

Adquisición de datos e información espacial.

Introducción

En este curso ya hemos visto diversos temas: desde la comprensión de los datos, la información y el conocimiento; la parte conceptual e histórica de los SIG; los sistemas de referencia desde la visión de lo SIG; los modelos geográficos, de representación y de almacenamiento; la componente espacial y temática, cada uno lo vimos en forma independiente.

Unidad: euros/m ²	2004	2005	2006	2007	2008	2009	2010	2012
TOTAL NACIONAL:								
Andalucía	345,5	411,0	454,9	535,3	534,7	490,8	369,6	293,9
Aragón	666,0	714,3	733,6	788,3	712,8	677,6	426,5	386,7
Asturias (Principado de)	413,7	352,6	379,4	454,4	507,6	437,0	352,7	287,6
Baleares (Illes)	641,1	664,3	628,9	625,2	672,0	586,2	413,0	331,7
Canarias	651,0	464,7	690,2	566,9	676,6	565,1	375,1	235,6
Castilla	279,2	291,6	391,3	358,3	373,8	432,1	645,2	339,7
Castilla y León	279,2	291,6	391,3	358,3	373,8	432,1	645,2	339,7
Castilla-La Mancha	437,5	514,9	605,1	493,9	507,9	419,7	348,2	234,6
Cataluña	949,0	1.120,1	1.116,8	1.034,9	911,7	821,7	646,8	607,7
Comunidad Valenciana	548,7	682,9	717,3	788,9	791,4	661,1	389,4	396,8
Extremadura	272,6	334,8	376,0	468,9	468,9	474,6	394,3	387,0
Galicia	371,0	404,0	382,4	465,6	467,9	488,2	274,3	198,2
Madrid (Comunidad de)	1.148,3	1.286,9	1.270,9	1.111,7	1.038,2	862,8	646,7	738,4
Murcia (Región de)	284,3	339,4	435,9	385,1	374,6	493,1	249,9	199,1
Navarra (Comunidad Foral de)	512,9	548,8	n.s.	n.s.	747,7	726,6	525,3	728,8
País Vasco	579,2	643,2	720,1	760,2	862,1	822,0	600,5	318,1
Riña (La)	806,4	797,8	n.s.	917,3	972,7	n.s.	297,2	89,2
Ceuta y Melilla	496,3	629,9	429,2	422,3	388,9	392,3	345,1	275,1



En este nuevo módulo, y ya llegando al final de este curso, veremos y estudiaremos los **distintos métodos de creación del dato y la información geográfica**, y los **distintos orígenes** desde los cuales podemos aplicar todos los conocimientos adquiridos en la parte teórica y la práctica.

También veremos las **principales fuentes existentes de datos geográficos**, **los fundamentos que lo sostienen y que tipo de dato podemos obtener a partir de cada una de ellas**.



Como iremos viendo progresivamente, **existe una amplia gama de orígenes posibles que nos permiten incorporar los datos geográficos a los GIS, y que permiten llevar a cabo distintos tipos de análisis.**

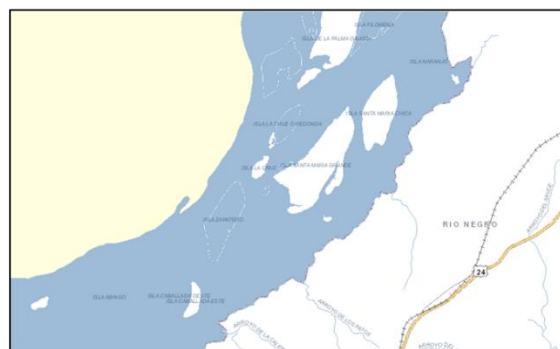
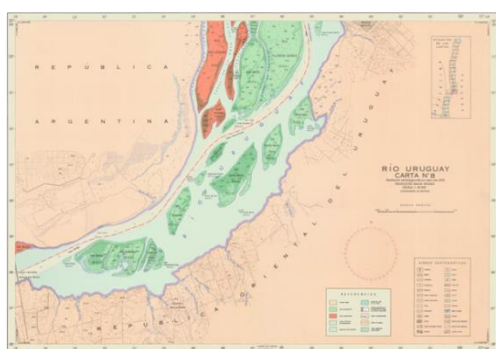
Como primer paso, destacaremos que las distintas metodologías de recolección de información geográfica nos condicionan en dos cosas:

- La **forma en que estos datos llegan a nuestro poder.**
- Las **operaciones que debemos realizar para adaptarlos a un proyecto GIS en particular.**

Históricamente, toda la información geográfica que se manejaba dentro de un SIG tenía su origen en un mapa en papel. **Estos mapas debían procesarse para adaptarse a la naturaleza propia del SIG.**

Tal como vimos en el primer módulo, en el desarrollo de los SIG primero apareció el **software** y luego vinieron los **datos geográficos**. Estos primeros datos geográficos de los que se disponía no se encontraban en **formato digital** (no eran adecuados para su uso dentro de un SIG), sino que eran datos analógicos.

Se comienza, pues, a transitar los distintos **procesos de digitalización** de la cartografía, es decir a **la conversión de los datos en formato impreso en datos en formato digital que un SIG pueda manejar.**



En los orígenes de los SIG, los datos disponibles eran variados en su naturaleza, pero no es su formato. Estamos hablando de formatos papel tipo: **mapas, cartas de navegación, fotografías aéreas**, etc.

Con el avance tecnológico los datos se comienzan a producir directamente en **formato digital**. Tal es caso de los relevamientos con GNSS, la topografía digital, las imágenes satelitales, los datos LIDAR. **Esta nueva realidad es especialmente valiosa considerando los SIG como herramientas básicas de manejo de datos geográficos.**

Pero, en pleno siglo XXI, todavía existen datos geográficos en formato impreso (y las técnicas que se emplearon en su creación) siguen siendo válidas. Tal fue el caso de la tercera entrega del presente curso

Datos digitales y datos analógicos

Como ya hemos comentado, casi hasta el agotamiento, los sistemas de información geográficos requieren de datos digitales para poder llevar a cabo las tareas propias.

Siguiendo la línea presentada en la introducción, buena parte de los datos actuales se generan en formato digital.



Veamos en primer lugar, las ventajas de los datos digitales frente a los analógicos.

- **Facilidad de actualización:** La cartografía digital es editable y esto simplifica la introducción cambios. Las capas geográficas permiten actualizar los datos a distintos ritmos ya que existen independencia entre capas.
- **Facilidad de distribución:** Los avances tecnológicos en temas de comunicación (Internet) hacen más sencillo y menos costoso distribuir cartografía digital que analógica.

- **Espacio de almacenamiento:** Hoy se producen más cantidad de datos y se actualizan más rápidamente. Un soporte digital permite almacenar una enorme cantidad de datos ocupando una fracción del espacio físico. A modo de ejemplo, las imágenes satelitales. Estas son generadas cada pocos días para una misma zona. Antiguamente era necesario esperar varios años para realizar vuelos aéreos y restituirlos con técnicas fotogramétricas para poder contar con información de calidad. Esto repercute en dos cosas: los datos son actualizados con mayor frecuencia y requieren más espacio de almacenamiento digital.
- **Facilidad y precisión de análisis:** Los datos digitales permiten hacer con los datos geográficos digitales cosas que no eran posibles con los analógicos. A modo de ejemplo: a partir de los datos digitales podemos analizar el "índice verde" de una zona geográfica, cosa que era imposible de hacer a partir de datos analógicos.
- **Facilidad de mantenimiento:** Aunque no se introduzcan modificaciones y no se actualicen los datos, el formato digital hace más fácil su conservación. No existe **degradación del dato digital**, aunque si del soporte digital. Por otro lado, los datos digitales pueden replicarse y respaldarse con facilidad; esto garantiza su persistencia a cargo plazo y a un menor costo que la de los datos analógicos.



Fuentes primarias y secundarias

Historicamente, los datos analógicos (generalmente) no han sido tomados pensando en su utilización en un SIG; pero nos sirven de base para obtener otros que sí pueden emplearse. El advenimiento de esta disciplina, acompañado por el avance tecnológico han hecho que los datos geográficos sean una necesidad imperiosa en diversos tipos de proyectos, tal como ya hemos presentado en otros módulos.

Pero también existen otra variedad de datos que ya han sido recogidos considerando su utilización dentro de un SIG, y la forma en la que se presentan ya es adecuada para incorporarlos en este y trabajar con ellos, con una necesidad mínima de conversión y procesamiento previo.

Veremos una clasificación inicial de las fuentes a partir de este criterio:

Esto da origen a una clasificación por su fuente:

- **Datos primarios:** son aquellos que podemos emplear en un SIG y que ya son susceptibles de ser sometidos a las operaciones de manejo y análisis de los SIG. *Comprenden: las imágenes digitales (de varios tipos) o los datos obtenidos con dispositivos GNSS.*
- **Datos secundarios:** derivan de algún otro tipo de dato previo, el cual no es adecuado para su empleo en un SIG, y que deben ser manipulados previamente para poder operar ser analizados. *Comprenden: las versiones digitales de los mapas clásicos, los datos procedentes de un muestreo o levantamiento tradicional, datos provenientes de cartografía impresa, etc.*

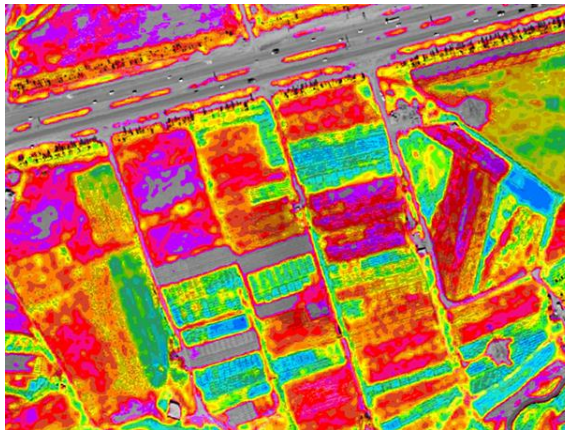
Fuentes primarias para la captura de datos ráster

En primer lugar veremos el caso de la **Teledetección**, esta refiere a la adquisición de información a pequeña o gran escala de un objeto o fenómeno, ya sea usando instrumentos de grabación o instrumentos de escaneo en tiempo real inalámbricos o que no están en contacto directo

con el objeto (como por ejemplo *aviones, satélites, astronave, boyas o barcos*).

La teledetección consiste en recoger información a través de diferentes dispositivos de un objeto concreto o un área.

A modo de ejemplo, la observación terrestre o los satélites meteorológicos, las boyas oceánicas y atmosféricas, las imágenes por resonancia magnética (MRI en inglés), la tomografía por emisión de positrones (PET en inglés), los rayos-X y las sondas espaciales son todos ejemplos de teledetección.



Actualmente, el término se refiere de manera general al uso de tecnologías de sensores para adquisición de imágenes, incluyendo: instrumentos a bordo de satélites o aerotransportados, usos en electrofisiología, y difiere en otros campos relacionados con imágenes como por ejemplo en imagen médica. Este tema se verá en forma exhaustiva en curso de IPR del Departamento de Geomática.

La gran cantidad de satélites existentes nos permite disponer de una cantidad de datos de imagen sin precedentes en manos de los analistas espaciales para su uso en aplicaciones que cubren prácticamente todos los aspectos de la actividad humana. Cubren órbitas terrestres bajas, medias y altas. Están controlados por agencias gubernamentales (como la NASA y la Agencia Espacial Europea) y por compañías privadas (como Digital Globe y Airbus). Cubren todos los segmentos del espectro electromagnético, desde el ultravioleta y el color natural al infrarrojo cercano, medio y térmico, así como sensores de microondas activos como el radar.

