

PROCESAMIENTO INFORMÁTICO DE LOS MATERIALES DEL ATLAS LINGÜÍSTICO-ETNOGRÁFICO DE COLOMBIA: SISTEMA DE INFORMACIÓN GEOGRÁFICA

Johnatan Estiven Bonilla
johnatan.bonilla@caroycuervo.gov.co

Daniel Eduardo Bejarano
daniel.bejarano@caroycuervo.gov.co

Julio Alexander Bernal Chavez
julio.bernal@caroycuervo.gov.co

Ruth Yanira Rubio
ruth.rubio@caroycuervo.gov.co

Andrea Lizeth Llanos
andrea.llanos@caroycuervo.gov.co

Grupo de Investigación en Lingüística de Corpus - Instituto Caro y Cuervo

RESUMEN

Esta ponencia presenta el procesamiento informático, digitalización y sistematización de los datos léxicos, etnográficos y materiales suplementarios del Atlas Lingüístico Etnográfico de Colombia (ALEC) en un Sistema de Información Geográfica (SIG). En la primera parte, se presenta un acercamiento descriptivo a los materiales que componen el ALEC. La segunda parte expone generalidades conceptuales sobre el manejo de datos espaciales, metadatos y el desarrollo de bases de datos con posibilidades de aplicación específica para datos lingüísticos y etnográficos en la dimensión diatópica. La tercera parte muestra la propuesta metodológica que incluye los criterios tenidos en cuenta para el tránsito de la información del soporte análogo a digital, la elección de los metadatos para la descripción y organización de la información y el modelamiento de la base de datos espacial y, por último, se presenta el Sistema de Información Geográfica del ALEC y sus posibilidades de uso, dando cuenta de su funcionalidad hasta el momento.

Palabras Clave: Bases de Datos Espaciales, Sistema de Información Geográfica, Atlas Lingüístico-Etnográfico de Colombia

1. CARACTERÍSTICAS DEL ALEC

El Atlas Lingüístico-Etnográfico de Colombia (ALEC) (Instituto Caro y Cuervo, 1982) es una de las investigaciones más destacadas en el campo de la lingüística colombiana. Se constituye como una de las dos obras más representativas del

Instituto Caro y Cuervo, siendo la primera el Diccionario de Construcción y Régimen de Rufino José Cuervo (1987). El ALEC fue creado con el objetivo de investigar los dialectos y subdialectos del español hablado en el territorio colombiano (Flórez, 1961), y actualmente representa uno de los pilares fundamentales para el estudio del español del país desde la gramática, la fonética y la dialectología.

Para la recolección de datos del ALEC fueron llevadas a cabo encuestas que se realizaron entre 1956 y 1978 con base en un cuestionario de 1500 preguntas. Participaron un total de 23 encuestadores y 2234 informantes, distribuidos a lo largo de 262 puntos del territorio colombiano¹. Como resultado fueron publicados seis tomos, cada uno de 50 x 35 cm, que contienen 1696 láminas con 1523 mapas de información lingüística, etnográfica o mixta, adiciones de texto, material fotográfico e ilustraciones. Los mapas de tipo lingüístico registran las diferentes formas en las que los informantes denominan los conceptos por los que fueron preguntados, mientras que los de tipo etnográfico especifican áreas de difusión geográfica de “cosas” o elementos propios de la vida popular y rural del país. De manera adicional, se publicó un manual que recoge información referente a aspectos metodológicos de la labor, la organización y explicación de la lectura de mapas y elementos de descripción de encuestadores, localidades y encuestados; así como un suplemento del Tomo III con muestras de habla espontánea y dos discos que contienen audios de juegos y cantos de velorio de las costas caribe y pacífica colombianas. Se cuenta también con más de 16000 fotografías y 1112 audios inéditos, ya digitalizados y en poder del Instituto Caro y Cuervo.

