Geospatial Data: How to find it, how to assess it, how to teach about it, and why it matters

Joseph J. Kerski
Esri

The willingness to experiment with innovative approaches to GIS has yielded benefits for numerous field-based teaching and research activities across the university.

– *Peter Knoop | University of Michigan*

Purpose of this presentation

1. Empower you with skills to find geospatial data for your teaching, research, or project.

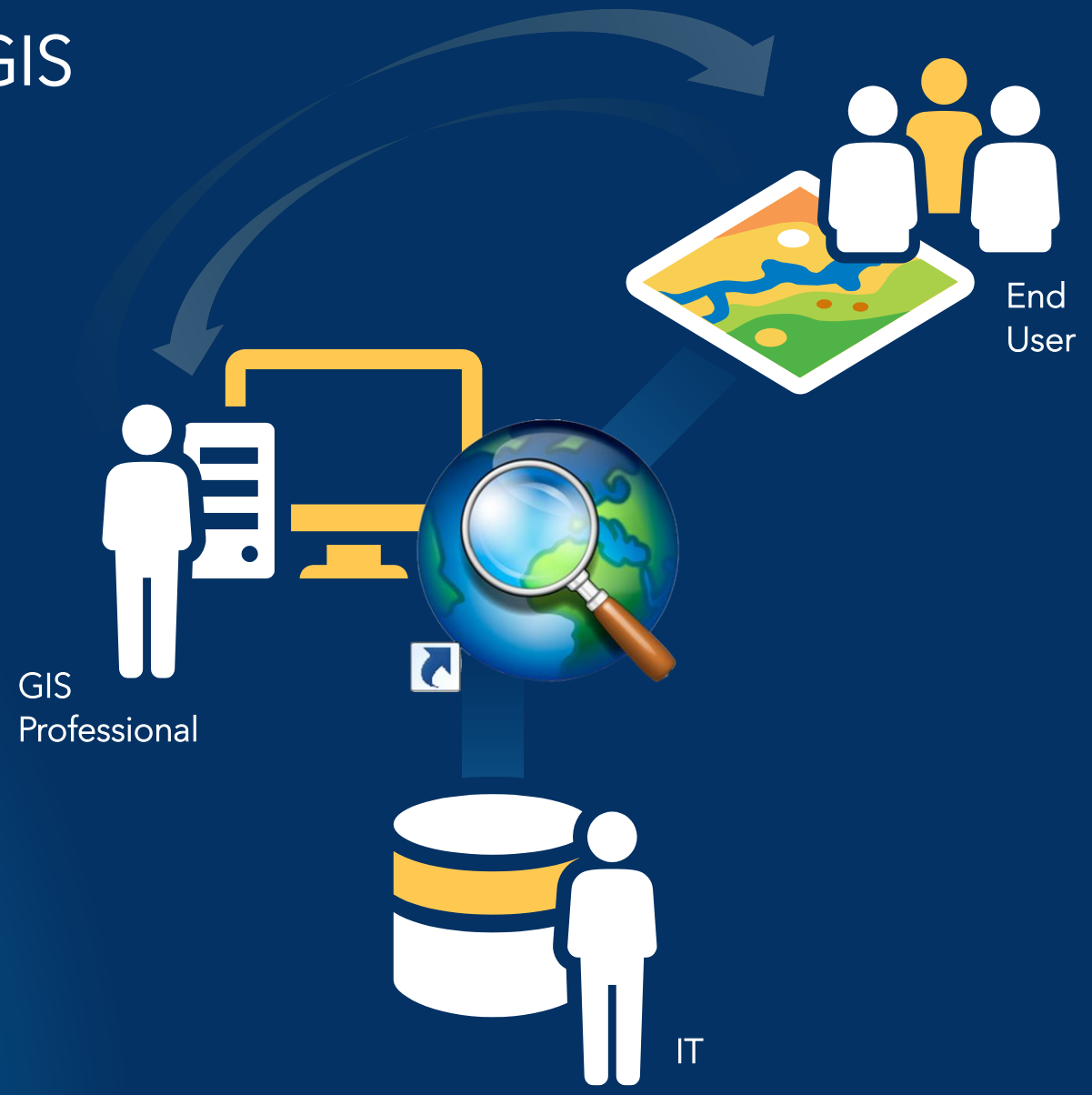
2. Enable you to be able to make wise decisions about whether your data is appropriate for your needs.

3. Provide you with resources to teach societal issues surrounding data, such as location privacy, data quality, copyright, and ethics.

4. Connect you to available Esri support and resources.

Data is a key component of GIS

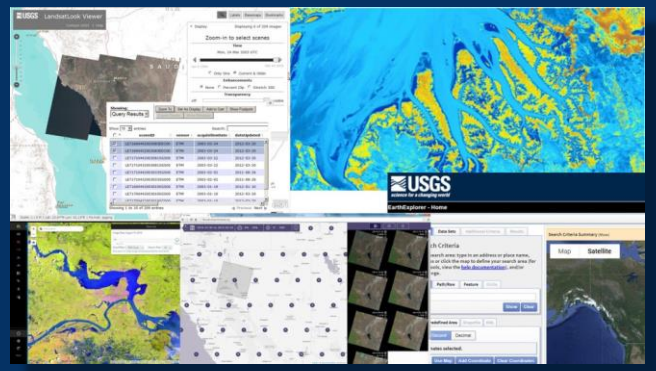
- Software
- Information Products
- **Data** ←
- Hardware
- People
- Process



Sources: Roger Tomlinson, Caitlin Dempsey

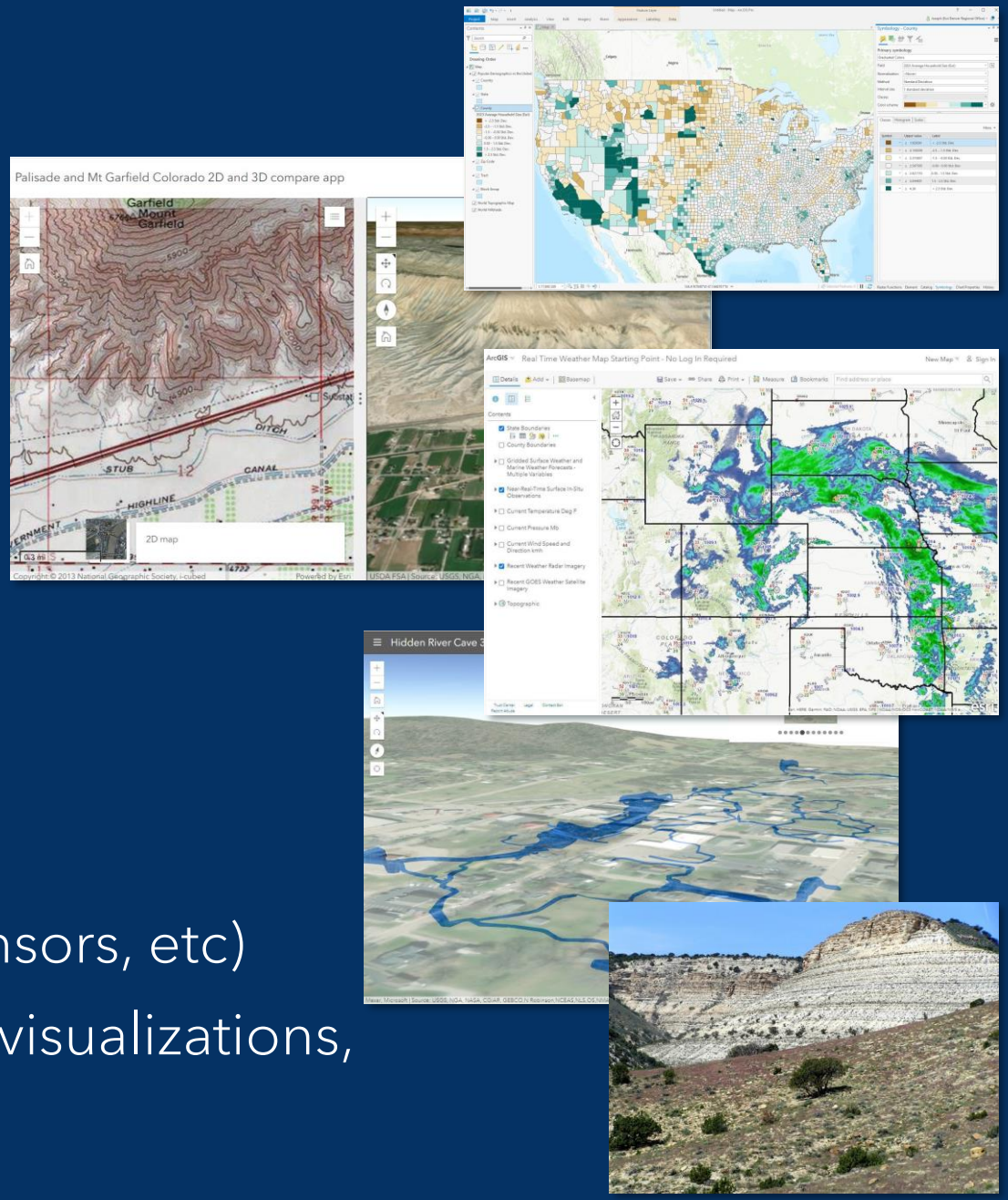
Geospatial Data Characteristics

- Is tied to a location or a set of locations.
- Often is combined with attribute information.
- Often contains a temporal component.
- Can be generated in a widening variety of ways: Crowdsourced surveys, Machine Learning/Artificial Intelligence techniques, feeds from the Internet of Things (IoT), UAVs and other expanding remote sensing techniques, unstructured data, tabular data, and more.

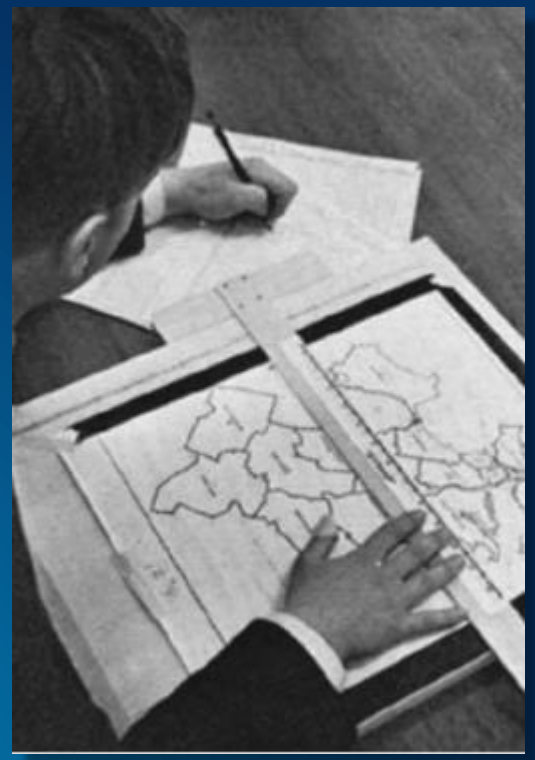


Geospatial Data Sources

- Tables and spreadsheets
- Field survey results
- Field observations
- Imagery from aircraft, UAVs, satellites
- Grids and Tiles
- GeoJSON and other formats
- Shapefiles and geodatabases
- Ground photographs
- IoT feeds (stream gauges, traffic cams, sensors, etc)
- Augmented reality, 2D and 3D maps and visualizations, more.



In the past, it was difficult to store, serve, find, and access geospatial data.



… for data providers *and* data users.

Today, geospatial data is offered with a greater volume, variety, and accessibility than ever before. It also frequently includes a temporal aspect.

Today, geospatial data can be accessed as downloadable files and as streaming services. Data as services is an **enormous leap forward**!

