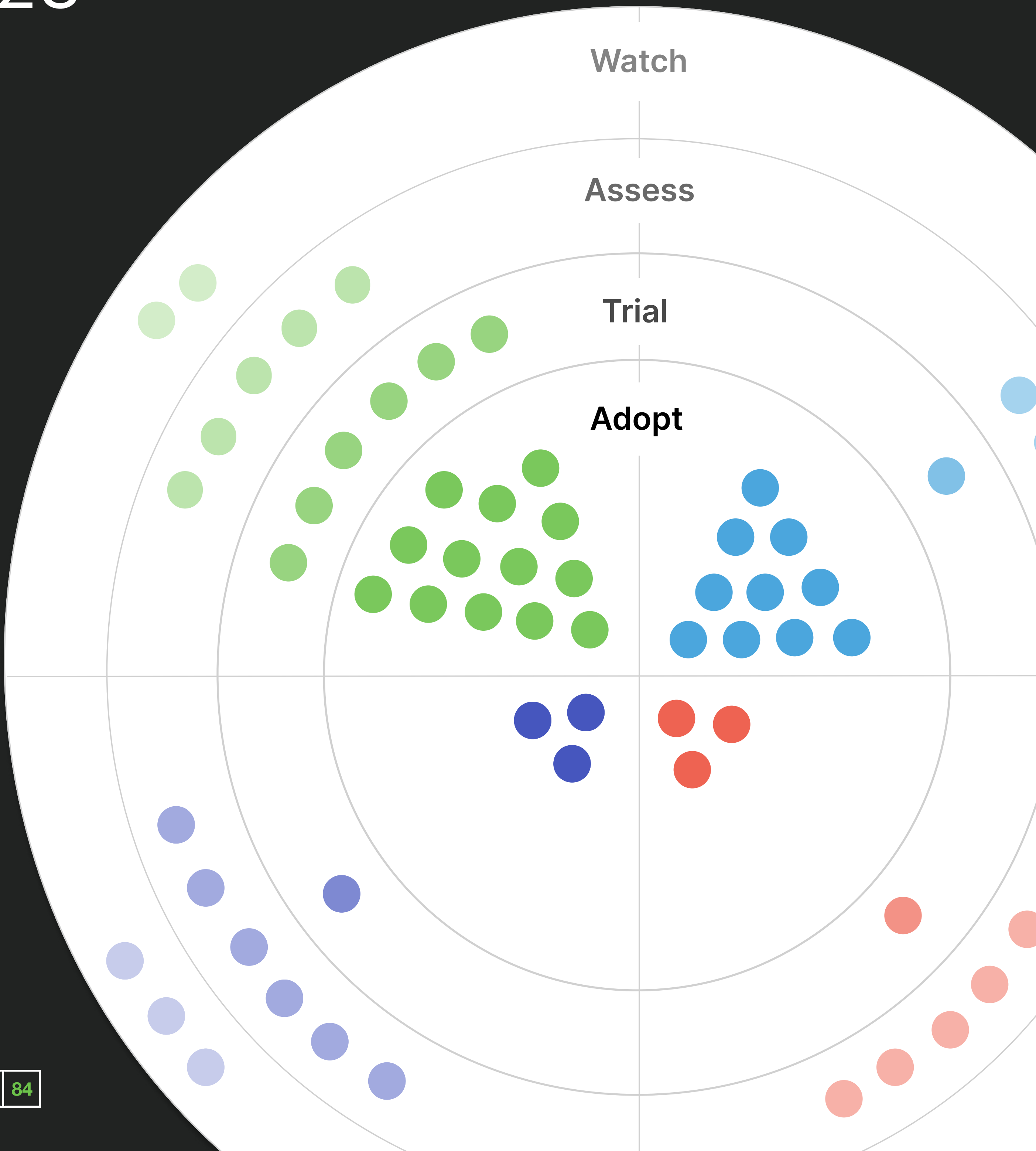
Geospatial Technology Radar

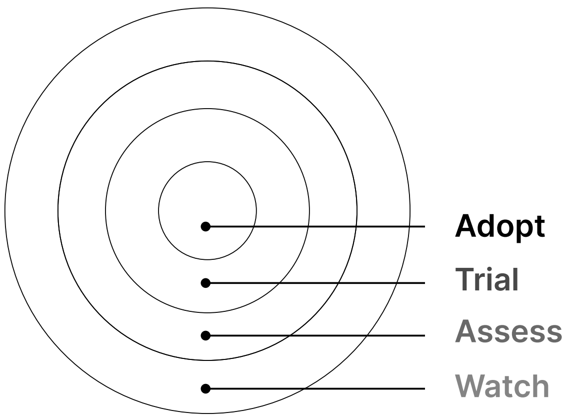
2023



Element 84

We’re glad you’re here

Element 84 launched the Geospatial Technology Radar to facilitate knowledge sharing in the geospatial community.



Here’s how to read it

Each ‘blip’ or item on the radar recieved a categorization, from the highest recommendation Adopt to Watch. Here are how we define these categories:

Ring definitions:

Adopt
Blips that we think you should seriously consider using. Something where there's no doubt that it's proven and mature for use.

Trial
Blips that we think are ready for use, but not as completely proven as those in the Adopt ring. We think you should use these on a trial basis, to decide whether they should be part of your toolkit. Typically we've used trial blips in production, but we realize that readers are more cautious than us.

Assess
Things to look at closely, but not necessarily trial yet - unless you think they would be a particularly good fit for you. Blips in the Assess ring are things that we think are interesting and worth keeping an eye on.

Watch
Things that haven’t happened yet, but we can make confident forecasts about their forthcoming developments. For example, we know a satellite or a constellation that is planned for launch that will add new capabilities.

Our Geospatial Tech Radar

Initially inspired by the ThoughtWorks Tech Radar, we have been thinking about creating a geospatial tech radar for almost a decade. Back in 2014, we published our first take on a geospatial version on our blog. It proved to be a successful post, and we loved seeing the community benefit from the information we shared. Ahead of this year's FOSS4G North America gathering, we're rolling out a 2023 edition of our Geospatial Tech Radar idea, which we plan to update annually to track what's up-and-coming in geospatial techniques, standards, data, and tools/platforms.

Why make a Geospatial Tech Radar?

Although the idea of a geospatial tech radar is unique, our inspiration for this project stems directly from the ThoughtWorks Tech Radar. In the same way that the facilitators of the radar at ThoughtWorks aim to catalyze the spread of information within the tech community, we feel that the idea of a geospatial tech radar is aligned with our dedication to developing and supporting an open knowledge ecosystem for geospatial technology. Element 84 is much smaller than ThoughtWorks, but our software engineering, UX, devops, and machine learning teams comprise one of the larger geospatial professional services firms and represent decades of experience within this space. As leaders on open standards like STAC and STAT, managers and contributors to several open source projects (stac-utils, Raster Vision, Cirrus, GeoTrellis, etc.), and teams that support an array of government, commercial, and nonprofit customers, we are exposed to a significant cross section of geospatial technology, and we believe we are in a unique position to contribute to the ever-evolving conversation around the tools and techniques used in our industry.

