**ALTERNATIVA LIBRE PARA LA RECOLECCIÓN DE DATOS DE CAMPO PROPUESTA PARA EL GRUPO DE INVESTIGACIÓN GEOMATICA**

El presente trabajo de investigación es elaborado por Daniel Carvajal Patiño y dirigido por Gustavo Diaz en el marco del 8° Diplomado DUNTI.

**RESUMEN**

Es bien sabido que para realizar una buena investigación debe realizarse una buena toma de datos, los datos erróneos o mal tomados pueden llevar la investigación por mal camino, hay una gran variedad de metodologías o protocolos para realizar esta tarea, sin contar los miles herramientas que ayudan en la misma. El presente se trabajó plantea la búsqueda, elección y configuración de una alternativa software para ArcGIS, un software de recolección de datos utilizado actualmente por el grupo de investigación Geomatica. El cual presenta una alta tasa de elaboración de proyectos que requieren de trabajo de campo, por ende, necesita una metodología eficiente para la recolección de los datos que se usaran en el desarrollo de los proyectos. Para Geomatica ‘ArcGIS’ es una herramienta que presenta características poco favorables, por esto se desea cambiar dicho software por uno que permita un mayor control.

**PALABRAS CLAVE**

Recolector de Datos, Software Libre, Datos.

**INTRODUCCIÓN**

Para cualquier grupo de investigación la recolección datos es un proceso muy importante1, ya que el análisis, uso o tratamiento de los mismos ayudará en el alcance de los objetivos de las investigaciones que se realicen. Por ende, deben tener una metodología o herramientas que permitan la captura de datos de una manera eficiente. Cabe resaltar que no todas las investigaciones manejan una misma forma de recolección, las investigaciones sobre el espacio y las investigaciones en microbiología exigen métodos y herramientas distintas. En lo que a este trabajo compete y dado que Geomatica es un grupo orientado a hacer proyectos con “Trabajo de campo”2, se hablará e investigará sobre herramientas software para la recolección de datos en campo. Respecto a metodologías3 no se indagará mucho dado que para este tipo de proyecto se debe salir a campo a hacer la recolección, ya sea recolectando llenando formularios, realizando encuestas, entre otros. Actualmente Geomatica usa un software instalado en dispositivos portátiles que facilita la recolección. Lo que requiere es encontrar una alternativa que les presente más autonomía y control dado que ArcGIS no es una herramienta libre lo que representa un control muy básico, casi nulo sobre la herramienta y sus componentes. Para Geomatica seria ventajoso adaptar dichas funcionalidades a sus necesidades. Otra característica desfavorable de ArcGIS es que requiere de un pago, el cual Geomatica preferiría usar en desarrolladores que mejoren la herramienta libre que se planea buscar.

**OBJETIVOS**

**OBJETIVO GENERAL:**

* Encontrar un software que sirva como alternativa a ArcGIS Collector, el programa actual que usa Geomatica para la recolección de datos en las salidas de campo.