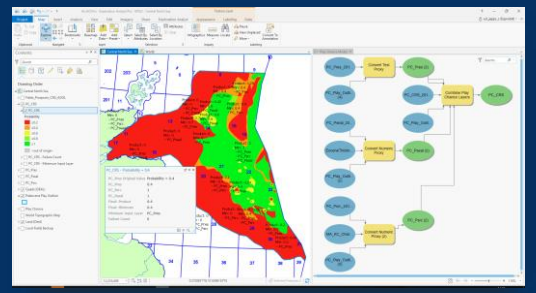
From field surveys

From Infographics

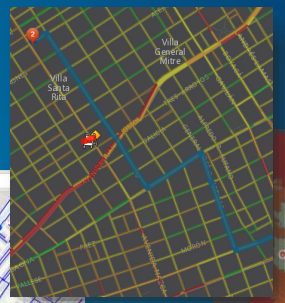
From portals

From geodatabases and shapefiles

Shared in models

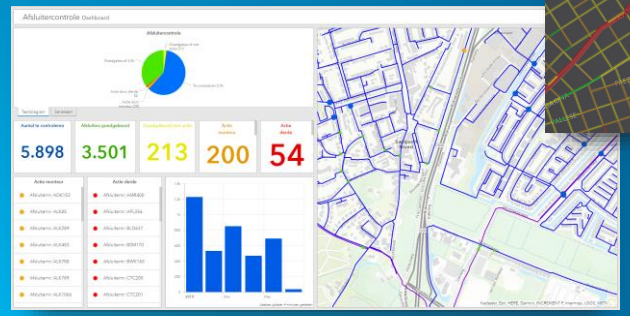


From vector and rasters

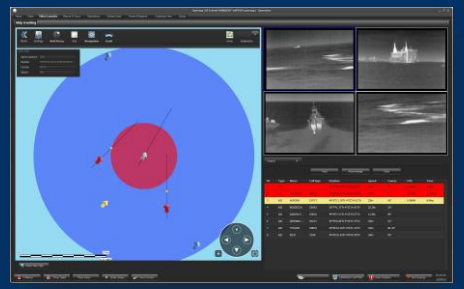


From Images

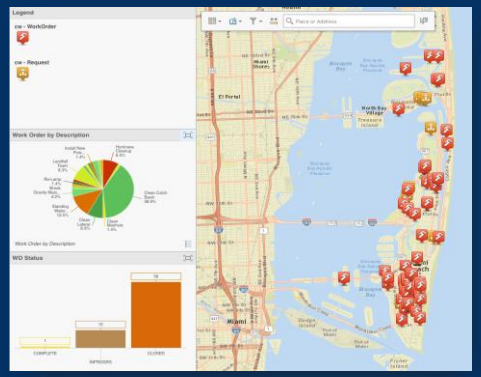
From Dashboards



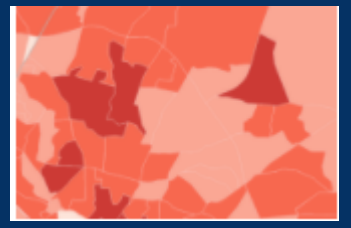
From real-time feeds



From streaming services



Example 1 of traditional downloading method vs new streaming method of accessing data: US Census Demographics



Traditional Method

Download TIGER files and demographic files for desired geographic unit. Perform attribute join to connect demographics to census geography.

Disadvantages: Long multi-step process. Only works for 1 level of geography at a time. Data is static. Cumbersome field names need to be translated. Must repeat for different countries. Desktop GIS.

New Method

Stream demographics from ArcGIS Living Atlas of the World directly to your GIS project.

Advantages: Reduces data acquisition time from hours to minutes. No processing required. Can be used in ArcGIS Online or ArcGIS Pro. Multi-scale. Works for many countries. Data often updates automatically.

Example 2 of traditional downloading method vs new streaming method of accessing data: Elevation

Traditional Method

Download elevation data from nationalmap.gov or other portal to GIS software, create hillshades, slope, aspect, and other derivative products.

Disadvantages: Long process to download individual blocks of terrain and create products one at a time. Edgematching issues. Desktop GIS. Must repeat for multiple countries.

New Method

Stream elevation service from ArcGIS Living Atlas of the World directly to your GIS project.

Advantages: Reduces data acquisition time from hours to minutes. No processing required. No edges to match. Stream data to match boundaries of your project. Can be used in ArcGIS Online or ArcGIS Pro. Multi-scale. Derivative products also available as services. Some data sets are global.

But even this is just the **beginning.**

One could argue that geospatial data was always “big data” – as the GIS community has always been keen on accessing and using large amounts of vector and raster data. But truly big **geospatial “Big Data” is starting to arrive.**

Working with these large data sets requires data warehousing, distributed computing, analytics in the cloud, libraries and binaries, and tools such as the Esri Geospatial Cloud, GeoEvent Server, ArcGIS GeoAnalytics Server, ArcGIS Image Server, and other enterprise solutions. It will require people who understand GIS analytics from a systems perspective and with web tools.

For more, see this Esri article on real time big data analytics and these other articles: https://www.esri.com/arcgis-blog/?s=#geoanalytics

Data skills are in high demand–in GIS positions and non-GIS positions alike

Data Gathering, Conversion, Processing and Management

- Work with a variety of geospatial datasets or systems from sources within and external to the federal government.
- Follow data requirements, data standards, access rules, and business rules.
- Identify and document reference data sources, integration process and domain values as information is collected.
- Coordinate and conduct the review of geospatial data products as directed by client
- Digitize map and geographical feature data into various GIS layers.
- Assist with data collection, consolidation, sharing, and general data management activities.
- Acquire data from multiple remote sensing / photogrammetric sources
- Process data acquired from multiple sources (e.g., Orthorectification) and analyze data for specific management questions
- Develop geospatial products from data, including orthomosaics, elevation models, remote sensing products, and land cover classifications.
- Assist with data sharing and GIS/data management activities
- Assist with the management of geospatial files, file geodatabases and enterprise geodatabase (SQL Server)
- Generate digital map products

From a job posting of GIS specialist in landscape biology/ecology on geosearch.com

What data skills are in demand?

- Generate data
- Acquire data
- Convert data
- Assess data
- Manage data and users
- Catalog and document data
- Curate data
- Serve and publish data
- Communicate with data

Indeed, **Data Science** has emerged in the job market and is offered as a discipline in many university programs.

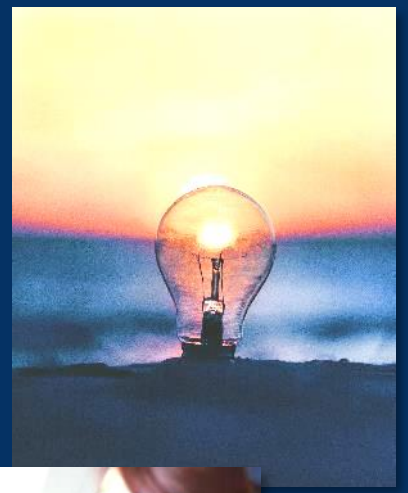
Using a variety of data makes your educational program vibrant, interesting, and relevant to a wider diversity of students.

Innovation

Dynamic Curriculum

Diversity of Faculty Experience

Cutting Edge Technology



Employment Opportunities

What is a Modern GIS?

- Geospatial Cloud
  - Software & Apps
  - Information Products
  - Data
  - Hardware
- People
- Process

Automation

Application configuration

Enterprise/Cloud

Requirements gathering

Geospatial Data is integral to Modern GIS

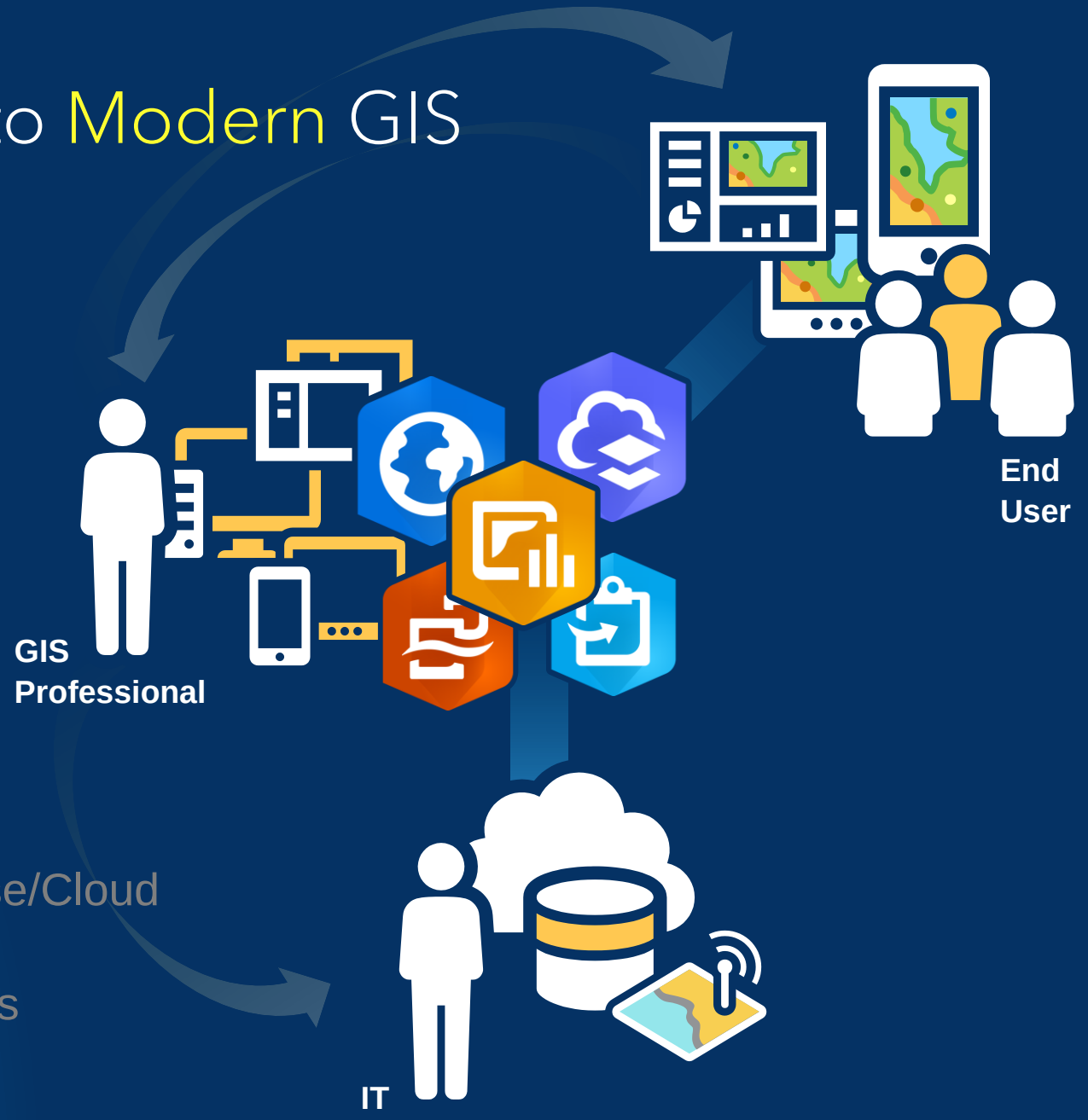
- Geospatial Cloud
  - Software & Apps
  - Information Products
  - Data ←
  - Hardware
- People
- Process