Further, our commitment to open technology is rooted in a desire to work with our geospatial peers in service of the greatest possible impact.

What makes our Geospatial Radar different?

When we set out to develop our take on the tech radar concept, we made several departures from the ThoughtWorks Tech Radar in terms of both the organization and the graphical presentation of the information.

If you are familiar with the Thoughtworks example, you may know that each technology entry is known as a "blip" and the blips are organized into quadrants (categories) and rings (levels of confidence in the technology). In the Thoughtworks version, the four quadrants are Platforms, Languages, Tools, and Techniques. In order to better align our commentary with our experience and expertise, we made the decision to replace Platforms and Languages with two new categories: Data and Standards. We also combined Tools with Platforms as they felt difficult to distinguish. Data sources, both commercial and open, are a key element of working with geospatial technology. And we have been leaders in both developing new standards and specifications and leveraging standards in the work we do for our customers, so the Standards category made sense on those grounds.

Second, we made some changes to the "rings". The rings are nested such that the newest technology often starts in an outer "Assess" ring and then moves inward to "Trial" and then "Adopt". Thoughtworks also has an outer ring called "Hold" that is used to designate technologies with which they have developed experience, but they recommend against using.

When we developed our draft list, we had very few blips that fell into the Hold category; I think we are generally more interested in highlighting what's new and interesting or where significant changes have taken place than in warning people off things. Instead of "Hold", we've made the outermost ring "Watch" for technologies that are approaching but may not yet exist. We felt this was both a better fit for the radar metaphor and aligned with our interest in data and tooling that we can foresee will have an impact but with which we cannot yet develop real world experience.

The Thoughtworks radar also has some additional elements. Because it's released every six months, they include some directionality for the blips, indicating if a technology is "moving" or maturing toward the Adopt status. Since this release is our first radar since 2014, we did not add this directional information. However, if we are able to turn this into an annual endeavor, we may add this to future editions. We also removed all numbers from our radar to ensure it is clear to readers that we are not ranking elements against each other. Rather, anything in the same bucket is equal.

Finally, we asked our UX team to take a critical eye to the graphic design and presentation of the radar. In addition to removing the numbers, we have added shading to indicate the distance of a blip from the innermost Adopt ring. In the online interactive version, we worked to make it possible for users to stay in one place when reviewing the detailed info rather than jumping to a different section of the page. In addition to the shift in viewing location, we also made the names of everything viewable at once.

Finally, in the PDF version, we have changed from the circular radar presentation to a gridded version at the top of each section.

How did selection for our Geospatial Tech Radar work?

Our Geospatial Tech Radar is not intended to be a comprehensive ranking of every concept and product in the geospatial technology space. Similar to ThoughtWorks, we made a deliberate decision to make our Geospatial Tech Radar an opinionated one based on actual experience, rather than trying to be a comprehensive overview of everything in the ecosystem. When you read, we're pretty confident, you'll have a lot of thoughts along the lines of "What about _____? Why wasn't XYZ technology included?" Essentially, the answer is that either a) we don't feel like we have enough experience with it to have a strongly held opinion; or b) it's so well accepted that it's not new information or nothing has changed in the recent past that we believe it's worth highlighting. For example, you might think that PostGIS is a glaring omission. We have a lot of experience with it, and it's our default database for vector geospatial data, but we believe its capabilities are so well established that we don't think we have a lot to say about it. The same goes for core tools like GDAL or PROJ. Finally, we applied some clear biases in favor of open source, open standards, and open data; we're fundamentally in the tank for cultivating an open knowledge ecosystem for doing geospatial work.

In terms of our process, we gathered our key technology leaders in the organization to outline the concept and solicit input. We then compiled a list of technologies with which we had experience and we felt we might have something useful, interesting, or otherwise compelling to say.

Once our list of blips was solidified, engineering and design leads took to writing their thoughts on each one.
By pulling direct insight from all sides of the organization, we were able to compose a snapshot of each blip from folks who interact with these principles and products on a daily basis.

What's next for the Geospatial Tech Radar?
In a future version of the Geospatial Radar, I'd like to make several changes. First, I'm pretty confident that we probably have some omissions that simply didn't occur to us on this first version. Second, we are interested in digging into the "direction of travel" question outlined above. Third, I'd like to be able to articulate some general themes that we can extract from what we have chosen to highlight.

If you're stumbling on this post soon after its publication, you can learn more about the process and inspiration behind our radar at FOSS4G-NA 2023 in Baltimore October 23-25. At this point, we're excited to get the word about this idea, and we're excited to start brainstorming for next year's edition. Check out the interactive radar or download it as a PDF. Also, we would love to hear from you! Get in touch with our team if you're interested in learning more about the radar, or about any of the specific elements we cover in the radar itself!

Contents

Tools

ADOPT

TRIAL

ASSESS

WATCH

Dask

Dask is a Python tool for parallelizing and distributing computation across a cluster of machines.

It is useful when you have workflows that take a long time or data that does not fit into memory. Dask mimics the interface of familiar scientific Python libraries numpy, scikit-learn but instead of eagerly computing the result, it constructs and optimizes a task graph before distributing tasks to its workers. It is straightforward to get started using Dask. Some libraries (like xarray) automatically integrate Dask, so you can benefit from parallelization without even knowing about it. Since Dask is 100% Python, all stack-traces and debugging are also in Python. Since it is entirely open source there are a myriad of ways to deploy distributed Dask clusters to various HPC and cloud computing environments as well as several companies that provide managed solutions: Coiled, Domino Data Labs, and Saturn Cloud.

Figma

Figma has improved dramatically in the past year, particularly with the introduction of a new dev mode, variables, and robust community plugins.

When Figma was first acquired by Adobe a ripple of fear spread through the design and developer community - would our beloved tool become overly complex and inevitably slow? Figma has squashed those fears fully. It remains efficient yet feature-rich and remains our one and only tool for mocking up designs. From a geospatial perspective, we've reviewed the kits from ARCGIS which we found to be strong, but not necessarily more tailored to geospatial than other UI kits. The World Map by Matt Wright and Placemark by Tom MacWright have been useful community plugins and we love pulling real data from Google Sheets into our designs, and this write-up from Atomic Object on Google Sheet Sync by Dave Williames covers that well. Covering all of what makes Figma an incredibly powerful tool could be its own radar in and of itself - especially given how rapidly it is changing. Our takeaway is that switching from Figma is nearly unimaginable at this point in time.

GeoPandas

GeoPandas gives you all the familiarity of pandas with geo functionality sprinkled on top!

You can read data from old crusty data formats (shapefiles, geojson, geopackage...) or from new sparkly ones (geoparquet, flatgeobuf...). Once your data is read in, GeoPandas integrates beautifully with with rest of the scientific Python libraries; you can use pandas indexing to subset your data, you can plot using matplotlib (or just use `.explore()` to get a quick interactive map), and you can even speed up analysis by using dask-geopandas.

Hugging Face

Hugging Face is a community and data science platform that supports use of machine learning models based on open source technology.