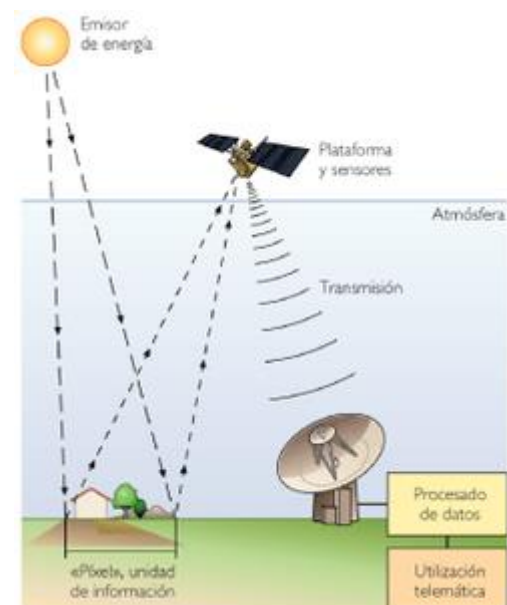
La aplicación <https://maps.esri.com/rc/sat2/index.html> muestra la ubicación actual de cerca de 14.000 objetos creados por el hombre que orbitan la Tierra. Los datos proceden de Space-Track.org, una organización que promueve la seguridad en los vuelos espaciales, la protección del medio espacial y el uso pacífico del espacio compartiendo información situacional con los operadores de satélites internacionales y de Estados Unidos y con otros organismos.

Volviendo a la teledetección, la información geográfica se deriva de las mediciones de la **cantidad de radiación** electromagnética reflejada, emitida o dispersada desde los objetos físicos. Es posible identificar los siguientes elementos:

- **Emisor de energía:** Un sensor de imágenes pasivo captura la energía reflejada o emitida desde la escena que visualiza. La luz del sol reflejada es la fuente más común de energía electromagnética medida por los sensores pasivos. Estos sensores permiten obtener observaciones globales de la Tierra y de la atmósfera. Por otro lado, un sensor activo es un instrumento que

emite energía y detecta la radiación reflejada desde la superficie de la Tierra o desde otro objetivo. Se usa en diversas aplicaciones relacionadas con la meteorología y la atmósfera, como el radar para medir los ecos de ciertos objetos (por ejemplo, las nubes de lluvia), el LIDAR para capturar valores de elevación detallados de la superficie y el sonar para medir la profundidad del fondo marino.

- **Objetos que interaccionan con la radiación o la emiten.**
- **Atmósfera.**
- **Plataforma y sensores que recoge la radiación.**
- **Receptor de datos.**
- **Unidades de procesamiento.**
- **Utilización telemática.**



Actualmente, las imágenes se capturan desde una amplia gama de altitudes que van desde el nivel del suelo a más de 35000 kilómetros sobre la Tierra. Las imágenes que proceden de cada altitud ofrecen distintas ventajas para cada aplicación. Este diagrama desglosa algunas de las altitudes de sensor más utilizadas:

- **Geosíncrono** (22.236 millas): Los satélites que siguen la rotación de la Tierra parecen estacionarios en el cielo para los observadores situados en el suelo. Aunque son los más utilizados para las comunicaciones, los satélites con órbitas geo-sincrónicas como el GIFTS hiper-espectral son útiles para monitorear fenómenos

cambiantes como las condiciones meteorológicas. El Syncom de la NASA, lanzado a principios de la década de 1960, fue el primer satélite de "vuelo a gran altura".



- **Heliosíncrono** (375-500 millas): Los satélites de esta órbita mantienen el ángulo de la luz solar en la superficie de la Tierra del modo más coherente posible, lo que significa que es posible comparar imágenes de la misma estación a lo largo de varios años, como sucede con las imágenes de Landsat. Esta es la zona primordial para los sensores que observan la Tierra.



- **Satélite atmosférico** (100.000 pies): También conocidos como pseudo-satélites, estos vehículos no tripulados rozan los límites más altos de la atmósfera detectable. La nave experimental Helios de la NASA medía las llamaradas solares antes de precipitarse sobre el océano Pacífico cerca de Kauai.



- **Reactor** (90.000-30.000 pies): Los aviones reactores que vuelan a 30.000 pies y alturas superiores pueden recorrer las áreas

afectadas por alguna catástrofe en un tiempo muy corto; esto los que los convierte en una buena plataforma para ciertos tipos de aplicaciones ópticas y de imágenes multi-espectrales.



- **Aparatos de aviación generales** (100-10.000 pies): Las aeronaves pequeñas que pueden volar a baja velocidad y baja altitud son desde hace tiempo los vehículos idóneos para la fotografía aérea y la orto-fotografía de alta calidad. Estas aeronaves, que van desde las avionetas Cessnas a los ultraligeros y los helicópteros, son las que se encargan de la obtención de la mayor parte de las imágenes ópticas urbanas.



- **Drones** (100-500 pies): Los drones son la incorporación más reciente a esta flota. Su capacidad para volar a baja altitud y con control remoto ofrece atractivas ventajas para la fotografía aérea, con resoluciones que llegan a menos de 1 pulgada. Los vehículos aéreos no tripulados militares pueden ser drones más pequeños o aviones.



- **Instalado o manipulado en el suelo** (Nivel del suelo): Cada vez más, las imágenes captadas en el nivel del suelo se abren camino

hasta los flujos de trabajo de los SIG. Cosas como Google Street View, las imágenes a nivel de calle de HERE y Mapillary; los dispositivos de mano para la obtención de imágenes multispectrales y otros sensores terrestres están encontrando aplicaciones en áreas como los conductos, la seguridad, el turismo, la propiedad inmobiliaria,



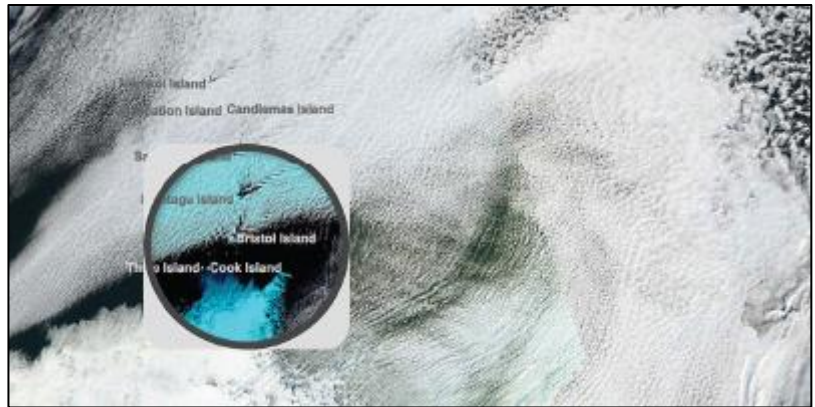
Como decíamos anteriormente existen dos tipos de sensores: activos y pasivos. Veamos los tipo de productos podemos obtener con cada uno.

- **Color Natural:** constituye el grueso de las imágenes. Estas son adecuadas para los análisis generales de las entidades terrestres y submarinas, para los estudios urbanos y para el reconocimiento, son las más utilizadas por los usuarios en general y, por lo tanto, las más fáciles de entender.

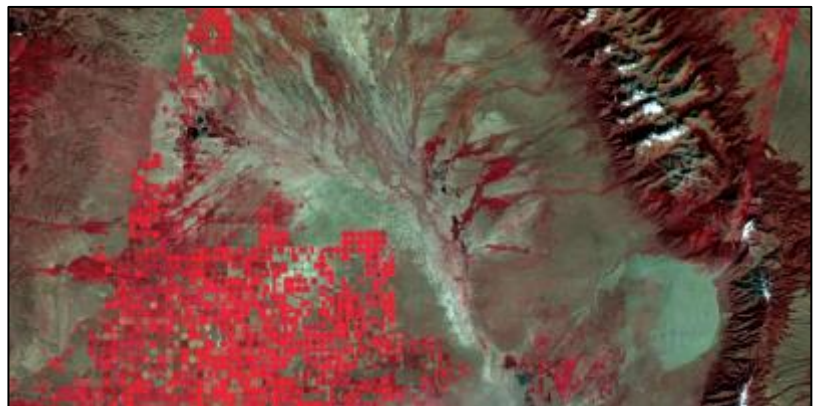


- **Falso color:** Las técnicas conocidas como falso color se utilizan en la tecnología de imagen (astronomía, las imágenes de satélite, imágenes médicas, exploración o minería), donde podemos aprovechar para poner de relieve pequeñas variaciones de color gris. Estas técnicas permiten asociar a un rango de gris dada un color específico que tiene el único propósito de hacer más visibles las áreas correspondientes. Esto es a veces directamente en la

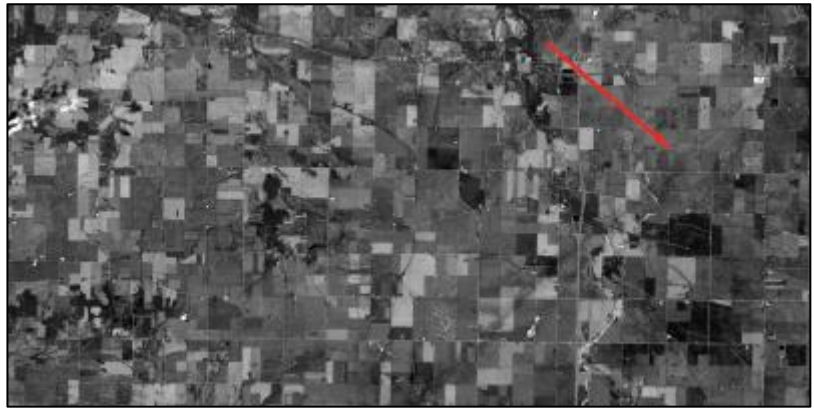
pantalla del dispositivo, sin tener que pasar por un programa informático, evitando así retraso en al percibir el resultado.



- **Infrarrojo:** Una imagen infrarroja, compuesta de energía de infrarrojo cercano permite resaltar la vegetación mas sana o de hoja ancha como tonor de color rojo intenso, mientras que los rojos más claros corresponden a las praderas o a las áreas con una vegetación más dispersa.

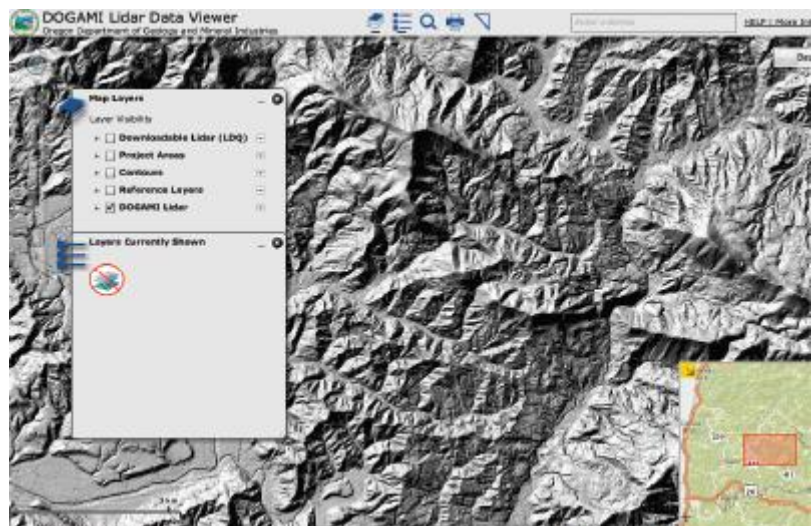


- **Pancromático:** Las imágenes pancromáticas con resoluciones más altas se crean cuando el sensor que obtiene las imágenes es sensible a un rango más amplio de longitudes de onda de la luz, que normalmente incluyen toda a parte visible del espectro que se almacena y se muestra como una imagen en escala de grises de una sola banda. Esto permite la creación de píxeles más pequeños en el sensor y produce una imagen más nítida.