Automation

Application configuration

Enterprise/Cloud

Requirements gathering



Data is integral to GaaS: Cloud Based GIS

Hubs

Notebooks

AI/ML

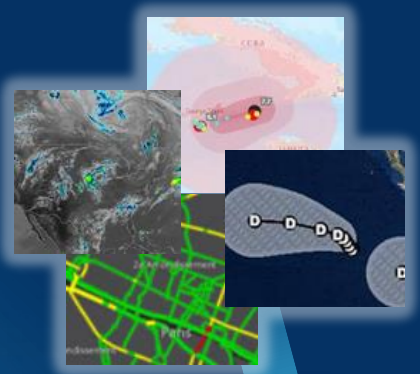
Big Data

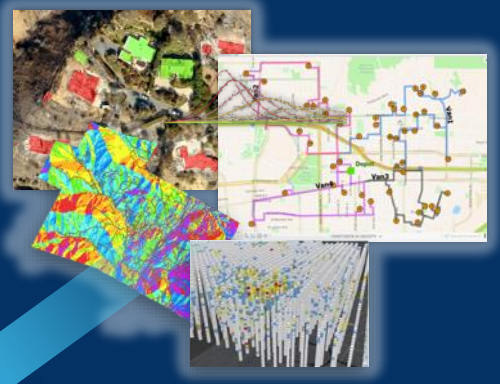
Imagery

3D

Drones

Mobility





Spatial Analytics

Data Management

Sharing & Collaboration

Content Discovery

Mapping & Visualization

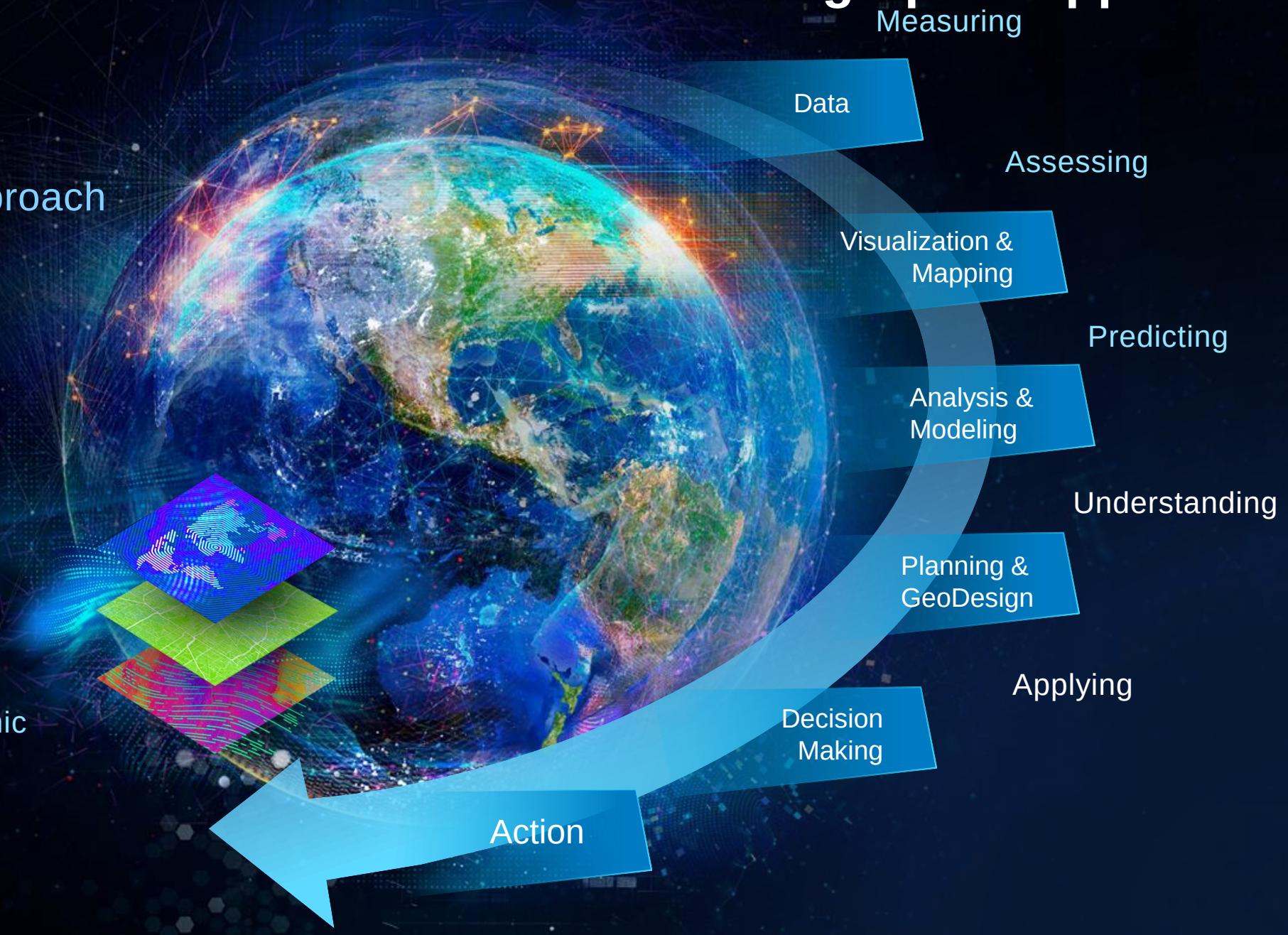
• Data provides the foundation for GIS and the Geographic Approach

GIS

Enables The
Geographic Approach

Providing a Process
and Framework . . .

. . . For Applying Geographic
Knowledge Widely



- **Data enables the holistic interdisciplinary view that is needed**

The Geographic Approach

A Way of Thinking and Problem Solving
That Integrates Geographic Science & Information
Into How We Understand and Manage Our Planet

Science Based

Integrated

Holistic

Spatial Understanding

Supporting Multi-Objective Solutions

Collaborative

. . . An Inclusive and Multi-Disciplinary Process Impacting Every Sector of Society

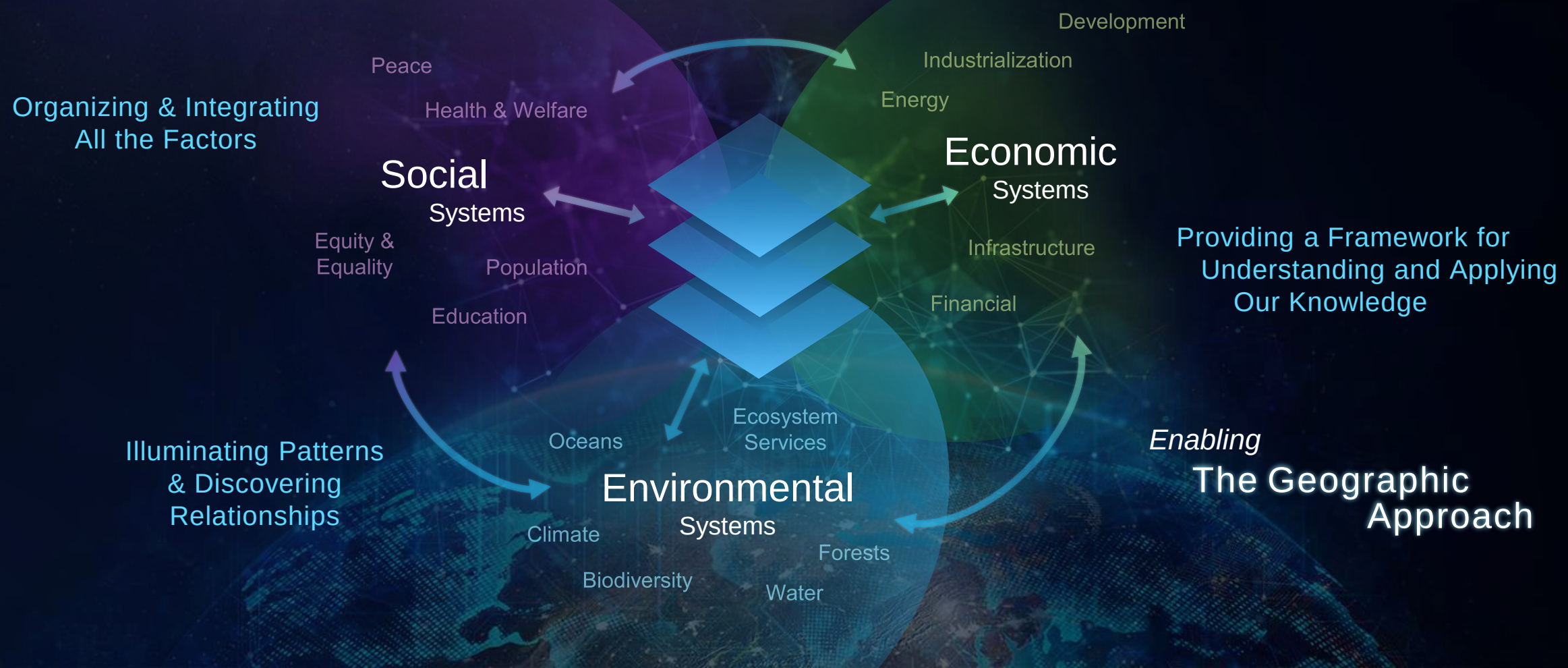
GeoScience and Understanding

- **Data enables GIS as the common language for solving problems**

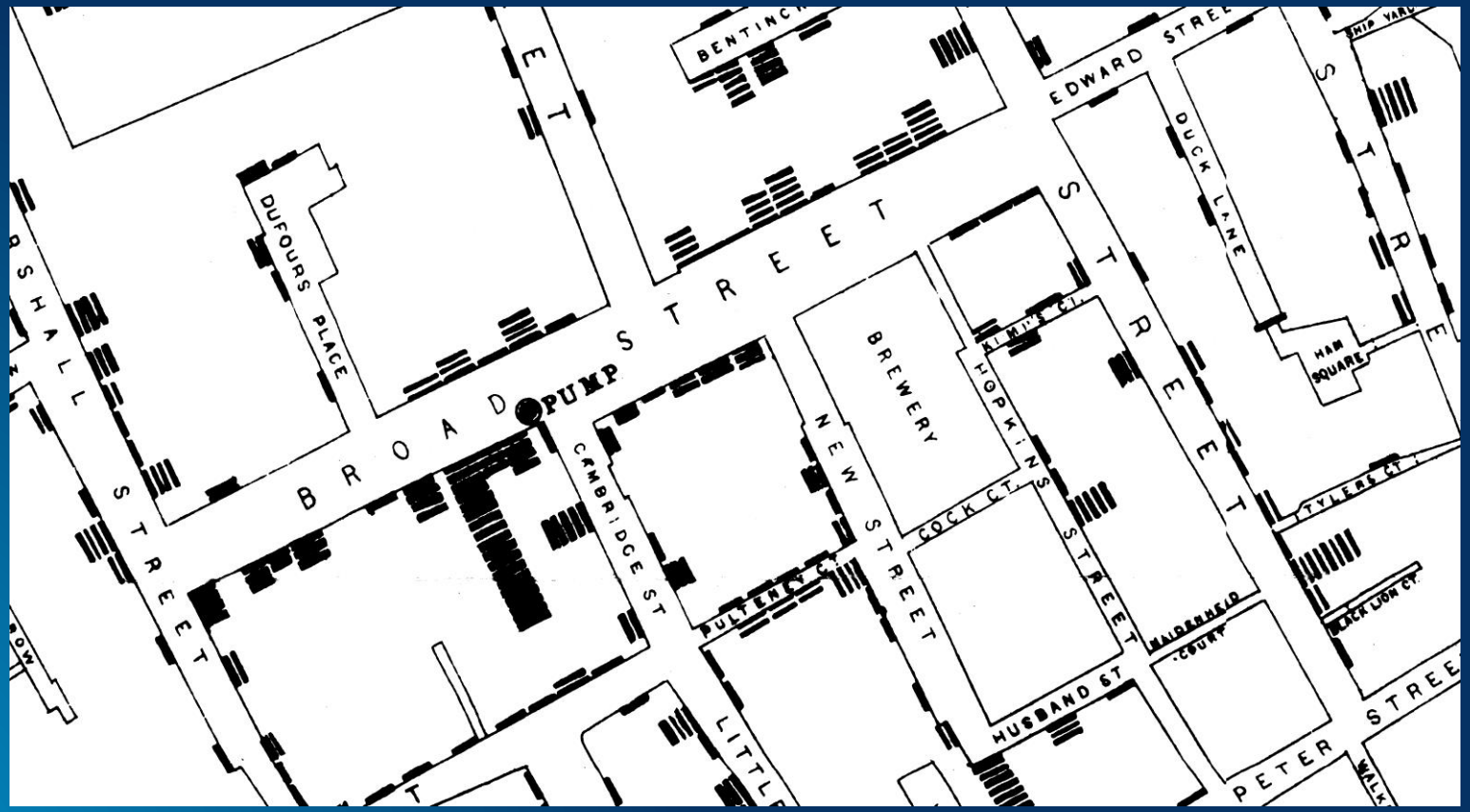
Sustainability Requires . . .

We See the World as One Single Ecosystem

Geography Provides the Science & Language to Do This



Data has always been important to spatial analysis



And today, everyone is not only a data consumer, but is a data **creator.** **What are the implications of this?**

Modern Strategies for Finding Geospatial Data

The challenge

(1) Data needs are often very specialized, in theme, scale, region, attributes, and other characteristics.

(2) Often it is difficult to find a similar need by someone else online.

(3) The ephemeral nature of the web makes it challenging to learn from others' strategies.

The benefits of the modern GIS paradigm for data

- Data are increasingly collected and available in a wider variety of formats AND as services.
- Data is increasingly being seen as an **asset**, not as something to charge for, but to serve for the benefit of all decision makers.
- Open data portal technology such as ArcGIS Hub make it easier for organizations to serve and curate data.

Multiple Sources

- Just as there is often no “single best source” for one specific topic on the web, no single “geospatial gold mine” exists that is the end-all site for spatial data.

- However, this is a good thing. Why? It forces the user to think critically about the data sets and choices. As is clear on the Spatial Reserves data blog, choices abound. Those choices require the user to investigate and evaluate each of them.

Words of Advice

- **Minimize** search time for geospatial data, so that you can **maximize** time available for data analysis.
- However, **don’t shortchange** your data searching by choosing only the data sets that are easy to obtain. All of your decisions will be based on the quality of the data that you acquire, so data quality matters!
- **Consider** the software you are using and seek to minimize the amount of data conversion you will have to do.