**OBJETIVO ESPECIFICO:**

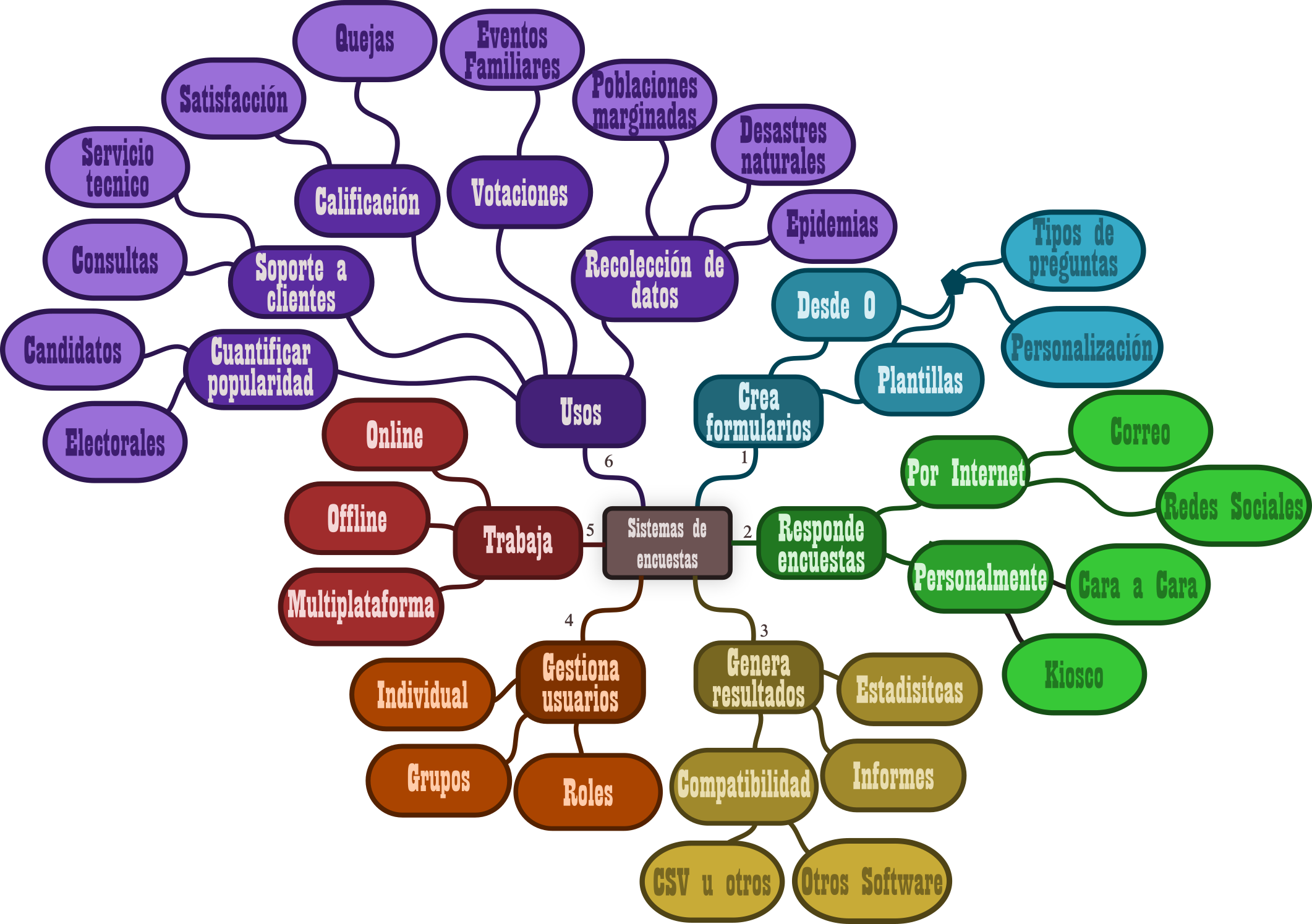
* Determinar los requerimientos que debe cumplir el software para ser una alternativa válida a ArcGIS Collector.
* Consultar las distintas herramientas existentes en el mercado para la recolección de datos a gran escala y que cumplan con los requisitos definidos.
* Seleccionar una de las distintas herramientas consultadas teniendo en cuenta las ventajas y desventajas presentada por cada una de ellas.
* Instalar la herramienta seleccionada en el servidor del grupo de investigación Geomatica.

**DESARROLLO**

**MARCO TEÓRICO**

Como se ha mencionado anteriormente los datos son una parte importante en cualquier tipo de proyecto o investigación, y cada tipo de investigación tiene sus propios métodos y herramientas para esta tarea. Para los proyecto que requieren trabajo de campo, es decir, proyectos o investigaciones que requieren de “acciones encaminadas a obtener en forma directa datos de las fuentes primarias de información, es decir, de las personas y en el lugar y tiempo en que se suscita el conjunto de hechos o acontecimientos de interés para la investigación”1,, es necesario tener herramientas que faciliten el trabajo pues en ciertas ocasiones se deben abarcar grandes cantidades de terreno y cargar con un montón de papel, como se solía hacer, es muy engorroso. Por fortuna y gracias a los avances de la tecnología, esta recolecta se puede realizar sobre los dispositivos móviles utilizando software especializado.