La información de los tomos está distribuida en dieciséis campos semánticos que describen la cotidianidad y realidad de los informantes y la vida del campo, a saber: Tiempo y espacio; Campo, cultivos, otros vegetales; Algunas industrias relacionadas con la agricultura; Ganadería; Animales domésticos; Reptiles, insectos, batracios, pájaros; Animales salvajes; Familia, ciclo de vida; Instituciones y vida religiosa; Festividades y distracciones; Vestido; Vivienda; Cuerpo humano; Alimentación; Oficios y empleo, transportes; Embarcaciones y pesca. En la parte final del tomo seis se muestran mapas con información sobre variación fonética y

¹ Es importante destacar que la totalidad del territorio colombiano no fue visitada para los trabajos consumados en el ALEC. Específicamente no se tuvieron en cuenta los departamentos de la zona suroriental del país (Guaviare, Guainía, Vichada, Vaupés) por el poco número de hablantes nativos de español en la época.

realizaciones gramaticales de palabras medidas por el uso del singular, el plural y el género, entre otras variables.

Teniendo en cuenta las características del ALEC, el tipo de información que contiene, las dimensiones de su publicación y la importancia que constituye para los estudios lingüísticos y la preservación del patrimonio lingüístico y cultural de Colombia, el Grupo de Investigación en Lingüística de Corpus ha trabajado desde el año 2015 en la digitalización de los materiales del atlas y así se ha establecido el propósito de acercarlo a la población interesada y convertirlo en una herramienta de consulta disponible en la web para usuarios espontáneos o especializados en los estudios de la lengua y la realidad etnográfica de la Colombia de la época.

2. EL PROYECTO DE EDICIÓN

2.1. Datos y datos espaciales

Una definición sencilla de dato es la de atributo que permite la descripción de una entidad determinada. En el ALEC, cada variante de una entrada puede ser considerada como un dato en tanto actúa de forma descriptiva sobre un referente específico. Por otra parte, los datos espaciales son grupos de datos que se relacionan con una ubicación sobre la superficie terrestre (Lamigueiro, 2014: 83). Teniendo en cuenta que el atlas es vivo ejemplo de la relación entre los estudios de la lingüística y su exposición a través de la representación cartográfica, la definición de dato debe enriquecerse para corresponder con la distribución geográfica de la información; es en ese sentido en que se debe pensar en datos espaciales que son referidos a través de coordenadas (que manejan una correspondencia unívoca con los datos) o convenciones (como sistemas de referencia de las coordenadas). La característica principal de los datos de naturaleza espacial radica en el dominio de la relación *información - ubicación*, en tanto un dato en bruto es indiferente a su distribución en el globo, contrario a lo que uno espacial sugiere.

2.2. Cuáles son los datos espaciales del ALEC

El ALEC constituye un gran banco de datos espaciales por ser un instrumento que expone elementos lingüísticos a través de su asociación con una localización específica. Cada variante presentada en un mapa, además de ser la muestra de la manera en la que una determinada palabra se utiliza y expande por el territorio

colombiano, está vinculada con una ubicación, dimensión y forma. Sus dimensiones y forma remiten a la representación gráfica de las variantes a través de símbolos; la ubicación permite darle un carácter de unicidad a cada variante en tanto la asocian con una coordenada específica, exacta y exclusiva.



Figura 1. Variantes y distribución parcial del mapa “CLARIFICAR EL GUARAPO” del Tomo I

Por otra parte, más allá de las variantes léxicas, las ilustraciones sobre objetos en los distintos puntos del país, las notas adicionales y distintos elementos que no necesariamente se encuentran en los mapas (como los del suplemento del tomo III) pueden constituir datos espaciales en el ALEC en tanto se asocian a puntos específicos dentro de la geografía nacional a través de una serie de códigos unívocos para cada punto explorado. A las localidades fueron asignados códigos compuestos por la grafía de máximo los dos primeros caracteres de cada departamento más un número designado según su ubicación sobre el espacio tal como muestra la tabla 1.

1	2
3	4
5	6

Tabla 1. Representación de las localidades en los mapas. (Flórez, 1983).