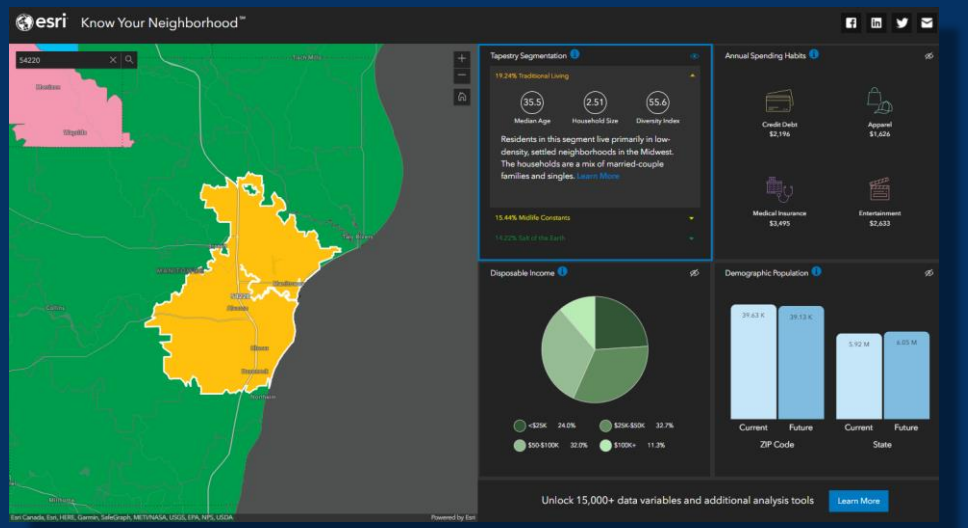
Strategy

1. Have a clear vision of the data you need, the scale, date, extent, resolution, and other characteristics before you spend time searching.

Set a limit on how much time you will devote to searching and processing before the analysis stage. Otherwise, data search can consume all of your project's available time!

More data is not always better!

Do you even need the raw data?



Would a static image, export, or PDF from a GIS or an online data service meet your needs?

If you don't need to clip the data to an area of interest, additional variables, or combine the data with other data sources, and just need the image for a report or another purpose, a static export may meet your needs.

Neither approach—a static map vs the actual data tables and GIS files) is necessarily "better"—it just depends on your objective for your investigation and your needs for customizing your investigation and further digging into the data's additional available fields, or a different geographic unit. Each approach requires a certain set of skills and so the decision is also bound up in the background of the user and the user's available time and resources.

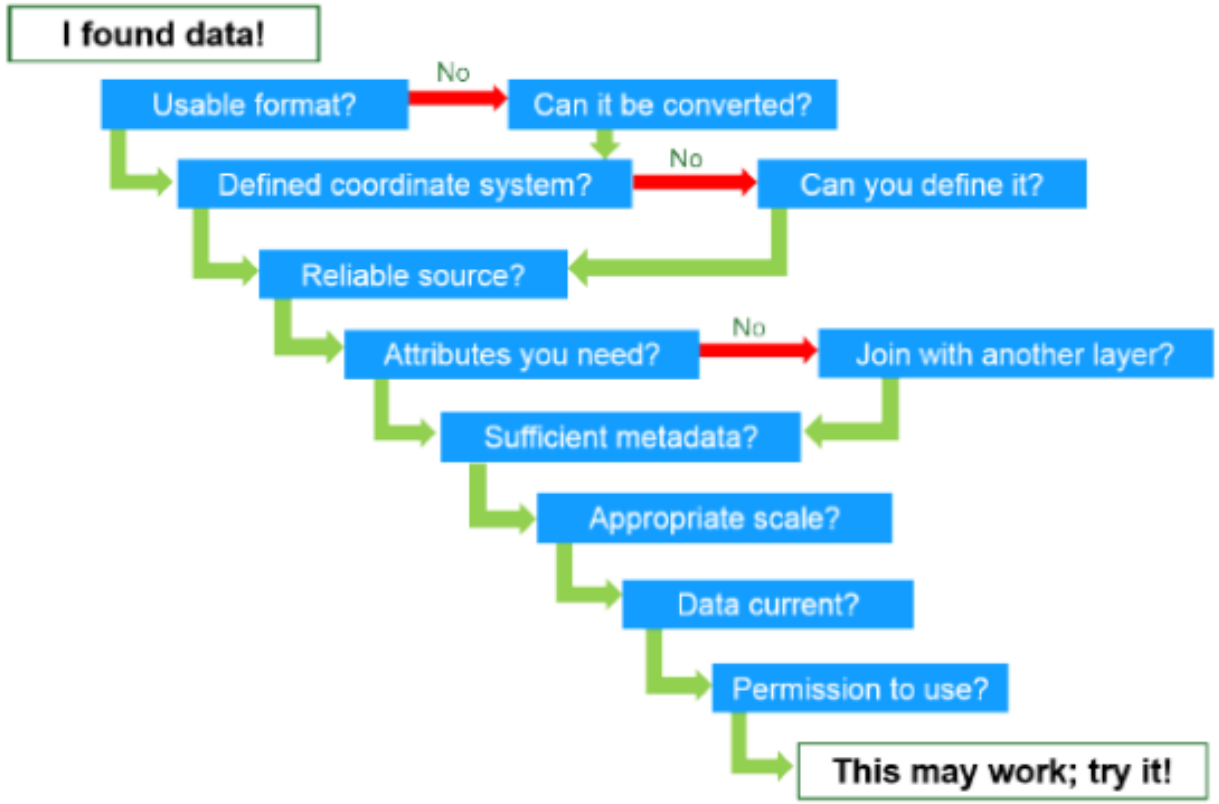
2. When you find a potential data source, examine it carefully so that you can decide if and how to use it.

To help you decide, see (1) this graphical aid to assist in deciding whether data can be used,

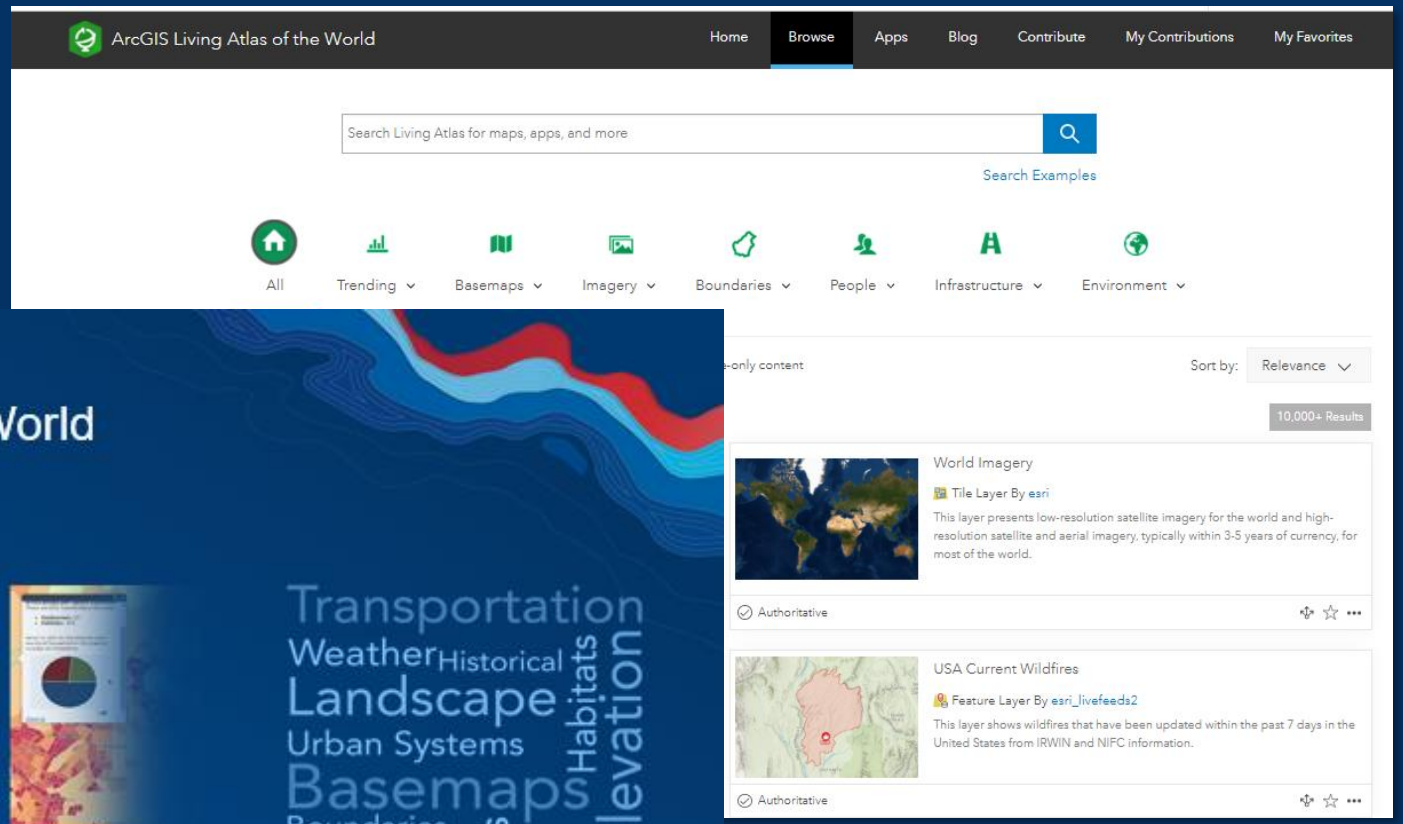
(2) And this set of 30 checks for data quality in a potential source.

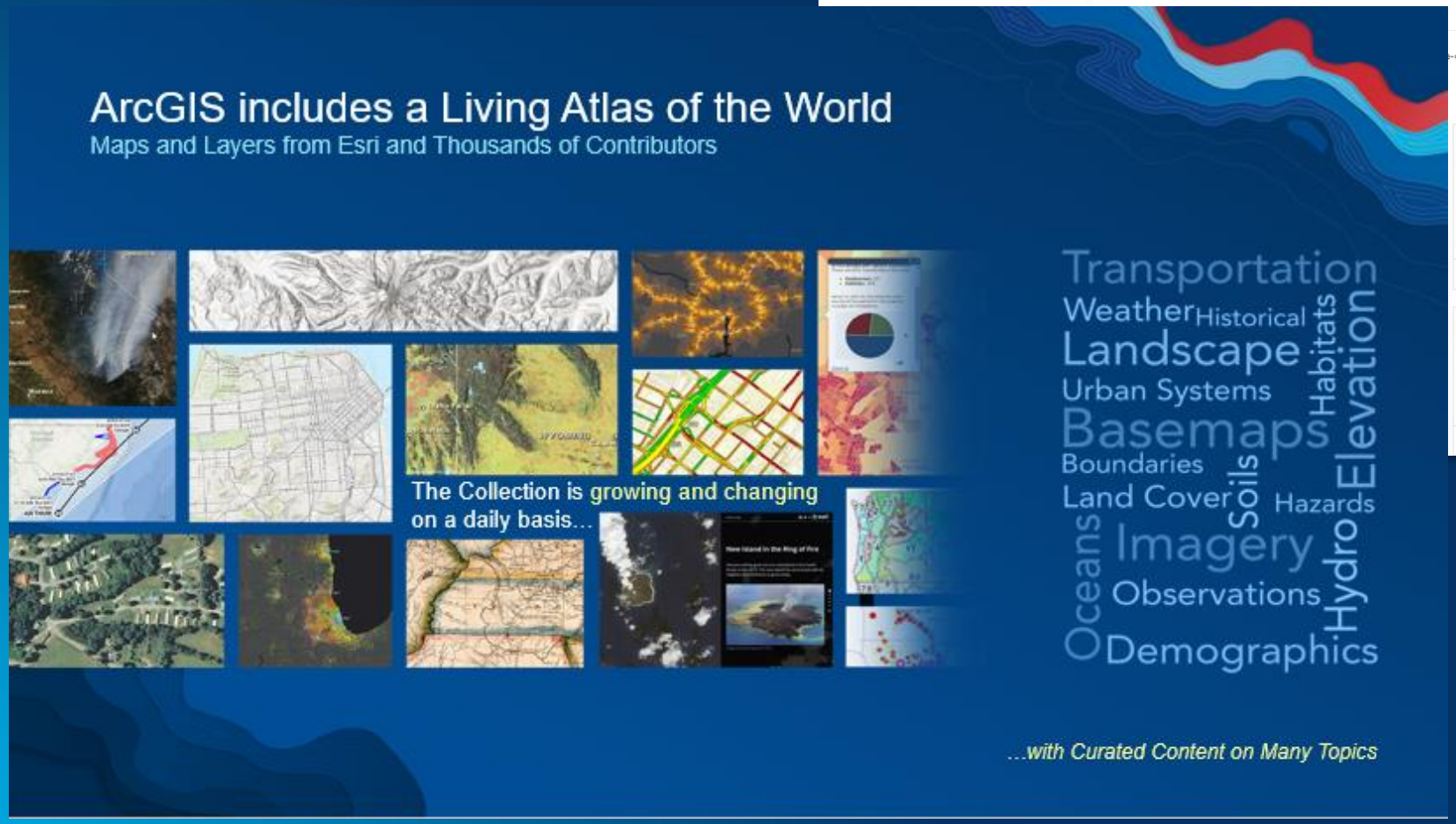
Considerations for choosing GIS data

Even when you find the data you need, it may not be beneficial for your project. Before you use a dataset you have found, you should investigate it to ascertain that it actually meets your needs.