It supports a range of data types and can be used to train and deploy a wide variety of geospatial models, including models for image classification, object detection, and segmentation. The Hugging Face Hub provides a number of pre-trained geospatial models that can be used for a variety of tasks, such as land cover classification, disaster response, and urban planning. For example, Google AI is using Hugging Face to train and deploy models for natural disaster response, and IBM and NASA are working with Hugging Face to make available a geospatial foundation model. We categorize HuggingFace as an “adopt” for geospatial users, especially the community and registry components. The suitability of the HuggingFace code library depends on the context of the particular task.

H3.js

H3.js, in conjunction with the H3 geospatial library, offers a robust and versatile solution for geospatial data visualization.

This tool is highly effective for aggregating and mapping data, providing a clear and concise representation of data points on a geographical map. The H3 library's grid-based approach simplifies the process of analyzing data across regions, enabling a wide range of applications, from location-based analytics to spatial data exploration. Whether used in business intelligence or urban planning, H3.js within the H3 library stands as a valuable asset for any project requiring insightful geospatial data visualization and aggregations.

MapLibre JS

MapLibreGL.js stands as a versatile and powerful tool within the realm of open source web mapping libraries.

Its capabilities empower developers to create interactive and feature-rich maps for a wide array of applications. Leveraging MapLibreGL.js, teams can seamlessly integrate maps into their web-based projects, offering users a visually appealing and data-rich experience. Thanks to its use of WebGL technology, this library excels in rendering performance, ensuring smooth and efficient map displays. With its flexibility and extensive feature set, this library has become an essential component for any endeavor requiring effective and engaging web-based mapping solutions. Its adaptability and support for geospatial data visualization make MapLibreGL.js a valuable asset in the toolkit of modern web developers.

OpenDroneMap

OpenDroneMap (ODM) is an open-source software package that offers robust photogrammetry capabilities, enabling users to convert drone-captured images into high-quality 2D and 3D maps, models, and orthomosaics.

Additionally, OpenDroneMap is user-friendly, making it suitable for both beginners and experienced professionals in fields like agriculture, construction, environmental monitoring, and more. Its active community and continuous development ensure that it stays up-to-date with the latest industry standards and technological advancements. By adopting OpenDroneMap, users can streamline their data processing, reduce costs, and enhance their ability to derive meaningful insights from drone-collected data.

PDAL

The Point Data Abstraction Library (PDAL) is the point cloud equivalent of the venerable Geospatial Data Abstraction Library (GDAL).

Led by Howard Butler at Hobu, PDAL has become the main game in town for point cloud format translation, analysis, and general processing. Its concept of pipelines provides a flexible framework for building performant re-usable data processing workflows, and its huge number of format readers and writers formats make it useful in almost any point cloud context. With an initial v1 release in July of 2015, it's been around long enough to go through several architecture and interface iterations to improve performance and user experience. It is stable, widely used, and actively supported; we categorize PDAL as Adopt.

PostGIS

PostGIS is a geospatial extension for PostgreSQL that includes raster, vector, and topological spatial data types, along with functions for working with and storing such data.

PostGIS enables users to leverage a single database and query language for both their standard relational data and spatial data in a highly performant manner, complete with all the expected PostgreSQL functionality around transactions and data integrity guarantees. Element 84 makes use of PostGIS on many of our projects because of these properties, and we regard it as something of a geospatial Swiss army knife; if you can think of a geospatial data processing problem, PostGIS can likely do it out of the box or provides a path to making it happen. The included function library enables everything from vector feature analysis via spatial joins and geometry operations to raster operations like map algebra. It even has built-in functions for less obvious things like vector tile generation. Another key feature is its support for spatial search operations against stored data via several different types of spatial indices. PostGIS is a fundamental tool for backing spatially enabled web applications, as well as a data science tool for performing spatial analyses. To learn more about PostGIS, one good place to start is the official "PostGIS Workshop". This hands-on self-paced course allows users to start from zero knowledge and effectively go as deep as they want into PostGIS concepts and how to use them practically.

Raster Vision

Raster Vision is an open source library and framework that makes it easy to apply deep learning to geospatial imagery and produce georeferenced outputs.

As a library, Raster Vision provides a full suite of utilities for dealing with all aspects of a geospatial deep learning workflow: reading geo-referenced data, training models, making predictions, and writing out predictions in geo-referenced formats. As a framework, Raster Vision provides a low-code way to quickly and repeatably configure experiments that execute a machine learning pipeline including analyzing training data, creating training chips, training models, creating predictions, evaluating models, and bundling the model files and configuration for easy deployment. Improvements in recent releases include: better STAC integration, support for temporal data, and support for exporting and using ONNX models.

TiTiler

TiTiler, an innovative project available on GitHub, offers a dynamic solution for those seeking to visualize raster data stored in the cloud.

TiTiler stands out by not only providing a fast modern framework built on the Python FastAPI library, but it is presented as a series of building blocks to create your own custom tile service API. Furthermore, its compatibility with Cloud Optimized GeoTIFF (COG) and STAC ensures seamless integration with cloud-native geospatial workflows. We recommend users trial TiTiler to test its flexibility, speed, and adaptability to various use-cases. It will require that other technologies are in use (such as COGs), and will require some work in customizing and deploying it depending on the application and cloud provider.

TorchGeo

TorchGeo is an open source library developed by the Microsoft AI for Good team that significantly overlaps with Raster Vision in terms of the capabilities that it provides.

TorchGeo is more tightly bound to PyTorch than Raster Vision, which has been designed to support other backends. TorchGeo also uses PyTorch Lightning for model training whereas Raster Vision was designed to use custom code for this functionality (though it is possible to use Raster Vision with Lightning as well). Both approaches have their pros and cons. For instance, TorchGeo has better support for acquiring and using data from well-known datasets. On the other hand, Raster Vision arguably provides a richer set of data manipulation tools, which allows for more customized inputs and outputs, as well as the Pipeline functionality, which allows Raster Vision to be used as a low-code framework. Users should choose between the two based on their use case.

Google Earth Engine

At Element 84, we have used Google Earth Engine quite a bit to do some simple land classification and light analysis.

It's a fine place to prototype our workflows and to get a quick view of a large amount of data. Where it is underutilized and should get more attention is as a source for vast, vast quantities of curated geospatial data that can be exported at very low cost. Want a cloud-free S2 mosaic over Azerbaijan in May of 2018? Easy. In our experimentation, it has been about as frictionless a path to data that is suitable for training and inference runs on ML models as anything else we have come across.

Trial x Tool

Kepler

Initially developed at Uber, kepler.gl is an open-source, WebGL-powered library designed specifically for the exploration and visualization of large-scale geolocation data sets.

With an intuitive user experience, low-coding requirements for basic visualizations, and easy integration with Jupyter notebooks, it simplifies the often-complex task of data-driven map creation. We believe that kepler.gl is a tool that's worth trialing. Its ability to handle large datasets without compromising on performance ensures that even teams with limited experience in geospatial analysis can produce meaningful, interactive visualizations.

Kubernetes

Kubernetes is a valuable technology to be familiar with, as some clients are coming to us with established infrastructure and need folks to manage it.