Existe una gran variedad de software especializado en la recolecta de datos, y como todos los softwares, unos más específicos que otros, unos con más o distintas funcionalidades que otros, unos más conocidos que otros. En general todos estos recolectores de datos presentan un árbol de funcionalidades de la siguiente manera.

****

**Figura 1.** Mapa general de funcionalidades para un sistema de encuestas o recolector de datos mediante encuestas.

**CONSULTA DEL MERCADO**

Se puede asegurar que existe una gran cantidad de software para la recolección de datos dado que se realizó una búsqueda que así lo demostró. Por mencionar algunos. Kobotoolbox, Portaldeencuestas, Surveymonkey, Eencuestas, Evalmaster, Google form, Quicktapsurvey, Suonper, ONA, Rotatorsurvey, Inty360, Moreapp, Formhub, Smap, Mfieldwork, ODK, Datawinners, ArcGIS Collector y más. La búsqueda se realizó sin ningún criterio de filtrado, todas las herramientas mencionadas poseen algunas de las funcionalidades mostradas en el árbol anteriormente descrito.

Pero no todas tienen las funciones que requiere Geomatica, y llegados aquí surge una inquietud, ¿Que requiere Geomatica? Mediante distintas reuniones con ingenieros del grupo se obtuvieron los siguientes requisitos.

1. Recolección de datos de manera offline.
2. Observación de los datos recogidos en tiempo real, para evitar errores en los mismos.
3. Toma de posición de los datos y la visualización de los datos en un mapa.
4. Poseer control sobre la aplicación. Con control se refiere a poder ingresar al código de la aplicación y cambiarlo si es necesario, para que de esta manera se pueda optimizar el proceso.

Teniendo en estos criterios y en especial el 4°, se realizó un proceso de selección entre las distintas herramientas consultadas.

**COMPARACIÓN**

Como primer filtro se usaron los 3 primeros requisitos, para ser más específicos aquellas herramientas que no cumplieran con el primer requisito, ya que, al tratarse de trabajos de campo, la funcionalidad offline es lo más importante. Con este primer filtro quedan las siguientes herramientas: ONA, Rotatorsurvey, Inty360, Moreapp, Formhub, Smap, Mfieldwork, ODK, Datawinners, Kobotoolbox. Se prosigue entonces a hacer otra comparación entre ellas.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Software | Encuestas | Offline | Análisis | Open Source | Funciones de Pago | Pago | Plataformas | Lenguaje |
| ONA | **✔** | **✔** | **✔** | **BSD** | **✔** | **Mensual** | **WEB, Linux, Mac** | **Python** |
| Rotatorsurvey | **✔** | **✔** | **✔** | **✘** | **✔** | **Anual** | **Windows, Android** | **✘** |
| Inty360 | **✔** | **✔** | **✔** | **✘** | **✔** | **Mensual** | **Windows, Mac, Android, iOS** | **✘** |
| Moreapp | **✔** | **✔** | **✔** | **✘** | **✔** | **Créditos** | **WEB** | **✘** |
| Formhub | **✔** | **✔** | **✔** | **BSD** | **✘** | **✘** | **Linux** | **Python – JS** |
| Smap | **✔** | **✔** | **✔** | **GPL** | **✘** | **✘** | **Linux** | **JS – Java** |
| Mfieldwork | **✔** | **✔** | **✔** | **✘** | **✔** | **Créditos** | **WEB** | **✘** |
| ODK | **✔** | **✔** | **✔** | **Apache** | **✘** | **✘** | **Linux, Mac, Windows** | **Java – JS** |
| Datawinners | **✔** | **✔** | **✔** | **✘** | **✔** | **Anual** | **WEB** | **✘** |
| Kobotoolbox | **✔** | **✔** | **✔** | **AGPL** | **✘** | **✘** | **Linux** | **Python** |

**Tabla 1.** Comparación de las funcionalidades de varias herramientas recolectoras de datos.

Al ser requerido un control sobre la herramienta y este control solo puede ser posible si la herramienta es “Libre”4, la primera selección queda de la siguiente manera.

