

# NAUTIA

Guillermo Sánchez Gutiérrez-Cabello

Abril 2020

## 1 Plataforma UPM sobre refugiados y NAUTIA

A pesar de que Naciones Unidas establezca en 2018, en la Agenda 2030, los Objetivos del Desarrollo Sostenible (ODSs), solo en ese mismo año hubo 70.8 millones de personas forzadas a abandonar sus lugares de residencia.

El contexto histórico actual arroja datos alarmantes, pues 37.000 personas se ven obligadas diariamente a huir. De ellas, más de 26 millones de personas son reconocidas como refugiados. Según el Alto Comisionado de la Naciones Unidas para los Refugiados (UNHCR, por sus siglas en inglés) en 2019, la mayoría de esas personas eran menores de 18 años y 28.000 eran niños desacompañados. Todas ellas se han visto obligadas a emigrar por diversas causas como la pobreza, la violación de los derechos humanos, el cambio climático, desastres naturales, por fuertes crisis políticas y sociales o por guerras. Paradójicamente, muchos de estos desplazamientos se realizan hacia poblaciones más pobres que de las que provienen y, desgraciadamente, la mayoría de las comunidades de acogida para estos refugiados también están envueltas en crisis económicas y sociales. Por todo ello, resulta imperativo dar soporte a la integración entre campos de refugiados y comunidades de acogida.

Con la integración entre ambas comunidades como idea principal, se desarrolla la Declaración de Nueva York sobre Refugiados, la cual pone en marcha un Marco Integral de Respuesta de Refugiados (CRRF, por sus siglas en inglés). En el ámbito de esta idea **integrada** nace la Alianza Shire, primera alianza multiactor de acción humanitaria en España en la que participan diversas entidades del sector público, privado y universitarios. Entre ellas, se encuentra la Plataforma UPM sobre Refugiados, que surgió como un grupo interdisciplinar para, precisamente, dar apoyo a la Alianza Shire.

Los objetivos de la plataforma se encuadran **entorno** a la mejora de la calidad de vida de los refugiados mediante la intervención con proyectos tecnológicos y del desarrollo de una metodología interdisciplinar que permite realizar el diagnóstico de soluciones integrales. Así nace NAUTIA (Need Assessment under a Technological Interdisciplinary Approach). NAUTIA es una metodología **dedicada** a identificar de manera cuantitativa las necesidades básicas de los habitantes residentes en los campos de refugiados y comunidades de acogida para poder mejorar su calidad de vida a través de propuestas basadas en la tecnología.

Esta metodología se puso en práctica en el campo de Refugiados de Shimelba, en la región de Tigray (Etiopía), en octubre de 2018 y en dos campos saharauis: Smara y Aiun.

La aplicación de esta metodología ha puesto en evidencia lo complejo que es realizar dicha aproximación integral e interdisciplinar y, por lo tanto, lo necesario de implementar una herramienta que automatice dicho proceso. En concreto, en este trabajo, se explica el desarrollo de una gran **parte la segunda** versión de la herramienta, la cual está centrada en la automatización de la recogida de datos y detección de necesidades de campos y comunidades con características similares a Shimelba y Tigray. Por lo tanto se aborda el desarrollo casi completo del *backend*, dejando sin desarrollar toda la parte de aplicación web y móvil para que pueda ser usable por el usuario. Esta herramienta se denomina como NAUTIA Tool Kit y sentará las bases para que en una versión posterior evolucione y pueda ser escalable a cualquier contexto, no solo africano, sino a campos de refugiados con localización en Europa, por ejemplo.

## 2 Descripción de datos de origen y definición de requisitos

Como se ha especificado anteriormente, la herramienta ya cuenta con una primera versión. Esta versión, si bien no está **para nada** automatizada, sirve muy bien para entender el propósito y necesidad de automatizar la metodología NAUTIA. La manera de trabajar con esta novel versión de *NAUTIA Tool Kit* sigue los siguientes pasos:

1. Proceso de entrevistas, cuestionarios y trabajo de campo para levantar datos en campos de refugiados y comunidades de acogida. Todo se realiza utilizando el entorno *Open Data Kit (ODK)* y, de manera más concreta, la aplicación móvil *ODK Collect*.
2. Proceso de pasar a mano los resultados obtenidos del primer paso a una hoja excel relativamente estructurada y con determinados cálculos ya automatizados. Es importante darse cuenta que este paso significa duplicar el trabajo, porque cada tipo de dato se replica por un lado para campos de refugiados y por otro para la comunidad de acogida.
3. Proceso de **calcula** de indicadores utilizando otra hoja excel de características similares a la anterior.
4. Proceso de generación de informe y gráficos de manera manual con todos los resultados obtenidos previamente.

Como se puede observar en este breve resumen de utilización de la **herramienta** para aplicar la metodología, es todo bastante ineficiente y muy dependiente de la acción humana, la cual suele estar sujeta frecuentemente a errores, lo que conlleva de manera inevitable no poder asegurar la ausencia de resultados erróneos y alejados de la realidad. Además, más allá de la necesidad de

que fuera una persona la que realizase el proceso entero, por la naturaleza de la metodología y los tipos de datos que se recogen y se quieren obtener, al carecer de automatización, es necesario que la persona que realiza este proceso tenga un perfil técnico y un conocimiento profundo de la metodología y la herramienta.

Bajo esta problemática, Susana Muñoz Hernández, profesora de la Escuela Técnica Superior de Ingenieros Informáticos de la UPM, directora del grupo de cooperación TEDECO y colaboradora en la Plataforma UPM sobre refugiados, propone a la plataforma el desarrollo de una beca de colaboración para poder desarrollar una herramienta informática que realmente automatice todo lo que sea susceptible de ello con respecto a la primera versión de la herramienta. Dicha beca la disfruté yo y a continuación detallo todo el proceso para entender los datos de origen y los requisitos necesarios para poder implementar la herramienta.

Es importante saber que la primera versión de la herramienta no fue desarrollada por profesionales informáticos. Aunque esta primera versión estaba bien implementada y recogía todo lo necesario para poder aplicar la metodología, tenía muchas carencias a la hora de afrontar un desarrollo informático. La trazabilidad de los datos entre distintas capas de la herramienta era baja, la estructuración de los datos en la hojas Excel en ningún momento aseguraba la integridad de los datos y, por último, no existía ningún documento de requisitos.

Con todo esto sobre la mesa, se aplicó ingeniería inversa haciendo un estudio de dichas hojas Excel y manteniendo reuniones, alguna de ellas superior a las dos horas, con las dos becarias de la plataforma que habían desarrollado de manera integral la primera versión.

Los datos nacen de los siguientes cuestionarios, los cuales se anexan a este documento, realizados con la aplicación para móviles android *ODK Collet* en terreno:

1. Busines survey.
2. Local leaders survey.
3. Entities survey.
4. General form.
5. General citizen focus group survey.
6. Women focus group survey.
7. Household survey.
8. Sanitation infrastructure survey.
9. Water infrastructure survey.
10. Waste management survey.
11. Energy infrastructure survey.

12. Shelter survey.
13. Public space survey.
14. Priorities Survey.
15. Farmyard and crops survey.
16. Comunal services survey.
17. Transport services survey.

La aplicación permite realizar y almacenar los resultados de todas estas encuestas sin necesidad de internet. Posteriormente, una vez se tiene acceso a la red, la aplicación permite subir los resultados a un servidor de ODK (*ODK aggregate*) donde éstos se pueden exportar en formato CSV o XML.

Adicionalmente a esos 17 cuestionarios, existe otra fuente de datos, la cual también se anexa al la memoria. Ésta es una bibliografía en formato de tabla Excel. Es importante destacar que en ninguno de los 17 cuestionarios, ni en la bibliografía, existe un campo donde se identifique de manera inequívoca, ni de ninguna manera posible, que comunidad o campo se está analizando. Por lo tanto, siendo este quizá el dato más importante, quedará totalmente fuera de la ETL y se tendrá que introducir a mano o mediante código. Quedando de esta manera como una de las propuestas clave para que en la siguiente versión también se modifiquen los cuestionarios.

Antes de continuar analizando la gran hoja excel que es el centro de la herramienta y de donde se extraen prácticamente todos los requisitos, es necesario comentar todos los cambios de formato que se han hecho en esta bibliografía para asegurar la ausencia de errores o duplicidad en los datos (el fichero con el registro de cambios realizados de manera detallada también se anexa):

- Reordenación de filas correspondientes a población disgregada por sexo para que tenga un orden coherente.
- Renombre de todas las filas duplicadas para hacerlas únicas y así evitar duplicidad de índices a la hora de trabajar con *DataFrames* durante el proceso ETL.
- Restricción de ciertas entradas de datos para limitar su rango. Ejemplos: campos de sí o no, campos con formato fecha o campos de regiones climáticas para evitar dos referencias distintas a la misma zona.

Posteriormente, toda esta información recabada se vuelca en en una hoja de cálculo Excel cuyo nombre es "20190822\_Edición Listado de Datos.xlsx" y se estructura de la siguiente manera

- El archivo cuenta con 12 hojas diferentes. La primera de ellas es "INSTRUCCIONES" y, la segunda, un índice que referencia al resto

- Hoja "0. Generales": Recoge todos los datos del contexto general del país con el objetivo de contextualizar la situación en el ámbito estatal.
- Hoja "0. Generales Campo Refugiados": Es la única con datos exclusivos del campo del refugiados y éstos son de carácter general.
- Hoja "1. Sociales y Económicos": Recoge, por duplicado, datos de demografía, culturales, seguridad personal, economía y de género.
- Hoja "2. Gobernanza": Recoge, por duplicado, políticas públicas (no se trata en este trabajo) y actores políticos involucrados.
- Hoja "3. Físicos y Medioambientales": recoge, por duplicado, datos de Identificación geográfica, topografía y recursos naturales.
- Hoja "4. Urbanísticos": Recoge, por duplicado, datos sobre información general, uso del suelo, viales y espacio público urbano.
- Hoja "5. Infraestructuras": Recoge por duplicado, datos de agua, saneamiento y drenaje, gestión de residuos, energía (de manera triplicada, pues se divide en los sectores residencial, comercial y servicios comunitarios) y movilidad y transporte.
- Hoja "6. Servicios": Recoge, por duplicado, datos educativos (de manera triplicada en escuela primaria, secundaria y profesional), sanitarios, socioculturales, deportivo/recreativos, religiosos, sobre cementerios y sobre tecnologías de la comunicación (TIC).
- Hoja "7. Alojamiento": Recoge, por duplicado, datos generales, de calidad constructiva y sobre el grado de satisfacción.
- Hoja "8. Seguridad Alimentaria": Recoge, por duplicado, datos de antecedentes, comida, fuentes de alimentación, ubicación física de cultivos y ganados y sobre la continuidad del acceso a los alimentos y tecnología.

Antes de continuar definiendo esta capa de la herramienta, me gustaría aclarar por que se hace inca pié en la recogida duplicada o triplicada de los datos. Pues en vez de haber un lugar donde se ponga un identificador único de la comunidad, independientemente de que sea de acogida o de refugiados, para una fila de datos existen dos columnas, una para el campo y otra para la comunidad de acogida. De esta manera, la herramienta queda totalmente limitada a que la relación de cardinalidad entre comunidades es 1:1, lo cual queda muy alejado de la realidad. Por otra parte, cuando se hace referencia, por ejemplo a los datos triplicado sobre cuestiones educativas, en vez de existir un identificador único para los centros y un campo de datos sobre que tipo de educación es, se replican las filas de datos por cada tipo de escuela. Estos son algunos de los ejemplos que ponen de manifiesto la complejidad de uso de la herramienta y la necesidad de escalarla a una con fundamentos informáticos fuertes.

Continuando con la definición de esta hoja Excel, las tablas contenidas en ella se estructuran de la siguiente manera:

- La primera columna es la categoría fuente, donde se especifica de donde provienen los datos, la mayoría de las veces es un cuestionario específico. Otras especifica "trabajo de campo" el cual engloba un conjunto de formularios distintos y, dependiendo del contexto, se refiere a unos o a otros. Pero, en ningún caso, se especifica a que cuestionario se refiere según el contexto. Por último, también existe una categoría fuente que es un cálculo, es decir, es un dato derivado de dos o más datos ya existentes.
- La segunda columna es el número de dato de contexto, su código.
- La tercera columna es la descripción del dato de contexto a recopilar.
- La cuarta columna es información del dato con respecto al campo de refugiados. A veces, de manera irregular, esta columna se divide en otros campos.
- La quinta columna es exactamente igual que la cuarta pero en referencia a la comunidad de acogida.
- La sexta columna son comentarios
- De la séptima en adelante es una guía sobre la metodología para la recogida del dato concreto.