Figura 2. Ilustraciones para el mapa "PICO" del tomo III

3. LA HERRAMIENTA INFORMÁTICA

3.1 Diseño de la base de datos espacial del ALEC

Con el fin de desarrollar una base de datos flexible y consistente se identificaron los metadatos principales para el establecimiento de entidades y sus respectivas relaciones. Los metadatos son datos descriptivos sobre el contenido o la localización de otros con los que se relacionan (Pérez Matos, 2006). El metadato aporta información mínima necesaria para dar cuenta o identificar un evento o dato que sea objeto de análisis. Nadkarni (2011) también aporta a la definición al proponer que los metadatos son tipos de datos que tienen el objetivo de describir o definir elementos similares. Para la digitalización de los materiales del ALEC fue necesaria la elaboración de relaciones de metadatos de manera que se pudiera contar con una descripción robusta de la información consignada en la obra, evitando al máximo la pérdida de datos reflejo de la realidad lingüística del país.

El primer grupo de metadatos describe la entidad denominada *localidades* y contienen la geometría espacial (latitud, longitud), que permite su posterior visualización mediante el servidor de mapas, y datos descriptivos como el departamento, actividades económicas, fecha de encuesta, altura sobre el nivel del mar, número de habitantes, (en la fecha de la encuesta) , temperatura, año de fundación y vías de acceso. En segundo lugar, encontramos la entidad encuestadores sobre los cuales conocemos el nombre, años de vinculación y las

localidades donde realizó encuestas. Finalmente, en tercer lugar, tenemos la entidad denominada *informantes* de la cual contamos con los metadatos nombre, localidad, nivel educativo, ocupación, procedencia, edad, género y la procedencia de sus padres y cónyuge².

Los metadatos utilizados para el registro de la información de los mapas lingüísticos y etnográficos están divididos en las tablas Mapa, Variantes, Variantes Cartografiadas, Variantes No Cartografiadas, Adiciones, Adiciones Georreferenciadas, Suplementos y Suplementos Georreferenciados.

En Mapa se registran los datos para etiquetar la lámina como el título, sus traducciones al francés y al inglés, la referencia a otros atlas lingüísticos (si la hay) y el campo semántico al que pertenece (especificado con un id). En Variantes se enlistan las variantes cartografiadas y no cartografiadas y se les asigna un número de identificación y un código que especifica el símbolo con el que aparecen sobre el mapa. La hoja Variantes Cartografiadas relaciona los id de las localidades y los asocia con la ocurrencia las variantes e información adjunta que pueda tener lugar por informantes o localización.

La información de datos lingüísticos y etnográficos que no se encuentra ligada a una localización precisa del mapa en el momento de la encuesta se tipifica en las tablas Adiciones y Suplementos. Adiciones recoge el etiquetado de datos adicionales a los que las variantes o la información de la lámina remiten y que no es cartografiada pero sí que se asocian a alguna localidad en la hoja Adiciones Georreferenciadas. Finalmente, en Suplementos se etiquetan datos relacionados con el Suplemento del Tomo III que son llevados a la relación con una localidad del mapa en la tabla Suplementos Georreferenciados.

De lo anterior obtenemos el siguiente modelo relacional de la Base de Datos Espacial del ALEC que se encuentra instalado en un servidor de base de datos Postgresql y se conecta a un Sistema de información Geográfica basado en Geoserver mediante PostGIS.

² Para la recolección de la información del ALEC, se intentó que los informantes cumplieran con una serie de requisitos mínimos como ser nativos del sitio de la encuesta o haber pasado la mayor parte de su vida allí, tener una edad entre los 40 y los 60 años y ser campesinos, preferiblemente (Flórez, 1983: 45-49).