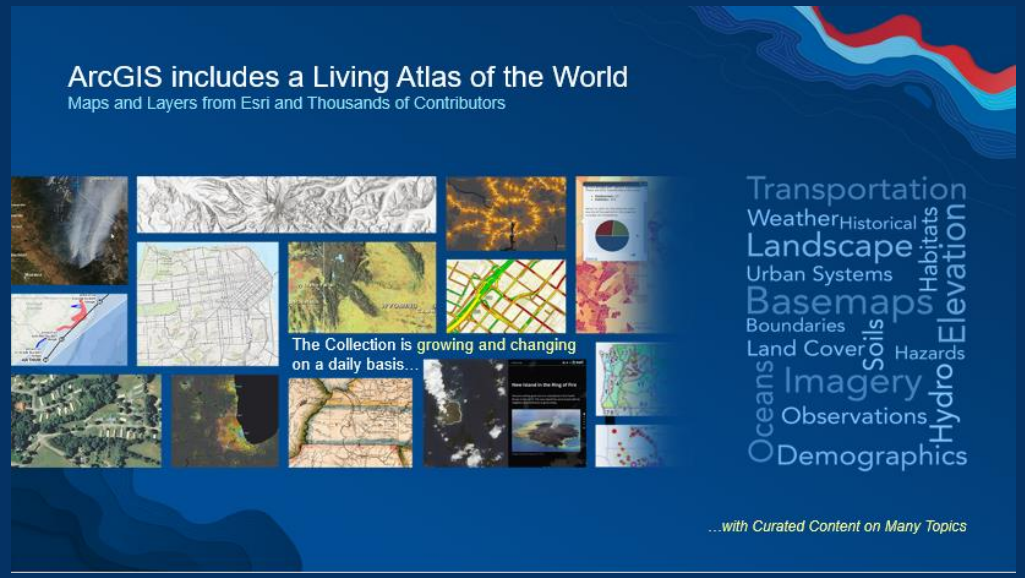


3. Start your search with the [ArcGIS Living Atlas of the World.]





Why start with the ArcGIS Living Atlas of the World?



(1) Content: The ArcGIS Living Atlas of the World is the closest thing to a one-stop shop: > 8000 curated and living data layers, maps, and apps.
(2) Ease of use: The maps, apps, and data layers in the Living Atlas are ready for immediate use in ArcGIS Online, ArcGIS Pro, and in other applications.
(3) The content is rich in metadata and is hosted by authoritative sources: NOAA, USGS, UNEP, nonprofits, academia, local to international government agencies, and private companies.

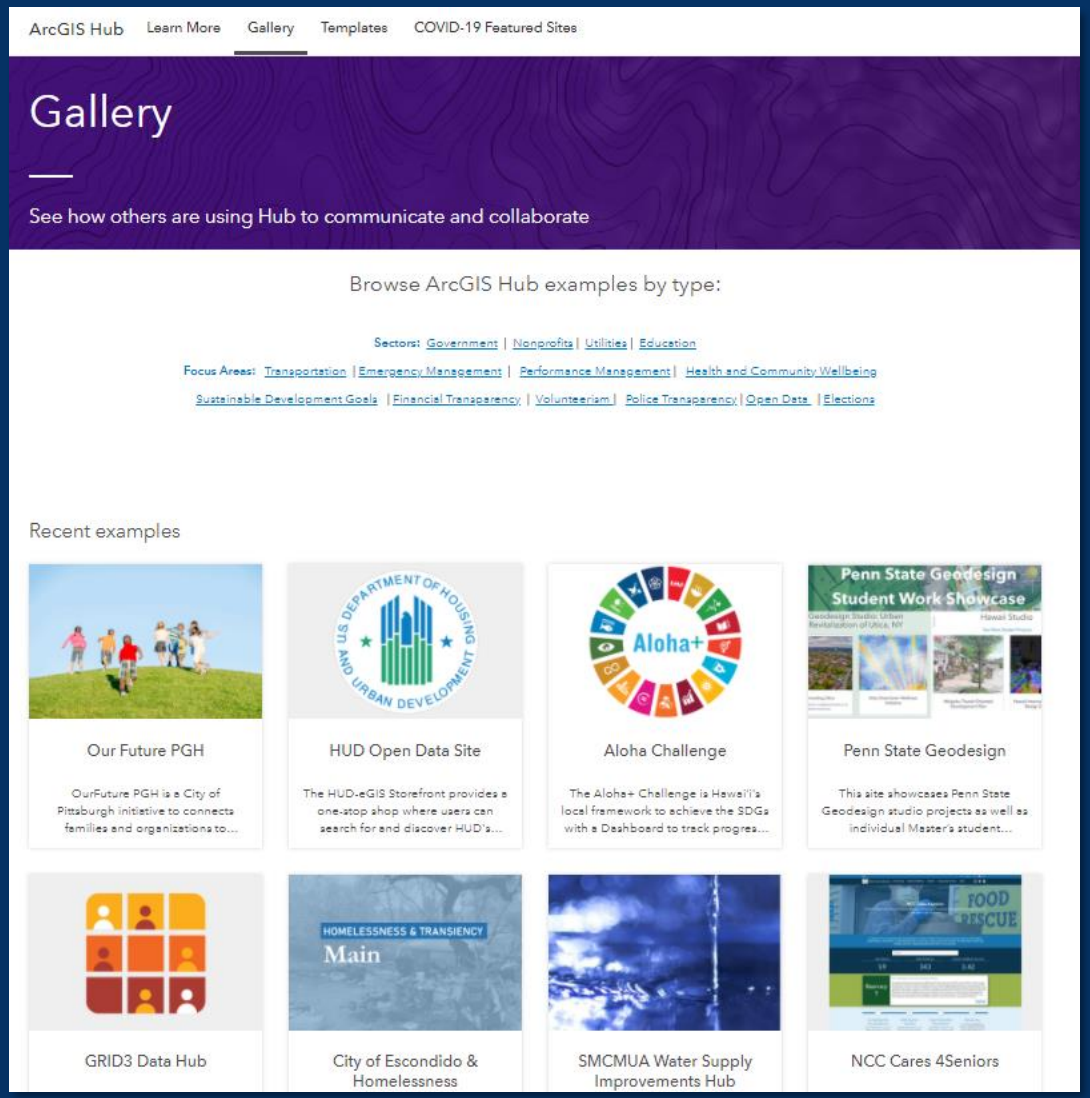
Living implies 2 things – (1) the content is continually being re-analyzed and curated, and – the content reflects our dynamic planet—many of the layers are updated in real time or near real time—weather, storms, earthquakes, volcanoes, traffic, stream gauges, and much more.

4. Investigate ArcGIS Hub sites. ArcGIS Hubs have emerged as a viable and easy way for organizations to host data, and as a result, thousands of them exist.

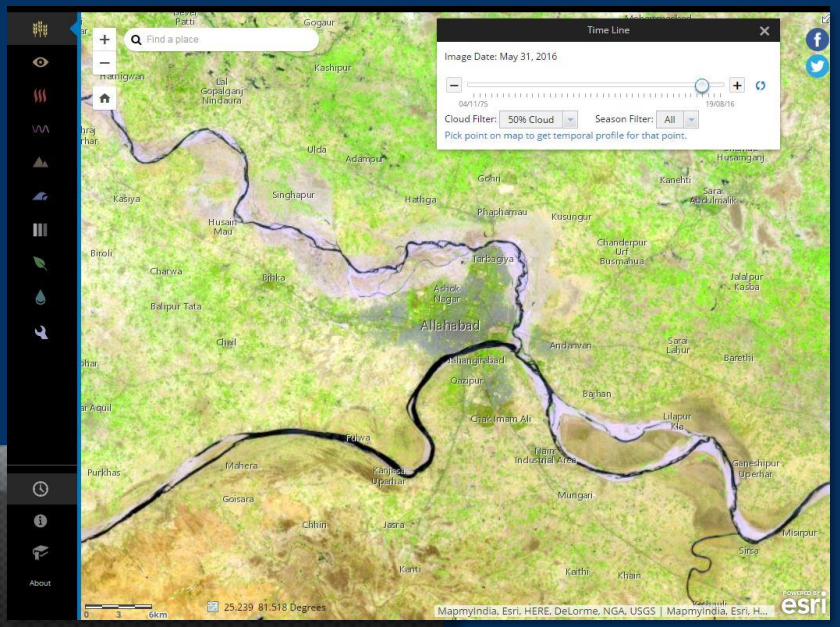
Plus, Hub sites usually offer an array of download and streaming data options. Starting points include searching on “Your area of interest” and then “GIS data portal” or “ArcGIS Hub”.

As one of hundreds of examples, here is the Boulder County ArcGIS Hub.

There are also useful starting points such as ArcGIS Hub search tools and galleries.



5. Investigate selected “top 10” (such as satellite imagery, Landsat, and vector data portals) lists that we have reviewed and compiled on the Spatial Reserves data blog.



Spatial Reserves

A guide to public domain spatial data

Spatial Reserves | About the book | About the authors | Exercises and data for the book

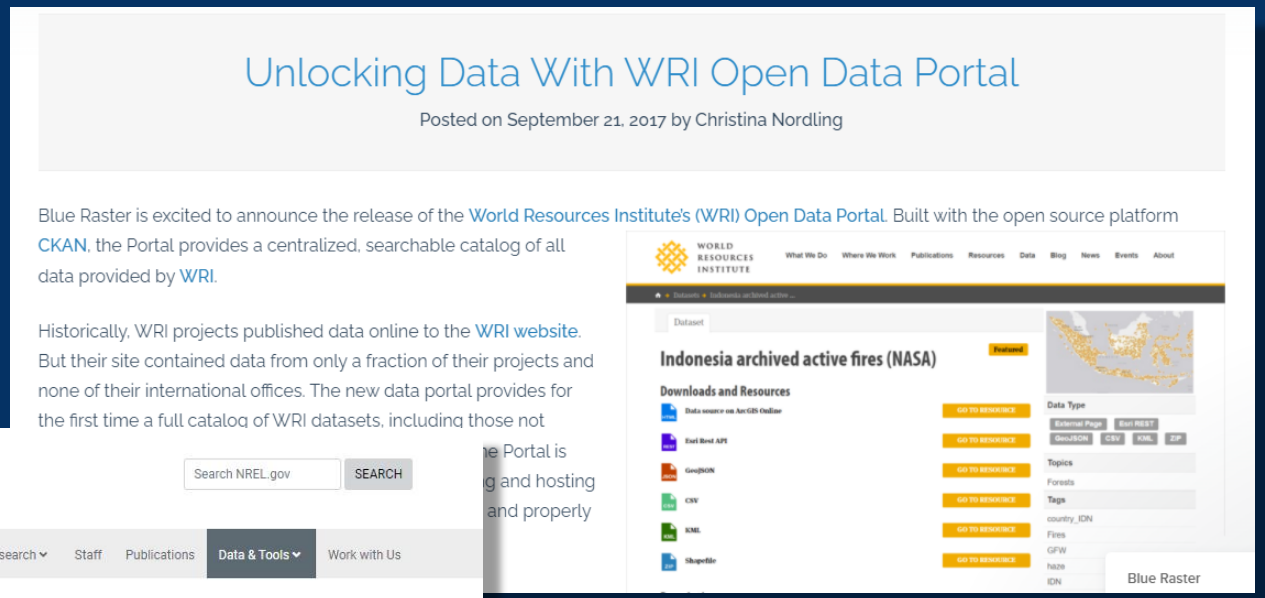
Home > Public Domain Data > The Top 10 Most Useful Geospatial Data Portals, Revisited