Una vez están recogidos y representados todos los datos de esta capa de la herramienta, se realiza el cálculo de indicadores, los cuales representan la situación real y las necesidades de las comunidades. Representando, además, las brechas existentes entre comunidad de acogida y campo de refugiados.

Estos indicadores se representan en otra hoja Excel: "(NO MODIFICAR) Edición Indicadores V1.xlsx", que se estructura de la siguiente manera: **CUANDO SE DESARROLLE ESTA PARTE SE DOCUMENTARÁ**

## 2.1 Definición de requisitos

De toda esta descripción completa de la primera versión de la herramienta es de donde se extraen los requisitos de la nueva versión. Pues, de manera generalizada, esta nueva versión no incluye ninguna funcionalidad, ningún dato ni ninguna información nueva. No obstante, todo el entorno de esta nueva versión de *Nautia Tool Kit* es mucho más grande y al nivel más alto de análisis se compone de dos grandes partes: El desarrollo de la aplicación web y móvil necesario para que sea asequible a cualquier usuario y el desarrollo de la parte *backend* dedicada a todo el proceso ETL (Extracción, transformación y carga de datos en sus siglas en inglés) y muestra de indicadores, así como el diseño de APIs para que puedan ser usadas por la aplicación web. En este trabajo se aborda esta segunda parte. Los requisitos funcionales resultantes de la ingeniería inversa y referentes a la parte desarrollada en este trabajo, se expresan de la siguiente manera RFX.Y. En este código la R hace referencia a Requisitos, la F a Funcionales, la X tiene un rango de 0 a 9 y representa los grandes grupos de

datos de la herramienta y la Y para los diferentes requisitos contenidos en X. La relación entre X y grupos de datos queda de la siguiente manera:

- 0 hace referencia a Generales Campo Refugiados.
- 1 hace referencia a Generales Campo Refugiados.
- 2 hace referencia a Sociales y Económicos.
- 3 hace referencia a Gobernanza.
- 4 hace referencia a Físicos y medioambientales.
- 5 hace referencia a Urbanísticos.
- 6 hace referencia a Infraestructuras.
- 7 hace referencia a Servicios.
- 8 hace referencia a Alojamiento.
- 9 hace referencia a Seguridad Alimentaria.

De esta manera, los requisitos funcionales quedan definidos de la siguiente manera:

1. RF0.1: La herramienta permitirá almacenar el nombre del país donde se encuentran la comunidad de acogida y el campo de refugiados.
2. RF0.2: La herramienta permitirá almacenar la cantidad de población disgregada por edad y sexo de la siguiente manera. hombres y mujeres en los siguientes grupos de edad: menores de 5 años, entre cinco y 17 años, entre 18 y 59 años, mayores de 60.
3. RF0.3 La herramienta permitirá almacenar la tasa de crecimiento.
4. RF0.4: La herramienta permitirá almacenar el porcentaje de población refugiada.
5. RF0.5: La herramienta permitirá almacenar el Índice de Desarrollo Humano (IDH).
6. RF0.6: La herramienta permitirá almacenar la esperanza de vida al nacer
7. RF0.7: La herramienta permitirá almacenar las etnias existentes en el país
8. RF0.8: La herramienta permitirá almacenar las religiones existentes en el país.
9. RF0.9: La herramienta permitirá almacenar las lenguas habladas en el país.

10. RF0.10: La herramienta permitirá almacenar la población activa disgregada en los siguientes sectores: agricultura, ganadería, industria y sector servicios.
11. RF0.11: La herramienta permitirá almacenar la cantidad de población que vive en condiciones de pobreza.
12. RF0.12: La herramienta permitirá almacenar el coeficiente GINI.
13. RF0.13: La herramienta permitirá almacenar el PIB *per cápita*.
14. RF0.14: La herramienta permitirá almacenar la línea de pobreza.
15. RF0.15: La herramienta permitirá almacenar la moneda local.
16. RF0.16: La herramienta permitirá almacenar el tipo de cambio entre USD y la moneda local.
17. RF0.17: La herramienta permitirá almacenar el régimen de gobierno.
18. RF0.18: La herramienta permitirá almacenar el asentamiento de la población disgregado por población rural y población urbana.
19. RF0.19: La herramienta permitirá almacenar la densidad de población disgregada por población urbana y población rural.
20. RF0.20: La herramienta permitirá almacenar el acceso a agua potable disgregado por acceso urbano y acceso rural.
21. RF0.21: La herramienta permitirá almacenar el acceso a saneamiento mejorado disgregado por acceso rural y acceso urbano.
22. RF0.22: La herramienta permitirá almacenar el acceso a la electricidad disgregado por acceso urbano y acceso rural.
23. RF0.23: la herramienta permitirá almacenar la tarifa eléctrica nacional expresada en moneda local/kWh.
24. RF0.24: La herramienta permitirá almacenar el Mix eléctrico de generación.
25. RF0.25: La herramienta permitirá almacenar el voltaje de media tensión expresado en kV.
26. RF0.26: La herramienta permitirá almacenar el voltaje de la red de distribución expresado en V.
27. RF0.27: La herramienta permitirá almacenar la tasa de analfabetismo.
28. RF0.28: La herramienta permitirá almacenar la tasa de acceso a internet.
29. RF0.29: La herramienta permitirá almacenar la cantidad de población que vive en *slum*.



30. RF1.1: La herramienta permitirá almacenar la fecha de estabilización del campo.
31. RF1.2: La herramienta permitirá almacenar las razones de desplazamiento.
32. RF1.3: La herramienta permitirá almacenar la cantidad de movimientos secundarios (personas/año que abandonan el asentamiento).
33. RF1.4: La herramienta permitirá almacenar el la integración con la comunidad de acogida, expresada en tipos de relaciones.
34. RF1.5: La herramienta permitirá almacenar la región climática.
35. RF1.6: La herramienta permitirá almacenar la cantidad de amenazas naturales sufridas en los últimos años disgregadas por inundaciones, sequías, sismos, huracanes o tornados y plagas a cultivos.
36. RF1.7: La herramienta permitirá almacenar la cantidad de hectáreas deforestadas los últimos 10 años.
37. RF1.8: La herramienta permitirá almacenar si existen áreas de valor ambiental a menos de un día andando.
38. RF1.9: La herramienta permitirá almacenar la vegetación natural autóctona.
39. RF1.10: La herramienta permitirá almacenar los cultivos autóctonos.
40. RF1.11: La herramienta permitirá almacenar la la temperatura disgregada por temperatura máxima, temperatura mínima y temperatura media expresada de °C.
41. RF1.12: La herramienta permitirá almacenar la humedad relativa.
42. RF1.13: La herramienta permitirá almacenar la pluviometría anual disgregada en pluviometría máxima y pluviometría mínima expresada en mm.
43. RF1.14: La herramienta permitirá almacenar la irradiancia expresada en Kw/m<sup>2</sup>/día.
44. RF1.15: La herramienta permitirá almacenar la velocidad del viento expresada en m/s
45. RF1.16: La herramienta permitirá almacenar la cota de rios o lagos en un radio de 20 km expresada en m.
46. RF1.17: La herramienta permitirá almacenar el coste del litro de diesel expresado en moneda local.
47. RF1.18: La herramienta permitirá almacenar el coste de la bombona de gas expresado en moneda local.

48. RF1.19: La herramienta permitirá almacenar el coste del litro de kerosene expresado en moneda local.
49. RF1.20: La herramienta permitirá almacenar el coste del litro de etanol expresado en moneda local
50. RF1.21: La herramienta permitirá almacenar el coste aproximado de la leña/mes expresado en moneda local.
51. RF1.22: La herramienta permitirá almacenar si en el campo de refugiados existe permiso de movilidad exterior al campo.
52. RF1.23: La herramienta permitirá almacenar la distancia máxima a la que se pueden mover los refugiados.
53. RF1.24: La herramienta permitirá almacenar la posibilidad de mejoramiento de la vivienda.
54. RF2.1: La herramienta permitirá almacenar la población total disgregada por los rangos de edad menores de 5 años, entre 5 y 17 años, entre 18 y 59 años y de 60 años en adelante.
55. RF2.2: La herramienta permitirá almacenar si existen personas con capacidades reducidas.
56. RF2.3: La herramienta permitirá almacenar las religiones existentes en las comunidades.
57. RF2.4: La herramienta permitirá almacenar las distintas lenguas que se hablan en las comunidades.
58. RF2.5: La herramienta permitirá almacenar la composición familiar disgregada por número de mujeres adultas, número de hombres adultos, número de jóvenes y número de niños.
59. RF2.6: La herramienta permitirá almacenar la frecuencia de aseo personal.
60. RF2.7: La herramienta permitirá almacenar el material de limpieza de excreta.
61. RF2.8 La herramienta permitirá almacenar la sensación de seguridad femenina en el asentamiento disgregada por número de respuestas positivas en la calle durante el día, en la calle durante la noche, en zonas de aseo, en letrinas de noche, en espacios abiertos, durante la recogida de leña y en el puesto de trabajo.
62. RF2.9: La herramienta permitirá almacenar la sensación de seguridad femenina en el interior de la parcela o la vivienda disgregada por número de respuestas positivas durante el día y durante la noche.

63. RF2.10: La herramienta permitirá almacenar la identificación de lugares conflictivos para las mujeres.
64. RF2.11: La herramienta permitirá almacenar si existen comités de seguridad.
65. RF2.12: La herramienta permitirá almacenar si existen patrullas femeninas de seguridad.
66. RF2.13: La herramienta permitirá almacenar la iluminación de las letrinas públicas.
67. RF2.14: La herramienta permitirá almacenar el acceso al empleo disgregado por cantidad y sexo.
68. RF2.15: La herramienta permitirá almacenar el coste de la canasta básica expresado en moneda local/mes.
69. RF2.16: La herramienta permitirá almacenar el número de personas por principal fuente de ingresos.
70. RF2.17: La herramienta permitirá almacenar la los ingresos medios mensuales por principal fuente de ingresos expresados en moneda local.
71. RF2.18: La herramienta permitirá almacenar los ingresos medios mensuales por hogar expresados en moneda local.
72. RF2.19: La herramienta permitirá almacenar es gasto familiar mensual disgregado por gasto en alimentación, en ropa, en energía, en agua en servicios de transporte, en educación y en salud o sanidad expresado en moneda local.
73. RF2.20: La herramienta permitirá almacenar las prioridades de la población expresada en número de personas por prioridad.
74. RF2.21: La herramienta permitirá almacenar si existe mujeres en puestos de liderazgo.
75. RF2.22: La herramienta permitirá almacenar el número de hombres, mujeres y niños que realizan labores de recogida de leña.
76. RF2.23: La herramienta permitirá almacenar el número de hombres, mujeres y niños que realizan labores de cocinado en el hogar.
77. RF2.24: La herramienta permitirá calcular la sensación de seguridad en el asentamiento.
78. RF2.25: La herramienta permitirá calcula la sensación de seguridad en el interior de la parcela o vivienda.
79. RF2.26: La herramienta permitirá calcular la tasa de letrinas publicas que se separan por sexo.

80. RF2.27: La herramienta permitirá calcular la tasa de letrinas públicas iluminadas.
81. RF2.28: La herramienta permitirá calcular el ingreso medio de un hogar al mes.
82. RF2.29: La herramienta permitirá calcular el ingreso *per cápita* por día y por fuente de ingreso.
83. RF2.30: La herramienta permitirá calcular el número de personas con ingresos *per cápita* de 1,25 al día.
84. RF2.31: La herramienta permitirá calcular el gasto promedio de un hogar al mes.
85. RF2.32: La herramienta permitirá calcular la capacidad de ahorro de un hogar al mes.
86. RF3.1: La herramienta permitirá almacenar documentos con políticas públicas.
87. RF3.2: La herramienta permitirá almacenar los actores con actividad local en la zona.
88. RF4.1: La herramienta permitirá almacenar la localización geográfica del asentamiento en forma de latitud, longitud y altitud.
89. RF4.2: La herramienta permitirá almacenar la altura del nivel freático expresada en m.
90. RF4.3: La herramienta permitirá almacenar la cota superior del asentamiento expresada en m.
91. RF4.4: La herramienta permitirá almacenar la cota inferior del asentamiento expresada en m.
92. RF4.5: La herramienta permitirá calcular el desnivel del asentamiento.
93. RF4.6: La herramienta permitirá calcular el desnivel freático.
94. RF4.7: La herramienta permitirá calcular el desnivel de los lagos y ríos de la zona.
95. RF5.1: La herramienta permitirá almacenar si existe un plan urbanístico.
96. RF5.2: La herramienta permitirá almacenar la vigencia del plan urbanístico expresada en años.
97. RF5.3: La herramienta permitirá almacenar los tipos de límites del asentamiento.