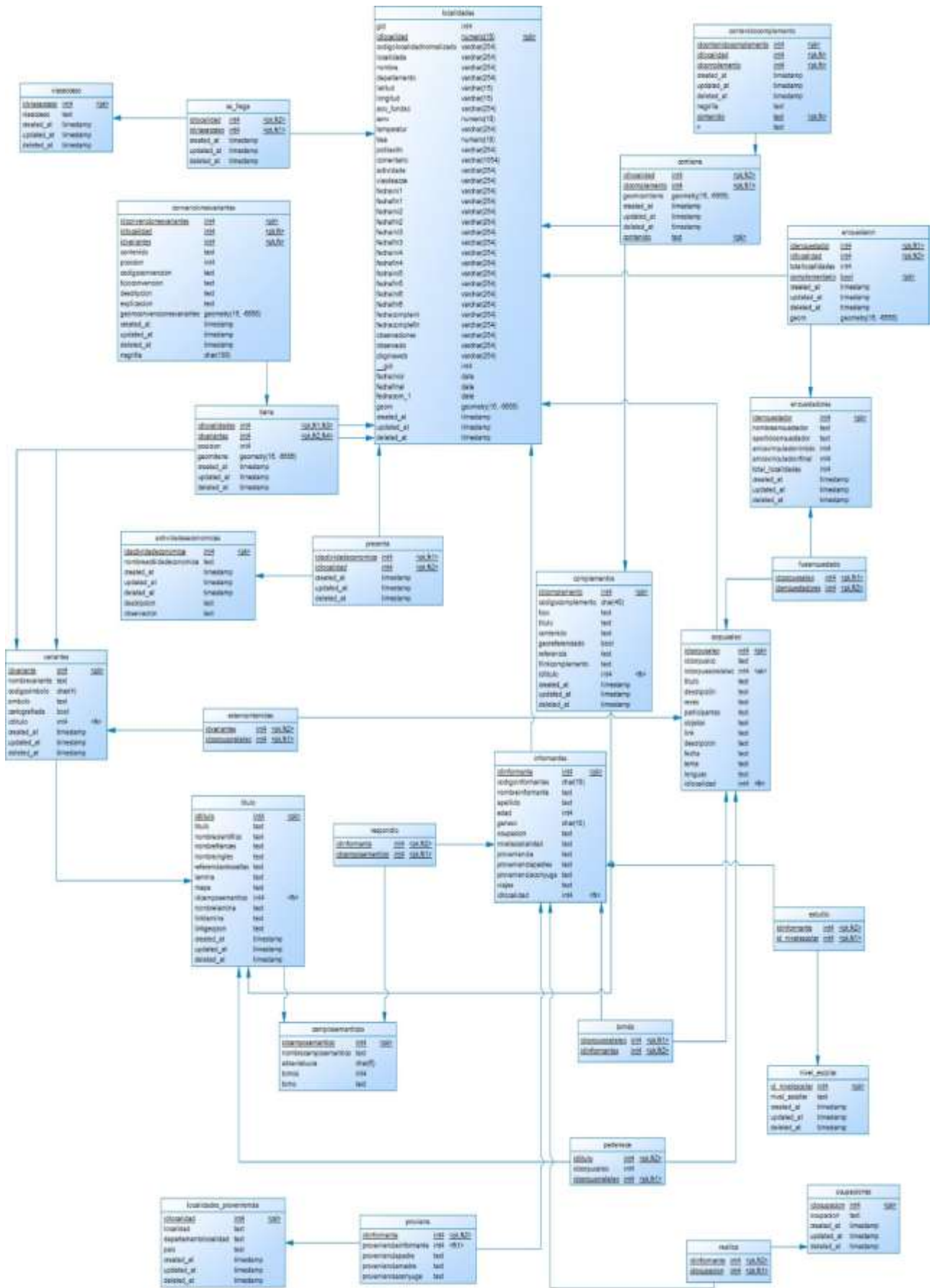


Figura 3. Modelo Relacional de la Base de Datos Espacial del ALEC

4. Sistema de Información Geográfica (SIG) para el ALEC

4.1. ¿Qué es un SIG?

Un SIG se puede definir como un conjunto de herramientas para la captura, el almacenamiento, la consulta, el análisis o la presentación de información geográfica determinada y la posterior toma de decisiones, según sugiere Burrough (1986). Asimismo, este tipo de sistemas sobrepasan la mera producción de mapas y permiten la recuperación y combinación de datos para la creación de nueva cartografía, así como ponen a disposición de los interesados herramientas de análisis y simulación (Rigaux, Scholl, Voisard, 2001).

Un SIG permite el almacenamiento de datos espaciales, la carga de nueva información y facilita la creación de mapas fiel a los expuestos por el atlas o nueva en tanto se puede basar en los datos que ya contiene la obra y representarlos según los intereses del usuario. En ese sentido, un SIG para el ALEC permite responder a cuestionamientos puntuales sobre la naturaleza o las características de sus datos lingüísticos y etnográficos en tanto está en la facultad de identificar datos, localizarlos en un punto dentro de una geografía específica, establecer relaciones entre unos y otros tipos de datos, tendencias sobre información lingüística diatópica o generacional y sugerir modelos sobre las posibilidades de interactuar con los datos de manera variada.

