

Estrategia CLARIAH-IATEXT 2026-28

(Aprobada en la Comisión CLARIAH-IATEXT el 16/03/2026)

(Ratificada en la Comisión Ejecutiva del IATEXT el 16/04/2026)

1. Presentación y contexto institucional

El Instituto Universitario de Análisis y Aplicaciones Textuales (IATEXT) de la Universidad de Las Palmas de Gran Canaria (ULPGC) participa en la infraestructura europea de investigación CLARIAH-ES como nodo activo del consorcio español. CLARIAH-ES es la infraestructura nacional de investigación para las Humanidades Digitales, integrada en las ERIC (*European Research Infrastructure Consortium*) [CLARIN-ERIC](#) y [DARIAH-EU](#).

El IATEXT está reconocido como [K-Center](#) (Centro de Conocimiento) de CLARIN desde el 20 de octubre de 2018 y ha solicitado su reconocimiento como [B-Center](#) de CLARIN el 19 de enero de 2026. A esto se une la solicitud de implantación de un Máster Universitario en Humanidades Digitales Inteligentes, remitido para su tramitación al Vicerrectorado de Titulaciones, Calidad e Innovación el 18 de febrero de 2026; un Máster que responde en sus especialidades a competencias vinculadas a cada una de las ERIC que aglutina CLARIAH-ES.

Cumplidos estos objetivos, se hace necesario desarrollar un plan que permita tanto consolidar, como ampliar la actividad investigadora del IATEXT en CLARIAH y, por ello, se presenta una estrategia que define las prioridades, proyectos y acciones del nodo CLARIAH-IATEXT para el trienio **2026-2028**. Esta estrategia se sustenta en las propuestas transversales formuladas por los investigadores del IATEXT que participan de la Comisión de CLARIAH-ES del IATEXT, el mapa estratégico presentado por la dirección del IATEXT a partir de estas propuestas y los acuerdos adoptados por la Comisión CLARIAH-ES en su sesión del 16 de marzo de 2026. Así pues, esta estrategia refleja tanto el consenso institucional alcanzado sobre las líneas de actuación prioritarias, como la visión a medio plazo del IATEXT como referente en Humanidades Digitales Inteligentes.

2. Diagnóstico Estratégico

El IATEXT afronta este período desde una posición de consolidación interna, con un equipo multidisciplinar y una trayectoria reconocida en el análisis textual y patrimonial. No obstante, el diagnóstico previo a la planificación identifica un conjunto de oportunidades y retos que condicionan la estrategia.

Dimensión	Oportunidades	Retos
Tecnológica	Madurez de los LLMs y sistemas RAG; avances en HTR/OCR	Necesidad de aumentar la capacitación técnica de los equipos investigadores.

Dimensión	Oportunidades	Retos
Patrimonial	Vasto acervo documental canario sin digitalizar y dispersión de sitios webs de patrimonio canario (aprox. 500)	Barreras administrativas para el acceso a fondos archivísticos. Y desarrollo de herramientas para fomentar y facilitar la digitalización.
Internacional	Posición geoestratégica de Canarias entre Europa, África y América	Identificar la especificidad de un <i>Atlantic K-centre</i> de CLARIN y alcanzar la colaboración con instituciones africanas para acciones de rescate patrimonial.
Institucional	Apoyo de la ULPGC; Apoyo del Gobierno de Canarias como palanca de financiación	Dependencia de colaboraciones interinstitucionales (cabildos, ayuntamientos, archivos, instituciones culturales privadas, etc.)

3. Misión, Visión y Objetivos Generales

Misión: Desarrollar, mantener y difundir herramientas tecnológicas y recursos lingüísticos de alta calidad para la investigación en Humanidades Digitales Inteligentes, con especial atención al patrimonio lingüístico, histórico y cultural canario y del atlántico latino hablante y africano.

Visión: Consolidarse como el nodo de referencia en el ámbito atlántico dentro de la infraestructura CLARIAH-ES, liderando la integración de la Inteligencia Artificial en el tratamiento del patrimonio textual y proyectando su actividad hacia África y América Latina.

Objetivos Generales:

- 1 Desarrollar herramientas tecnológicas de impacto para el tratamiento automatizado de documentos históricos y patrimonio cultural.
- 2 Posicionar al IATEXT como Centro de Conocimiento (K-Centre) Atlántico dentro de la red CLARIN.
- 3 Fortalecer la capacidad investigadora del equipo mediante formación continua en Humanidades Digitales Inteligentes.
- 4 Ampliar la red de colaboraciones internacionales hacia África.