98. RF5.4: La herramienta permitirá almacenar el tipo de gestión del suelo del asentamiento.
99. RF5.5: La herramienta permitirá almacenar si existe un plan de crecimiento del asentamiento.
100. RF5.6: La herramienta permitirá almacenar el número de años que prevé el plan de crecimiento.
101. RF5.7: La herramienta permitirá almacenar si existen instrumentos para una gestión de riesgos en planificación urbana.
102. RF5.8: La herramienta permitirá almacenar los tipos de desastre que afectan a la planificación urbana.
103. RF5.9: La herramienta permitirá almacenar el tamaño medio de la parcela expresado en .
104. RF5.10: La herramienta permitirá almacenar si las parcelas están delimitadas.
105. RF5.11: La herramienta permitirá almacenar la superficie de suelo privado expresada en  $m^2$ .
106. RF5.12: La herramienta permitirá almacenar si existen areas inundables o con riesgo de deslabe.
107. RF5.13: La herramienta permitirá almacenar la identificación geográfica de las areas inundables y con peligro de deslabe expresada en latitud, longitud y altitud.
108. RF5.14: La herramienta permitirá almacenar cantidad de superficie urbana inundable o con peligro de deslabe expresada en  $m^2$ .
109. RF5.15: La herramienta permitirá almacenar la longitud de los viales expresada en metros lineales.
110. RF5.16: La herramienta permitirá almacenar el sistema de drenaje de pluviales.
111. RF5.17: La herramienta permitirá almacenar el acondicionamiento del firme de los viales.
112. RF5.18: La herramienta permitirá almacenar la identificación geográfica de las áreas de esparcimiento expresada en latitud, longitud y altitud.
113. RF5.19: La herramienta permitirá almacenar la superficie de las áreas de esparcimiento expresada en  $m^2$ .
114. RF5.20: La herramienta permitirá almacenar la cantidad de áreas de esparcimiento acondicionadas disgregadas por tipo de área.

115. RF5.21: La herramienta permitirá calcular el estado del mantenimiento de los viales.
116. RF5.22: La herramienta permitirá calcular la tasa de areas de esparcimiento acondicionadas.
117. RF6.1: La herramienta permitirá almacenar la cantidad de puntos de agua en el asentamiento.
118. RF6.2: La herramienta permitirá almacenar la cantidad de puntos de agua en funcionamiento durante todo el año en el asentamiento.
119. RF6.3: La herramienta permitirá almacenar la calidad del agua durante todo el año en el asentamiento.
120. RF6.4: La herramienta permitirá almacenar los tipos de sistema de potabilización del agua.
121. RF6.5: La herramienta permitirá almacenar el consumo diario *per cápita* de agua expresado en Litros/persona.
122. RF6.6: La herramienta permitirá almacenar los tipos de fuente que permiten acceso a agua.
123. RF6.7: La herramienta permitirá almacenar el tiempo diario dedicado al abastecimiento de agua.
124. RF6.8: La herramienta permitirá almacenar los distintos tipos de sistemas de riego disgregados por operativos y no operativos.
125. RF6.9: La herramienta permitirá almacenar la potencia de la bomba del sistema de riego expresada en W.
126. RF6.10: La herramienta permitirá almacenar la fuente de energia de la bomba del sistema de riego.
127. RF6.11: La herramienta permitirá almacenar si la defecación se hace al aire libre o no.
128. RF6.12: La herramienta permitirá almacenar si las letrinas son compartidad o individuales.
129. RF6.13: La herramienta permitirá almacenar la cantidad de letrinas disgregadas por letrina con losa, letrinas sin losa y letrinas con losa y ventiladas.
130. RF6.14 La herramienta permitirá almacenar la cantidad de letrinas disgregadas por letrinas sin estructura estable, letrinas con techo y paredes estables, letrinas con techo, paredes y puerta con cierre correcto y letrinas con techo paredes, cierre e iluminación.

- 131. RF6.15: La herramienta permitirá almacenar la cantidad de puntos de recogida de residuos colectiva.
- 132. RF6.16: La herramienta permitirá almacenar la cantidad de servicios de recogida de basuras colectiva al mes.
- 133. RF6.17: La herramienta permitirá almacenar la distancia del vertedero a las viviendas.
- 134. RF6.18: La herramienta permitirá almacenar la distancia del vertedero a los recursos hídricos.
- 135. RF6.19: La herramienta permitirá almacenar si el vertedero ocupa zonas fértiles.
- 136. RF6.20: La herramienta permitirá almacenar si existe servicio eléctrico a través de la red eléctrica.
- 137. RF6.21: La herramienta permitirá almacenar si existe previsión de ampliar la red eléctrica.
- 138. RF6.22: La herramienta permitirá almacenar los beneficiarios de la demanda objetivo a cubrir disgregados por servicios y cantidad de ellos.
- 139. RF6.23: La herramienta permitirá almacenar la cantidad de cortes diarios de electricidad.
- 140. RF6.24: La herramienta permitirá almacenar si existe curva de demanda.
- 141. RF6.25: La herramienta permitirá almacenar la distancia aproximada del centro poblacional a la red eléctrica expresada en Km.
- 142. RF6.26: La herramienta permitirá almacenar los distintos tipos de sistemas de generación disgregados por sector residencial, sector comercial y servicios comunitarios.
- 143. RF6.27: La herramienta permitirá almacenar la capacidad de los sistemas de generación.
- 144. RF6.28: La herramienta permitirá almacenar las distintas fuentes de generación disgregadas por sector residencial, sector comercial y servicios comunitarios.
- 145. RF6.28: La herramienta permitirá almacenar el número de sujetos conectados a cada fuente de generación.
- 146. RF6.29: La herramienta permitirá almacenar la cantidad de horas de acceso a cada fuente de generación.
- 147. RF6.30: La herramienta permitirá almacenar los electrodomésticos disgregados por sector residencial, sector comercial y servicios comunitarios.

- 148. RF6.31: La herramienta permitirá almacenar la cantidad existente por tipo de electrodoméstico.
- 149. RF6.32: La herramienta permitirá almacenar la potencia de los electrodomésticos.
- 150. RF6.33: La herramienta permitirá almacenar la electricidad consumida por día de los electrodomésticos.
- 151. RF6.34: La herramienta permitirá almacenar el tipo de cocina predominante en el asentamiento.
- 152. RF6.35: La herramienta permitirá almacenar el consume promedio de leña
- 153. RF6.36: La herramienta permitirá almacenar el número de mujeres que cocinan dentro de casa y el número de personas que cocinan fuera de casa.
- 154. RF6.37: La herramienta permitirá almacenar las horas promedio que dedican las mujeres a cocinar diariamente.
- 155. RF6.38: La herramienta permitirá almacenar el combustible predominante en el campo.
- 156. RF6.40: La herramienta permitirá almacenar la cantidad de veces por semana que las mujeres van a recoger leña.
- 157. RF6.41: La herramienta permitirá almacenar el tiempo que tardan las mujeres en ir a recoger leña.
- 158. RF6.42: La herramienta permitirá almacenar el número de mujeres cuya salud se ve afectada por cocinar con leña.
- 159. RF6.43: La herramienta permitirá almacenar el plano del alumbrado público.
- 160. RF6.44: La herramienta permitirá almacenar el tipo de tecnología del alumbrado público.
- 161. RF6.45: La herramienta permitirá almacenar si funciona o no el alumbrado público.
- 162. RF6.46: La herramienta permitirá almacenar la localización geográfica de las luminarias expresada en latitud, longitud y altitud.
- 163. RF6.47: La herramienta permitirá almacenar la distancia entre los puntos de luz.
- 164. RF6.48: La herramienta permitirá almacenar la cantidad de mujeres que ven influenciada su seguridad por el alumbrado público.
- 165. RF6.49: La herramienta permitirá almacenar la distancia a los puntos de trabajo.



166. RF6.50: La herramienta permitirá almacenar la localización geográfica de los puntos de movilidad expresada en latitud, longitud y altitud.
167. RF6.51: La herramienta permitirá almacenar las formas básicas de transporte interno y transporte externo.
168. RF6.52: La herramienta permitirá calcular el acceso a agua y el acceso a saneamiento saneamiento.
169. RF6.63: La herramienta permitirá calcular la calidad del sistema de saneamiento.
170. RF6.64: La herramienta permitirá calcular la calidad de la letrina.
171. RF6.65: La herramienta permitirá calcular la calidad de la edificación de la letrina.
172. RF6.66: La herramienta permitirá calcular la localización del vertedero.
173. RF6.67: La herramienta permitirá calcular el número total de sujetos con acceso a la electricidad disgregado por sector residencial, sector comercial y servicios comunitarios.
174. RF6.68: La herramienta permitirá calcular la energía consumida de promedio por un hogar, por un negocio y por servicios comunitarios en un día.
175. RF6.69: La herramienta permitirá calcular el gasto promedio en electricidad por un hogar.
176. RF7.1: La herramienta permitirá almacenar el número de centros educativos disgregados por guarderías, centros de educación primaria, centros de educación secundaria y centros de educación superior.
177. RF7.2: La herramienta permitirá almacenar el número de alumnos por cada cuetro.
178. RF7.3: La herramienta permitirá almacenar el número de puestos de trabajo por cada centro.
179. RF7.4: La herramienta permitirá almacenar el número de profesores por cada centro.
180. RF7.5: La herramienta permitirá almacenar el material predominante en cada escuela.
181. RF7.6: La herramienta permitirá almacenar la carga educativa de cada escuela.
182. RF7.7: La herramienta permitirá almacenar el tipo de educación demandada por cada escuela.
183. RF7.8: La herramienta permitirá almacenar la hora de comienzo y la hora de finalización de cada escuela.

184. RF7.9: La herramienta permitirá almacenar si la cubierta del edificio de cualquier servicio tiene filtraciones.
185. RF7.10: La herramienta permitirá almacenar la seguridad estructural del edificio de cualquier servicio.
186. RF7.11: La herramienta permitirá almacenar el confort térmico del edificio de cualquier servicio.
187. RF7.12: La herramienta permitirá almacenar las amenazas climáticas a las que está expuesto el edificio de cualquier servicio
188. RF7.13: La herramienta permitirá almacenar las causa de no recibir educación.
189. RF7.14: La herramienta permitirá almacenar los tipo de centro sanitarios.
190. RF7.15: La herramienta permitirá almacenar la superficie del asentamiento que queda fuera de un radio de 2km de los centros de atención primaria.
191. RF7.16: La herramienta permitirá almacenar la superfice del asentamiento que queda fuera de un radio de 5km de los centros de atención pirmaria.
192. RF7.17: La herramienta permitirá almacenar la identificación geográfica de los centro sanitarios expresada en latitud, longitud y altitud.
193. RF7.18: La herramienta permitirá almacenar la cantidad de camas por hospital.
194. RF7.19: La herramienta permitirá almacenar la cantidad de personas con acceso a tratamientos y medicinas.
195. RF7.20: La herramienta permitirá almacenar la identificación geográfica de los centros socioculturales, de los centros deportivos, de los mercados, de las tiendas, de los centros religiosos, de los centros de comunicación e informática y de los cementerios expresada en latitud, longitud y altitud.
196. RF7.21: La herramienta permitirá almacenar la superficie del campo sin acceso a centros socioculturales, a los centros deportivos, a los mercados, a las tiendas y a los centros religiosos.
197. RF7.22: La herramienta permitirá almacenar la superficie de los cementerios.
198. RF7.23: La herramienta permitirá almacenar si existe sistema de drenaje en los cementerios.
199. RF7.24: La herramienta permitirá almacenar si la cota inferior de los cementerios está 1,5 metros por encima del nivel freático.
200. RF7.25: La herramienta permitirá almacenar los sistemas de telecomunicación que permiten la disponibilidad de acceso a datos.