Por otro lado, un sensor activo es un instrumento que emite energía y detecta la radiación reflejada desde la superficie de la Tierra o desde otro objetivo. Se usa en diversas aplicaciones relacionadas con la meteorología y la atmósfera, como el radar para medir los ecos de ciertos objetos (por ejemplo, las nubes de lluvia), el LIDAR para capturar valores de elevación detallados de la superficie y el sonar para medir la profundidad del fondo marino.

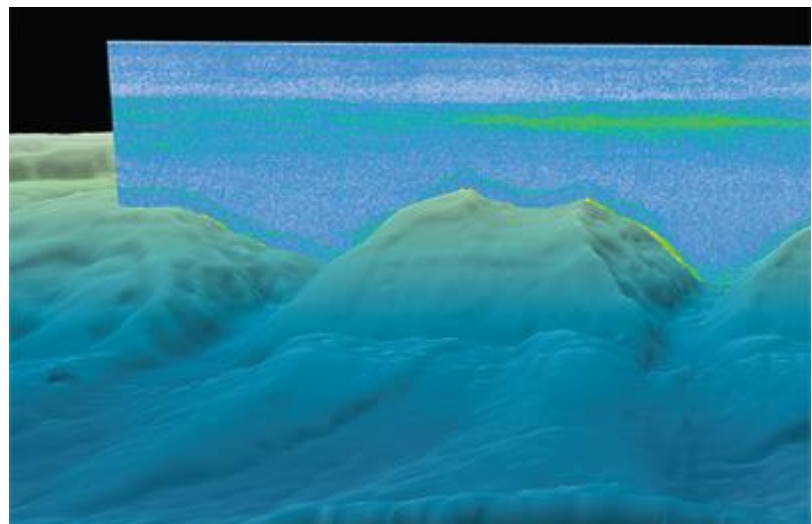
- **Tecnología LIDAR:** Las capturas de estos datos se realiza con sistemas montados en aviones que recogen hasta 500.000 puntos por segundo y crean grandes conjuntos de puntos de elevación densos y exactos para un área extensa.



- **Radar:** Esta tecnología tiene dos ventajas principales: funciona en la oscuridad y puede ver a través de las nubes. Esto lo convierte en la herramienta ideal para recopilar inteligencia y monitorizar la climatología.



- **Sonar:** En una iniciativa global liderada por la NOAA, los datos de sonar de columnas de agua se capturan con tecnología acústica (o de sonar) activa desde cerca de la superficie hasta el fondo marino. Como los barcos se mueven durante la captura, los datos terminan pareciendo una cortina o una lámina, como puede verse aquí. La NOAA y las otras flotas académicas e internacionales participantes ponen los datos a disposición de los investigadores y el público de todo el mundo a través de esta aplicación.



Un concepto importante (pero que veremos brevemente) trata sobre la resolución de este tipo de dato. Existen tres tipos de resoluciones principales en la teledetección.

- La **resolución espacial** es una característica física fundamental de los sistemas de teledetección y refiere al tamaño del objeto que puede ser resuelto y la medida más habitual es el tamaño del píxel.
- La **resolución espectral** se refiere a las partes del espectro electromagnético que se miden.
- La **resolución temporal** (ciclo de repetición) describe la frecuencia con que se recogen imágenes de la misma zona.

Para abandonar este tema en el presente curso, dejamos unas reflexiones de **Kass Green** (fundadora de una de las primeras compañías comerciales dedicadas a procesar datos de Landsat):

“Como las grandes obras de arte, las imágenes revelan su carácter y su estructura de formas complejas: siempre sorprendentes, a veces sutiles, en ocasiones desconcertantes. Primero llega el asombro por su belleza: los gélidos glaciares de Groenlandia, las delicadas ramificaciones de perfil LIDAR de una secuoya, el borde dentado de una falla en el radar, los verdes intensos de los trópicos, las líneas marcadas del impacto humano, la resurrección de los bosques San del monte Santa Helena, las ondulaciones de las tierras de cultivo de Asia y África, las nieves perdidas del Kilimanjaro. Cada imagen nos anima a descubrir más, a mirar una y otra vez.

Tras la primera mirada, empezamos a explorar. ¿Qué es lo que crea esa respuesta espectral única? ¿Por qué están los árboles en las pendientes orientadas al norte y los arbustos en las pendientes orientadas al sur en esta área? ¿Tienen las ubicaciones de diferentes especies de árboles relación con la pendiente y la elevación? ¿Por qué ha ardidado esta casa mientras que las llamas ni siquiera han tocado la de al lado? ¿Cuántas personas viven en este pueblo? ¿Qué cosechas se cultivan aquí? ¿Habrá suficiente comida para alimentar a esta gente? ¿Cómo ha cambiado el paisaje de una forma tan radical? ¿Quién lo ha cambiado?

A continuación, gracias al SIG, descubrimos las conexiones. Si tenemos suerte, viajamos al campo con nuestras aplicaciones móviles para ver por nosotros mismos cómo varía el paisaje en relación con las imágenes y con otras capas SIG. Usamos herramientas de escritorio para organizar y corregistrar las capas de información y extraemos de ellas las variables que son más predictivas. Aprendemos a obtener de las imágenes y los datos SIG información sobre la ubicación, la altura, la forma, la textura, el contexto, la sombra, el tono y el color de cada objeto. Y a continuación creamos mapas: inventariamos los recursos y monitorizamos sus cambios a lo largo del tiempo.

Las imágenes han sido mi billete de entrada al mundo. Gracias a ellas, he viajado por todo el planeta, he oído historias asombrosas y he conocido a personas fascinantes y llenas de pasión por sus proyectos y sus comunidades. Soy muy afortunada por haber descubierto la belleza de las imágenes y, a través de ella, haber encontrado el trabajo que sin duda estaba destinada a hacer."



Fuentes primarias para la captura de datos vectoriales

Dentro de las fuentes primarias de captura para datos vectoriales encontramos la **topografía tradicional**. Esta se basa en el principio de ubicar tridimensionalmente de cualquier punto. Esta ubicación se lleva a cabo a partir de determinar los ángulos y distancias desde otros puntos conocidos.



La topografía clásica utiliza equipos como los teodolitos y estaciones totales; estos últimos pueden definir puntos hasta con una precisión de 1 milímetro. Esta metodología de relevamiento de datos vectoriales se lleva a cabo directamente o sobre el terreno, y lleva mucho tiempo y elevados costos, pero sigue siendo una de las mejores maneras de obtener ubicaciones de los puntos de alta precisión.

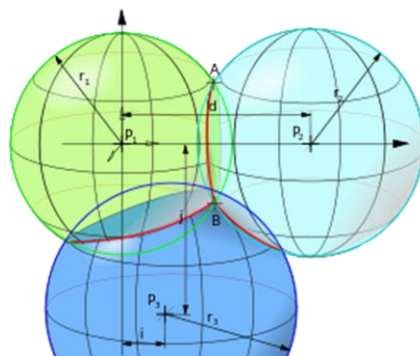
Otra fuente de datos primaria surge a partir de la aparición de los **Sistemas Globales de Navegación por Satélite (GNSS)**; estos permiten la obtención de coordenadas geográficas de un modo inmediato, con las consecuencias que esto tiene para su uso en actividades como la elaboración de cartografía. En esta tecnología, los errores son del orden de unos pocos metros o menos, pudiendo llegar a centímetros con el equipo adecuado.



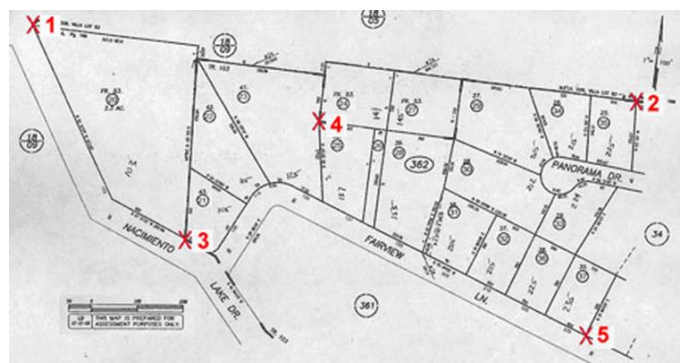
Se basa en el cálculo de una posición sobre la superficie terrestre midiendo las distancias de un mínimo de tres satélites de posición conocida. Un cuarto satélite aportará, además, la altitud. La precisión de las mediciones de distancia determina la exactitud de la ubicación final.

En la práctica, un receptor capta las señales de sincronización emitida por los satélites que contiene la posición del satélite y el tiempo exacto en que ésta fue transmitida. La posición del satélite se transmite en un mensaje de datos que se superpone en un código que sirve como referencia de la sincronización.

Cada medida de la distancia coloca al receptor en una cáscara esférica de radio la distancia medida. Tomando varias medidas y después buscando el punto donde se cortan, se obtiene la posición. Sin embargo, en el caso de un receptor móvil que se desplaza rápidamente, la posición de la señal se mueve mientras que las señales de varios satélites son recibidas.



Los sistemas GNSS se utilizan para la captura de construcciones, los límites de tierras y propiedades, perforaciones, etc. También empleado para obtener marcas de referencia para su uso en otros proyectos de captura de datos.



El formato de intercambio GPX

El formato de archivos GPX establece una forma estándar para el intercambio y almacenamiento de información de mapas en dispositivos GNSS, teléfonos inteligentes y computadoras. Gracias a este formato GPX, los softwares GIS pueden leer datos creados en diferentes equipos GNSS.

El formato GPX está basado en el standard **XML (Lenguaje de Marcado Extensible)**, es decir, es un archivo de texto, Un usuario experimentado puede leer un archivo GPX manualmente usando simples programas de edición de texto; sin necesidad de software especializado para decodificar la información. A modo de ejemplo:

```
<?xml version="1.0" encoding="UTF-8"?>
<gpx xmlns="http://www.topografix.com/GPX/1/1" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
creator="byHand" version="1.1" xsi:schemaLocation="http://www.topografix.com/GPX/1/1
http://www.topografix.com/GPX/1/1/gpx.xsd">
  <wpt lat="39.921055008" lon="3.054223107">
    <ele>12.863281</ele>
    <time>2005-05-16T11:49:06Z</time>
    <name>Cala Sant Vicenç - Mallorca</name>
    <sym>City</sym>
  </wpt>
</gpx>
```

Es posible acceder a información crítica de mapas incluso si no hay software o equipos especializados disponibles.

El formato GPX acomoda una cantidad determinada de puntos que pueden indicar una ruta o camino, así que los archivos GPX pueden tener cualquier tamaño. Los archivos simples con pocos elementos ocupan menos de 100 Kb.

Estos archivos son auto-contenidos, esto quiere decir que se requiere sólo un archivo para describir un conjunto completo de datos geográficos.



