercuartílico, para comparar distintas escalas y representar desviaciones respecto de los valores centrales.

En 8 municipios se imputó el endeudamiento por habitante mediante la mediana de su estrato poblacional.

Un total de 10 entidades sin cuentas rendidas de 2024 se analizaron separadamente; las 101 restantes se incorporaron al análisis multivariante.

Análisis de componentes principales (PCA).

Se aplicó la técnica estadística PCA por separado a los indicadores socioeconómicos y financieros para reducir su dimensionalidad mediante componentes no correlacionados, conservando la máxima variabilidad posible (Jolliffe y Cadima, 2016).

El análisis socioeconómico generó un componente (PCS1) que sintetiza los indicadores de renta y pobreza relativa y explica el 92,72 % de su varianza conjunta.

El análisis económico-financiero generó 2 componentes: PCF1, asociado principalmente al gasto y la inversión por habitante, y PCF2, al endeudamiento por habitante. Ambos explican conjuntamente el 87,58 % de la varianza de los indicadores financieros transformados.

Así, los 5 indicadores se sintetizaron en 3 componentes principales. En futuras aplicaciones debería ampliarse el número de variables consideradas.

Agrupamiento mediante K-means.

Se incorporaron 3 componentes al algoritmo K-means, técnica de aprendizaje automático no supervisado que agrupa entidades según la semejanza de sus características, mediante la minimización de la suma de las distancias euclídeas al cuadrado entre cada entidad y el centroide de su conglomerado.

Para equilibrar la contribución de las dimensiones socioeconómica y económico-financiera al agrupamiento mediante K-means, los 3 componentes se estandarizaron y se asignó una ponderación de 1 al componente socioeconómico (PCS1) y de $1/\sqrt{2}$ a cada uno de los 2 componentes financieros (PCF1 y PCF2).

El número de conglomerados se seleccionó combinando un mínimo de 4 entidades por grupo y la maximización del coeficiente de silueta (Rousseeuw, 1987). Aunque K=6 obtuvo el mayor coeficiente de silueta (0,4175), se seleccionó K=5 (0,4047), dado que la mejora de 0,0128 puntos era reducida y la solución de 6 conglomerados no cumplía el criterio de tamaño mínimo establecido.

Se identificaron 5 conglomerados de 13, 48, 32, 4 y 4 entidades, respectivamente, cuyos perfiles pueden orientar la planificación fiscalizadora, sin que su pertenencia implique necesariamente riesgos o irregularidades en la gestión pública.

La estabilidad de los conglomerados ante cambios en las ponderaciones y la imputación de datos podrá evaluarse en futuras aplicaciones.

3 Pobreza relativa: porcentaje de población con ingresos por unidad de consumo inferiores al 60 % de la mediana nacional de renta equivalente (INE, 2024). Adicionalmente, se utilizó el índice de Gini como indicador descriptivo de la desigualdad de renta, sin incorporarlo al PCA.