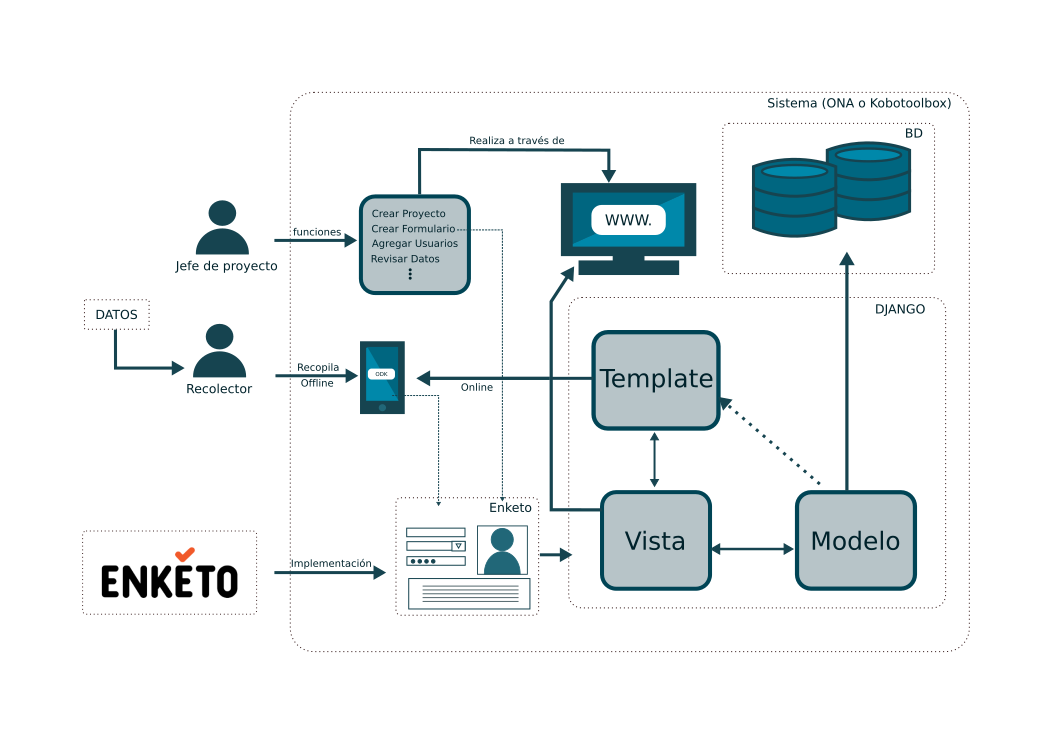
|  |  |  |  |
| --- | --- | --- | --- |
| Repositorio | Lenguaje | GitHub | Ultimo commit |
| [ONA](https://github.com/onaio/onadata) | Python - JS | ✔ | Hace 3 días |
| [Formhub](https://github.com/SEL-Columbia/formhub) | JS - Python | ✔ | 15-abr-15 |
| [Smap](https://github.com/smap-consulting/smapserver) | JS - Java | ✔ | Hace 4 días |
| [ODK](https://github.com/opendatakit/collect) | Java - JS | ✔ | Hace pocas Horas |
| [Kobotoolbox](https://github.com/kobotoolbox) | Python - JS | ✔ | Hace pocas horas |

**Tabla 2.** Lista de herramientas recolectoras de datos libres. **Hoy:** 17 – Marzo - 2017

En este punto se realizó un filtro basado en la vigencia del software y en su lenguaje. Primero dado que un proyecto vigente presenta mayor compatibilidad con las nuevas tecnologías que uno abandonado. Como lenguaje de la herramienta se escogió Python, dado que es un lenguaje que ha tomado gran importancia en los últimos años y por qué varios de los ingenieros de Geomatica prefieren este lenguaje. De esta manera quedan dos herramientas a escoger ONA o Kobotoolbox. Cabe resaltar ambas herramientas cumplen con los 4 requisitos que solicita Geomatica.