Some new clients will benefit from using it over AWS services, but unlike AWS services, the management and upgrade of k8s-deployed services will need to be provided by e84 engineers, so succession planning for contract EOL must be considered. To deploy for new clients or in-house use, we need a reference deployment; this will establish a base cluster and deploy ArgoCD or its equivalent, with which we can use Gitflow as our management strategy. We need some trial time with this deployment style to learn it well, but automatic sync of cluster state with a repository of manifests and Helm charts is superior in practice to, e.g., Terraform deployments. Experimentation with Crossplane to manage AWS resources through the same infrastructure is advised to further reduce friction.

PySAL

PySAL provides Python implementations of spatial algorithms.

It is used primarily with vector data and has substantial educational resources. The PySAL maintainers even have a book: Geographic Data Science with Python. PySAL integrates and builds off of the rest of the scientific Python stack (geopandas, matplotlib...) and shares maintainers with other core scientific Python libraries.

AtlasIcons

Fontawesome continues to be our go-to icon set, but this new open source MIT-led kit is strong, and has a great set of ecological and a few unexpected map-related icons.

As our work continues to grow in the world of sustainability we are looking towards kits like this. We love that they are open source, accept suggestions, and enable building custom icons. They also built a Figma plugin to drop the icons right into your designs. From a geospatial perspective, we were surprised and impressed by their monument icons, which we could see being useful for playful landmarks, and are more globally minded than ones we’ve seen elsewhere. Their electric charging station icons are also superior, and we hope to be seeing more of those on maps. For a kit just about the 10th of the size of Fontawesome, we’re impressed but not ready to fully switch.

AWS Sagemaker Geospatial

SageMaker Geospatial is a new AWS service released for general availability in May 2023.

The service shares a brand name with the more general AWS SageMaker service, but it offers specialized geospatial-related functionality, such as datasets and earth observation jobs. It is currently a notebook-only capability (no batch capability) that is available only in the us-west-2 AWS region; as a notebook-based capability, it can be compared to offerings like Microsoft's Planetary Computer and Google Earth Engine. Our current categorization of SageMaker Geospatial is Assess. SageMaker Geospatial has the potential to be very exciting, but we recognize that it is still under development and we believe that it is only suitable for exploratory work at this time.

DuckDB

DuckDB is an embedded analytical database management system with exceptional query performance, especially for analytical workloads, due to its vectorized execution engine.

DuckDB is lightweight and easy to embed in applications, making it a great choice for projects with limited resources. Its support for SQL and compatibility with popular database interfaces like PostgreSQL and SQLite allow for seamless integration. Furthermore, DuckDB excels in handling complex queries and analytical operations, making it suitable for data analytics and reporting. However, DuckDB is less suitable for transactional workloads and concurrent write-heavy applications, as it prioritizes read-heavy use cases. Additionally, its ecosystem and community support, while growing, may not be as extensive as that of more established database systems, limiting the availability of resources and third-party tools.

Open Source Large Language Models (LLMs)

LLMs have received a lot of attention for their utility in generating text and images through tools like ChatGPT, BARD, Claude, and many others, but LLMs also have the potential to revolutionize the way we interact with geospatial data.

For example, LLMs can be used to write natural language queries for geospatial data, which can make it easier for users to access and analyze the data they need. LLMs can also be used to read, translate, summarize, and understand geospatial metadata, which can help users to better understand the meaning and context of their data. Open source LLMs, such as Llama and Falcon, are particularly attractive for geospatial applications because they are freely available and customizable. This means that users can fine-tune LLMs to their specific needs, such as the type of geospatial data they are working with or the specific tasks they need to perform.

OpenTofu

Recently, Hashicorp re-licensed Terraform under the Business Source License (BUSL) which has led to uncertainty around Terraform usage due to possibly ambiguous language in the license.

As a result, the OpenTofu project was created to maintain a fork of Terraform (from version 1.5) under the Mozilla Public License 2.0. Though any fork of a major technology raises concerns about future development and maintenance, OpenTofu has a large active community with multiple corporate sponsors, and sits within The Linux Foundation. Though OpenTofu has feature parity with Terraform 1.5, we expect there to be drift as each project continues. Thus we recommend organizations concerned with using BUSL-licensed software begin to trial OpenTofu.

Segment Anything Model (SAM)

SAM is an open source AI model released by Meta that can be used to identify and “cut out” any object from an image.

SAM is a powerful tool for geospatial segmentation (assigning each pixel in an image to a class or object). We categorize the underlying technology as a strong “adopt”. However, the off-the-shelf version's limitation to three bands (red, green, and blue) is a drawback, as it does not allow users to take advantage of the information found in additional bands from multispectral or hyperspectral imagery. Geospatial imagery often includes bands such as near-infrared (NIR) and shortwave infrared (SWIR), which can provide valuable information about features such as vegetation, water, and soil. By incorporating these additional bands into SAM, users could achieve more accurate and informative segmentation results. To fully harness the power of geospatial imagery, users could either train a custom SAM model with additional bands or employ a hybrid approach (using SAM unchanged to produce preliminary results together with another model to produce final results). While the existing code supports the latter, a more integrated solution would probably necessitate reimplementing the model from scratch and extensive training.

Stamen X Stadia Tiles

In the evolving landscape of web mapping libraries, the migration of Stadia Tiles from Stamen Maps represents a significant shift.

For more than a decade, Stamen provided free, custom-designed basemaps, catering to a wide range of users, thanks to their dedication to open data and cartographic expertise. As time went on, maintaining and modernizing these maps became increasingly challenging, necessitating a transition. Stadia Maps has emerged as a valuable partner in this endeavor, offering a reliable tile service that not only preserves the iconic map styles, but also enhances them. Utilizing modern open-source projects like OpenMapTiles and MapLibre, Stadia Maps employs vector tile technology to rebuild and host updated data, ensuring more stability and support. This transition brings several advantages, including vector map options that offer smoother zooming, reduced bandwidth usage, and greater compatibility with popular mapping libraries. While the shift is undoubtedly significant, the legacy of Stamen's map tiles will continue to thrive, thanks to Stadia Maps and their commitment to accessible, free mapping solutions for the community.

Felt

Felt's mapping application emerges as a versatile tool in the landscape of web mapping tools.

This application is designed to empower users of all backgrounds, offering a platform for a diverse range of mapping endeavors, from apartment searches to brainstorming sessions. Felt simplifies the process of map creation, enabling users to bring their ideas to life, incorporate data, or engage in creative drawing. Felt makes collaboration effortless, as colleagues and friends can join in on projects or explore the final creations. Felt's customization options allow users to tailor every aspect of their maps, from pixel-level details to a wide spectrum of styles. Additionally, the tool supports the transformation of static maps into interactive interfaces that tell engaging stories. With Felt's data integration capabilities, users can effortlessly incorporate datasets such as political districts and transit routes, turning them into visually appealing assets. The resulting detailed maps can be shared and embedded, providing a personalized and ad-free alternative to traditional mapping services. Felt encourages collaboration and empowers users to address local challenges by involving communities in map-based solutions.