A modo de ejercicio



1. Descargue la aplicación **Trails para el iPhone** o **BackCountry Navigator para Android y Windows Mobile**.
2. **Utilice esta App para hacer un relevamiento de una zona geográfica de su interés.** A modo de ejemplo: paradas de ómnibus, depósitos de basura, arbolado de la ciudad, infraestructura urbana, etc.
3. **Descargue a su computadora el archivo generado con la App.**
4. **Analice la forma de cargar los datos GPX en el software GIS de su interés.** Tanto ArcGIS como QGIS permite llevar a cabo esta tarea.

Fuentes secundarias

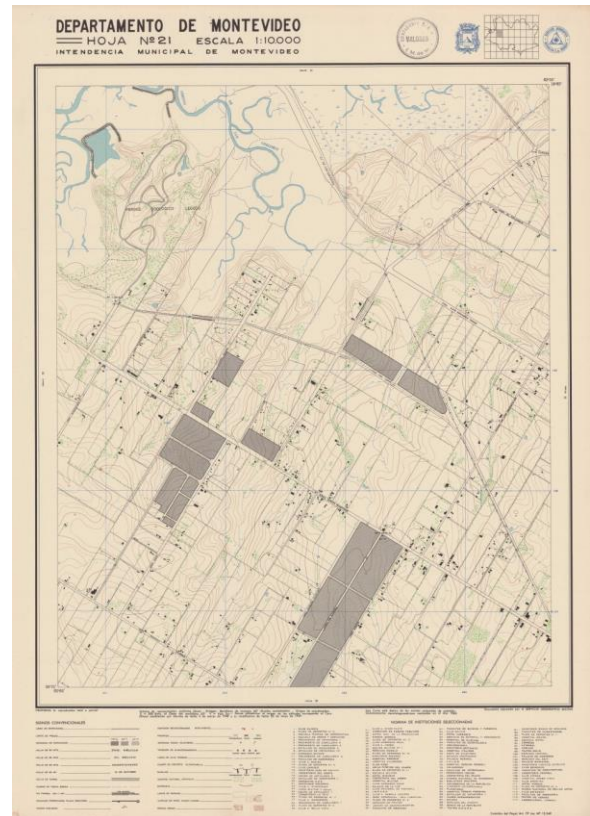
En esta sección veremos algunas de las principales fuentes de datos secundarias, es decir, **aquellas que derivan de algún otro tipo de dato previo, el cual no es adecuado para su empleo en un SIG, y que deben ser manipulados previamente para poder operar ser analizados.**

Veremos tres fuentes secundarias que entendemos fundamentales en este curso: el **escaneado de cartas y su posterior digitalización en pantalla**, la **digitalización a través de tabletas** y la **geocodificación a partir de tablas de direcciones.**

Como complemento, veremos también el reconocimiento de imágenes con datos alfanuméricos mediante herramientas de OCR y la utilización de esta información en los softwares GIS.

Escaneado de cartas y la digitalización en pantalla

Tal como hemos visto en la tercera entrega obligatoria del presente curso, una de las fuentes de datos que podemos utilizar son las clásicas cartas en formato papel.



La **cartografía análoga** (también llamada **cartografía tradicional**) refiere a los mapas/planos/cartas obtenidos por procedimientos gráficos, y por extensión a todo el conjunto de operaciones destinadas a la obtención de la información en **soporte gráfico con la que se diseña el mapa**.

En la cartografía análoga el propio mapa impreso (o los positivos fotográficos) que se emplean para su obtención constituyen los documentos finales donde queda guardada la información cartográfica. Estos productos cartográficos reciben comúnmente el nombre de **copias duras**.

Cualquier corrección, modificación o puesta al día de la cartografía análoga se realiza actuando sobre los documentos gráficos que contienen la información.

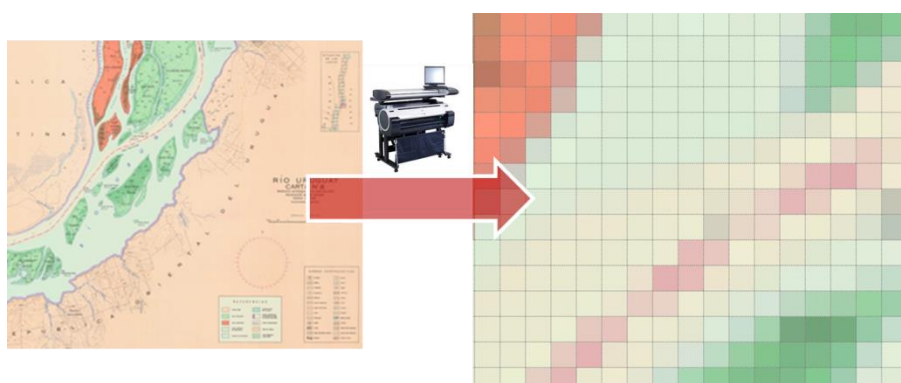
Para que estas copias duras puedan ser utilizadas en el ámbito de los GIS, requieren ser **escaneadas**. Estos documentos papel son escaneados con el objeto principal de **reducir el desgaste y deterioro, mejorar el acceso de los usuarios, proporcionar almacenamiento de base de datos** y la **creación de índices geográficos**.

Los negativos fotográficos, los mapas de papel, las fotografías aéreas y las imágenes se escanean y se geo-referencian para proporcionar el contexto geográfico de otros datos, ser objetos de análisis en sí mismos, u oficiar de soporte para la generación de nueva información.

El escaneado es el proceso habitual antes de la digitalización en pantalla o vectorización.

Veamos estos dos conceptos:

- **Escaneado:** refiere al **proceso de digitalización que convierte una imagen impresa (analógica) en una imagen digital**. El resultado de este proceso es, por tanto, y desde el punto de vista de un SIG, una **capa ráster** (el escaneo no arroja una matriz de pixels). Una vez digitalizado el dato puede utilizarse dentro de un SIG o cualquier otro software tal como, por ejemplo, un software de tratamiento de imágenes.



- **Digitalización en pantalla o vectorización:** implica la **creación manual de entidades asistida por la capacidad de alinear (manual o automáticamente) a las celdas ráster**, y la utilización del trazado ráster y herramientas de reconocimiento de forma.

Un **escáner** o **scanner de computadora** es un dispositivo que se utiliza para copiar (mediante el uso de la luz), imágenes impresas o documentos a formato digital (a color o a blanco y negro). Nace en 1984 cuando Microtek

crea el MS-200, el primer escáner blanco y negro que tenía una resolución de 200dpi (200 dots per inch / 200 puntos por pulgadas).

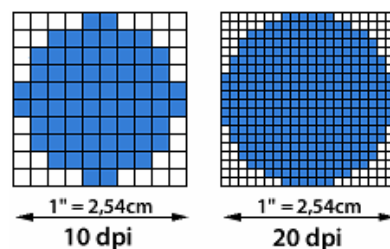


Al obtenerse una imagen digital, es posible corregir defectos, recortar un área específica de la imagen o también digitalizar texto mediante técnicas de OCR. Estas funciones las puede llevar a cabo el mismo dispositivo o aplicaciones especiales.

Los parámetros básicos que definen las características de un escáner son:

- **Resolución espacial:** Se mide en puntos por pulgada (DPI, "dots per inch") y nos indica el número de pixels que el sensor es capaz de tomar por cada unidad de longitud (pulgadas) sobre el papel.

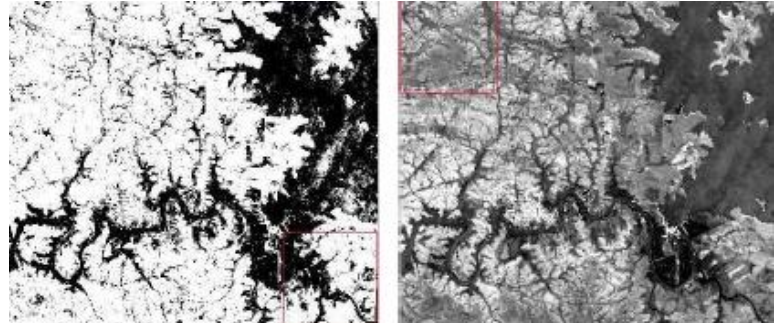
Los **DPI** es una unidad de medida para resoluciones de impresión y escaneado, concretamente, el número de puntos individuales de tinta que una impresora o tóner puede producir en un espacio lineal de una pulgada, o la cantidad de líneas que un scanner puede leer por pulgada. Las resoluciones habituales son entre 100 y 2500 DPI; para una misma carta, un scanner de mayor resolución espacial, genera imágenes digitales más grande que uno de menor tamaño.



- **Resolución radiométrica:** Resolución o sensibilidad radiométrica hace referencia al número de niveles digitales utilizados para expresar los datos recogidos por el sensor. En general, cuando

mayor es el número de niveles mayor es el detalle con que se podrá expresar dicha información.

Un scanner de baja resolución radiométrica levantará imágenes de menor calidad en la definición de los elementos del mapa.



Para determinados tipos de trabajo es mejor utilizar un scanner de baja resolución radiométrica. Veamos un ejemplo de esto, para digitalizar un plano topográfico, donde predominan los colores blanco y negro (sin niveles de grises) es mejor utilizar un scanner de baja resolución radiométrica.

Al scanear una carta o plano, **es importante tener en cuenta el tamaño del pixel sobre la imagen en relación al tamaño del píxel sobre el terreno.**

Para esto, es fundamental tener dos datos: en primer lugar, **la resolución (en DPIs) de la imagen resultante**; en segundo lugar, **la escala original de la carta**. A modo de ejemplo: tenemos una carta a escala **1/50000** scanneada a una resolución de **300 DPI**.

Para comenzar, tenemos que conocer la cantidad de filas por centímetro lineal que tiene nuestra imagen. Sabiendo que **1 pulgada** corresponde a **2.54 centímetros** tenemos que:

$$300 \text{ DPI} = \frac{300 \text{ Filas}}{1 \text{ pulgada}} = \frac{300 \text{ Filas}}{2.54 \text{ cm}} = 118.11 \text{ Filas/cm}$$

Esto significa que nuestra imagen scanneada tiene **118.11 filas en un centímetro lineal**. Veamos ahora a cuanto corresponde efectivamente una línea. Esto se hace a través del cociente de 1cm entre 118.11.

$$\frac{1 \text{ cm}}{118.11} = 0.08466 \text{ mm}$$

Teniendo en cuenta el factor de escala, debemos ver a cuanto corresponde **0.08466 mm** sobre la superficie de terreno. Esto se hace multiplicando este

último valor por **50.000**.

$$0.08466 \text{ mm} \times 50000 = 4.23 \text{ metros}$$

Es decir, cada píxel representa **4.23 metros** sobre la superficie terrestre. ¿Y qué significa esto? Es sencillo, ninguna entidad geográfica podrá tener mayor precisión planimétrica que 4.23 metros.