**SELECCIÓN DE HERRAMIENTA**

Se debe escoger alguna de estas dos herramientas para su instalación. Para ello se realizó una consulta minuciosa en cada uno de los repositorios de dichas plataformas, ambos disponibles en GitHub. Se revisaron los contribuidores, la licencia, los Issues, el método de instalación, las dependencias, la comunidad, los commits, entre otras cosas. De esto se concluye que ambos proyectos están muy relacionados entre sí y esto hace difícil la decisión a tomar. Ambas están siendo desarrollados muy de la mano, pues se encontraron conexiones y relaciones entre ellas, por ejemplo, el recolector de datos de Kobotoolbox está basado en onadata, el cual es el recolector de datos de ONA, ya que ONA es la empresa detrás de onadata. En los repositorios de ONA existen Forks de los repositorios de Kobotoolbox lo que podría significar algún tipo de contribución de ONA a Kobotoolbox, también ambos proyectos comparten contribuidores. Respecto a Issues y commits la cosa no ayuda pues en ciertos aspectos onadata presenta ventaja, mientras que en las otras Kobotoolbox es quien gana, incluso ambas herramientas funcionan en una instalación típica o en Dockers. Cabe resaltar, nuevamente, que esta información fue tomada de lo que mostraban los repositorios de GitHub de ambos proyectos.

Ya que el desarrollo de ambas herramientas va muy de la mano queda comparar su estructura, instalación y funcionamiento. Al estar desarrolladas en Python, más específicamente en Django, ambas cuentan con la siguiente estructura.

**Figura 2.** Estructura del funcionamiento de onadata y Kobotoolbox.

Es necesario hacer énfasis en una de las tantas similitudes que presentan onadata y kobotoolbox y es el uso de Enketo,ya que esta librería presenta una gran variedad de opciones y robustez a la hora de hacer toma de datos offline.

**INSTALACION**

Hay que comprobar la facilidad de instalación, y la documentación al respecto. Onadata presenta un proceso de instalación complicado y una documentación incompleta y desactualizada para dicho proceso, a tal punto que existen varias guías de terceros con pequeñas modificaciones en cada caso, e incluso se realizaron pequeñas modificaciones no documentadas en el proceso de instalación. Para la instalación de onadata se tuvieron en cuenta las recomendaciones de las distintas guías disponibles5. En caso contrario Kobotoolbox6 es muy sencilla de instalar un par de líneas en la terminal, una buena conexión a Internet, un par de minutos o una hora y listo. Las instalaciones anteriormente mencionadas se realizaron en local, ya que al servidor solo se montará una de ellas.

**FUNCIONAMIENTO**

Comparando el funcionamiento y experiencia de usuario, Kobotoolbox es la mejor opción. El funcionamiento de onadata se puede probar creando una cuenta en la página web [https://odk.ona.io](https://odk.ona.io/) y el funcionamiento de Kobotoolbox se puede comprobar creando una cuneta en la página web [https://kf.kobotoolbox.org](https://kf.kobotoolbox.org/). Las cuentas creadas en las plataformas web son cuentas limitadas. Onadata Classic (que es la herramienta instalada, pues existe una opción con muchas más funcionalidades disponible en la web [https://ona.io/](https://ona.io/home/), pero no es útil para Geomatica dado que tendrían que incurrir a gastos, no de licencia pero si de servicios) presenta una gran desventaja frente a Kobotoolbox puesto que no deja de ser un CRUD para encuestas XLSFORM hechas a mano en Excel o cualquier hoja de cálculo. La estructura mencionada anteriormente la cumple onadata (ona.io), pero no la cumple onadata classic (odk.ona.io). En cambio, kobotoolbox si presenta las mismas funcionalidades tanto la versión que se instaló como la versión en la plataforma oficial, a excepción de un cambio en la GUI.