GitHub Copilot

GitHub Copilot is impressive and the ergonomics of an in-editor assistant remain uniquely valuable.

However, the rapid pace of large model advances in terms of the cross-fertilization of information from different domains and emergent reasoning capabilities has diminished the relative utility of Copilot significantly. One could argue that an OpenAI subscription, even at double the cost of Copilot ($20 vs $10), is nevertheless a far better deal. Copilot can write functions that the user describes in comments or add comments to existing code, but GPT enables engineers to engage in back and forth about theoretical considerations, API design, and other matters which seem to be a bigger impediment to productivity than merely writing a block of code (at which it is quite good as well). There have been concerns raised about privacy, but in our experience, prompts tend to be abstract enough to avoid anything confidential or project-specific, and these concerns seem to be fully obviated by the "enterprise" offering of ChatGPT.

OS-Climate

OS-Climate aims to assemble data and create software tools and a platform for measuring climate risk and making climate-aligned decisions in order to "dramatically boost global capital flows into climate change mitigation and resilience".

The past several years have seen hundreds of newly funded startups proposing new technologies related to climate change, and many of these - Treefera, Pachama, Floodbase, Vibrant Planet, and Impact Observatory, to name just a few - involve geospatial Earth observation data. Many of these systems rely on some combination of public and private data, and public Earth observation infrastructure provided as public open data - Sentinel-5P, GOSAT, Landsat, Sentinel-2, PRISMA, ENMAP, and MODIS are some of the notable ones - plays a key role enabling new climate-related products and services. These startups are attempting to address an array of climate adaptation and climate mitigation challenges: flood and wildfire response as well as greenhouse gas monitoring, deforestation, biodiversity, and land management. However, while many of them rely on public data and open source tooling, few of them are building business models that support and extend open source software. OS-Climate is an exception. Hosted by the Linux Foundation, OS-Climate aims to assemble data and create software tools and a platform for measuring climate risk and making climate-aligned decisions in order to "dramatically boost global capital flows into climate change mitigation and resilience".

The project has launched several initiatives, including: transition analysis, physical risk & resilience, portfolio alignment, and a data commons. The initial members are mostly large finance and global consulting firms like CapGemini, BNP Paribas, Goldman Sachs, EY, Allianz, BNY Mellon, Amundi, and S&P Global. These are not the organizations that we typically associate with open source projects, but, at least in the marketing materials, they are recognizing the power of open source and open data to support technology collaboration at scale. And while many of the members are competitors, the OS-Climate efforts are aimed at building common infrastructure and tools as a "pre-competitive layer" (that's a new term for me). Further, they seem to be working with NGOs that are leaders in this space, such as World Resources Institute, Open Earth, Climate Policy Institute, and World Benchmarking Alliance. However, if you are thinking about becoming a member, unless you are an NGO or academic, you have to already be a Linux Foundation member, and then OS-Climate dues are $30,000/year for General membership and $100,000year for Premium. Alas, that's a bit spendy for a small business. The working groups appear to mostly be off the ground, but there isn't a lot of tooling yet, so we are assessing this as a Watch. If you would like to follow along, there is an OS-Climate Community Hub on GitHub.

Data

TRIAL

ASSESS

WATCH

Earth on AWS

Earth on AWS hosts a diverse range of public datasets, including well-known collections like Landsat, Sentinel-2, NAIP, GOES, as well as specialized regional datasets such as Southern California Earthquake data or Elevation data for the Arctic.

If you are already utilizing AWS and require access to any of these datasets, they can be a valuable resource. However, we categorize their usability as “Trial” because each dataset varies significantly in format and scope, necessitating separate evaluation. E84's Earth Search offers a STAC API for some of the more popular data collections, but there is no unified, consistent distribution method for all datasets. Each dataset may be under the management of different third-party entities, with no guarantees regarding their quality, consistency, or ongoing availability. Therefore, it is crucial to carefully assess each dataset to determine its suitability for your specific application. One notable exception to this assessment process is Landsat data, as the collection of Landsat data on AWS is the authoritative source managed by USGS, ensuring reliability and continuity.

Planetary Computer

The Planetary Computer is Microsoft's petabyte-scale catalog of geospatial data and associated services.

It includes the "foundational" earth observation products, including Landsat and Sentinel 2, as well as a huge number of other datasets around a variety of domains including climate, oceanography, and terrain. These data are provided free-of-charge, though they do use an authentication layer to prevent abuse of their systems. Alongside these datasets, Microsoft hosts several customizable compute environments which enable data scientists to do their work alongside their data, thereby avoiding expensive and slow cross-region transfers. We categorize the Planetary Computer as trial because we have seen its use in production systems, but we have not yet observed its use at the scale necessary to give us confidence in building a long-term business plan around their services.

Planetary Variables

Planet Labs has developed a new family of products, called Planetary Variables, that represent what we believe is the most sophisticated effort yet to create new products that leverage both their data as well as open data from other Earth observation instruments.

Commercial Earth observation satellite constellation operators have often sought to break out of a business model that amounts to “selling pixels”. The theory goes that if there are lots of pixel sellers, the imagery products will become more difficult to differentiate and it will be a competitive race to the lowest price, so a satellite operator should try to create value-added products that represent a more sophisticated (and valuable) product. While selling pixels is currently a viable business model (and some companies, like Umbra, are saying they are just going to focus on being the best pixel-sellers around), many commercial providers aspire to move up the value chain and provide analytics and services that are built on top of the data they collect. Some satellite operators, like Maxar, have developed or acquired professional services teams that support their customers, develop[ing both analytics and applications. Planet Labs, while also has its own professional services team and is mostly selling pixels right now, has developed a new family of products that represent what we believe is the most sophisticated effort yet to create new products that leverage both their data as well as open data from other Earth observation instruments. Planet is calling these new products “Planetary Variables” (which is a pretty cool brand).
The first set of these Planetary Variables was developed through an acquisition of VanderSat, which had developed several novel data analytics products for computing soil moisture, land surface temperature, vegetation optical depth, and biomass, by combining data from Planet, NASA, ESA, and JAXA satellites. The initial Planetary Variables are correspondingly related to soil moisture, crop health, land surface temperature, forest carbon, and forest structure. This early focus on Earth systems is a great example of how Planet is pursuing its mission to not only image the Earth’s landmass every day but to also make global change visible, accessible, and actionable. However, while these initial variables are clearly focused on natural systems that support agriculture, forestry, and climate change, Planet has more in the wings, including a variable that monitors building and road construction. This doesn’t map building footprints and road centerlines. Rather, it uses machine learning to look for recent changes on the land surface that indicate new buildings and roads are under construction and flags those locations. If you buy this Roads and Buildings product, you subscribe to receiving regular updates for the areas in which you are interested. These updates are delivered as GeoJSON highlighting the detected changes as well as access to base maps that can support your own analysis of the results. While this is not open source software, Planet has committed to publishing peer-reviewed articles that outline their methodology in a transparent way. We believe that Planetary Variables will likely still require integration into a customer’s workflows, applications, and infrastructure, but they are a compelling new approach to getting advanced analytics into the hands of those that need them.