En síntesis, el desarrollo de un SIG para el ALEC es un paso de destacada importancia en tanto:

- Brinda la posibilidad de capturar, editar, consultar, analizar, interactuar y crear nueva información desde los datos del ALEC a través de herramientas de la tecnología y la geografía.
- Favorece la creación de productos para la difusión del material consignado en el ALEC como el Atlas Web.
- Posibilita la investigación relacionada con la geolingüística, la dialectología o lingüística en general y cualquier tema de interés del Instituto Caro y Cuervo.
- Se establece como un elemento destacado dentro de la voluntad por conservar y divulgar el patrimonio lingüístico y cultural de Colombia, en formato digital.

4.1. Ingreso y creación de mapas



Figura 4. Ingreso de mapas mediante hojas de cálculo

El SIG del ALEC como herramienta que administra la base de datos espacial permite el ingreso de los mapas de dos maneras. La primera es mediante la carga de hojas de cálculo que contengan los metadatos anteriormente mencionados en un orden establecido (Figura 4). En segundo lugar, es posible el ingreso de la información directamente en el aplicativo mediante la función añadir Título (Figura 5) que lleva al usuario paso a paso hasta completar la relación de variantes con localidades y el ingreso de información adicional como notas de texto, imágenes o audio.

Como novedad se destaca la posibilidad de ingresar datos nuevos a la base de datos, es decir, en cada una de las entidades es posible la creación de información que no necesariamente pertenezca al ALEC. En caso de contar con nuevos datos lingüísticos es posible generar nuevos tomos y por tanto ingresar toda la información concerniente a nuevas localidades, informantes, encuestadores y, por supuesto, variantes lingüísticas y etnográficas con todas las herramientas de consulta y análisis.



The screenshot shows the 'Titulo' section of the SIG ALEC-ADMIN system. At the top, there is a navigation bar with 'SIG ALEC-ADMIN' and a user profile 'Javier Fernandez'. Below the navigation bar, the title 'Titulo' is displayed, along with a '+ Crear Titulo' button and a search bar labeled 'Buscar Titulo'. A vertical sidebar on the left contains icons for home, document, user, and grid. The main content area features a table with the following data:

Id Titulo	Titulo	Nombre Científico	Nombre Francés
3001	CORTEJAR, PRETENDER		courtiser, faire la cour
3002	ANILLO DE BODA		alliance, anneau de mariage
3003	INDUMENTARIA DEL NOVIO EL DÍA DE LA BODA		tenue du fiancé le jour du mariage
3004	INDUMENTARIA DE LA NOVIA EL DÍA DE LA BODA (I)		tenue de la fiancée le jour du mariage
3005	INDUMENTARIA DE LA NOVIA EL DÍA DE LA BODA (II)		tenue de la fiancée le jour du mariage

Figura 5. Ingreso de láminas en línea

4.2. Consulta y análisis de información del ALEC

Como se observa en la Figura 5 el SIG del ALEC permite en principio la consulta de los mapas e información relacionada guardando las convenciones y distribución de la versión impresa. La búsqueda se puede realizar por Tomo, Campo Semántico, Nombre o Número de Lámina o directamente escribiendo la Variante de interés del investigador. En el mismo orden, se cuenta con la posibilidad de añadir nuevas capas con nuevos mapas o variantes del interés del investigador y activar o desactivar su visualización sobre el croquis.



Figura 5. Visualización del Mapa TEJO

Por último, en cuanto a consultas, la información de carácter adicional que acompaña los mapas como texto, audios e imágenes se puede buscar, descargar y visualizar en la plataforma (Figuras 6 y 7) mediante los iconos que aparecen encima del listado de variante al llamar a cada uno de los mapas.