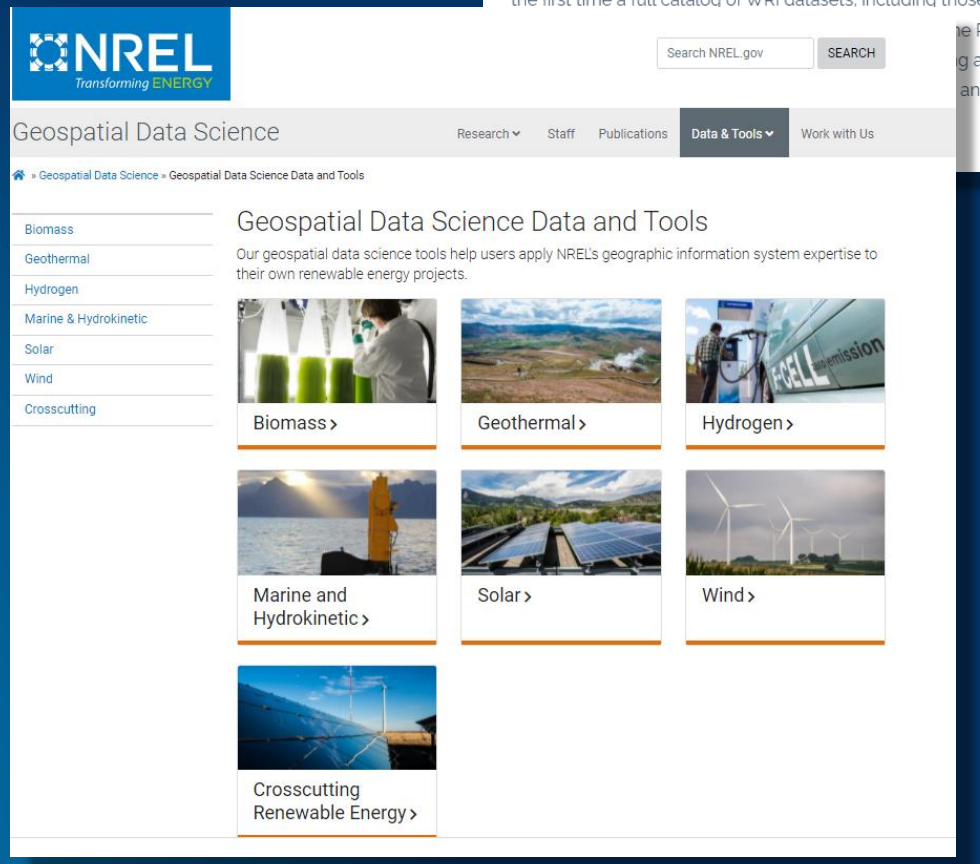
The Top 10 Most Useful Geospatial Data Portals, Revisited

February 18, 2019 josephkerski Go to comments Leave a comment

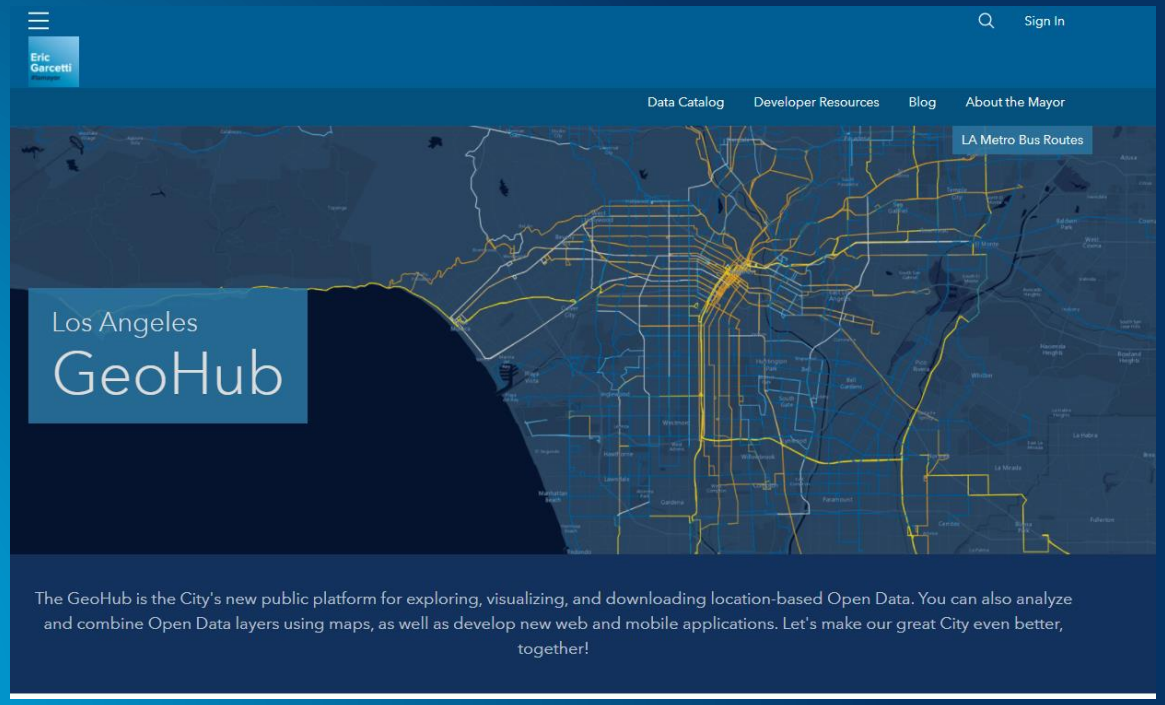
We have been writing this geospatial data column for 7 years now, beginning when our book *The GIS Guide to Public Domain Data,* was published. Over those years, in addition to keeping issues such as data quality, copyright, privacy, and fee vs. free at the forefront of the conversation, we have tested and reviewed many geospatial data portals. Some of these portals promise more than they deliver, some have been frustrating, but many have been extremely valuable in GIS work. Back in 2017 we listed 10 of those that we have found most useful, rich with content, easy to use, and with metadata that is available and

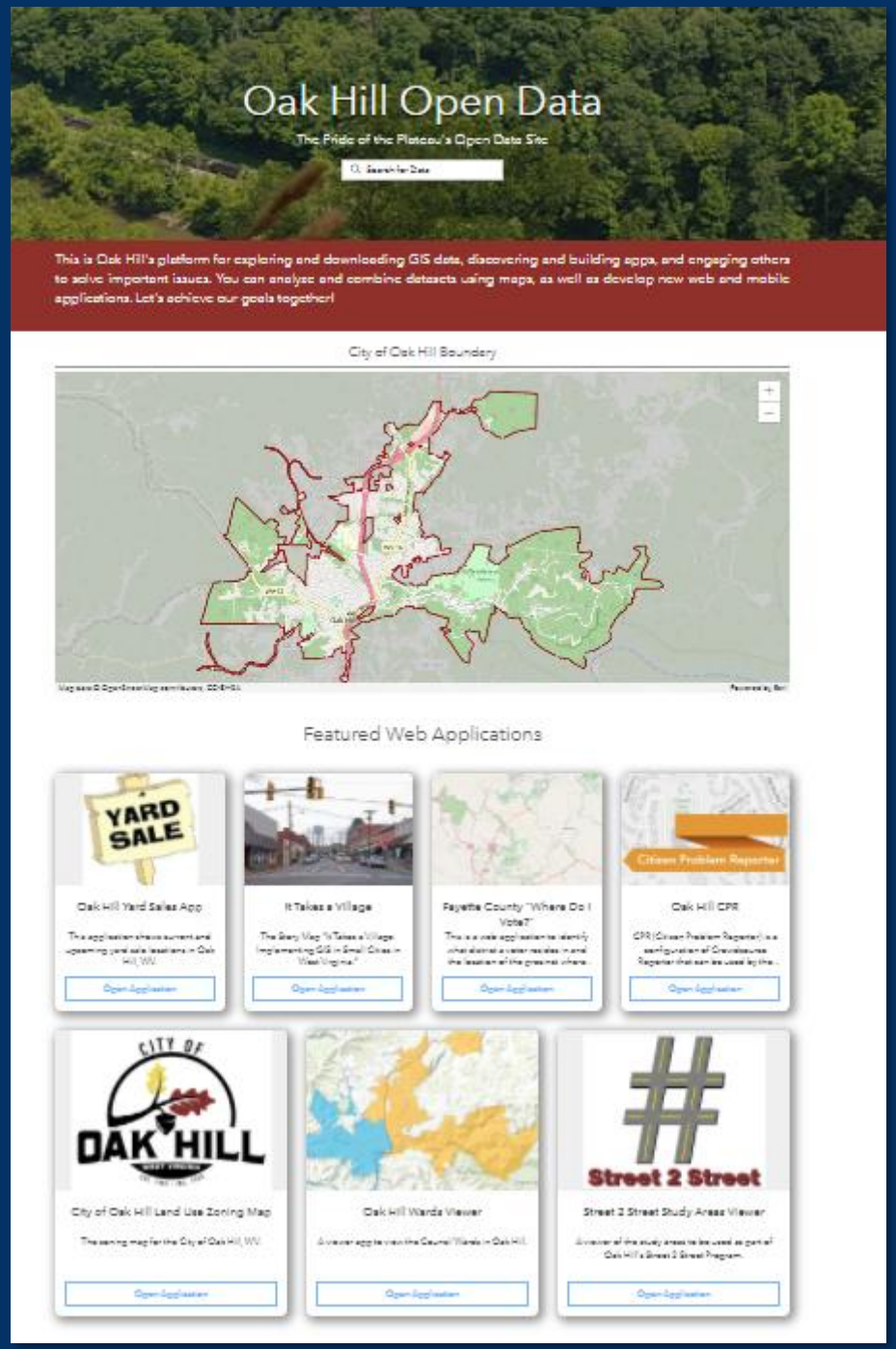
6. Investigate portals organized by theme that cover your topic (such as WRIs on land use and natural resources) or NREL's renewable energy data.



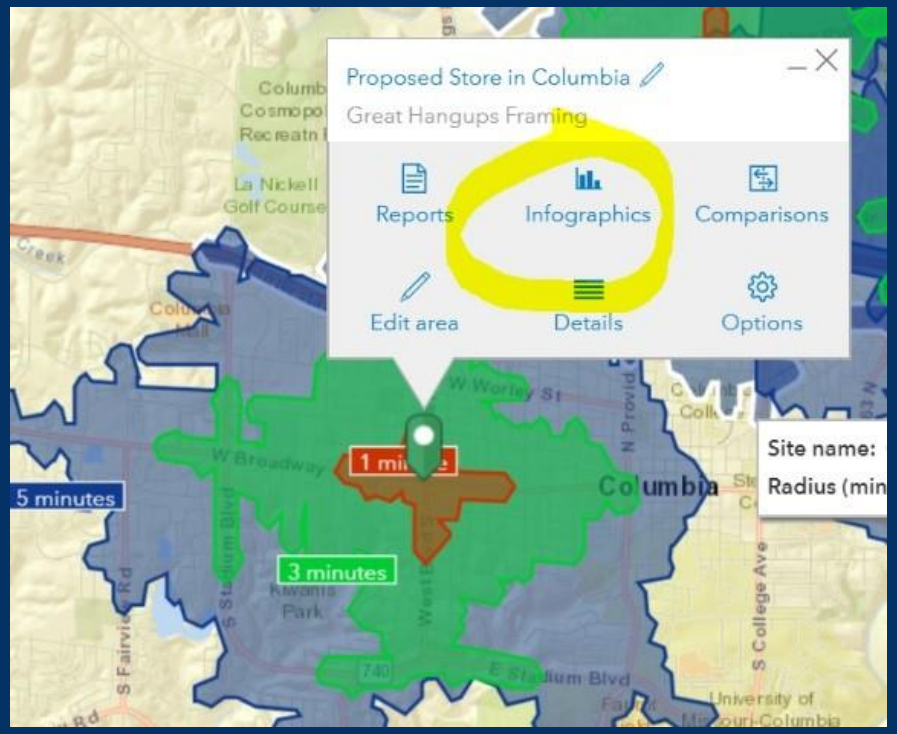


7. Investigate portals organized by location, that would cover at least in part, your study area (such as the Ordnance Survey for the UK, and state, province, region, or city (examples here for Oregon, here for the large city of Los Angeles, here for the small city of Oak Hill West Virginia.

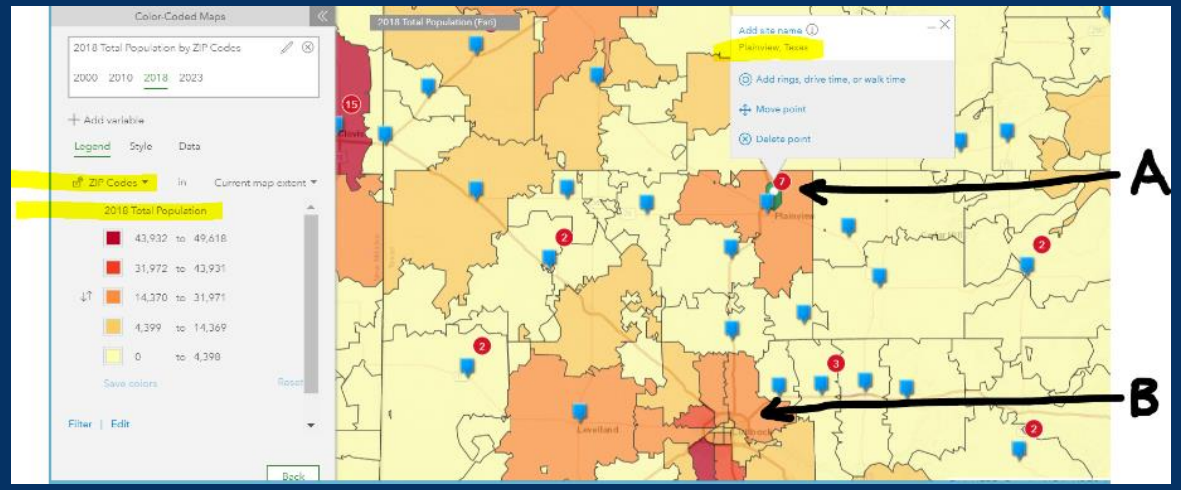




8. Use Business Analyst Web as a data source—for multiple countries, for hundreds of demographics, business locations, and behavior variables (health, expenditures, lifestyle, and much more). Updated and with projections into the future as well.