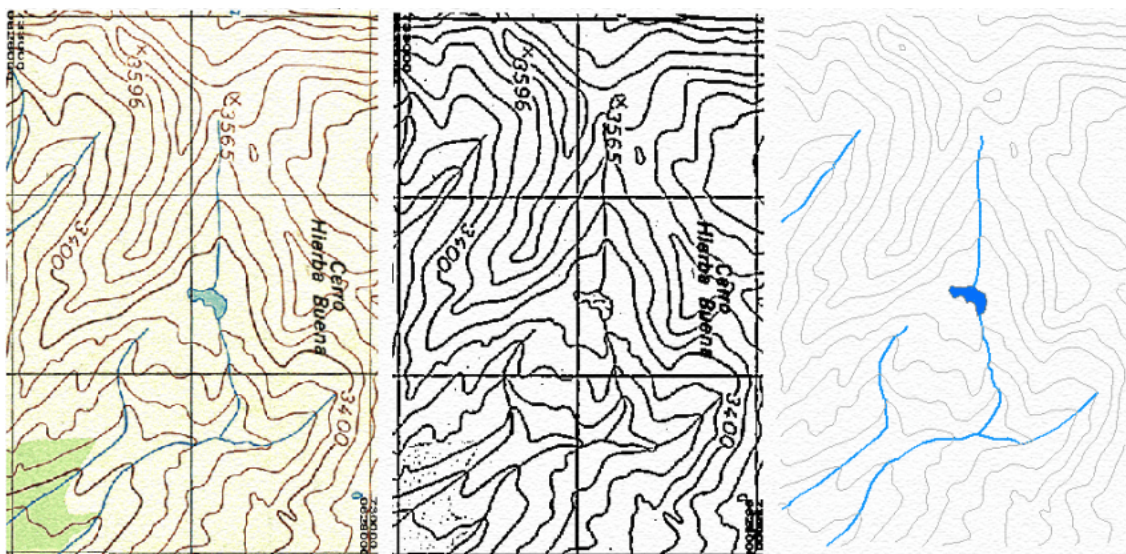


Tal como se ve, es simplemente una cuestión de multiplicar y dividir. El estudiante podrá llevar todos estos cálculos a una sola ecuación matemática.

Ejercicios como el que acabamos de presentar pueden ser propuestos durante el parcial.

Veamos ahora el tema de la vectorización o digitalización en pantalla. Tal como ya hemos mencionado, existe una gran cantidad de documentos impresos como mapas, planos, cartas topográficas; **esta información requiere ser digitalizada (o vectorizada) para poder integrarla dentro de un SIG.**

Uno de los principales problemas a enfrentar es el gran esfuerzo para efectuar dicho trabajo, todo este **proceso de integrar una copia dura (documento impreso) que ha sido scaneado a un medio digital de forma manual se conoce como digitalización.**



Existen dos mecanismos de digitalización:

- **Digitalización manual:** Para digitalizar o vectorizar manualmente, se debe contar con una imagen de referencia sobre la cual se

pueden discernir elementos de interés como infraestructura, coberturas vegetales, usos del suelo, cuerpos de agua. Estas imágenes de referencia pueden ser cartas escaneadas, imágenes satelitales, fotografías aéreas. Previamente se deben haber creado los sets de datos geográficos para digitalizar cada aspecto de interés por separado. Este proceso tiene la ventaja de que se hace con precisión y según el criterio de la persona que lo está realizando. Vectorizamos lo que vemos, utilizando las herramientas propias de cada uno de los softwares, cosa que hemos visto en la segunda parte de la Entrega 3.

- **Digitalización automática:** La vía automática requiere en primer lugar, **clasificar el valor de los píxeles de la imagen**. Tal como ya comentamos, los valores de cada píxel viene de la mano de la resolución radiométrica con que cuenta el sensor para discriminar o separar los valores que recibe espectralmente de la carta o superficie terrestre. Pero estos valores no son los apropiados para separar por ejemplo una cobertura de bosque de una de pastos o de rastrojos. Por ello se deben, analizar la imagen, establecer los rangos de los valores radiométricos que determinan un tipo específico de cobertura. Una vez sabido esto, se pasa a hacer una reclasificación de los píxeles de la imagen cambiando los valores que traen por defecto, a los que deseamos que tengan, según tengamos codificados nuestros criterios.

A partir de esto, se genera una nueva imagen con los píxeles valorados según nuestros criterios. Lo que sigue es hacer la vectorización automática mediante las herramientas apropiadas. **El proceso general de vectorización automática nunca es limpio**, es decir, se generan polígonos que, o son muy pequeños o no corresponden a los usos que definimos, por lo tanto es necesario hacer un chequeo de los resultados y ajustarlos mediante las herramientas de edición vectorial.

Antes de terminar podemos hacer una aclaración, **no siempre scannear un mapa implica una posterior vectorización del mismo**. Muchas veces las

cartas son digitalizadas para oficiar de mapa base o simplemente para evitar el deterioro.

Vectorización manual en Tableta Digitalizadora

Esta metodología de digitalización no requiere el uso de un scanner; pero utiliza un dispositivo llamado **tableta digitalizadora**.



Estos dispositivos son una superficie plana (como una mesa inclinada), sobre la cual se sitúa la copia dura a digitalizar; sobre este **se van dibujando las distintas entidades con un cursor especial**. Este cursor registra los movimientos del operario, convirtiendo las posiciones del cursor en coordenadas reales, que son las que van a constituir las distintas entidades.

Esté método de digitalización permite generar **entidades vectoriales**, mediante el trazado de la forma de esta o, en punto, sencillamente indicando su localización.

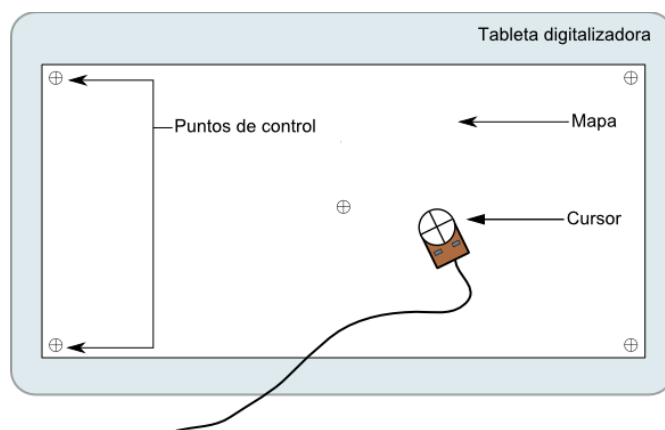
Existen varias versiones de tabletas digitalizadoras, fundamentalmente asociado al tamaño de la misma, o visto de otra forma, al tamaño máximo de la copia dura que se puede digitalizar.



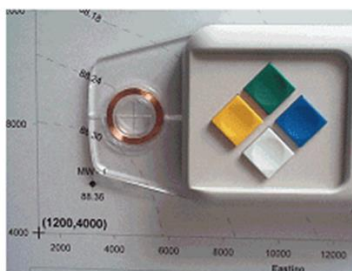
La persona que digitaliza, sigue con el cursor las formas de las distintas entidades (como si las estuviera calcando) para indicar al sistema las geometrías que se quieren definir.

En general el proceso de digitalización con este dispositivo es muy sencillo y conlleva tres pasos:

1. **Registro.** Esta etapa garantiza que las coordenadas de las entidades digitalizadas sean correctas. El mapa se fija a la tableta, y se señalan los **puntos de control** de coordenadas conocidas.



2. **Digitalización.** Se calcan las entidades puntuales, lineales y poligonales utilizando el puntero de la tableta.



3. **Asignación de la componente temática.** A cada una de las entidades digitalizadas se le añaden sus correspondientes atributos temáticos. Este paso no se realiza ya con la tableta digitalizadora.

Reconocimiento óptico de caracteres

El reconocimiento óptico de caracteres (OCR), es **un proceso dirigido a la digitalización de textos, los cuales identifican automáticamente a partir de una imagen símbolos o caracteres que pertenecen a un determinado**

alfabeto, para luego almacenarlos en forma de datos digitales. Así podremos interactuar con estos mediante un programa de edición de texto o similar.

Existen un gran volumen de datos en papel; esta información es solamente alfanumérica (texto). Por otro lado, se generan continuamente grandes cantidades de información escrita en todo tipo de soportes.

Muchas veces es necesario poder automatizar la introducción de caracteres evitando la entrada por teclado; sea esto por temas de ahorro de recursos humanos y/o temas de tiempo. En los últimos años la digitalización de la información ha devenido un punto de especial interés para la sociedad. En el caso concreto de los textos, existen y se generan continuamente grandes cantidades de información escrita, tipográfica o manuscrita en todo tipo de soportes. En este contexto, **la automatización de la introducción de caracteres evitando la entrada por teclado implica un importante ahorro de recursos humanos y un aumento de la productividad**, al mismo tiempo que se mantiene, o hasta se mejora, la calidad de muchos servicios.

A modo de ejemplo, podemos citar los siguientes sitios WEB para que los estudiantes hagan pruebas del uso de herramientas OCR online:

<https://www.onlineocr.net/>

www.newocr.com/

www.free-ocr.com/

Procesos de Geo-referenciación de un modelo ráster

Si bien la geo-referenciación se puede aplicar al modelo vectorial y al modelo ráster, en este curso nos centraremos en el proceso de geo-referenciación del datos en el modelo ráster.

Como ya hemos comentado en este y otros módulos, los datos ráster se obtienen, comúnmente, al escanear mapas o recopilar fotografías aéreas e imágenes de satélite. **Algunos de estos modelos rásters tienen definidos, de origen, el sistema de referencia; tal es el caso de las imágenes de satélites y de las fotografías aéreas digitales. Pero otros, no; tal es el caso**

de las cartas escaneadas. Por lo general, no contienen información de referencia espacial.

También puede ocurrir que, en las fotografías aéreas y las imágenes de satélite, a veces la información de ubicación que se entrega con ellas no es adecuada y los datos no se alinean correctamente con otros datos que ya tiene.

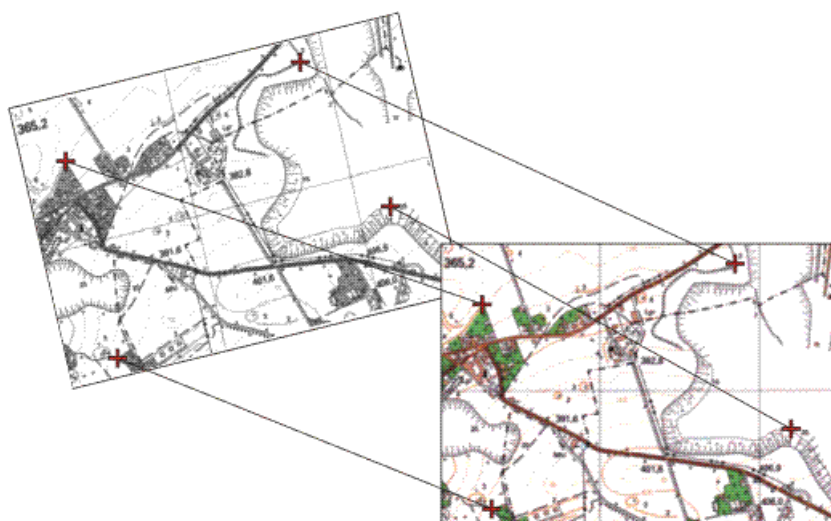
Muchas veces, para utilizar algunos datos en modelo ráster junto con los otros datos geográficos, es necesario alinearlos o georreferenciarlos en un sistema de referencia particular.

Como hemos visto en el segundo módulo, un sistema de referencia se define mediante una proyección de mapa (un método por el cual la superficie curva de la tierra se representa en una superficie plana).