Por estas razones se elige instalar Kobotoolbox en los servidores de Geomatica. Y aquí surgen nuevos inconvenientes. Kobotoolbox presenta tres maneras para ser instalada en un servidor, dockers, manualmente o usando una máquina virtual con vagrant(Usada para las pruebas). La instalación de Dockers y la Manual se descartan debido a su complejidad, dado que para ambas se requiere de conocimientos profesionales en servidores y en Linux. La instalación mediante Vagrant y máquina virtual es las más sencilla de realizar, pero presenta un inconveniente y es que kobotoolbox ya no le da soporte a esta metodología. La documentación y foros sobre esta metodología en la actualidad son nulas.

**COMPARACION CON ARCGIS**

Igual que con todas las demás herramientas ArcGIS cuenta con una plataforma Web donde se pueden observar una breve explicación de sus funcionalidades. Para empezar ArcGIS es todo un “ecosistema” de aplicaciones, Apis, funcionalidades, cursos, personas, etc. que comparados con Kobotoolbox, se queda pequeño. Para empezar ArcGIS cuenta con una mayor cantidad de personas desarrollando y dando soporte, y esto se puede ver en el GitHub de ArcGIS. Como segundo punto de comparación. Kobotoolbox y ArcGIS tienen un enfoque diferente. Kobotoolbox es una herramienta para la recolección de datos y ArcGIS es una herramienta cartográfica, cabe resaltar que en el mundo de ArcGIS existe también una herramienta recolectora de datos, la cual es libre de usar, ArcGIS Collector. En otras palabras, ArcGIS no cobra por la recolección de los datos en sí, si no por el tratamiento o uso de estos datos en su plataforma y con las distintas funcionalidades sobre mapas que presenta, pues como herramienta cartográfica presenta una amplia cantidad de funciones sobre mapas. ArcGIS en el mercado es comparado con Google Maps, ya que con ambas herramientas se pueden crear distintos tipos de mapas.

**INSTALACIÓN EN SERVIDOR**

KoboToolbox debe estar disponible para su uso por personal de Geomática, por ende, se realiza su instalación en uno de los servidores de Geomática. Actualmente existen dos maneras de instalar KoboToolbox en servidores. La primera es configurando los repositorios de KoboToolbox uno a uno, la segunda es usando Docker.

La instalación directa de los repositorios presenta una alta dificultad debido a la falta de documentación respecto a las herramientas, frameworks o tecnologías que utiliza para su desarrollo. La instalación con Docker7 resulta más sencilla, incluso no se requieren conocimientos amplios sobre Docker para conseguir una instalación funcional. Esto se hace posible gracias a la buena documentación que hay respecto al uso de KoboToolbox con Docker, incluso existe una wiki8 en donde se expone una guía de instalación. Por su facilidad y soporte se usa la instalación mediante Docker.

KoboToolbox queda disponible en la red de la Universidad industrial de Santander, exactamente en la dirección **XXXX**, donde se puede acceder y consumir la aplicación.

**PERSONALIZACIÓN**

La personalización de la instancia de KoboToolbox en la red de la UIS, como software libre que es, sigue siendo posible a pesar de estar sobre Docker, obviamente esto requiere un conocimiento más amplio sobre dicha herramienta dado que hay que usar comandos, modificar archivos de configuración (*compose.yml*) entre otras tareas específicas de la herramienta Docker.