- 201. RF7.26: La herramienta permitirá almacenar la cantidad de antenas repetidoras próximas al sitio.
- 202. RF7.27: La herramienta permitirá almacenar la cantidad de personas con conocimiento de tecnologías disgregada por usuario movil llamadas, uso de internet en el móvil y/o tablet, uso de PC o portátil y programación y administración de sistemas.
- 203. RF7.28: La herramienta permitirá almacenar la cantidad de personas por tipo de aplicación en uso.
- 204. RF7.29: La herramienta permitirá almacenar la cantidad de personas por tipo de aplicación que sea una necesidad.
- 205. RF7.30: La herramienta permitirá calcular la cantidad de alumnos por profesor en las escuelas.
- 206. RF7.31: La herramienta permitirá calcular el porcentaje de escuelas con buena calidad constructiva.
- 207. RF7.32: La herramienta permitirá calcular el acceso a educación.
- 208. RF7.33: La herramienta permitirá calcular el porcentaje de centros sanitarios con buena calidad constructiva.
- 209. RF7.34: La herramienta permitirá calcular la localización correcta de los cementerios.
- 210. RF8.1: La herramienta permitirá almacenar la cantidad de viviendas en el asentamiento.
- 211. RF8.2: La herramienta permitirá almacenar las estancias independientes de cada casa visitada.
- 212. RF8.3: La herramienta permitirá almacenar por cada casa si la cubierta es estanca, si tiene seguridad estructural, si el sistema constructivo es seguro ante amenazas climáticas, si todas las estancias tienen ventanas, si las paredes no tienen humedades, si no contienen animales y si la temperatura interior es menor a 30°C.
- 213. RF8.4: La herramienta permitirá almacenar el coste de la construcción en el asentamiento.
- 214. RF8.5: La herramienta permitirá almacenar las mejoras de la vivienda identificadas por la cabeza de la familia.
- 215. RF8.6: La herramienta permitirá almacenar el tipo de construcción existente en el asentamiento.
- 216. RF8.7: La herramienta permitirá calcular la cantidad de personas por vivienda.

- 217. RF8.8: La herramienta permitirá calcular la cantidad de viviendas con hacinamiento.
- 218. RF8.9: La herramienta permitirá calcular el hacinamiento de la población.
- 219. RF9.1: La herramienta permitirá almacenar la causa de la inseguridad alimentaria.
- 220. RF9.2: La herramienta permitirá almacenar los grupos de población afectados por la inseguridad alimentaria.
- 221. RF9.3: La herramienta permitirá almacenar la cantidad de personas con acceso por tipo de alimento.
- 222. RF9.4: La herramienta permitirá almacenar la cantidad de personas por cantidad de veces que se come al día.
- 223. RF9.5: La herramienta permitirá almacenar la cantidad de personas por tipo de comida principal.
- 224. RF9.5: La herramienta permitirá almacenar una lista que almacene por defecto las Kcal/100g de los platos típicos.
- 225. RF9.6: La herramienta permitirá almacenar la cantidad de personas por cada plato típico del día.
- 226. RF9.7: La herramienta permitirá almacenar la ingesta, estableciendala por defecto en 70g.
- 227. RF9.8: La herramienta permitirá almacenar la cantidad de personas por fuente principal de suministro de alimento.
- 228. RF9.9: La herramienta permitirá almacenar los meses de cultivo del asentamiento.
- 229. RF9.19: La herramienta permitirá almacenar la identificación geográfica de huertos y corrales expresada en latitud, longitud y altitud.
- 230. RF9.20: La herramienta permitirá almacenar si los huertos y corrales son públicos o privados.
- 231. RF9.21: La herramienta permitirá almacenar las técnicas de pastoreo predominantes.
- 232. RF9.22: La herramienta permitirá almacenar el sistema predominante de drenaje de corrales.
- 233. RF9.23: La herramienta permitirá almacenar el tipo de riego.
- 234. RF9.24: La herramienta permitirá almacenar si se utilizan sistemas de fertilización.

235. RF9.25: La herramienta permitirá almacenar los meses del año en los cuales el auto abastecimiento es posible.
236. RF9.26: La herramienta permitirá almacenar los tipos de alimentos de cultivo propio.
237. RF9.27: La herramienta permitirá almacenar la duración de los alimentos no perecederos.
238. RF9.28: La herramienta permitirá almacenar las técnicas de conservación de alimentos.
239. RF9.29: La herramienta permitirá almacenar las técnicas de secado de granos.
240. RF9.30: La herramienta permitirá almacenar la potencia del motor de los molinos de granos.
241. RF9.31: La herramienta permitirá almacenar si el molino de grano funciona o no.
242. RF9.32: La herramienta permitirá almacenar si el molino de grano está bien ubicado.
243. RF9.33: La herramienta permitirá calcular el número de comidas promedio al día.
244. RF9.34: La herramienta permitirá calcular la cantidad de calorías ingeridas por persona y día.
245. RF9.35: La herramienta permitirá calcular el porcentaje de población respecto del total por fuente principal de suministro de alimentos.
246. RFX.Y: *QUEDAN CIERTOS RF REFERENTES A LOS INDICADORES QUE SE ESPECIFICARÁN CUANDO SE DESARROLLE*

En cuanto a los requisitos no funcionales referentes a la parte desarrollada, se expresarán con el código RNF.X, siendo la X el número de requisito no funcional. Éstos son:

1. RNF.1: Precisión: La herramienta asegurará que los datos se almacenan y se calculan de manera precisa, correspondiéndose con la realidad.
2. RNF.2: Integridad de los datos: La herramienta asegurará la integridad de los datos evitando duplicidad de ellos y posibles definiciones distintas de un mismo dato.
3. RNF.3: Eficiencia: La herramienta deberá poder funcionar en dispositivos con bajos recursos.
4. RNF.4: Mantenibilidad.
5. RNF.5: Modificabilidad.
6. RNF.6: Reusabilidad.

### 3 Proceso de diseño de la base de datos

De la misma manera que se hizo para comprender el funcionamiento de la herramienta así como sus requisitos, se mantuvieron reuniones con las dos becarias que trabajaron en la primera versión de *NAUTIA Tool Kit* para entender como quedaban divididos los datos y de que manera se querían almacenar. Aunque la hoja Excel que guardaba todos los datos de contexto organizaba muy bien los grupos de datos, había algunos de ellos que no estaban bien representados. Lo más importante era darse cuenta que las comunidades de acogida y los campos de refugiados no eran dos entidades distintas con una relación de cardinalidad 1:1, pues eso es lo que se extraía de la primera versión. Ambas comunidades son, en definitiva, comunidades. Si bien, los campos de refugiados tienen una características extras que no tienen los campos de refugiados. Por lo tanto, de cara al proceso de diseño de la base de datos se propondrá una entidad padre llamada *Community* de la que heredaran dos entidades: *HostCommunity* y *Camp*. Y esta entidad padre será prácticante el epicentro de la base de datos, de la cual saldrán la gran mayoría de relaciones con todo el resto de entidades que definen la base de datos. Pues a diferencia de la primera versión donde los datos se representan por duplicado de manera ineficiente, en esta versión desaparece la necesidad de hacer dos tablas por cada infraestructura, por cada servicio, por cada urbanismo, etc., por ejemplo.

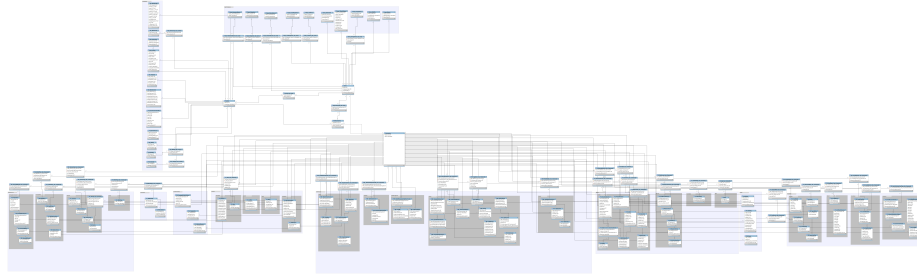
A parte de lo comentado, hay datos exclusivos del campo de refugiados, los cuales, de manera evidente solo se pueden relacionar con la entidad *Camp* y así queda reflejado en el modelo. Y, por último, existen dos datos (Religión y Lengua) que en la primera versión se representaban por duplicado tanto en los referentes a los generales del país como a datos socio económicos, aun proviniendo estos de la misma fuente de datos. Esto es un error, pues los datos socio económicos se relacionan directamente con su comunidad y esta se relaciona directamente con su país, que está relacionado con lo datos generales de contexto, donde se ha dejado de manera exclusiva los datos Religión y Lengua en esta nueva versión.

Una vez entendidas estas peculiaridades, además de la definición de los requisitos, los cuales no solo evidencian la cantidad de datos que son necesarios almacenar, sino la necesidad de que estos se ajusten a la realidad, se ha decido utilizar una base de datos relacional y, concretamente, se ha decidido desarrollarla con el gestor MySQL utilizando el editor MySQL Workbench. Los motivos han sido que es un potente software libre soportado por una entidad como Oracle y, por lo tanto, nos aporta seguridad de que lo podremos mantener en el tiempo sin problemas.

De cara al diseño, de la misma manera que la hoja Excel está dividida por grupos de datos, los datos en el modelo lógico también se dividirán por grupos y, como se verá más adelante, cada entidad en el modelo empezará por las iniciales correspondientes a su grupo de datos con el objetivo de facilitar su comprensión por parte de terceros, pues el modelo resultante consta de 153 tablas.

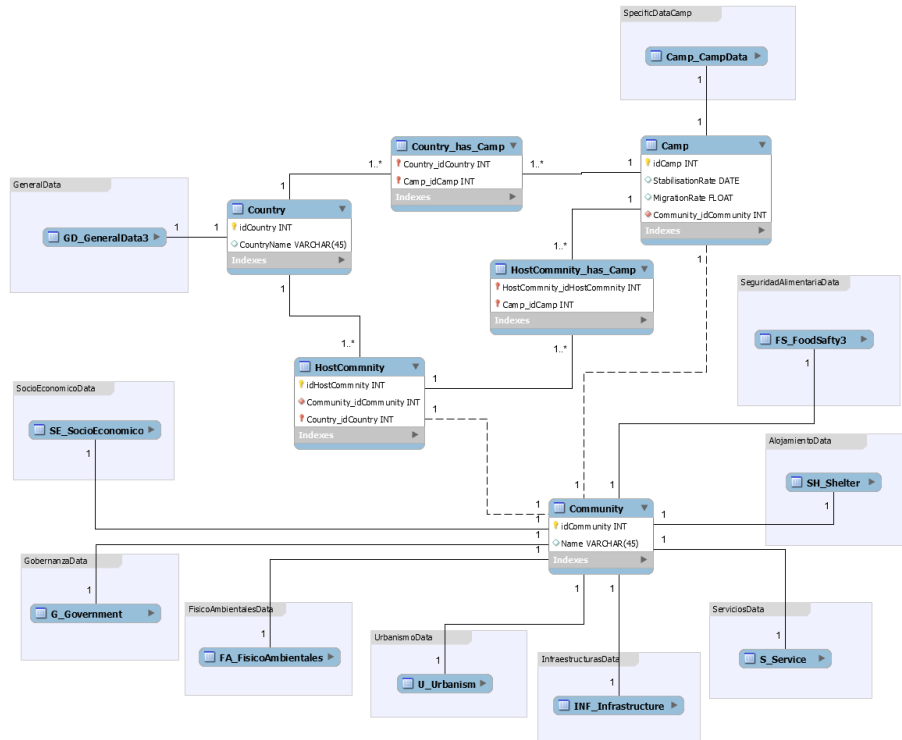
## 4 Diseño lógico propuesto de BD

Resulta imposible explicar de una manera clara el diseño lógico propuesto de una sola vez, pues el tamaño es muy grande y no se aprecia ningún tipo de detalle al verlo en una sola imagen como se puede comprobar:



No obstante, el modelo completo se anexa a este documento y tiene la suficiente calidad como para poder hacer el zoom necesario. Para poder explicarlo, se hará una abstracción de este modelo dividiéndolo en grandes conjuntos de datos y, posteriormente, profundizando grupo por grupo.