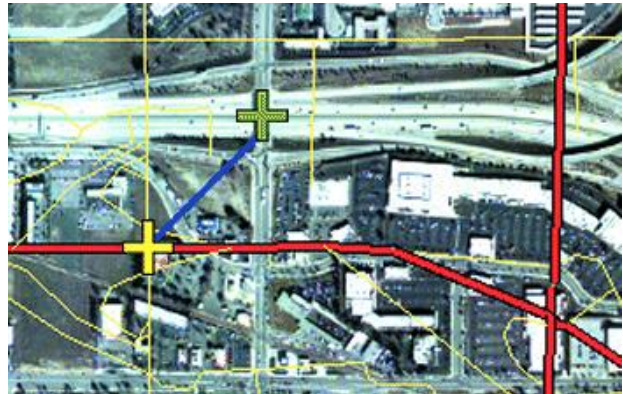
El proceso de georreferenciación de datos en el modelo ráster, permite definir su ubicación mediante coordenadas de mapa y asigna el sistema de coordenadas del marco de datos. La georreferenciación de datos de ráster permite visualizarlos, consultarlos y analizarlos con otros datos geográficos.

En síntesis, **la geo-referenciación es el proceso de asignación de coordenadas geográfica a elementos existentes.**

Se basa en conocer coordenadas de algunos lugares de la imagen. Estas coordenadas pueden ser obtenidas de muchas formas: por trabajos de campo o a partir de una fuente de datos espaciales anterior.



Estas coordenadas conocidas son elegidas como **puntos de control**. Hacen que la imagen original se deforme para adoptar la forma y caber dentro del sistema de coordenadas elegido.



El proceso implica la identificación de una serie de puntos de control del terreno, conocidos como coordenadas X, Y, que vinculan ubicaciones del modelo ráster con ubicaciones de los datos relacionados espacialmente (datos de destino).

Los puntos de control son ubicaciones que se pueden identificar con precisión en el ráster y en coordenadas del mundo real. Se pueden usar: intersecciones de corrientes o caminos, afloramientos de roca, el extremo de una punta de tierra, la esquina de un campo establecido, esquinas de calles o la intersección de dos cercos. En definitiva, cualquier elemento (de la naturaleza o humano) que sea fijo e identificable, puede ser candidato para un punto de control. **Generalmente se eligen un conjunto de puntos de control para llevar a cabo la geo-referenciación, y otros que se utilizan para evaluar la calidad del proceso.**

Otro elemento a tener en cuenta es la elección del conjunto de puntos de control. **Estos deben estar uniformemente distribuidos en el ráster a efectos de evitar deformaciones producto de la concentración de los puntos.**

Los puntos de control se utilizan para generar una transformación que desplazará el ráster desde su ubicación existente a la ubicación espacialmente correcta.

Entendemos como **vínculo (o vector, o link)** a la **conexión entre un punto de control del ráster (el punto de partida) y el punto de control correspondiente de los datos de destino alineados (el punto de destino).**



La cantidad de vínculos que son necesarios dentro del proceso de geo-referenciación crear depende de la complejidad de la transformación que necesitamos utilizar para transformar el modelo ráster en coordenadas de mapa.

Sin embargo, no siempre agregar más vínculos resultará necesariamente en un mejor registro. Como ya comentamos, es una buena práctica expandir los vínculos por todo el ráster en lugar de concentrarlos en un área. **Normalmente, tener al menos un vínculo junto a cada esquina del dataset ráster y unos cuantos por el interior genera los mejores resultados.**

Generalmente, cuanto mayor sea la superposición entre el dataset ráster y los datos de destino, mejor será la alineación, porque tendrá puntos más espaciados con los que georeferenciar el dataset ráster. Por ejemplo, si sus datos de destino solo ocupan un cuarto del área de su dataset ráster, los puntos que podría utilizar para alinear el dataset ráster se confinarían a esa área de superposición. Así, no es probable que las áreas situadas fuera del área de superposición se alinearan correctamente.

Otro elemento a tener en cuenta es que los datos geo-referenciados son menos precisos que los datos con los que se alinean. Para minimizar errores, debemos geo-referenciar a datos de la máxima resolución y la escala más grande para nuestras necesidades.

Transformación polinómica

Una vez que hayamos creado suficientes vínculos, podremos transformar, o combar, el ráster para combinar de forma permanente las coordenadas de mapa de los datos de destino.

Es posible utilizar distintos tipos de transformación, basadas en polinomios, o un SPLINE, o transformación proyectiva para determinar la ubicación correcta de las coordenadas de mapa para cada celda en el ráster.

La transformación polinómica utiliza un polinomio basado en puntos de control y un algoritmo de adecuación por mínimos cuadrados.



La transformación polinómica está optimizada para la precisión global, pero no garantiza la precisión local.

La transformación polinómica produce dos fórmulas:

- Una fórmula para calcular la coordenada X de salida de una ubicación de entrada (X, Y)
- Otra fórmula para calcular la coordenada Y de una ubicación de entrada (X, Y).

El objetivo del algoritmo de adecuación por mínimos cuadrados es derivar una fórmula general que se pueda aplicar a todos los puntos, normalmente a costa de un ligero movimiento de las posiciones de los puntos de control.

Como veremos a continuación, la cantidad de puntos de control requeridos para este método depende del orden de la transformación polinómica. Los polinomios del menor orden tienden a generar un error de tipo aleatorio, mientras que los polinomios de mayor orden tienden a generar un error de extrapolación.

La cantidad de puntos de control viene dada por la ecuación:

$$Cantidad = [(orden + 1) * (orden + 2)] / 2$$

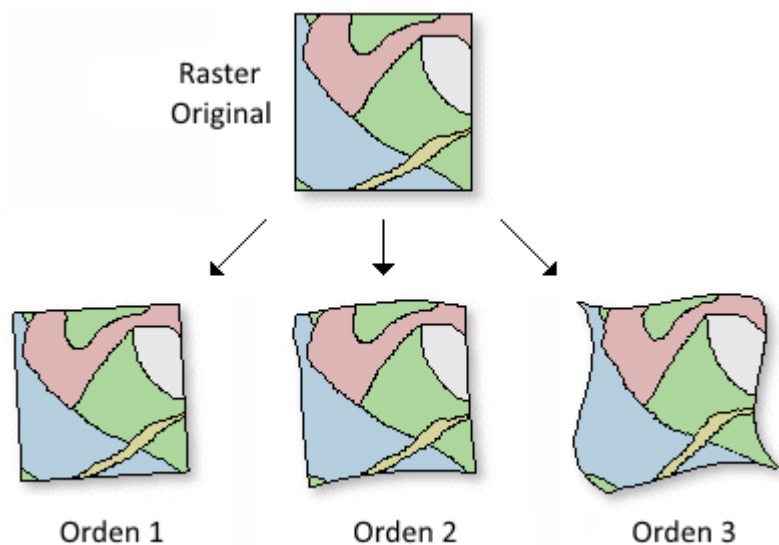
Es decir, la cantidad de puntos de control (o de vínculos) requeridos para este método debe ser de 1 para un cambio de orden cero, 3 para un primer orden, 6 para un segundo orden y 10 para un tercer orden.

Una transformación polinómica de orden cero se utiliza para desplazar los datos. Esto se utiliza generalmente cuando los datos ya están georreferenciadas, pero un pequeño cambio alinea mejor los datos. Solo se necesita un vínculo para realizar un cambio con polinomio de orden cero. Podría ser una buena idea crear unos cuantos vínculos y luego elegir el que parezca ser el más preciso.

Por otro lado, **la transformación afin de primer orden para desplazar, escalar y girar un modelo ráster**. Esto normalmente da lugar a líneas rectas en el ráster original representadas como líneas rectas en el ráster resultado. Así, los cuadrados y los rectángulos del dataset ráster se suelen cambiar a paralelogramos de ajuste de escala arbitraria y orientación angular.

Con un mínimo de tres vínculos, la ecuación matemática utilizada con una transformación de primer orden puede representar exactamente cada punto ráster en la ubicación de destino. **La presencia de más de tres vínculos introduce errores, o errores residuales, que se distribuyen por todos los vínculos**. Sin embargo, debemos agregar más de tres vínculos, porque si uno se coloca mal, tiene un impacto mucho mayor en la transformación. Así, aunque el error de transformación matemática pueda aumentar al crear más vínculos, la precisión general de la transformación también aumentará.

Cuanto mayor sea el orden de la transformación, más compleja será la distorsión que se puede corregir. Sin embargo, **raramente se necesitan transformaciones de más de tercer orden**. Las transformaciones de orden superior necesitan más vínculos y, por tanto, implican más tiempo de procesamiento. En general, si es necesario extender, escalar y girar un modelo ráster, alcanza con una transformación de primer orden. No obstante, si hay que “doblar” o “curvar” el modelo ráster, debemos utilizar una transformación de segundo o tercer orden.



Transformación SPLINE

La transformación por SPLINE es un verdadero método de deformación elástica vectorial y **optimiza para la exactitud local, pero no para la global.**

Se basa en una función por SPLINE, un polinomio por partes que mantiene la continuidad y suavidad entre polinomios adyacentes. Las transformaciones por SPLINE transforman los puntos de control de origen exactamente en puntos de control de destino; no se garantiza que los píxeles que están a una distancia de los puntos de control sean precisos.

Esta transformación es útil cuando los puntos de control son importantes y se necesita que se registren de forma precisa. Agregar puntos de control puede aumentar la precisión general de la transformación por SPLINE. **Ésta necesita un mínimo de 10 puntos de control.**

La transformación de ajuste optimiza el LSF global y la precisión local. Se basa en un algoritmo que combina una transformación polinómica y técnicas de interpolación de red irregular de triángulos (TIN). La transformación de ajuste realiza una transformación polinómica mediante dos conjuntos de puntos de control y ajusta los puntos de control de forma local para combinar mejor los puntos de control de destino mediante una técnica de interpolación de TIN. Necesita un mínimo de tres puntos de control.

Error cuadrático medio

Cuando se obtiene la fórmula general y aplica al punto de control, se devuelve una medida del error (el error residual). A modo de ejemplo, en una transformación polinómica, cada punto de control va a tener errores, o sea una diferencia entre las coordenadas ingresadas (donde queremos que el punto esté localizado) y las coordenadas finales (donde el punto resulta al final).

El error es la diferencia entre dónde terminó el punto de partida, frente a la ubicación real especificada (la posición del punto de destino). El error

total se calcula mediante la suma cuadrática media (RMS) de todos los errores residuales para calcular el error RMS.

$$RMS = \sqrt{\frac{\sum_{i=0}^n a_i^2}{n}}$$

En la formula a_i es el error en cada punto de control.

Este valor describe el grado de coherencia de la transformación entre los distintos puntos de control (vínculos). Cuando el error es especialmente grande, podemos quitar y agregar puntos de control para ajustarlo.

Aunque el error RMS es una buena evaluación de la precisión de la transformación, no debemos confundir un error RMS bajo con un registro preciso. Por ejemplo, la transformación todavía puede contener errores considerables debido a un punto de control mal introducido. Cuantos más puntos de control de calidad equivalente se utilicen, más preciso será el polinomio a la hora de convertir los datos de entrada en coordenadas de salida. Normalmente, las transformaciones por spline y de ajuste dan un error RMS prácticamente cero o cero; sin embargo, esto no significa que la imagen se georreferencia perfectamente.

El residual hacia adelante le muestra el error en las mismas unidades que la referencia espacial del marco de datos. El residual inverso muestra el error en las unidades de píxeles. El residual inverso hacia adelante es una medida de qué tan cercana está su exactitud, medida en píxeles. Todos los valores residuales más cercanos a cero se consideran más precisos.

Calidad de la Información geográfica

Si bien el tema de la información geográfica forma parte de un curso completo, veremos algunos conceptos que son necesarios a la hora de cerrar el este módulo.