4. Ejes estratégicos: proyectos prioritarios

La Comisión CLARIAH-ES del IATEXT, en sesión del 16 de marzo de 2026, aprobó por asentimiento el siguiente orden de prioridad en los objetivos estratégicos del nodo para el período 2026-2028. La estrategia se articula en torno a tres ejes principales.

4.1. Eje 1 — Automatización del Tratamiento Documental: HTRLingSearch

"Desarrollar la herramienta HTRLingSearch que debe automatizar en lo máximo posible el proceso completo de creación de corpus de documentos escritos a mano, transcripción,

curación, etiquetado, asignación de metadatos y creación de una web de consulta sobre el contenido de los corpus." — Acuerdo de la Comisión CLARIAH-ES, 16 de marzo de 2026.

Este eje constituye la **primera prioridad estratégica** del nodo y responde a una necesidad crítica en el campo de las Humanidades Digitales: la existencia de una ingente cantidad de documentación histórica manuscrita que permanece inaccesible para la investigación por carecer de tratamiento digital adecuado. La herramienta *HTRLingSearch* abordará el ciclo completo de trabajo documental, desde la imagen digitalizada hasta el buscador semántico con inteligencia lingüística.

El proceso que HTRLingSearch automatizará comprende las siguientes fases encadenadas:

Fase	Descripción	Tecnología implicada
1. Digitalización	Captura y preprocesado de imágenes de documentos manuscritos	Escaneo de alta resolución, normalización de imagen
2. Transcripción (HTR)	Reconocimiento automatizado de texto manuscrito	Handwritten Text Recognition (HTR) / OCR avanzado
3. Curación textual	Corrección, validación y limpieza del texto transcrito	Revisión asistida por IA
4. Etiquetado lingüístico	Anotación morfosintáctica, semántica y de entidades	NLP, modelos de lenguaje especializados
5. Asignación de metadatos	Descripción archivística normalizada (Dublin Core, EAD)	Estándares de metadatos humanísticos
6. Web de consulta	Interfaz de búsqueda con inteligencia lingüística sobre el corpus	Motores de búsqueda semántica, RAG

La implementación inicial se abordará sobre dos fondos documentales en abierto, uno del Archivo Histórico Provincial de Las Palmas (AHPLP), en particular sobre actas de ayuntamientos, y el otro, sobre Cabildos Atlánticos Digitales en la Edad Moderna (CADIEM). Una vez validada, la herramienta estará diseñada para ser escalable a cualquier serie documental a nivel nacional e internacional, convirtiéndose en un recurso de la infraestructura CLARIAH-ES con proyección europea.

4.2. Eje 2 — Inteligencia Artificial Aplicada al Patrimonio Canario

"Desarrollo de una web de consulta sobre todo el patrimonio cultural, material e inmaterial, de Canarias basado en un sistema de Generación Aumentada por Recuperación (RAG) sobre un modelo grande del lenguaje de Inteligencia Artificial." — Acuerdo de la Comisión CLARIAH-ES, 16 de marzo de 2026.

La **segunda prioridad estratégica** consiste en el desarrollo de un agente conversacional de Inteligencia Artificial especializado en el patrimonio canario. En la actualidad, la información sobre el patrimonio material e inmaterial de Canarias se encuentra dispersa en aproximadamente 500 sitios web de diversa procedencia institucional y calidad, lo que dificulta el conocimiento de su existencia y la consulta integrada por parte de investigadores, educadores y ciudadanos.

La solución propuesta se basa en la arquitectura **RAG (Retrieval-Augmented Generation)**, que combina la capacidad de recuperación de información de grandes bases de conocimiento con la generación de respuestas en lenguaje natural propia de los Modelos Grandes de Lenguaje (LLMs). Este enfoque garantiza que las respuestas del sistema estén fundamentadas en fuentes documentadas y verificables, evitando las "alucinaciones" características de los LLMs generativos sin base documental.

El sistema permitirá a los usuarios realizar consultas abiertas en lenguaje natural y mantener un diálogo contextualizado con el agente sobre cualquier aspecto del patrimonio canario. Sus potenciales usuarios abarcan desde investigadores académicos hasta turistas que visiten espacios culturales como el Museo Elder, pasando por estudiantes de distintos niveles educativos.

4.3. Eje 3 — Internacionalización y Liderazgo como Centro-K Atlántico

"Paralelamente a las dos propuestas anteriores, el IATEXT se debe postular como un Centro K Atlántico. Fomentar la internacionalización tanto hacia África como a Hispanoamérica." — Acuerdo de la Comisión CLARIAH-ES, 16 de marzo de 2026.