De esta manera, desde el más alto nivel, el modelo queda como sigue:



Antes de analizarlo, es necesario explicar que esta abstracción se ha desarrollado con la herramienta de MySQL para hacer modelos de entidad relación y que las cardinalidades expresadas entre tablas no siempre se ajustan a la re-

alidad, pues hay "entidades" que realmente son la representación de un gran conjunto de datos con sus peculiaridades.

Dicho esto, en esta abstracción del modelo si que hay entidades que se representan tal cual están en el definitivo. Una de ellas es la entidad *Community* que, como podemos observar, está altamente relacionada con casi todas las entidades del modelo. Como ya se ha especificado previamente, esta entidad hace la función de entidad padre de las entidades *Camp* y *HostCommunity* guardando una relación de cardinalidad 1:1 con cada una de ellas. Pues o una comunidad es un solo campo o es una solo comunidad de acogida. Entre estas dos entidades hijas, observamos que su relación de cardinalidad es de N:M. Se sabe porque la herramienta de MySQL genera una tabla intermedia llamada, *Cam.Has.Community*, con relación de cardinalidad 1:N entre *Camp* y *HostCommunity* y la entidad que las relaciona, respectivamente.

Este diseño soluciona uno de los mayores problemas que tenía la primera versión de la herramienta, pues ésta solo permitía almacenar datos de una comunidad de acogida por cada campo de refugiados, cuando es perfectamente posible que un campo tenga varias comunidades de acogida y una comunidad de acogida tenga varios campos.

Otra de las entidades que se representan en esta abstracción tal cual en el modelo definitivo es *Country*, la cual hace referencia al país al cual pertenece la comunidad de acogida por un lado, con una relación de cardinalidad 1:N, y a los países de los que provienen los refugiados de un campo de refugiados. Aunque no parezca natural una relación N:M entre un campo y un país se establece así precisamente porque la fuente de datos, que en este caso es el Excel de bibliografía, no separa los datos de procedencia por cada refugiado del campo, sino que lo generaliza a todo el campo, aunque haya refugiados de distintas nacionalidades. Aprovechando esta aclaración, es necesario decir que, si bien la herramienta representa la realidad de que los refugiados de un campo pueden proceder de diversos países, a la hora de ponerla en marcha va a ser imposible porque, en la definición de los datos de origen, lo cual queda fuera de este trabajo, eso no ocurre.

Ya han quedado explicadas y definidas todas las entidades de este modelo abstracto que se corresponden con el definitivo. En cuanto a las representaciones abstractas, si comenzamos desde la parte alta de este modelo y continuamos en sentido levógiro, en los extremos de la figuraX vemos que las entidades que nos quedan por definir a parte de representarse en formato de tabla, están contenidas en un *Layer*. Esto se hace con intención de representar los *layers* que constarán en el modelo final con el objetivo de una mayor organización y agrupación de los conjuntos de datos, así como una facilitación de la comprensión del modelo. Todas estos *layers* se corresponden con los grandes grupos de requisitos que ya han sido definidos en este trabajo y se relacionan de la siguiente manera:

- **SpecificDataCamp:** Conjunto de datos relacionados con *Camp*.
- **GeneralData:** Conjunto de datos relacionado *Country*.



- **SocioEconomicoData, GobernanzaData, FisicoAmbientalesData, UrbanismoData, InfraestructurasData, ServiciosData, AlojamientoData y SeguridadAlimentariaData:** Conjuntos de datos relacionados todos ellos con *Community*.

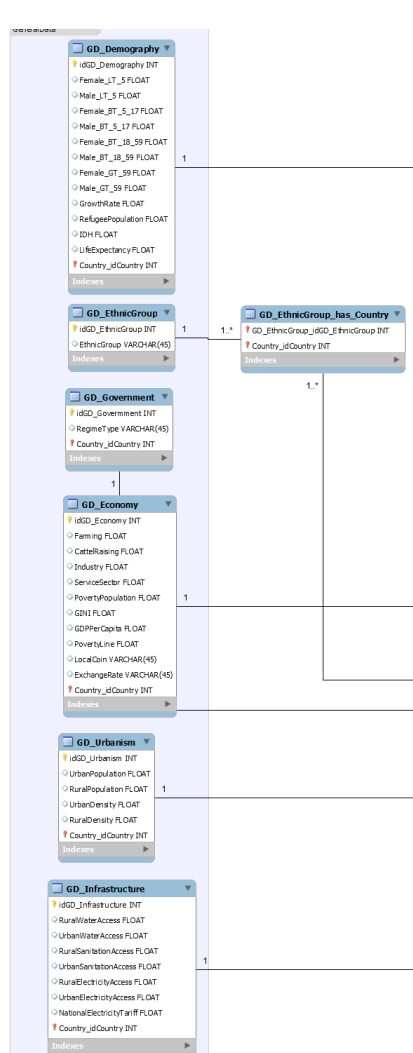
Todos estos grandes grupos son los que se irán explicando uno a uno e, incluso en la mayoría de veces, éstos se partirán por la mitad o por subconjuntos identificables de datos. En todas las siguientes imágenes veremos entidades cuyo nombre es auto contenido y sus respectivos atributos, los cuales guardan una relación directa con todos los requisitos que hacen referencia al almacenamiento de datos. Por lo tanto, lo que se explicará de las imágenes será con quien se relaciona cada tabla, pues en las figuras no va a aparecer, y la relación de cardinalidad existente entre ellas. Entre otras cosas, la gran cantidad de tablas resultantes en el modelo y la limitación de MySQL Workbench de no poder editar las flechas que unen tablas, hacen necesario explicar esta relación de cardinalidad, pues con las imágenes no queda claro.

A continuación, por lo tanto, se van a representar todos los conjuntos de datos y sus tablas. Todas ellas empezarán por un prefijo que hace referencia a su conjunto. En todo momento se especificará que conjunto y que parte del conjunto o subconjunto es. Todas las siguientes tablas se relacionan de manera directa con *Country*, *Camp* o *Community*, a las cuales las llamaremos tablas eje del modelo. Como ya se ha explicado, todas las del conjunto de datos GeneralData con *Country*, todas las de SpecificDataCamp con *Camp* y todas las restantes con *Community*.

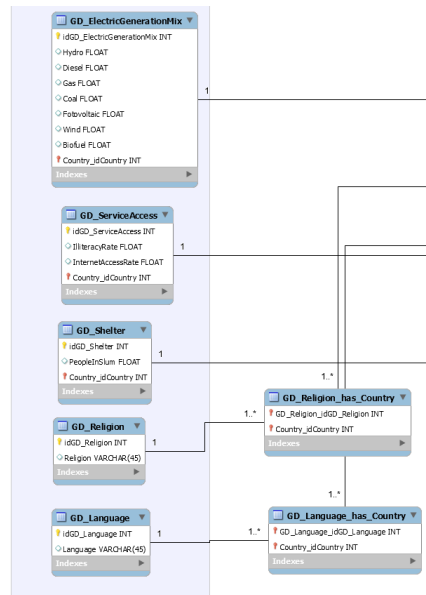
Existen 3 tipos de relaciones de cardinalidad distintas en este modelo. Cuando la relación entre la tabla en cuestión y la tabla eje sea 1:1 se podrá saber porque de la flecha naciente de dicha tabla se podrá ver un 1. Cuando la relación entre ambas sea 1:N (1 para la tabla eje y N para la tabla específica) se podrá saber porque de la flecha que nace de la tabla se verá 1..\*. Por último, cuando la relación sea N:M, esto se podrá saber porque en la imagen se verá una tabla extra con la siguiente estructura: **PREFIJO\_TablaEnCuestion\_has\_TablaEje**, siendo la relación entre la tabla en cuestión y esta intermedia 1:N y la de la intermedia con la tabla eje N:1, representando así la relación N:M y, cuando sea necesario, representando los atributos específicos de esta relación.

Si en algún caso se considera que alguna entidad o atributo necesita una explicación extra, esta será dada.

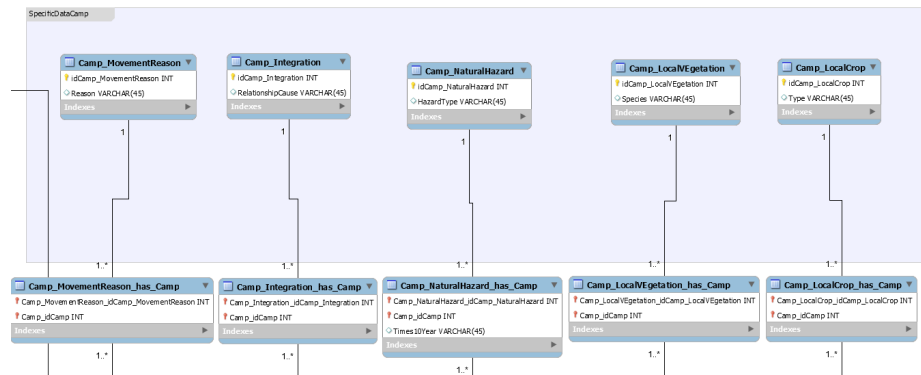
Primera parte del *layer* **GeneralData**:



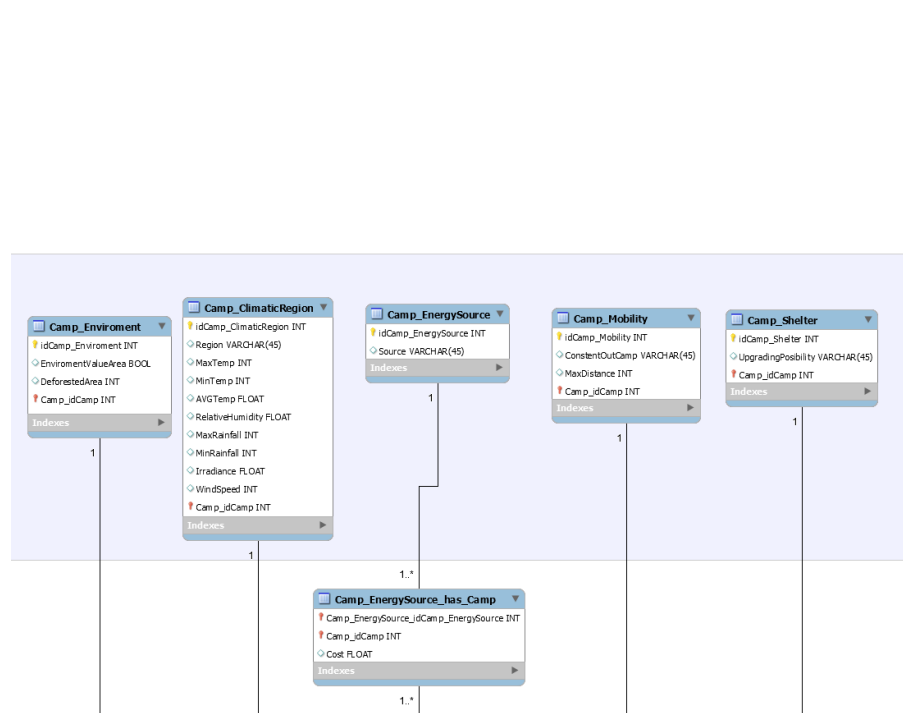
Segunda parte de **GeneralData**:



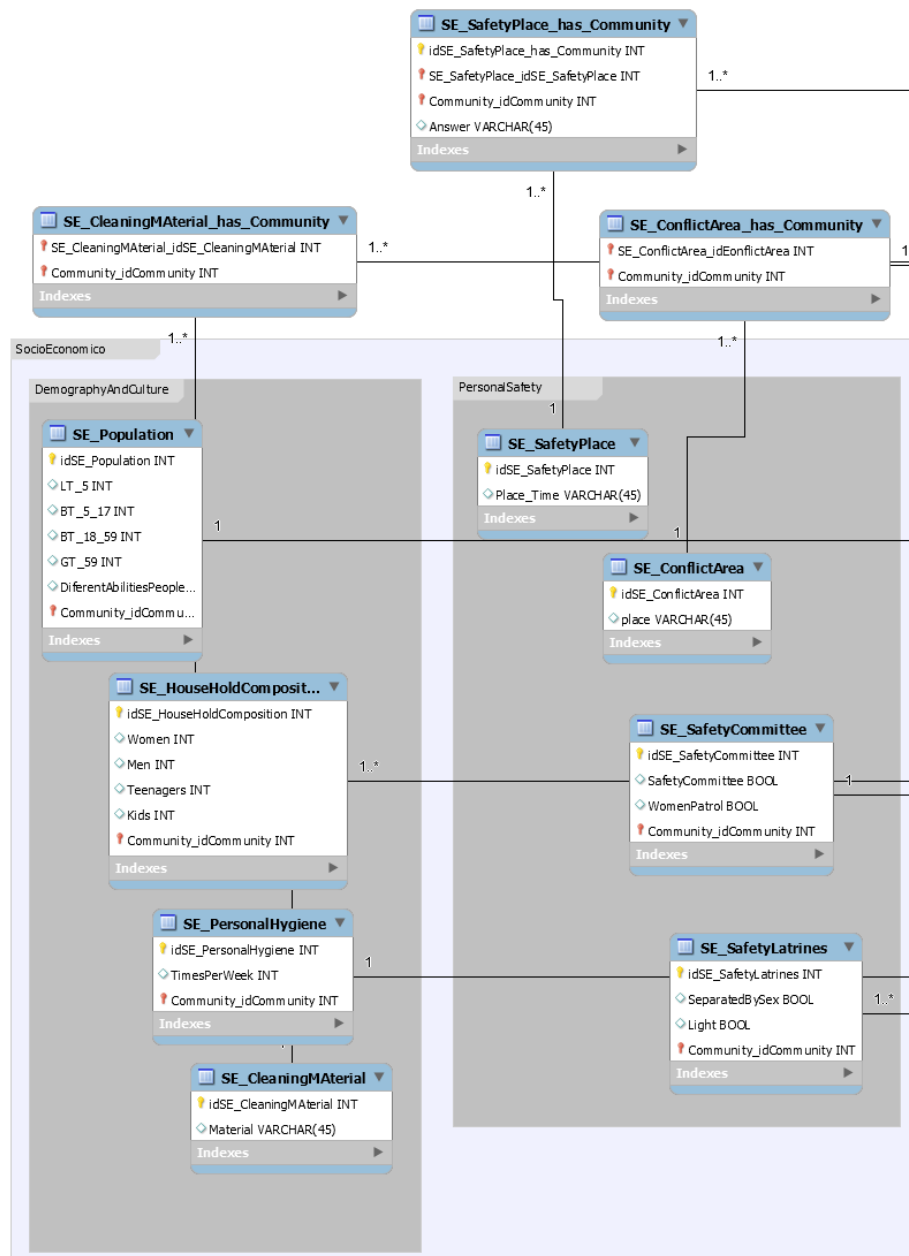
Primera parte del *Layer SpecificDataCamp*:



Segunda parte de *SpecificDataCamp*:

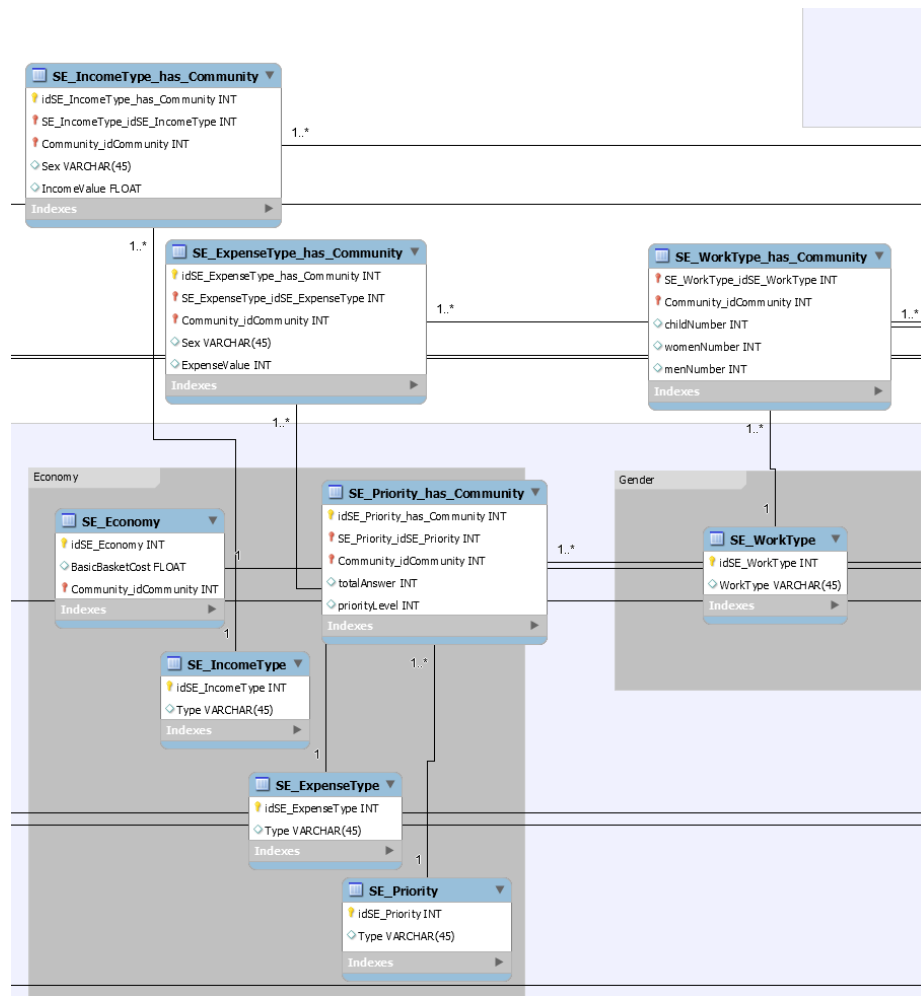


Primera parte de **SocioeconomicData**. se representan dos subconjuntos de datos: **DemographyAndCulture** y **PersonalSafety**:

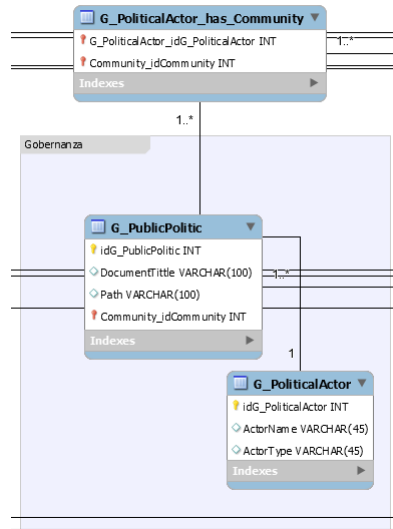


- SE\_Population: LT hace referencia a Less Than, BT a Between y GT a Greater Than.

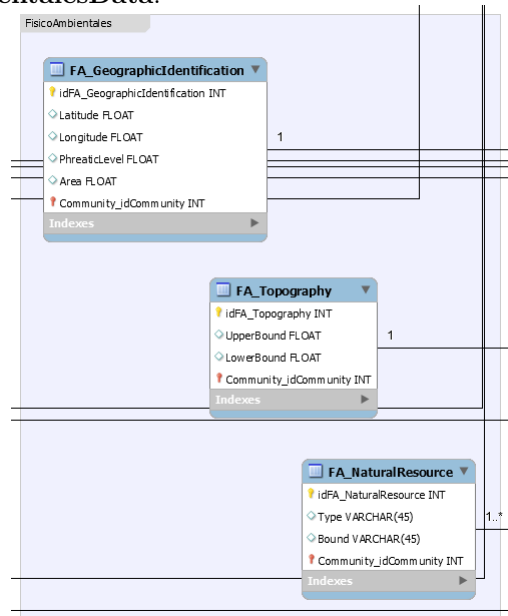
Segunda parte **SocioeconomicoData**. se representan dos subconjuntos de datos: **Economy** y **Gender**:



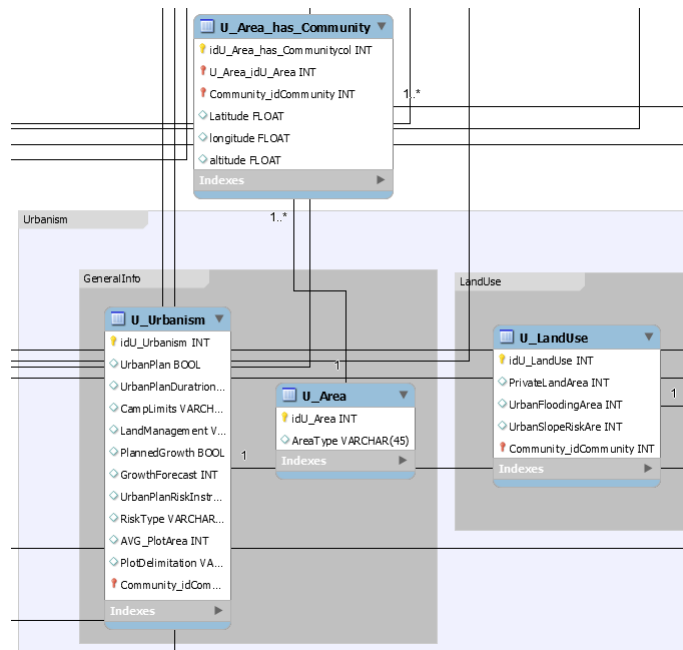
La siguiente figura son las entidades correspondientes al conjunto de dato de **GobernanzaData**:



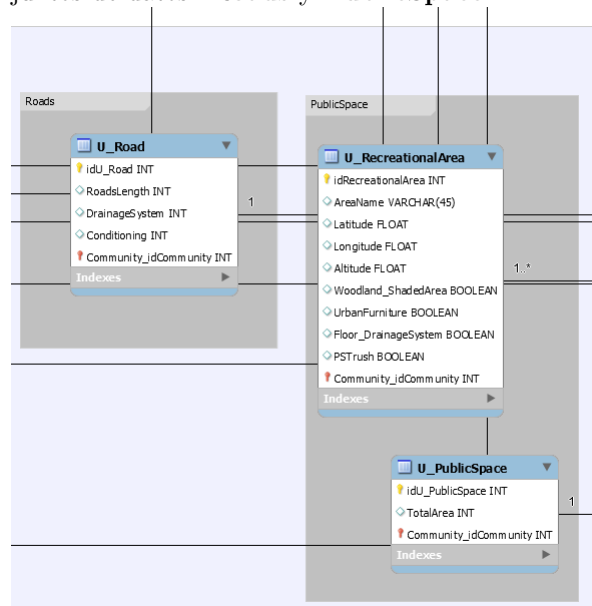
El siguiente conjunto de datos es el correspondiente al conjunto **FisicoAmbientalesData**:



Primera parte del conjunto **UrbanismoData**. Se representando dos sub-conjuntos de datos: **GeneralInfo** y **LandUse**:

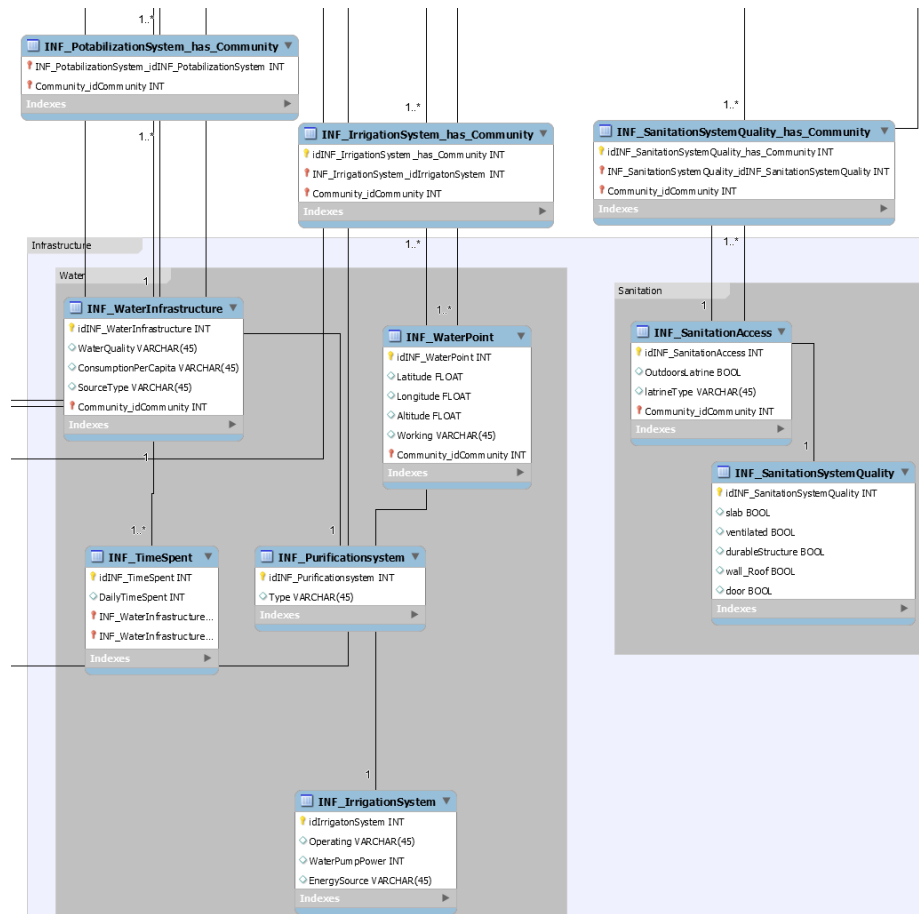


Segunda parte del conjunto **UrbanismoData**. Se representando dos subconjuntos de datos: **Roads** y **PublicSpace**:

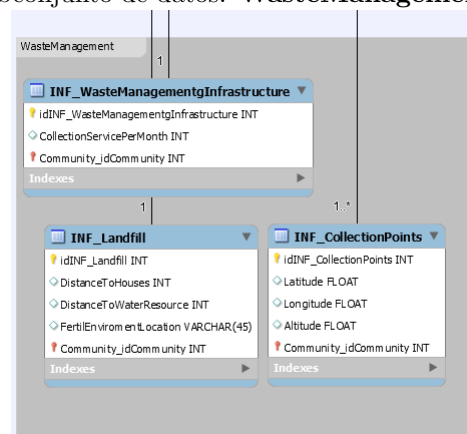


Primera parte de el conjunto **InfraestructurasData**. Se representan dos subconjuntos de datos: **Water** y **Sanitation**:



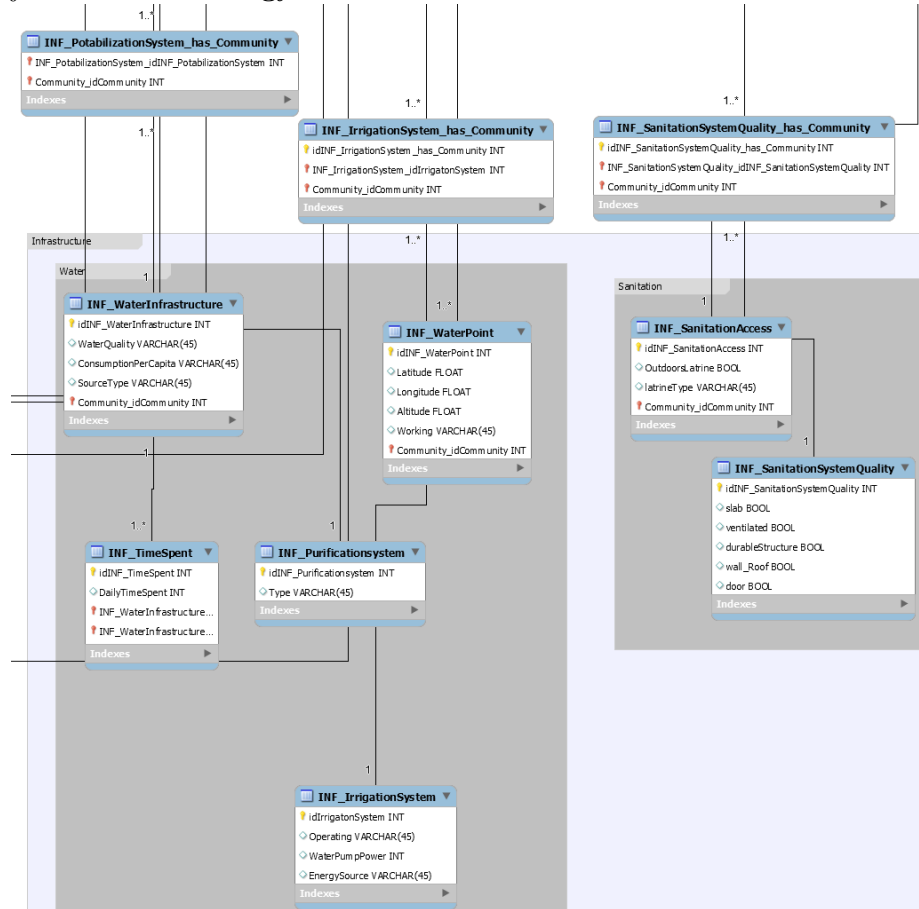


Segunda parte de el conjunto **InfraestructurasData**. Se representa un subconjunto de datos: **WasteManagement**:

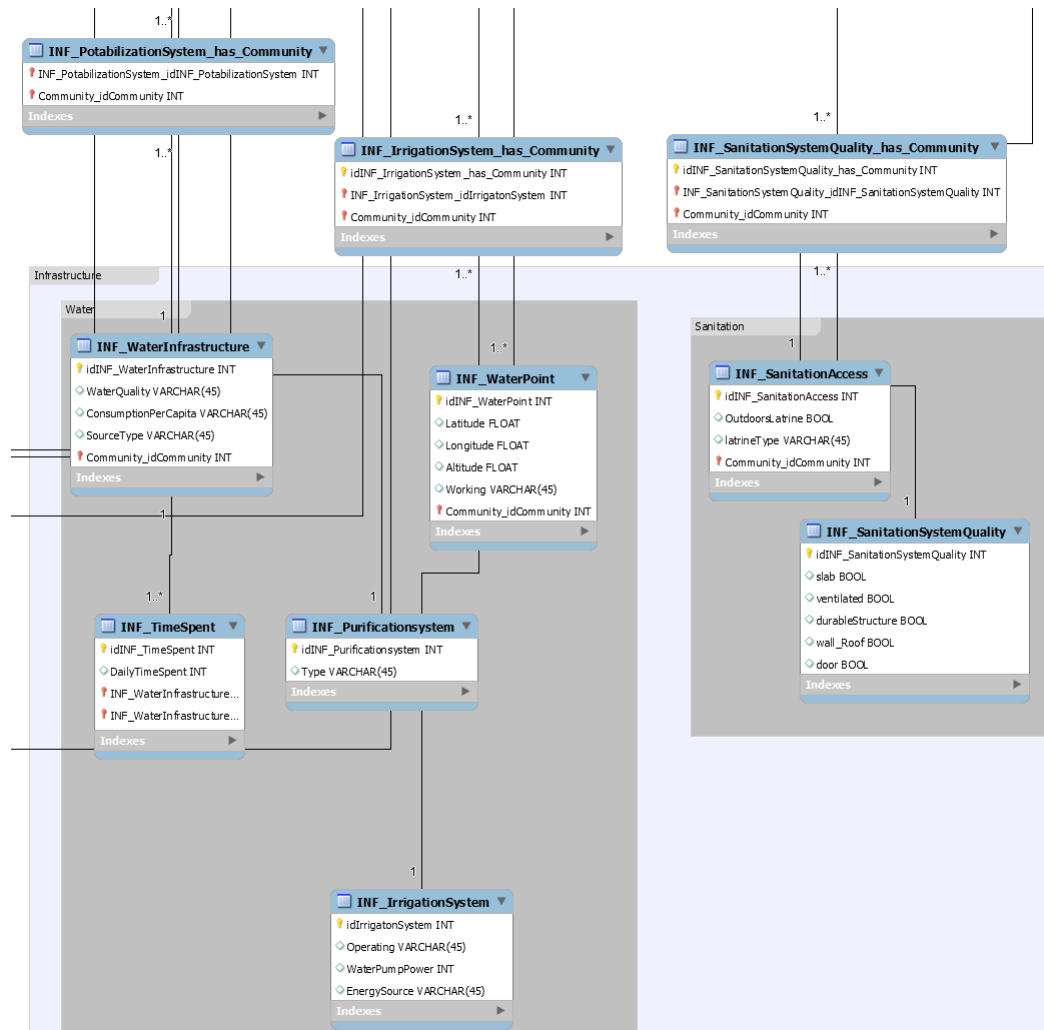


Tercera parte de el conjunto **InfraestructurasData**. Se representa un sub-

conjunto de datos: **Energy:**

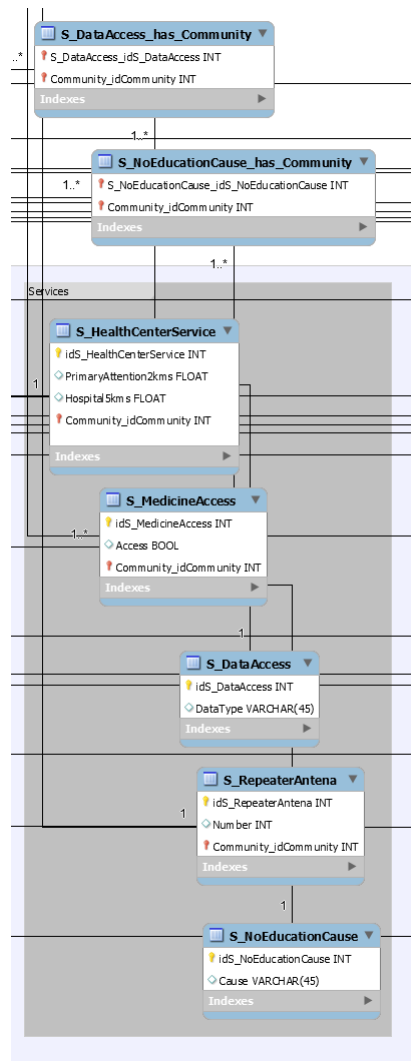


Cuarta parte de el conjunto **InfraestructurasData**. Se representa un sub-conjunto de datos: **Mobility:**

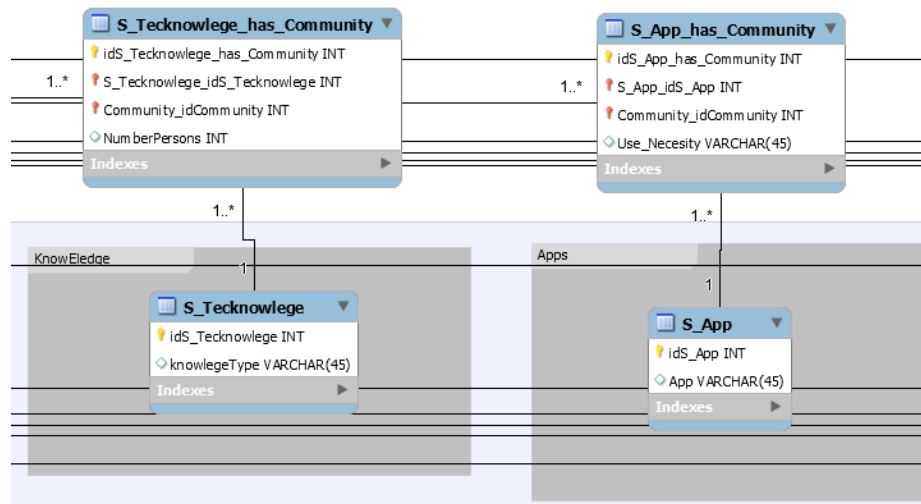


Primera parte de el conjunto **ServiciosData**. Se representa un subconjunto de datos: **Centers**:

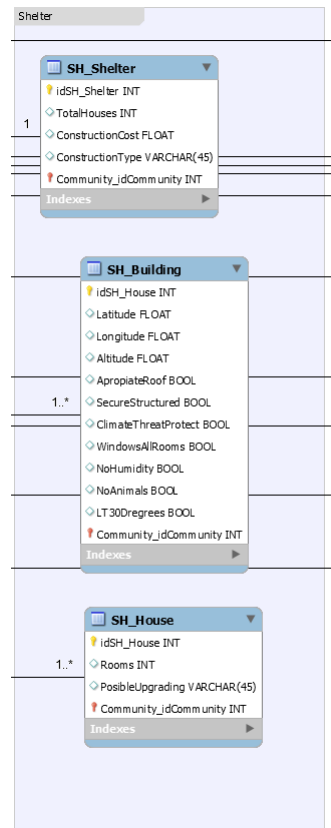




Tercera parte de el conjunto **ServiciosData**. Se representan dos subconjuntos de datos: **KnowEledge** y **Apps**:

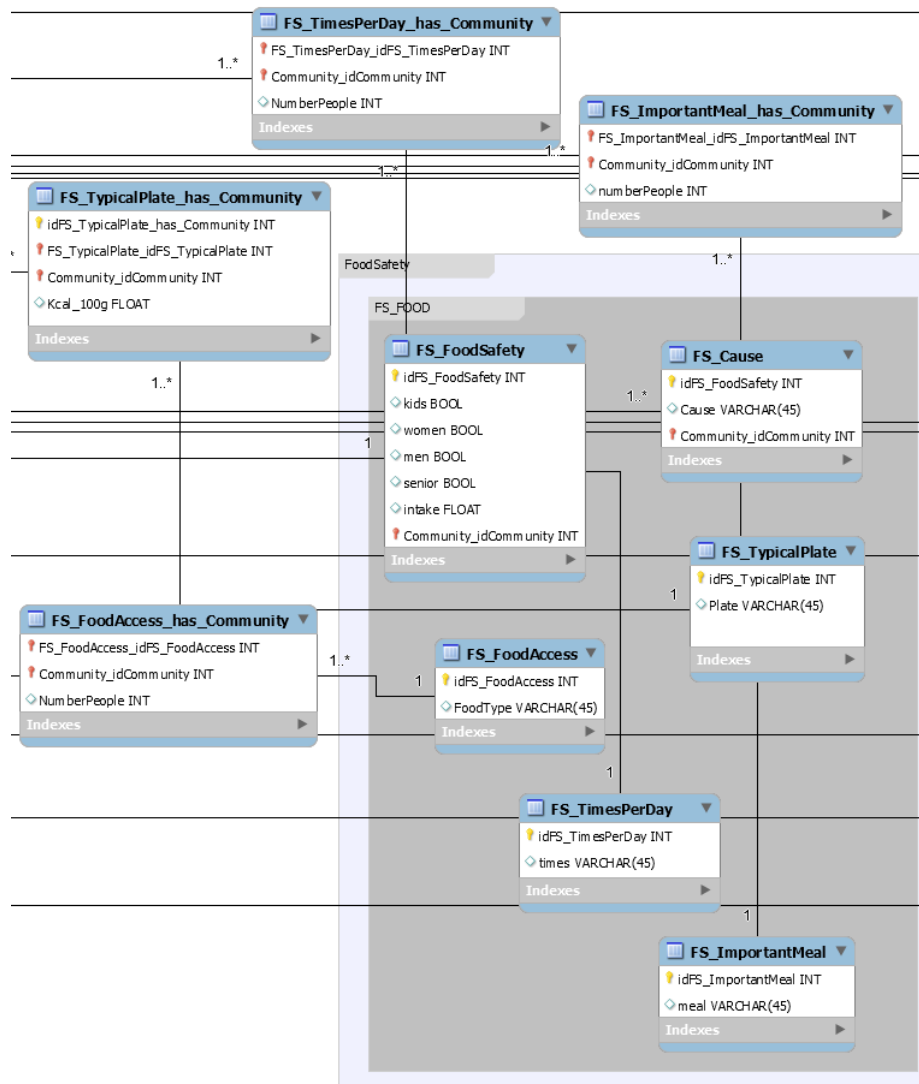


### Conjunto ServiciosData.

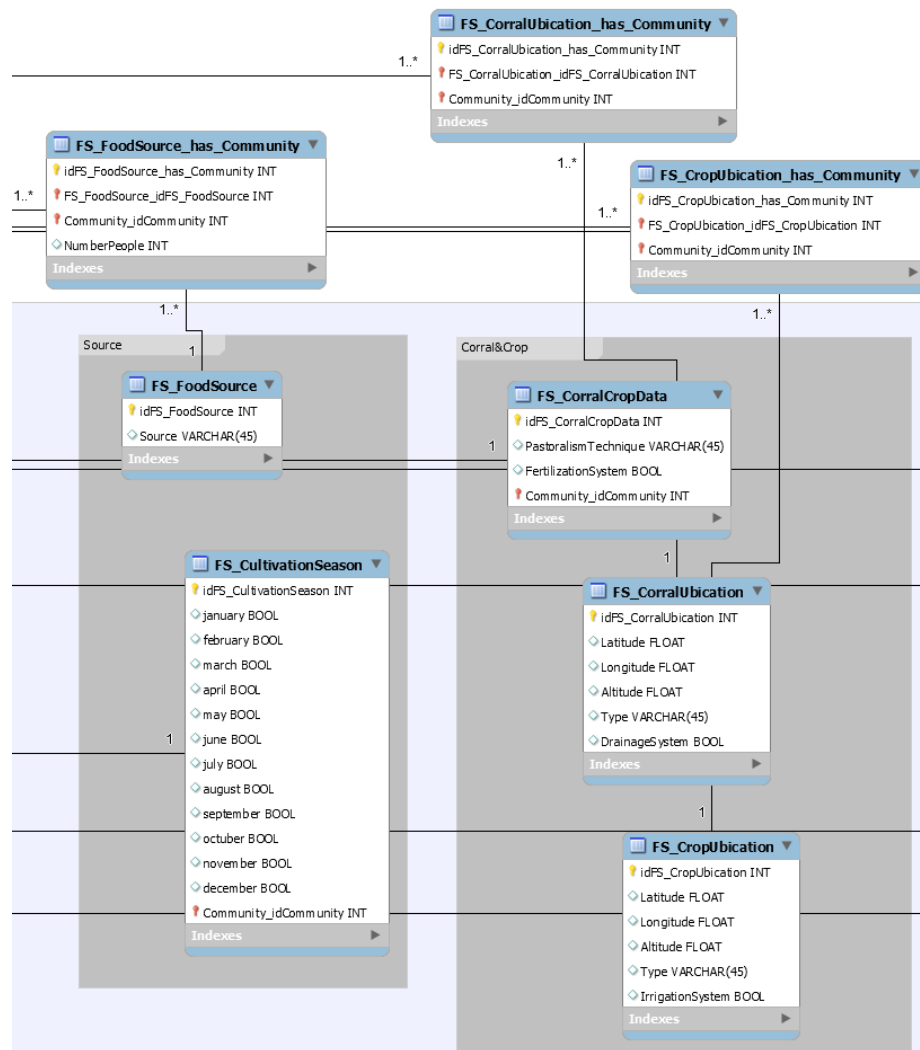


Primera parte de el conjunto **SeguridadAlimentariaData**. Se representa

un subconjunto de datos: **FS\_FOOD**:

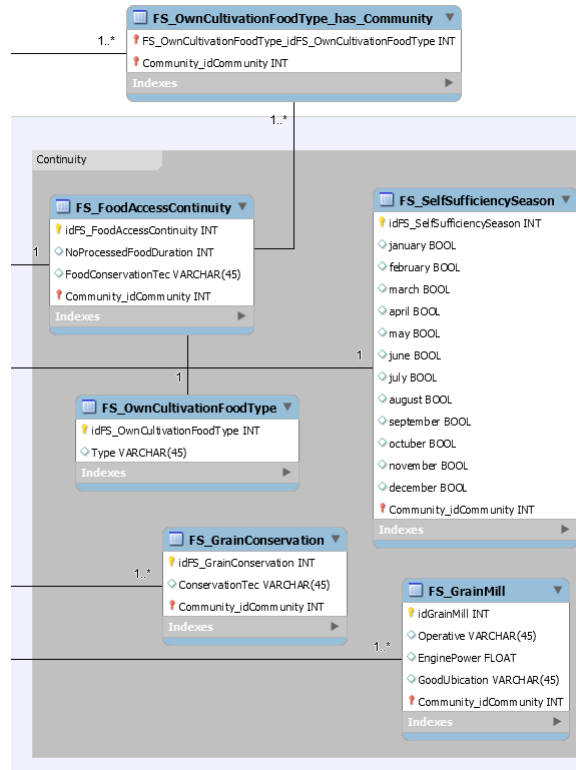


Segunda parte de el conjunto **SeguridadAlimentariaData**. Se representan dos subconjuntos de datos: **Source** y **Corral&Crop**:



Tercera parte del conjunto **SeguridadAlimentariaData**. Se representa un subconjunto de datos: **Continuity**:





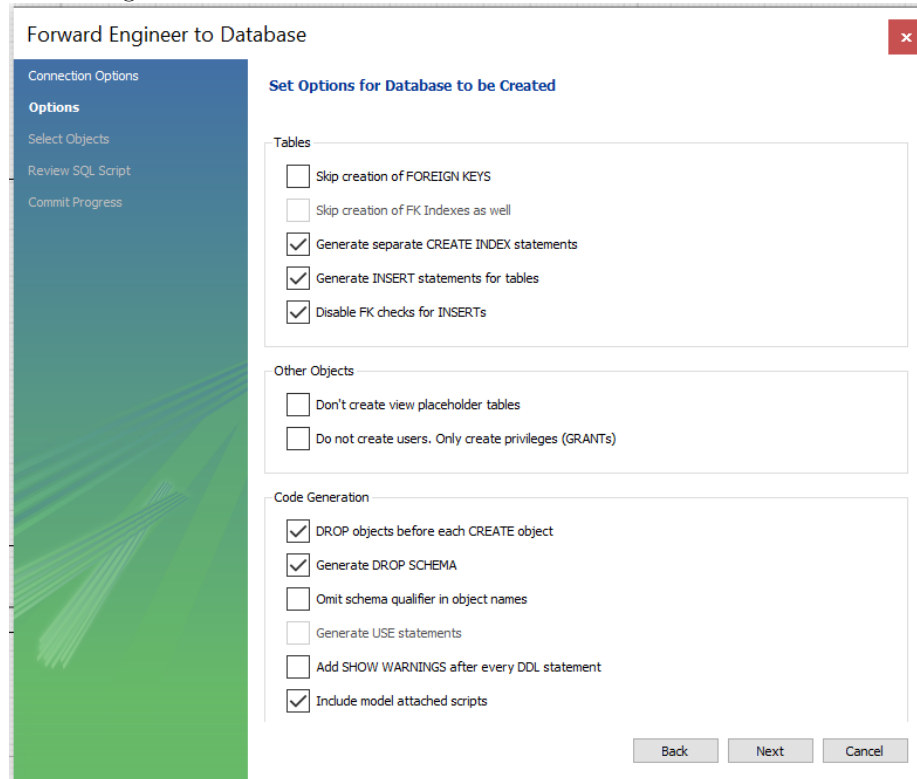
Estos son todas las entidades organizadas en conjuntos y subconjuntos de datos existentes en el modelo. Precisamente en la dos últimas imágenes, FiguraX y FiguraX, existen dos entidades peculiares. Estas son FS\_CultivationSeason y FS\_SelfSufficiencySeason. Ambas representan la temporada de un hecho durante el año y en un primer diseño se propusieron como una tabla con una sola columna llamada mes. De esta manera las tablas se relacionaban con *Community* con una cardinalidad N:M pero, por motivos de facilidad de desarrollo, comprensión de datos, facilidad de consultas a la base de datos y por considerarse que no afecta a la integridad de los datos, en una segunda versión se ha decidido desnormalizar ambas tablas aplicando la técnica conocida como vectorización de datos. Tal y como se puede observar en ambas figuras.

## 5 Creación del esquema de BD

Una vez desarrollado todo el modelo lógico de la base de datos, el editor Work-Bench de MySQL te permite realizar ingeniería directa sobre el modelo para poder obtener el *Script SQL* que al ser ejecutado genera la base de datos real con la que ya se puede trabajar.

El procedimiento es el siguiente. Dentro del entorno donde se ha representado el modelo se sigue la siguiente ruta *Database-Forward Engineer-*, aquí se

configuran los campos *Stored Connection*, *Connection Method*, *Hostname*, *Port* y *Username* de manera correcta y se presiona el botón *Next*. Entonces aparece una ventana en la que se seleccionan las opciones que se ven seleccionadas en la siguiente imagen:



Una vez seleccionado eso se presiona *Next* hasta el final y se guarda el Script generado en el directorio que corresponda.

En el caso de este trabajo, ese Script se llama "*CreateScript.sql*" y se anexa a este trabajo.

## 6 Proceso ETL (Programas Python y scripts SQL)

.

## 7 Proceso de muestra y cálculo de indicadores

.

## **8 Cuadro de mando**

.

## **9 Aproximación al diseño de Data Warehouse**

.

<https://blogs.upm.es/refugiadosupm/quienes-somos/> <http://www.itd.upm.es/alianzashire/quienes-somos/>