High Resolution SAR

Synthetic Aperture Radar (SAR) data is not new; the concepts date from the 1950's.

Unlike the camera used for optical data, a SAR sensor is mounted on a moving platform that emits radar pulses, and the distance traveled by the platform while the surface target is illuminated by the radar pulse is the "synthetic" antenna aperture. In most cases, the larger the aperture (distance traveled by the moving antenna), the higher the image resolution will be. This attribute means that SAR can create high resolution images from a relatively small physical antenna. SAR is interesting in other ways. While we often see SAR outputs as images (for example, displaying building or earthquake damage), the creation of a SAR data capture involves the transmission of successive pulses of radio waves to "illuminate" a target. The antenna records the echo of each pulse as the instrument moves through its orbit. To create a SAR image, successive pulses of radio waves are transmitted to "illuminate" a target scene, and the echo of each pulse is received and recorded. The pulses are transmitted and the echoes are received using a single beam-forming antenna, with wavelengths of a meter down to several millimeters. As the SAR device on board the aircraft or spacecraft moves, the antenna location relative to the target changes, and many overlapping echos of the same location are recorded. The result is a complex data structure from which we can extract an image of pixels representing elevation, for example, but there is often much more information in this phase history. This is really cool technology, but it's complicated to wrap your head around and we believe it remains largely underutilized for this reason.

We would like to see more open source tooling that will enable people to exploit these capabilities. While SAR is not new, we believe it is worth highlighting several recent developments. First, two commercial SAR constellation operators, Umbra and Capella, have launched satellites capable of much higher resolution captures. In the Umbra case, you can now purchase [16cm resolution SAR images. Capella has been flying longer, and they are now in the process of launching a new constellation, called Acadia, that offers high resolution captures as well. These two companies are also notable for their release of open data Umbra open data; [Capella open data for some locations. Umbra has also broken new ground with transparent pricing, data licensed under a Creative Common license (allowing others to reproduce and distribute its imagery as they see fit, so long as Umbra is credited), and a commitment to stay out of the analytics and professional services business. Finally, it's important to note that this high resolution SAR imagery has been possible for some time, and the key change was the expiration of regulatory limits on U.S. SAR companies by NOAA and the Commerce Department in August 2023. This change releases domestic SAR companies from most regulatory limits on what they can sell commercially, and we look forward to seeing the applications that this new data will enable.

Hyperspectral

Until recently, imaging satellites used cameras that capture the Earth in a few frequency bands.

Multispectral sources typically cover the optical red, green, and blue bands of visible light in addition to a few other frequency ranges (Sentinel-2 captures 13 bands and Landsats 8 and 9 and 11 bands). Hyperspectral imagery captures an entire block of the electromagnetic spectrum, enabling capture of up to 100's of spectral bands. Different materials can often be identified by their characteristic spectral signature. Hyperspectral imaging sensors cover a larger frequency spectrum in narrow, continuous frequency bands. They enable the capture and identification of spectral profiles. Hyperspectral imagery enables applications such as tree and plant species identification, monitoring of air quality (particulate and gaseous) and water quality (e.g. algal bloom detection), tree health observation, and more. Hyperspectral imagery is not new, and NASA JPL has periodically flown AVIRIS and AVIRIS-NG as airborne sensors for several years. The Italian Space Agency launched PRISMA in 2019 and Germany attached the DESIS instrument to the International Space Station in 2018 and then launched EnMAP in 2022. Demand for the latter is so high it's overwhelmed the queuing software. But this is just a taste of what's coming. PIXXEL plans a constellation of both VNIR and SWIR hyperspectral satellites over the next two years. Planet Labs will launch the first of two Tanager satellites in 2024.

Designed and manufactured by JPL, these satellites will initially focus on tracking point-source methane and CO_2 emissions, but they will be capable of much more. These will be joined by hyperspectral instruments from Esper, Wyvern, Orbital Sidekick, HySpec, and probably others. We are pretty excited about what will become possible in the next year. If you are interested in learning more about Hyperspectral imagery, we've created an explainer page.

OpenStreetMap for Foundation Model Training

OpenStreetMap (OSM) has been used extensively as a key global, geospatial data source for more than two decades.

The reason we are highlighting it here is its potential utility as a semantically rich data source for training geospatial foundation models. OSM text tags can be used to describe a wide range of geospatial features such as roads, buildings, natural features, and points of interest. One way to use OSM tags to train foundation models is to use a multi-modal approach similar to that used in the OpenCLIP model. In this approach, images and text tags would be used to train the model. The model would learn to associate the tags with the corresponding images, and this would allow the model to understand the meaning of the tags. We believe there is a great deal of promise in the use of OSM data for developing foundation models for geospatial data, and it is an active area of experimentation for us.

Overture Maps Foundation

The OpenStreetMap project is a longstanding success story for the construction of open, collaborative geospatial databases on a planetary scale.

It is unrivaled, and, alongside Wikipedia, it represents a landmark for what we can accomplish through collective action. However, significant contributions by large commercial organizations to the OpenStreetMap database have created tension within the community. In addition, many rivals to Google are likely feeling a sense of urgency to develop a viable alternative to GoogleMaps at a faster pace than the OpenStreetMap community is able to move. In 2022 Meta was joined by Microsoft, TomTom, and Amazon to launch the Overture Maps Foundation, a project under the aegis of the Linux Foundation's Joint Development Foundation (that's a lot of foundations in one sentence). Overture is motivated by four key aims. First, build a planetary scale map in a collaborative manner from disparate sources. Second, create a global entity reference system in order to better integrate multiple data sources that may reference real-world entities. Third, create a robust validation and error detection system. Fourth, implement a common, structured data schema. So far this year, Overture has signed up new contributing members and has dropped three data releases. Data releases are currently available in Parquet format from AWS and Azure. They are organized into five themes.

The Base Theme with land, land use, and water; a Buildings Theme; an Administrative Boundaries Theme; a Places (point-of-interest) Theme; and a Transportation Theme for roads, lanes, turn/access restrictions, and related data. Despite the implicit critique of OpenStreetMap represented by the global entity reference system (GERS) and a common data schema, Overture's data is designed to be complementary to OpenStreetMap and is released using OSM's ODbL license or the CDLA (for the Places Theme). Indeed, many of the themes integrate OSM data as a key component. There has been growing concern about the degree to which so much of the geospatial data infrastructure we rely on daily is in the hands of a single corporation in the form of Google Maps. The restrictive license applied to Google Maps users has attracted the attention of the Department of Justice Antitrust Division, and its possible that the competitive pressure represented by Overture may be encouraging Google to release some of its geospatial buildings data. However, despite substantial progress in a short period of time, the 2023 Overture data releases are marked as "alpha" and both GERS and the common data scheme are provisional. We recommend that interested organizations carefully assess the data before considering integration into a production system.

PLACE

Mapped data is critical to advancing economic, social, environmental, and other public good. When we consider the contemporary geospatial data landscape, we can observe that two distribution and intellectual property models have emerged over the past few decades.