El tercer eje estratégico responde a la oportunidad que representa la posición geoestratégica singular de Canarias —archipiélago atlántico situado en la encrucijada entre Europa, África y América— para proyectar la actividad del IATEXT a escala internacional. Este eje se desarrolla en dos líneas de acción paralelas.

Línea 3.1. Postulación como Centro-K Atlántico (CLARIN K-Centre): Los Centros de Conocimiento (*K-Centres*) de CLARIN son instituciones que ofrecen servicios especializados de asesoramiento, formación y acceso a recursos para la comunidad investigadora. El IATEXT, que participa actualmente en el *Spanish K-centre* sin un rol de liderazgo definido, aspira a constituir y liderar un K-Centre con vocación trasatlántica, que articule la colaboración entre instituciones europeas, africanas e hispanoamericanas en torno al análisis textual y las Humanidades Digitales, integrando, además, a otras instituciones públicas y privadas canarias implicadas en el desarrollo de herramientas digitales para la preservación e investigación en torno al acervo cultural de las islas.

Línea 3.2. Internacionalización del patrimonio canario en América: Esta iniciativa propone identificar y documentar el impacto de la "*canariedad*" en el patrimonio histórico y lingüístico

de países americanos como Cuba, Uruguay, Venezuela, República Dominicana y Estados Unidos. Se contempla la constitución de grupos de trabajo interdisciplinarios (histórico, lingüístico y de ciencias de la educación) que culminen en la creación de una exposición digital de carácter divulgativo.

Línea 3.3. Colaboración con África: Establecimiento de vínculos formales de colaboración con instituciones e investigadores africanos, abriendo una nueva dimensión geográfica para la actividad del nodo.

5. Hoja de Ruta 2026-2028

La siguiente tabla sintetiza la planificación temporal de las principales iniciativas del plan estratégico.

INICIATIVA	2026	2027 1 ^{er} semestre	2027 2 ^o semestre	2028
HTRLingSearch — Diseño y desarrollo del núcleo tecnológico	Inicio	Pilotos en AHPLP y CADIEM	Evaluación y propuestas de mejora	Escalabilidad nacional
Sistema RAG Patrimonio Canario — Desarrollo e integración de fuentes	Inicio	Versión beta pública		Versión estable y difusión
Centro-K Atlántico — Postulación y constitución	Postulación	Reconocimiento CLARIN		Operatividad plena
Internacionalización — Grupos de trabajo y colaboraciones	Contactos y acuerdos	Proyectos conjuntos		Exposición digital y publicaciones

6. Plan de formación y capacitación de los equipos investigadores

El desarrollo tecnológico previsto en los ejes estratégicos exige una actualización continua de las competencias de los equipos investigadores. Por ellos, se establece una acción formativa permanente que ofrecerá, al menos, dos cursos de formación cada semestre durante los tres años de la estrategia. Así, la Comisión CLARIAH-ES del IATEXT aprobó por unanimidad la celebración de una primera **Jornada Formativa del IATEXT** en el primer semestre de 2026, con dos cursos específicos

CURSO	FECHAS	PONENTE
Herramientas para la gestión de datos en Humanidades Digitales: Notepad++ y expresiones regulares	18-19 mayo 2026	F. J. Carreras Riudavets
Bases de datos para la Gestión de Datos en Humanidades Digitales	25-26 mayo 2026	F. J. Carreras Riudavets

Estos dos primeros cursos ofrecen bases metodológicas y técnicas necesarias para que los equipos investigadores puedan participar activamente en los proyectos estratégicos del nodo, especialmente y en primer lugar en lo relativo a la gestión de datos y el uso de herramientas digitales.

7. Conclusiones

La estrategia CLARIAH-IATEXT 2026-2028 traza una hoja de ruta ambiciosa y coherente para el desarrollo del nodo durante el próximo trienio. La estrategia combina el rigor tecnológico —con el desarrollo de herramientas de vanguardia basadas en HTR e Inteligencia Artificial— con una visión humanística y patrimonial que sitúa el acervo cultural y lingüístico canario en el centro de la actividad investigadora.

La apuesta por la internacionalización atlántica, materializada en la candidatura al Centro-K y en la expansión hacia África e Hispanoamérica, responde a la vocación histórica de Canarias como puente entre continentes y posiciona al IATEXT como un actor relevante en el ecosistema europeo de las Humanidades Digitales Inteligentes. La formación continua del equipo investigador garantiza, por su parte, la sostenibilidad del modelo a largo plazo.

En suma, esta estrategia constituye el marco de referencia para la toma de decisiones del IATEXT en el período 2026-2028, orientando los recursos humanos, técnicos e institucionales del nodo hacia objetivos de alto impacto científico, social y cultural.

Firmado a fecha de firma digital por el Secretario del IATEXT