Data available are here: https://doc.arcgis.com/en/business-analyst/web/data.htm



9. Because an increasing amount of research is tied to geospatial data (see the following posts on this topic: Part 1 and Part 2), check Google scholar, library databases, and other sources of peer-reviewed research, as well as conference proceedings from Esri, AGU, AAG, and other major events, to investigate the sources identified in the research.

Be tenacious despite the inherent challenges with obtaining data tied to scholarly research.

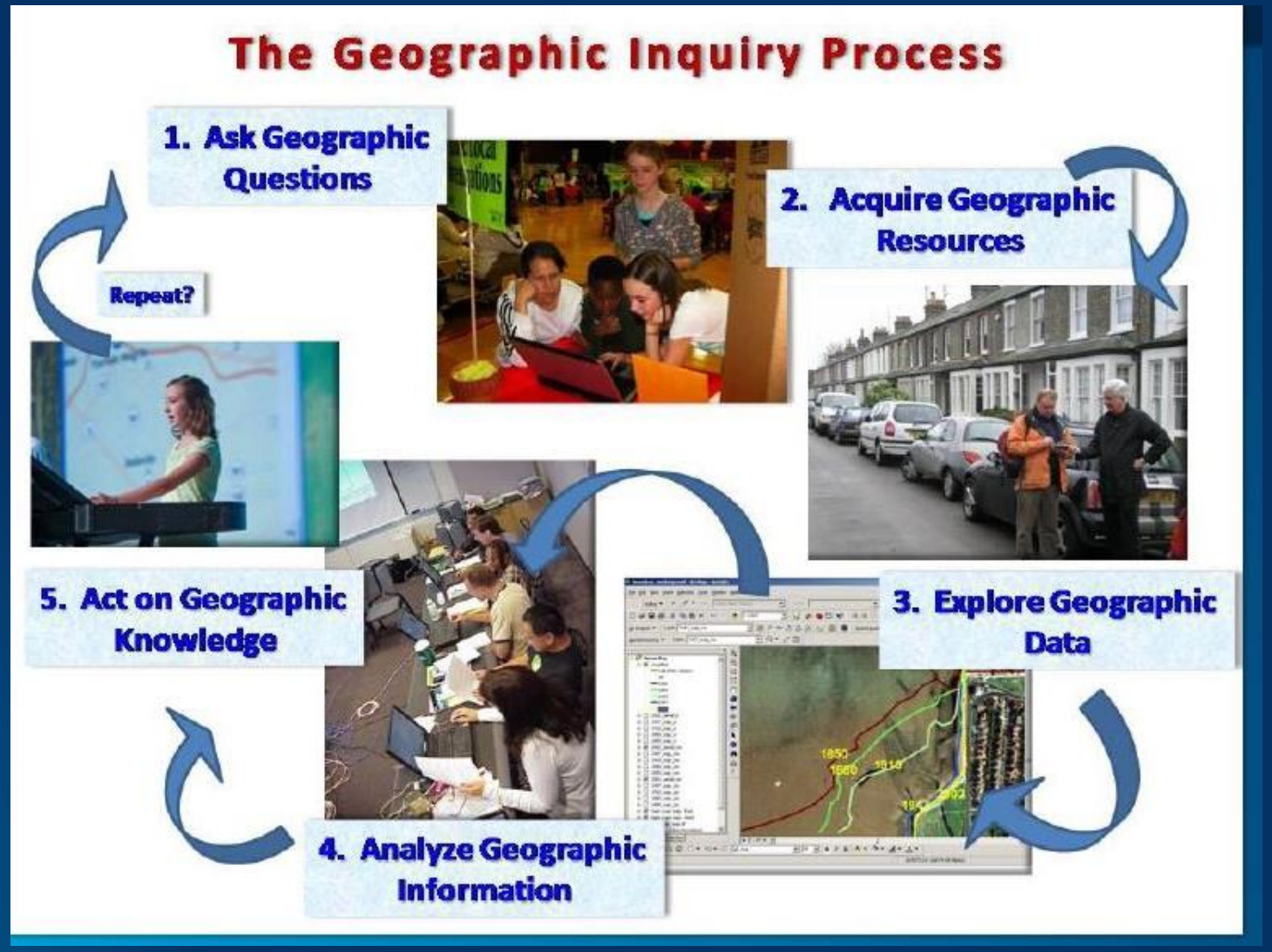
Coupling data with scholarly research

April 26, 2021 josephkerski Go to comments Leave a comment

How many times have you read an article or report and wish you had access to the data that the author used? You're not alone: A growing concern is being voiced in the research community that data are typically not included with scholarly papers. Why should all this matter? Hasn't this been the way publishing has always been? Well, in today's world where serious issues are growing, as recent global health and natural hazards challenges have made starkly clear, the provision of data could provide an immense leap forward for researchers and developers to provide solutions to solving these pressing issues. As we have written about for nearly a decade in this blog, despite crowdsourcing, the Internet of Things, sensors on, above,

The geographic Inquiry process…

… means that data acquisition is iterative.



Case study of these strategies in action

I recently received a request from an educator who needed to support a project involving drought in Colorado with the following data:

- Soil Moisture: Map of Soil Moisture for April 2013 – you have this source but the pixel size might be too large to be meaningful.
- Precipitation: Total for 2013 water year to encompass the 2013 flood.
- Streamflow: Time series of streamflow of Colorado River Oct 2012-Sep 2015 (with average streamflow marked).
- Groundwater: Is there a map of groundwater availability?
- Reservoir Levels: Time series of Lake Powell water level for Oct 2012-Sep 2015.
- Temperature – Need map of temperature in April 2013 for across Colorado?
- Agricultural Health: Is there an index for agricultural health? Would there be a time series or some data for 2013?
- Water use: Is there a data set that shows water use?

My investigation resulted in the following:

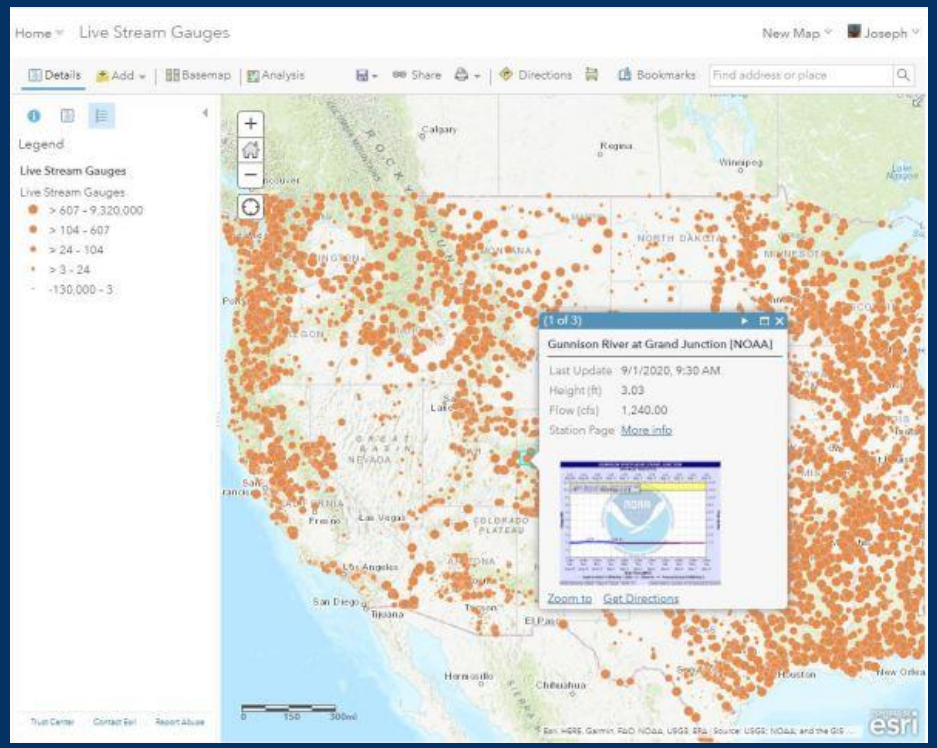
See portal at the Geospatial Centroid at Colorado State University, the Colorado Information Marketplace (data.colorado.gov), the Open Colorado data catalog, ColoradoView, and also portals hosted by individual state agencies such as the Department of Natural Resources, the Colorado Geological Survey, and the Department of Transportation.

The sources require close evaluation to determine if they meet your needs. Some are formatted as lists of data or map-based clickable links that require downloading the data, while increasingly, others provide streaming services. For example, this map service “DNR Viewer” from the Department of Natural Resources does not allow for the downloading of geodatabases, imagery, or shapefiles, but this data portal from the same agency does.

On the topic of streamflow: From the USGS, this Current Water Data resource has been a longstanding favorite of mine: https://waterdata.usgs.gov/nwis/rt.

Nowadays, however, it is also a live data feed in ArcGIS Online: The metadata is here: https://www.arcgis.com/home/item.html?id=4d8b9ff2d9e74ad18daecb2db7eaa87f

Examine the metadata and open in ArcGIS Online or in ArcGIS Pro. This is a premium service sign in to ArcGIS Online is required. It then looks like this at right) (there is also a global stream gauge layer, here):



On the topic of groundwater: See this report and sets of data. Reservoirs are available here at USGS and here in the ArcGIS Living Atlas of the World, but reservoir **levels** are more difficult to obtain: It may involve calling or emailing water resources agencies. Sometimes “20th Century methods” are best!

On the topic of Agricultural Health: There may not be a globally accepted measure for it analogous to the HDI for education-health-life expectancy from the UN. Here is a “agricultural stress index” map and measure, from FAO.

Contact state department of agriculture. Here is an agriculture-related story map entitled “Farming for the Future“. … with links to individual maps with data layers for the type of crop.

On the topic of temperature, I recently worked with an Esri colleague on the data team, and, good news!… he created this average global air temperature layer.

On the topic of precipitation: In conjunction with an educational project, I compiled this layer as well as other Colorado layers, and display them in this gallery. The precipitation data came from COCORAHS – the Community Collaborative Rain Hail & Snow Network. In terms of water use, this remains challenging, but this Living Atlas collection contains water related items. When I worked at USGS, I made extensive use of this water use report: https://waterdata.usgs.gov/nwis/wu.

From this report, you may need to go through this workflow: Query the system > obtain a CSV table > bring that table into ArcGIS Online or ArcGIS Pro to create a map from it.

How can you know if data is any good?

- Fitness for use: It is the data user's responsibility to determine whether a data set is fit for their project.

- Truth in Labeling: It is the data producer's responsibility to provide metadata and other information that describes the data accurately. This enables the data user to determine Fitness for Use.

- **→ Remember: With modern GIS capabilities and ease of publishing data in many different ways, every data user is now a potential data producer. Thus, everyone in GIS needs to pay attention to both fitness for use and truth in labeling.**

How can you know if data is any good? Data Quality Examples

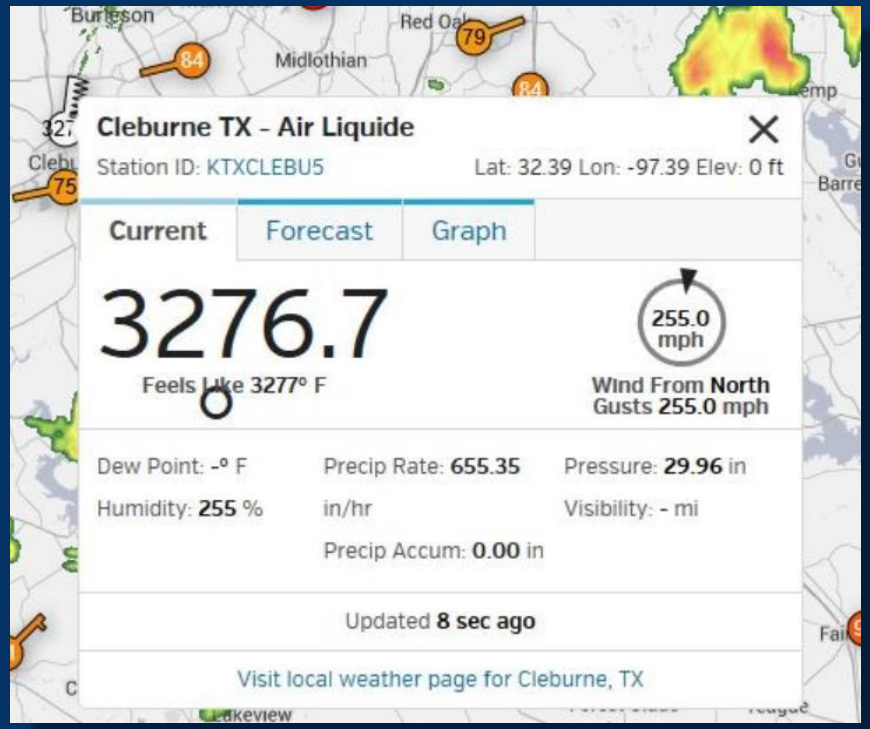
- Imagery: It is what it is. Well, not always.
- Why data quality still matters, now more than ever.
- Intentional offsets on web maps.
- Know your data! Example from mapping Lyme Disease.
- Faked Satellite imagery.
- Teaching about Spatial Data Quality.