When we consider the contemporary geospatial data landscape, we can observe that two distribution and intellectual property models have emerged over the past few decades. On one pole, commercial proprietary data licenses occupy a significant and powerful source of the data we use to do our work and live our lives. GoogleMaps, Planet, Pixxel, Maxar, EagleView, Esri, and a host of other firms sell or license the data they have collected. On the opposite pole, we have data released either into the public domain, mostly by government entities such as NASA, ESA, US Census, USGS, and so on, or through an explicit open data license, such as OpenStreetMap and Overture Maps Foundation. However, we should think of intellectual property as a spectrum of options, and there has been a relative poverty of alternatives to what feels like two polar opposites. Enter PLACE, a technology organization that emerged from the Omidyar Network (ON) in 2020. PLACE is founded on an observation that high-quality foundational mapping data is unequally collected and distributed and much of the reason for this can be traced back to the intellectual property models upon which the status quo rests. PLACE aims to create a "club goods" model for collecting and distributing geospatial data in the places where it is most needed.

PLACE serves as a trusted intermediary (and has literally created a legal data trust in the United Kingdom to hold data and licenses) between national governments and users of the data. PLACE negotiates agreements directly with national governments (currently in Africa and the Caribbean). All data produced in partnership with PLACE belongs to the government of each country. PLACE then receives an irrevocable, perpetual, royalty-free license from each government to copy of all data and its use by PLACE members through the PLACE Trust. This legal configuration is complex and has taken some time to set up, but PLACE is now off to the races in multiple countries, collecting high-resolution aerial and ground imagery in urban areas. They have iterated multiple aerial drone and camera configurations and have begun releasing prototype imagery in a few locations. They have also signed up the first batch of member organizations, and we're proud that Element 84 is one of them. Guided by a clear set of principles, PLACE is also working to catalyze local data collection and processing ecosystems in the low-income countries in which they operate. Despite the early success, however, PLACE has a lot of scaling to do before it will be able to claim broad data coverage. However, they represent a rare example of geospatial data collection efforts experimenting with new licensing and intellectual property models, and we encourage people to evaluate what they are offering.

Source Cooperative

Google, Microsoft, Apple, and large commercial firms have been developing, training, and using machine learning tools for mapping the Earth for several years.

Although OpenStreetMap and complimentary projects like Overture Maps Foundation, Humanitarian OpenStreetMap Team, and OpenAerialMap represent an open alternative, and projects like the Meta's Rapid Editor are applying AI in an open way, there remains substantial concern that the best machine learning models and training data for geospatial are being locked up behind the walls of commercial organizations. It is easy to forget that much of the contemporary progress in generative AI for text and imagery is grounded in rapid progress made possible by openly licensed imagery and text archives such as ImageNet. In the geospatial ecosystem, the CosmiqWorks initiative (now defunct) ran a series of ML competitions as SpaceNet. In December 2019, the Radiant Earth Foundation launched ML Hub as an open library of machine learning models, imagery, and training data. When Radiant Earth hired Jed Sundwall as its new Executive Director in 2022, he brought the experience of both having been the founder of an open data program in San Diego and creator and leader of the AWS Open Data program and the Registry of Open Data on AWS. Jed had an expansive vision for how the open geospatial data ecosystem could be improved, and since he joined the Radiant Earth team has folded the original ML Hub catalog as well as a broad array of other data into a new service, Source Cooperative, launched in June 2023.

Radiant Earth believes it should be much easier for both individuals and organizations to share and use open data. Source aims to be a neutral non-profit data utility that will enable organizations to share data without having to manage their own infrastructure. It is designed around sharing data using standard HTTP methods (rather than custom APIs or commercial cloud services). While Source isn't limited to geospatial, most of the initial datasets are geospatial and made available using cloud native geospatial standards and specifications. Source has already garnered significant traction and is now hosting data products from Maxar, VIDA, Streambatch, the Ecological Forecasting Initiative, the Clark University Center for Geospatial Analytics, and Protomaps. Critically, there is not only an array of data but developers are beginning to create new things using that data. Check out a couple of examples, including a mashup of building footprint data with FEMA flood hazards from Postholer, and this tutorial from VIDA showing how to use DuckDB to explore over 2.5 billion building footprints. While Source is in beta and under active development, we believe it is worth assessing for both data providers and application developers interested in leveraging cloud native geospatial open data. It is early days, but Radiant Earth has recently outlined some more details about their business model and roadmap for the coming year, and we encourage members of the geospatial community to consider how to leverage this important new platform.

NISAR

We wrote about High Resolution SAR in this section as well, but more developments are on the horizon for the SAR community.

The NASA-ISRO Synthetic Aperture Radar (NISAR) is expected to launch in 2024 (which is why we assessed it as Watch). Earth science is the primary objective for this joint NASA/ISRO mission, which will map the entire globe about every 12 days in order to collect a spatially and temporally continuous data set for understanding ecosystems, ice mass, biomass, sea level, groundwater, and natural hazards related to the Earth's surface, such as earthquakes, tsunamis, volcanoes, and landslides. NISAR will collect both L- and S- microwave band SAR data and because it will periodically revisit the same locations, it will be able to produce repeat-pass interferograms that can measure surface motion and change as small as 4mm per year. This technique is known as Interferometric SAR (InSAR). While NISAR is aimed at societal good and the data will be released as open data, we believe there will be opportunities for the commercial SAR satellite constellations collecting X-Band SAR to create data products that can leverage the continuous global coverage that NISAR will provide.

Planet Pelican

Planet Labs has two constellations: one that is made up of Doves and SuperDoves and continuously captures imagery of the Earth's land at medium resolution (3-5m), and a second that can be tasked to capture higher resolution images of specific locations.

The latter constellation dates back to Planet's 2017 acquisition of Skybox Imaging (aka Terrabella) from Google and was supplemented with new launches in 2017, 2018, and 2020, creating a total of about 20 SkyBox satellites today. Since then, Planet has been developing a successor constellation for the SkySat instruments, known as Pelican. The first demonstration satellite is expected to launch in November 2023 and is described as aiming to support "four 30's": up to 30 satellites, each capturing up to 30 images per day, supporting 30 min revisit cadence and 30 cm resolution. Because this doesn't exist yet, we are assessing Pelican as Watch, but this is going to be a significant capability and capacity enhancement for Planet, and we're excited about the new applications this will enable as well as representing a more significant competitive challenge for other high-resolution satellite operators like Maxar.

Techniques

ADOPT

TRIAL

ASSESS

WATCH

Accessible Design and Testing

Accessibility ensures that a tool can be used by all people, regardless of disability or ability.

Centering accessibility in software development starts with accessible UX design — by including disabled perspectives in user research and designing according to accessibility best practices (WCAG 2.1+). Writing accessible code requires training developers, as well as choosing component libraries and other tools that implement accessibility best practices. Making visualizations of complex datasets, such as maps, accessible is both complex and under-explored. Projects like Visa's Accessible Chart Components are steps in the right direction. Finally, accessibility testing ensures compliance, including Section 508 (Federal electronic content has to be accessible). In addition to automated testing during development, hybrid tools like Accessibility Insights for Web help identify and document issues. Manual testing is often required for screen reader accessibility. CANAXESS' approach creates consistency across screen readers and operating systems. We believe that after making initial improvements accessibility professionals and users with accessibility needs should be consulted for further testing and compensated for lending their expertise.