Figura 6. Consulta de una partitura perteneciente al mapa JUEGOS Y CANTOS DE VELORIO



Figura 6. Consulta de un audio perteneciente al mapa JUEGOS Y CANTOS DE VELORIO

5. CONCLUSIONES

A manera de conclusiones se pueden sugerir los siguientes puntos, divididos entre el conocimiento del Atlas Lingüístico-Etnográfico de Colombia (ALEC), la elaboración de una Base de Datos Espacial (BDE) y la creación de un Sistema de Información Geográfica (SIG).

Con respecto al conocimiento del Atlas Lingüístico-Etnográfico de Colombia (ALEC):

- Las labores realizadas por los investigadores durante la época de trabajo de campo sirven como evidencia de la riqueza lingüística del español popular en el territorio colombiano.
- Los materiales obtenidos y publicados en los que se consolidó el trabajo del ALEC se mantienen vigentes todavía en tanto posibilitan la investigación de la lengua y el desarrollo de herramientas amparadas en los avances tecnológicos.

Desde la elaboración de la Base de Datos Espaciales (BDE):

- Fue de suma importancia realizar la revisión de los materiales del ALEC para el establecimiento de los metadatos, la jerarquización de la información y la depuración de los elementos que constituirían una

base de datos consistente y sin elementos redundantes o negativos para la interrelación de la información.

Finalmente, y a partir de la creación del Sistema de Información Geográfica (SIG):

- Es importante destacar el trabajo interdisciplinar entre la lingüística y la geografía en su momento y la actual interacción entre los intereses lingüísticos y los desarrollos de las tecnologías.
- La creación de la base de datos espacial facilita la incorporación de datos al SIG al poseer la información organizada y digitalizada.
- A futuro el SIG permitirá realizar tareas de análisis, por lo tanto, se cuenta con una herramienta abierta a la actualización de la información y el desarrollo de nuevos estudios sobre el español de Colombia.
- El uso de software libre es beneficioso en tanto se trata de herramientas de fácil acceso y manejo, sin ningún costo de licenciamiento adicional.
- La información del SIG es fundamental en la creación del Atlas Web y el desarrollo de otros mapas interactivos o productos visuales que puedan implementarse en la herramienta de consulta abierta al público.

Referencias bibliográficas

Bonilla, J. (en prensa). Procesamiento informático de los materiales del Atlas Lingüístico-Etnográfico de Colombia: Modelamiento de la Base de Datos.

Burrough, P. (1986). *Principles of geographical information systems for land resources assessment*. Geocarto International.

Cuervo, R. J. Dapena, J. (1987). *Diccionario de construcción y régimen de la lengua española*. Instituto Caro y Cuervo, Bogotá.

Flórez, L. (1961). *Cuestionario para el Atlas Lingüístico Etnográfico de Colombia*. Instituto Caro y Cuervo, Bogotá.

Flórez, L. (1983). *Manual del atlas lingüístico-etnográfico de Colombia*. Instituto Caro y Cuervo, Bogotá.

García Mouton, P. (coord.). (2016), ALPI-CSIC [www.alpi.csic.es], edición digital de Navarro Tomás (dir.), *Atlas Lingüístico de la Península Ibérica*. CSIC, Madrid.

Güting, R. H. (1994). An introduction to spatial database systems. *The VLDB Journal - The International Journal on Very lArge Data Bases*, 3(4), 357-399.

Instituto Caro y Cuervo. (1982). *Atlas lingüístico-etnográfico de Colombia*. Instituto Caro y Cuervo, Bogotá.

Lamigueiro, O. P. (2014). *Displaying time series, spatial, and space-time data with R*. CRC Press.

Nadkarni, P. M. (2011). What Is Metadata. En: *Metadata-driven Software Systems in Biomedicine* (pp. 1-16). Springer London.

Pérez Matos, N. (2006). *De la asignación bibliográfica a la asignación de metadatos: un llamado al orden*. *Acimed*, 14(6).

Rigaux, P., Scholl, M., & Voisard, A. (2001). *Spatial databases: with application to GIS*. Morgan Kaufmann.

Silberschatz, A., Korth, H. F., Sudarshan, S., Pérez, F. S., Santiago, A. I., & Sánchez, A. V. (2002). *Fundamentos de bases de datos* (No. 04; QA76. 9. D3, S5y 2002.). McGraw-Hill.