Tal como hemos expuesto en las secciones anteriores y comentado ampliamente en clase, **los procesos de adquisición de la información**

geográfica (y su componente temática) deben tratar de ser lo más cercano posible a la calidad original de la información.

La digitalización no es por completo perfecta: **siempre existen deficiencias y fuentes de errores.**

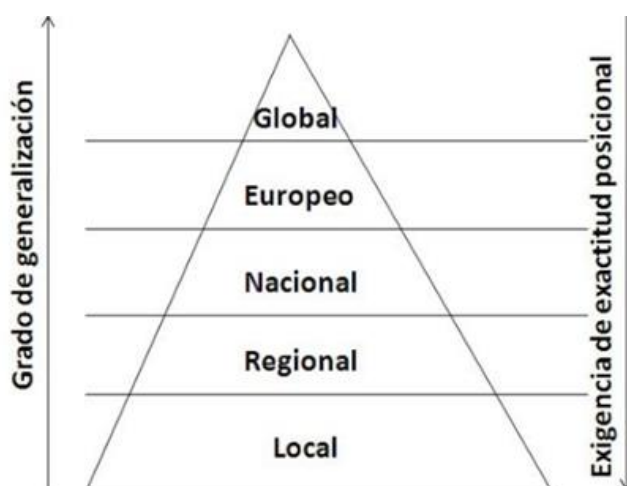
En este sentido tenemos que tener varias causas de errores y deficiencias, pero las dos principales son:

- **Errores del dato original**, es decir, de la fuente de datos.
- **Errores del proceso** de digitalización o captura de la información.

Por otro lado debemos tener en cuenta dos requerimientos fundamentales que afectan el uso de los datos e información geográfica:

- Necesidad de utilizar la información geográfica **procedente de distintas fuentes y con calidades distintas.**
- **Escalabilidad** de los datos.

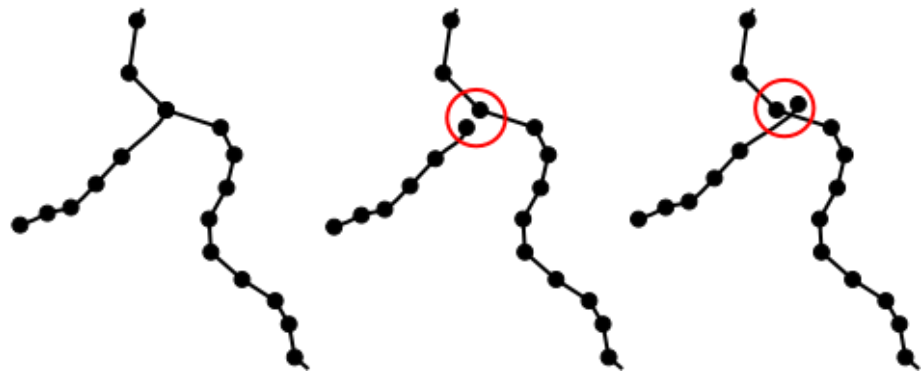
El siguiente cuadro describe la relación de **la interoperabilidad de entre los distintos niveles de creación y el uso de la información geográfica.** El mismo es extraído del curso de *"Introducción a la evaluación de la calidad de la información geográfica"*.



Calidad en los procesos de digitalización

Existen una amplia variedad de errores que son producto de los procesos de digitalización, sea esta en pantalla o en tableta. Un ejemplo de esto son

los llamados errores de conexión. Sean estos por exceso o por defecto, ambos se ilustran en la imagen que se presenta a continuación.



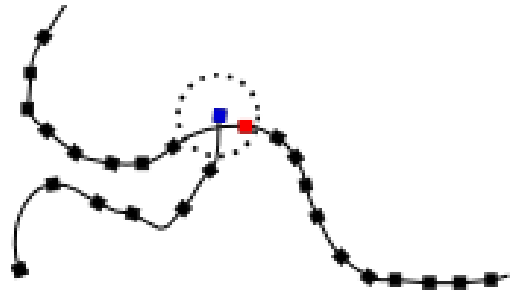
Progresivamente, los software SIG han ido incorporando funcionalidades que permiten evitar estos errores en el momento de la digitalización. Esto es una clara influencia de las herramientas CAD.

El modelo vectorial permite conocer las coordenadas de todos los elementos, gracias a esto podemos alcanzar una **exactitud** y **precisión perfecta**; algo imposible de lograr sin estas funcionalidades. Por ejemplo: Para garantizar la coincidencia entre los distintos vértices, podemos establecer **tolerancias** y **ajuste automático** en función de ellas (esto se conoce con el término inglés *snapping*).

Este tipo de funcionalidad de los GIS, permiten que los los polígonos adyacentes o líneas que se cortan en un punto dado lo hacen con total exactitud. De esta forma, los polígonos compartirán exactamente el mismo lado con las mismas coordenadas exactas, o se cruzan en el mismo e idéntico punto, y no únicamente pasan por un punto cercano (pero distinto) definido con la precisión con la que el operador haya podido ajustar ambas entidades visualmente. **Esto hace que las coincidencias no sean sólo visuales, sino que ahora son numéricas.**



En el siguiente ejemplo, el nodo azul representa el "nodo en edición". La "tolerancia" queda marcada por el círculo punteado. Como el nodo rojo de la línea existente se encuentra dentro de esa tolerancia, al añadir el nuevo nodo, este se situará en las coordenadas del nodo rojo.



Es importante conocer las herramientas de edición, las herramientas de snapping, los métodos de generación de cada tipo de entidad del software que utilizará en el futuro. Se debe tener en cuenta que, muchas veces, se debe decidir entre invertir un poco más de recursos económicos a la hora de seleccionar un software propietarios, o destinar los mismos recursos económicos para contratar recursos humanos que detecten y corrijan posibles errores.

Para finalizar: una muy breve introducción al informe de calidad de datos

Como ya comentamos el tema calidad de los datos e información geográfica será tratado en otro curso. Ampliaremos algunos de los conceptos detallados en secciones anteriores, e introduciremos algunos nuevos como parte final de este módulo.

Como ya comentamos, **todo dato espacial contiene (en mayor o menor medida) algún tipo de error**. Por lo tanto, es importante **conocer las razones por las cuales aparecen esos errores, para poder evaluar correctamente la validez del trabajo que realizamos con los datos y los resultados que obtenemos a partir de ellos**.

Apuntamos simultáneamente a cuatro cosas:

- 1) **Identificar** de la fuente de error;
- 2) **Detectar y medir** el error;

- 3) **Modelar la propagación** del error;
- 4) **Proponer estrategias** para la gestión y reducción del error.

¿Por qué esto es tan importante? Por la **aparición de los SIG** y por el **crecimiento del volumen de datos espaciales disponibles**, especialmente los derivados de satélites.

Solo recientemente, los usuarios y desarrolladores de los Sistemas de Información Geográfica han prestado atención a los problemas causados por el error, la exactitud y la imprecisión en el conjunto de datos espaciales. Ciertamente, existía la conciencia de que todos los datos contenían cierta inexactitud e imprecisión, pero su efecto en los problemas y soluciones de los SIG no habían sido considerados con gran detalle

Progresiva se ha ido incorporando el reconocimiento general de que el error, la inexactitud y la imprecisión pueden hacer fracasar algunos tipos de proyectos SIG. O sea, los errores no detectados, pueden dejar sin valor algunos de los análisis SIG



Lo irónico del presente planteo está en que el problema del error es inherente a uno de las grandes potencialidades de los SIG.

Gran parte de las soluciones aportadas por los SIG son posibles gracias a que cotejan y cruzan diversos tipos de datos con localización. Esto es particularmente útil al posibilitar integrar diversos conjuntos de datos discretos bajo un único sistema.

Desafortunadamente, cada vez que se importa un nuevo conjunto de datos, el SIG arrastrará el error inherente a los mismos.

¡La mezcla y combinación de errores puede llevar al conjunto de datos por caminos impredecibles!

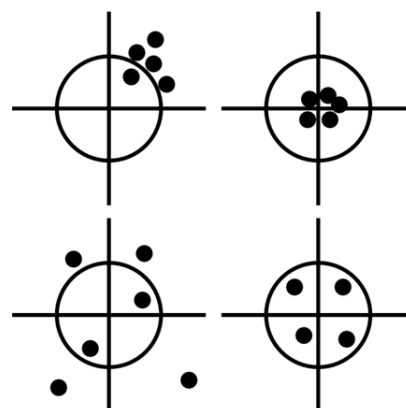
El punto central está en que a través del error se pueden alterar los análisis del SIG, que hay diversos caminos para reducir el error al mínimo mediante una cuidadosa planificación y métodos para estimar sus efectos en las soluciones SIG. El conocimiento del problema del error ha tenido como consecuencia beneficiosa el hacer sensible a los usuarios de los SIG de las potenciales limitaciones inherente al proceso para alcanzar exactitud y precisión en las soluciones.

Llegó el momento de definir dos conceptos que muchas veces pueden ser similares para los usuarios:

- **La Exactitud es el grado en el cual la información de un mapa o en una base de datos digital se muestra verdadera o con valores aceptables.** La exactitud es un asunto perteneciente a la cualidad de los datos y al número de errores contenidos en un conjunto de datos o mapa. Analizando una base de datos de un SIG, es posible considerar la exactitud horizontal y vertical con respecto a la posición geográfica, tanto atributiva y conceptual, como en la agudeza lógica.
 - El nivel de exactitud requerido puede variar enormemente de unos casos a otros.
 - Producir y compilar una gran exactitud en los datos puede ser muy difícil y costoso.
- **La Precisión hace referencia a la medida y exactitud de las descripciones en las base de datos de un SIG. Los atributos de información precisos pueden especificar las características de los elementos con gran detalle.** Es importante observar, no obstante, que los datos precisos - no importando el cuidado en su medida - pueden ser inexactos. Los topógrafos pueden cometer errores, o bien los datos pueden ser introducidos en las bases de datos incorrectamente.
 - El nivel de precisión requerido puede variar enormemente de unos casos a otros. Los proyectos de ingeniería como el de una carretera, y las herramientas de construcción, requieren una muy precisa medida, de milímetros a decenas de centímetros. Análisis demográficos de las tendencias del electorado pueden prescindir de esta precisión mediante un código postal o de circunscripción.
 - Obtener datos altamente precisos puede ser verdaderamente difícil y costoso. Relevar topográficamente

las localizaciones requiere de compañías específicas para la recogida de la información.

En síntesis; la precisión (indica el nivel de detalle con el que se recoge la información) y la exactitud (indica el grado en que los valores estimados se asemejan al valor real.).



El **error posicional** es el que más a menudo concierne a los SIG, pudiendo afectar a diferentes características de la información almacenada en una base de datos. Veremos algunos de ellos:

- **Exactitud y precisión posicional:** Es aplicable tanto a la posición horizontal como a la vertical. Exactitud y precisión están en función de la escala en la que ha sido creado el mapa (impreso o digital). Los mapas estándar empleados por el Servicio Geológico de los Estados Unidos (USGS) especifican que: *"se requiere una exactitud horizontal del 90 % en todos los puntos tomados que deben de estar entre 2,54 y 76.2 cm para mapas de escala superior a 1:20.000 y entre 2,54 y 127 cm para mapas de escala inferior a 1:20.000"*.