Profundizando un poco más en la personalización o modificación de KoboToolbox sorbe Docker, si se realiza la instalación tal cual la guía, es posible hacer modificaciones al aspecto visual de la aplicación. Si se entra al directorio *kobo-docker* y se revisan los directorios ocultos existe uno llamado *vols* en donde están los archivos “static” de KoboToolbox, es decir, los referentes los estilos, logos, imágenes, entre otros. Al modificarlos la imagen de KoboToolbox los reconocerá sin la necesidad de reconstruir dicha imagen. Sin embargo, el código fuente de KoboToolbox no se encuentra disponible para su modificación dentro de este directorio. Para poder lograr esto, antes de ejecutar el comando *build* en la instalación de KoboToolbox se debe modificar el archivo *compose.yml* descomentando la línea **# build: ../kobocat** (se deben descomentar 3 líneas en total, una por cada uno de los repositorios de KoboToolbox). Una vez descomentadas las líneas necesarias se prosigue con la instalación. De esta manera el código fuente de la aplicación queda accesible en el directorio oculto *vols.* KoboToolbox también presenta una herramienta para hacer “debug” al código fuente como se enuncia en su repositorio9.

Otra manera más profesional de hacer modificaciones al código de KoboToolbox es reconstruyendo las imágenes de cada Docker (kobocat, kpi, nginx, etc) con las nuevas modificaciones. Lo que implica un conocimiento muy amplio sobre Docker.

Cabe resalta que la modificación del código fuente implicara nuevas tareas debido en manejo de versiones con los repositorios de KoboToolbox excepto de que realice un desligue total con estos, lo cual tampoco es recomendado. También hay que tener en cuenta las cláusulas de la licencia de la aplicación respecto a modificaciones.

**CONCLUSIONES**

* Se realizaron varias entrevistas con ingenieros del grupo logrando una lista completa de requerimientos con los cuales fue posible escoger la herramienta más óptima para Geomatica.
* Se realizo la investigación y arrojo una lista de distintas alternativas para el software ArcGIS.
* Se decide usar Kobotoolbox debido a su ventaja sobre las demás herramientas consultadas.
* Se realizo la configuración e instalación de KoboToolbox en uno de los servidores de Geomática, estando disponible en la dirección **XXXX.**

**REFERENCIAS**

1. SOTO-LESMES Virginia Inés, DURAN María Mercedes de Villalobos. El trabajo de campo clave en la investigación cualitativa. AQUICHAN VOL. 10 No 3 - CHÍA, COLOMBIA - DICIEMBRE 2010 p. 253-266.
2. OMONTE Rivero Abraham. Ciencias Sociales y administrativas. Investigación académica. [en línea] <<http://www.mailxmail.com/curso-ciencias-sociales-investigacion-administrativas-academica/trabajo-campo-concepto-diseno>> [citado: 4-Nov-17]
3. UNIVERSIDAD nacional autónoma de México. Aspectos teóricos de la medición y métodos de recolección de datos. [en línea] <<http://www.psicol.unam.mx/Investigacion2/pdf/METO6F.pdf>> [citado: 4-Nov-17]
4. Free Software Foundation. ¿Qué es el software libre? [en línea] <<https://www.gnu.org/philosophy/free-sw.html>> [citado: 4-Nov-17]
5. Guais de Instalación onadata. <<https://github.com/badili/classic_ona/blob/master/installation-configuration.md>>, <<https://gist.github.com/oguya/8c7b81634d648500a8c00400abab327e>>, <<https://github.com/bibbox/app-azizi-amp>>, <<https://api.ona.io/static/docs/install.html>> [en línea] [citados: 04-Nov-17]
6. Guías de instalación Kobotoolbox [en línea] [citado: 4-Nov-17] <<http://support.kobotoolbox.org/customer/portal/articles/1691105-using-kobotoolbox-offline>>.
7. What is Docker [en linea] [citado: 26-Feb-18] <<https://www.redhat.com/es/topics/containers/what-is-docker>>.
8. Kobo Wiki [en linea] [citado: 26-Feb-18] <<https://www.kobowiki.org/index.php?title=KoBo_Toolbox_on_a_standalone_server>>
9. Debug Kobo [en linea] [citado: 26-Feb-18]< <https://github.com/kobotoolbox/kobo-docker#troubleshooting>>