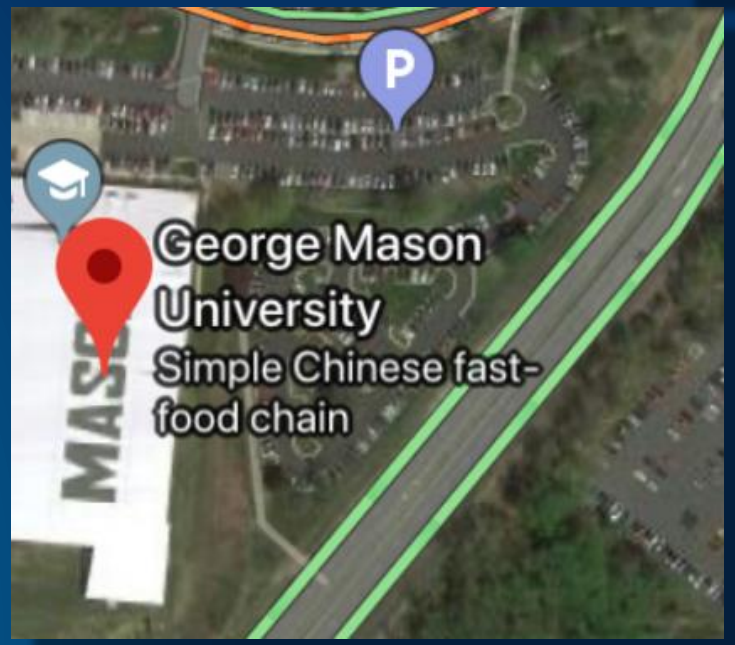
What matters in data quality

BE CRITICAL of the data. All data.
Including and perhaps especially – mapped data.

Don’t just accept it because it is a map.
Anyone can make a map.

Don’t just accept it because it is online!
Anyone can share content.





There is no shortage of good maps, *and* bad maps

Also pay attention to the surveyed N value of the maps you find!

For more bad maps, and what to do about them, see this presentation: https://sway.office.com/vFHqEIn3iGojHF29

All data are created in a societal context.

Selected Societal Issues Surrounding Data

- **Location Privacy**: Should I share my own location, the phenomenon that I am studying, and the location of people I am studying? What are the implications of doing so?

- **Copyright**. Am I allowed to use map or image content in my GIS project, story map, dashboard, or other means of communication? Where can I obtain permission to do so if I encounter copyrighted material? What sources can I regularly access that are not copyrighted?

- **Ethics**: How can GIS be used to benefit society and not harm society? What are everyday ethical decisions that need to be made in using GIS?

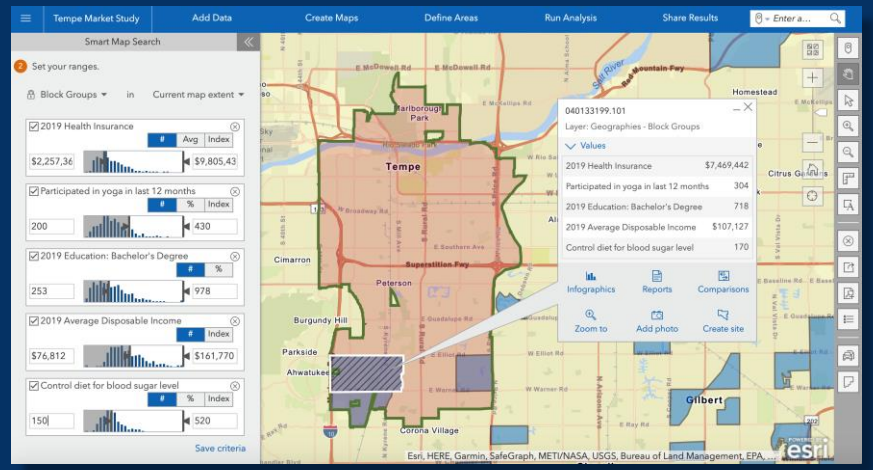
Ethics

Ethics matter in GIS because geospatial data are powerful and oft-used means of communicating, of influencing, and making decisions. Data is not created in a vacuum but rather within a value-laden societal framework. Implications include:

- (1) Maps are powerful means of communication, take the responsibility as map author seriously. Maps have an *aura of authenticity*–they tend to be believed. Be thoughtful about the projection, symbology, classification methods, and other techniques; do not intentionally mislead your audience.

- (2) Due to crowdsourcing and web GIS, everyone is now a potential map *producer*; no longer just a map *consumer*. More maps means maps of a wider variety of quality and purposes than ever before. Be thoughtful about the maps and data layers you select for your own purposes.

The Cartographer’s Code of Ethics



1. Have a straightforward agenda.
2. Strive to know your audience.
3. Do not intentionally lie with data.
4. Always show all relevant data when possible.
5. Don’t discard contrary data because it’s contrary.
6. Strive for accurate portrayal of data.
7. Avoid plagiarizing.
8. Symbol selection should not bias map.
9. Map should be repeatable.
10. Pay attention to different cultural values and principles.

Source: Cartography Thematic Map Design, sixth edition by Dent, Torguson, and Hodler.

Ethics: Deeper Dives

The cartographer's code of ethics.

The GIS certification institute's code of ethics.

Ways to teach about ethics.

Case studies in Ethics.

Teaching about and with data: Recommendations

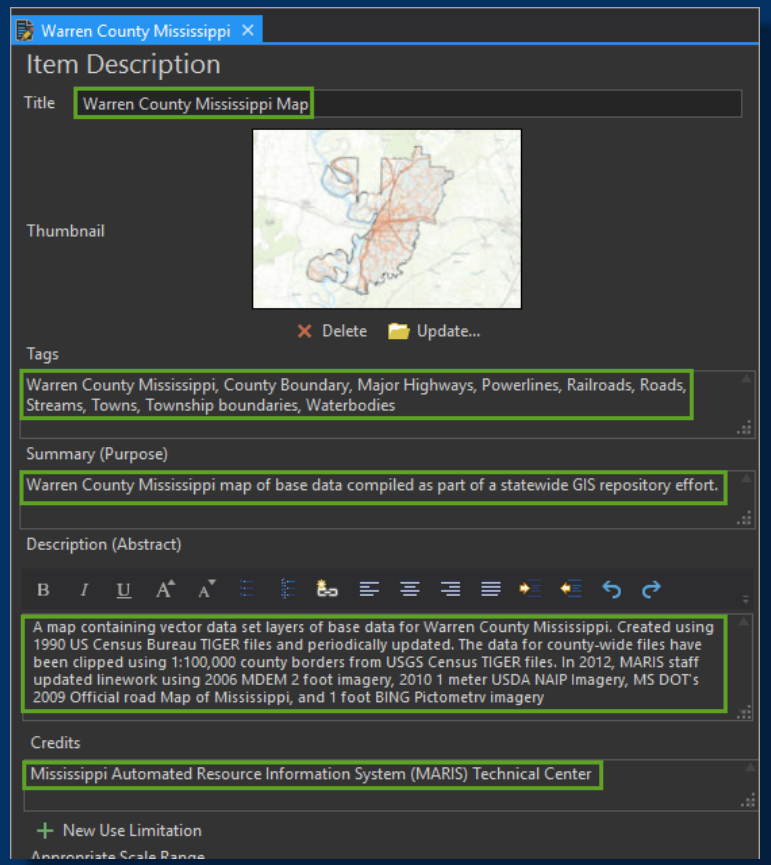
1. Include lessons where students must go to a wide variety of data portals at a variety of different scales and themes, practice accessing data from those portals and dealing with all the inherent challenges.
2. Include discussions about societal issues surrounding data throughout each course, rather than saving it for one class discussion at the end of the course.
3. Include at least a few lessons that are structured such that students face a problem, and have to determine which data would be appropriate, have to access and process that data, and use the data in a problem-solving environment with GIS tools.

Teaching about and with data: Recommendations II

4. Keep it interesting! Data discussions can be dry. For example, use this Gigapixel image to teach about Location privacy. Are any faces grayed out? How many people in a sample of 10 are on their smartphones?

5. Include at least 1 lesson where students have to populate their own metadata.

6. Include at least 1 lesson where students are faced with a map or data layer that looks valid and appealing but contains very little or no metadata. Discuss: Should the layer be used? If so, in what contexts?

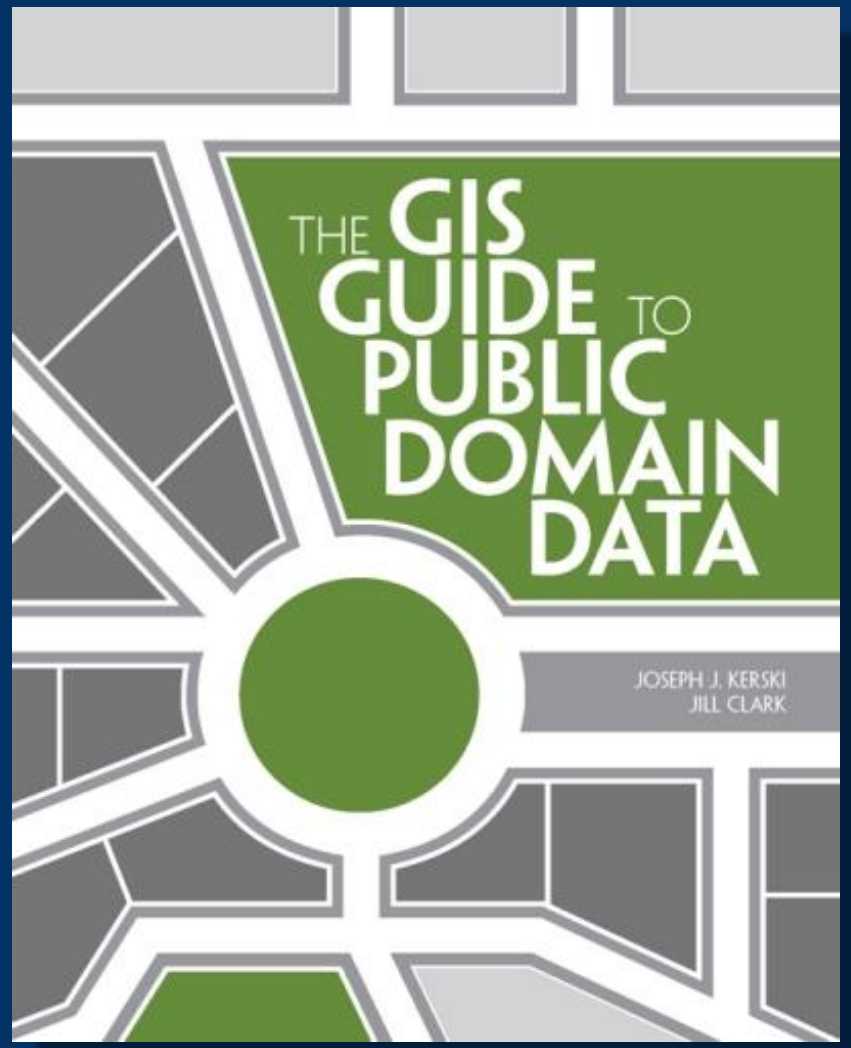


Continue your learning!

Geospatial Data book, blog, and 10 lessons that use public domain data in problem-solving scenarios: https://spatialreserves.wordpress.com

Put data to practical use via lessons in the Learn Lesson Library: https://learn.arcgis.com/en/gallery/#?q=data

Explore the ArcGIS Living Atlas of the World https://livingatlas.arcgis.com and revisit it often, as it is “living” and continually updated.



esri | THE SCIENCE OF WHERE