Tabla de precisiones estándar para algunas escalas de mapas	
1:1.200	± 1,015 m
1:2.400	± 2,033 m
1:4.800	± 4,063 m
1:10.000	± 8,467 m
1:12.000	± 10,159 m
1:24.000	± 12,192 m
1:63.360	± 32,187 m
1:100.000	± 50,800 m

Para escalas hasta 1:10.000, viene dada por la fórmula:

$$\text{precisión} = 0.0008 \times \text{escala} + 0.0004$$

A partir de 1:10.000 y hasta 1:100.000, viene dada por la fórmula:

$$\text{precisión} = 0.0005 \times \text{escala} + 2.6127$$

(Ambas fórmulas son aproximadas, pero dan una idea de la relación entre escala del mapa y la precisión. Siempre es preferible referirse a la tabla).

Esto significa que cuando vemos un punto en un mapa, tendremos esta probabilidad de que se encuentre dentro de una cierta área. Esto lo podemos hacer extensivo a las líneas.



Tal como hemos comentado desde el comienzo del curso, están los peligros de la falsa exactitud y de la falsa precisión, que son leídos en la información locacional desde los mapas con niveles de exactitud y precisión bajo los cuales han sido creados.

Esto es un verdadero peligro en las computadoras, ya que permite a los usuarios aumentar y reducir las vistas en un número infinito de escalas.

¡Exactitud y precisión están unidas a la escala original del mapa y no cambia aunque se use el zoom para aumentar o reducir la vista!

Estas operaciones pueden incluso hacer creer -falsamente- que la exactitud y la precisión son mejores.

- **Exactitud y precisión de los atributos:** Por su lado, la componente temática pueden ser inexacta o imprecisa. La inexactitud puede ser consecuencia de errores de distinto tipo. Los datos no espaciales pueden variar mucho también en precisión. La información precisa que los atributos describan fenómenos con gran detalle.

A modo de ejemplo, la descripción precisa de una persona que vive en una dirección particular puede incluir género, edad, ingresos, ocupación, nivel de educación y muchas otras características. Una descripción imprecisa puede incluir sólo los ingresos o sólo el género.

- **Exactitud y precisión conceptual:** Como ya hemos comentado, los SIG dependen sobretodo de la abstracción y la clasificación de

los fenómenos del mundo real; esto lo vimos al trabajar con los modelos geográficos, de representación y de almacenamiento.

Los usuarios determinan que cantidad de información debe usarse y como ésta debe ser clasificada en categorías apropiadas. En ocasiones, los usuarios pueden usar inapropiadas categorías o una clasificación errónea de la información.

Por ejemplo, la clasificación de ciudades por el comportamiento del voto electoral es una vía inadecuada para estudiar la fertilidad de las parejas; fallos en la clasificación de las líneas de alto voltaje puede limitar la efectividad en el diseño de un SIG en la construcción de las infraestructuras eléctricas.

Aun empleando correctas categorías los datos pueden estar mal clasificados. Un estudio de los sistemas de drenaje puede necesitar de una clasificación de las corrientes y ríos por su "orden", atendiendo su jerarquía al lugar donde una corriente particular desagua en el sistema tributario de la red. Los canales individuales pueden estar mal clasificados si los tributarios están mal localizados. Por ello, algunos estudios pueden no requerir un tipo preciso de categorización del orden de las corrientes. Todo lo más que pueden necesitar es la localización y el nombre de las corrientes fluviales, sin tener en cuenta el orden.

- **La lógica de la exactitud y precisión:** La información almacenada en una base de datos puede estar ilógicamente introducida, esto lo hemos problemas topológicos. Por ejemplo, los permisos necesarios para construir una subdivisión residencial en un plano de inundación pueden necesitar comparar la proposición con el mapa del plano de inundación. Por lo tanto, la construcción puede ser posible en algunas zonas del plano de inundación pero su uso no será conocido hasta que las variaciones de la inundación potencial hayan sido registradas y puedan ser usadas en la comparación. La cuestión es, pues, que la información almacenada en la base de datos de un SIG puede ser usada y cuidadosamente comparada, si produce resultados útiles. Los SIG están normalmente incapacitados para avisar a los usuarios cuando se

produce una inapropiada comparación o si los datos han sido utilizados incorrectamente. Algunas reglas topológicas de edición pueden ser incorporadas en el diseño de un SIG, como sucede con los "sistemas expertos".



Finalmente señalar, cometeremos una equivocación si creemos que una gran exactitud y una gran precisión de la información es necesario para todas las aplicaciones de los SIG.

La necesidad de exactitud y precisión puede variar radicalmente dependiendo del tipo de información codificada y del nivel de medida necesario para una particular aplicación.

Son los usuarios los que deben determinar el alcance de su trabajo.

Excesiva exactitud y precisión no sólo es costoso, sino también puede resultar muy engorroso.



Fuentes de error

A efectos de que el estudiante pueda ver lo complejo de la temática, veremos fuentes de error.

- **Antigüedad de los datos:** Las fuentes de datos pueden ser simplemente antiguas para ser usadas en un proyecto SIG. Las colecciones estándar del pasado pueden ser desconocidas, inexistentes o desfasadas.

Por ejemplo, los mapas donde se detallan las costas durante el siglo XVIII, contienen falta de precisión para ser utilizados hoy en día. Además, una parte de la información base puede, además, haber cambiado como consecuencia de la erosión, la deposición o cualquier otro proceso geomorfológico. Pese al poder de los SIG, la dependencia de datos antiguos puede tergiversar, sesgar o convertir en negativos los resultados.

- **Área de cobertura:** Los datos de una zona determinada pueden haber desaparecido completamente, o únicamente una parte de los niveles de información pueden ser utilizables en un proyecto SIG.

Por ejemplo, los mapas de vegetación o de suelo pueden estar incompletos en las zonas de transición o faltarles exactitud en la representación.

- **Escala del mapa:** La posibilidad de los mapas para mostrar detalles está determinada por la escala. Un mapa con una escala 1:1.000 puede ilustrar detalles más precisos que otro a una escala 1:25.000. La escala determina el tipo, calidad y cantidad de los datos. Debemos elegir la escala adecuada al nivel de detalles requerido en el proyecto. Transformar la pequeña escala de un mapa en otra más grande no aumentará el número de detalles o el nivel de precisión de dicho mapa.
- **Densidad de las observaciones:** El número de observaciones realizadas en un área es una guía de la verosimilitud del mapa y debe ser conocido por los usuarios del mismo. Un insuficiente número de observaciones puede no proporcionar el adecuado

nivel de resolución requerido para efectuar análisis espaciales y resolver los objetivos marcados en el proyecto SIG. En el caso de un punto, si las curvas de nivel poseen un intervalo de 120 cm, no es posible bajar el nivel de precisión. Las líneas de un mapa son una generalización basada en el intervalo de datos grabados, de este modo el más cercano al intervalo muestral, alcanza la mayor precisión de datos dibujados.

- **Relevancia:** Bastante a menudo, no es posible satisfacer los deseos de obtener datos de un lugar o de un área, y en su lugar se sustituyen estos datos por otros. Ha de existir una relación de validez entre los datos sustituidos y el fenómeno a estudiar, pero aun así pueden producirse errores en tanto en cuanto los fenómenos no han sido medidos directamente.
- **Formato:** Los métodos para transmitir, almacenar y procesar la información de forma digital, pueden introducir error en los datos. Las conversiones de escala y proyección, los cambios desde raster a vector y la resolución del tamaño y profundidad del píxel, son ejemplos de los posibles errores inherentes al formato de los datos. En ocasiones, los datos han de ser transmitidos y utilizados en múltiples SIG por lo que deben reformarse bajo un mínimo denominador común. Múltiples conversiones desde un formato a otro pueden crear un efecto similar a realizar copia tras copia en una máquina fotocopidora. Además, hay que tener en cuenta que los estándares internacionales para la transmisión, almacenamiento y recuperación de datos cartográficos no estén totalmente realizados.
- **Accesibilidad:** La accesibilidad de los datos es otra cosa. Lo que está disponible en un país puede no estarlo en otros. Antes de la desaparición de la Unión Soviética, no pocos mapas eran considerados como documentación clasificada y por lo tanto, imposibles de obtener por la mayor parte de la gente. Las restricciones militares, la rivalidad entre agencias, las leyes de privacidad y los factores económicos pueden restringir la validez de los datos o su nivel de exactitud.

- **Costo:** Extensos y veraces datos son, a menudo, demasiado caros de obtener o convertir. Iniciar una nueva colección de datos puede ser demasiado caro para los beneficios generados en un determinado proyecto SIG y sus diseñadores deben moverse entre su deseo de exactitud y el coste de la información. La verdadera exactitud es cara y puede ser inasequible.
- **Exactitud posicional:** La exactitud en la posición es una medida del desajuste entre los elementos del mapa y la verdadera posición de los atributos. Depende del tipo de datos usados u observados. Los cartógrafos pueden situar con exactitud objetos bien definidos, como carreteras, edificios, líneas divisorias y unidades topográficas discretas en mapas y en sistemas digitales, mientras que separaciones menos discretas como las existentes entre la vegetación o los tipos de suelo pueden ser el resultado de las estimaciones del cartógrafo. El clima, el relieve, los tipos de suelo, el drenaje y otros elementos faltos de una clara delimitación en la naturaleza, son susceptibles de ser interpretados. Defectos o trabajos parciales, errores de digitalización de mapas y de conversión en los mapas o en los escáner, pueden todos ellos producir mapas inexactos en un proyecto SIG.
- **Precisión en el contenido:** Los mapas deben de ser correctos y estar libre de presuposiciones. La precisión cualitativa hace referencia a la corrección en la clasificación y a la presencia de elementos específicos.

Por ejemplo, un bosque de pinos puede estar incorrectamente clasificado como un bosque de abetos, introduciendo de esta forma errores que no pueden ser conocidos o sospechados por el usuario de los datos o del mapa. Ciertos elementos pueden ser omitidos tanto desde el mapa como desde la base de datos espacial por descuido o intencionadamente.

- **Fuentes de variación de datos:** Las variaciones en los datos pueden realizarse por la introducción de errores de medida durante la observación, por la predisposición del observador o por falta de una adecuada calibración del equipamiento.

Por ejemplo, se puede no esperar precisiones submétricas con un GPS de mano sin corrector diferencial. Por otra parte, una incorrecta calibración en las formas de disolver el oxígeno puede producir valores incorrectos de concentración del mismo en una corriente.

- **Errores numéricos:** Diferentes ordenadores pueden no tener la misma capacidad para construir complejas operaciones matemáticas, pudiendo producir resultados significativamente diferentes desde un mismo problema.
- **Errores en los análisis topológicos:** Los errores lógicos pueden causar una incorrecta manipulación de los datos y de los análisis topológicos. Se pueden reconocer qué datos no son uniformes y están sujetos a variaciones.
- **Problemas de clasificación y generalización:** Para el entendimiento humano, la comprensión de una vasta cantidad de datos reside en su clasificación, y en algunos casos en su generalización.
- **Digitalización y errores geocodificados:** Los errores ocurridos durante el transcurso de las fases de manipulación de datos tales como la digitalización y la geocodificación, el recubrimiento y las intersecciones de los límites, y los errores de rasterización de un mapa vectorial.



Lecturas complementarias

Como habrán visto este módulo es un poco menos extenso que los anteriores; para finalizar dejamos esta lectura que analiza y compara algunos métodos de ingreso de datos espaciales.

- **Comparación de métodos de digitalización para el ingreso de información espacial a los Sistemas de Información Geográfica**
(paper de Carlos Pacheco, Alex Barrios y Juan López)

<http://www.saber.ula.ve/bitstream/123456789/24152/2/articulo4.pdf>