Adaptation Learning

Adaptation learning specializes in tackling domain shift, a problem that arises when the source and target datasets differ in distribution.

This technique is invaluable in the geospatial field, where an existing model for one type of sensor may not work well for another. Another motivating example is of a model that is effective on U.S. satellite images, but which falters on images of Africa due to varying vegetation and climate. Thus, we recommend the use of adaptation learning. No additional tools are needed beyond your existing machine learning framework, however, training data from the new domain is essential for effective adaptation.

Cloud Native Geospatial

The recommendation to adopt Cloud-Native Geospatial is rooted in its maturity and proven advantages.

Over the years, cloud-native technologies have demonstrated resilience, scalability, and adaptability across numerous domains, and geospatial is no exception. Cloud-native geospatial solutions facilitate dynamic, on-demand resource allocation optimizing costs and performance. The scalability of cloud-native architectures accommodates the growing volume and complexity of geospatial datasets, ensuring that users can harness the full potential of Earth observation data for wide-ranging applications utilizing multiple vast data streams. Moreover, this maturity brings with it a rich ecosystem of tools, best practices, and a large community of expertise that will only grow the knowledge base as more organizations make the shift.

Data Lakes

In the rapidly evolving data landscape, Data Lakes have emerged as a foundational component for organizations aiming to harness the power of their data.

These reservoirs allow organizations to store structured and unstructured data at scale, exposing a unified repository at query time. In the geospatial ecosystem, a key part of the Data Lake is large-scale ingestion of multiple data types from disparate platforms, sensors, resolutions, and spatiotemporal contexts, and optimizing query-side processing. Unlike traditional databases, Data Lakes offer flexibility in terms of storage, scale, and data ingestion, making them particularly well-suited for big data and real-time analytics. Their schema-on-read approach empowers users to define the structure of data at the time of reading, fostering a more agile and exploratory data analysis process. Over recent years Data lake maturity and the surrounding ecosystem of tools have significantly improved, alleviating earlier concerns of "data swamp" pitfalls. Given these advancements, we advocate for organizations to adopt Data Lakes.

Fine-tuning

Fine-tuning is closely related to adaptation learning and serves as a form of transfer learning.

It involves updating a pre-trained model with a smaller, new dataset to enhance its performance for a specific task. For instance, a model initially trained on ImageNet for object recognition can be fine-tuned using a dataset of satellite images to better identify specific land cover types. Another example would be refining a model pre-trained on diverse geospatial data to specialize in a particular biome of interest. Similar to adaptation learning, fine-tuning doesn't necessitate any specialized tools beyond your go-to machine learning framework; however, you will need new training data for the fine-tuning process.

Geospatial Foundation Models & Vision Transformers

Advances in self-supervised pre-training have revolutionized the training of generalist models, particularly in the geospatial domain where vast volumes of unlabeled satellite imagery exist.

These models can be fine-tuned for specific tasks with minimal labeled data. Concurrently, Transformer architectures, initially developed for NLP, have been successfully adapted to visual tasks, benefiting from their ability to scale in terms of parameters and training data. This adaptability makes them arguably more effective than convolutional architectures for handling large-scale datasets. Training multi-modal models that work across different sensor types or domains, like text and vision, is an emerging research avenue. Currently, we are engaged in research using OpenStreetMap data and tags together with imagery for developing a robust geospatial foundation model.

Notebooks

The concept of a notebook is general across several programming languages (for instance R Markdown).

Notebooks implement the concept of Literate Programming by putting narrative text, code blocks, and the output of those blocks together in one document allowing authors to tell a linear story composed of these elements. Notebooks are easy to write, easy to share, and are used extensively in education materials. Notebooks are definitely here to stay, and in Python the .ipynb format used in the Jupyter ecosystem has won - though it stores the output in the file itself generating massive diffs when you re-run it, even though the format is just kind of json-like. Notebooks can be executed locally (JupyterLab, VSCode, PyCharm ...), remotely on an HPC system (JupyterHub or any type of Juptyer server) or on the cloud (Google Colab, SageMaker, an instance of BinderHub). Even GitHub finally supports static rendering and diffing of `.ipynb` files (blog post with announcement).

SpaceX Smallsat Rideshare

The launch cadence for Falcon 9 rockets has continued to increase in 2023. After 62 launches in 2022, it is likely there will be more than 70 in 2023, and SpaceX has announced its plans to make 144 launches in 2024.

Most of the 2023 launches have been focused on expanding the Starlink internet access constellation, and most of the rest are dedicated to a single primary commercial or government payload. However, a few Falcon 9 launches each year are dedicated to lofting several dozen small satellites with a single launch vehicle for as little as $175K each. This Smallsat Rideshare service, known as Transporter, will launch four times in 2023 with Transporter 6, 7, 8, and 9 in January, April, June, and November, respectively. In addition to being relatively low cost, these dedicated smallsat rideshare launches are scheduled up to a year in advance, providing scheduling reliability, and can support a number of different size and weight configurations on a given launch. Further, because the launch is dedicated to smallsats, there is often a great deal of flexibility in terms of achieving the desired orbital parameters. However, the Transporter rideshares have had one key limitation: they have been limited to sun-synchronous orbit (SSO). SSO means exactly what it says on the tin: the satellite passes over each point on the Earth's surface at the same time of day, and this is usually a polar or near polar orbit. This is a great orbit for many applications that need consistent lighting conditions, such as imaging, reconnaissance, and weather satellites.

However, if you want frequent revisits for most of the earth at any time of day, a non-sun-synchronous orbit may be a better fit. For example, if you are operating a constellation of high-resolution SAR satellites, you don't care about lighting, and you may prefer a mid-inclination orbit and the higher revisit cadence this can provide. In these cases, smallsat operators have had to contract a dedicated launch from another launch service, such as Rocket Lab. In April 2024 SpaceX is planning its first mid-inclination smallsat rideshare launch to 45 degrees inclination and 550-600 km altitude. This new service will be called Bandwagon and will offer a new, lower-cost option for satellite operators that want alternatives to SSO configurations.

User Journey Mapping

One of the older methods in the design research space, User Journey Mapping continues to over-deliver on its value.

When we complete user research, our goal is not to simply understand the user, but also to understand their context of use and sequence of steps - as this simply makes decision making about interface design apparent. It also acts as a forcing function - as you work on a journey map you need to know more than you might otherwise get from interviews, so if you use this method it can hold you accountable for digging deeper in interviews. In the context of geospatial work, this is especially important because there are many possible ways that users interact with data, so often we're creating more than one journey map, or a map with multiple paths, to capture the variety of ways they interact in our tools. More than anything, we're checking to see if our tool begins and ends where their entire workflow begins or ends. Where possible, we want these the be mirrored - any time a user has to jump out of a tool and open another we're losing time to science, or time to insights. We are dogged in our efforts to ensure that we are creating flows that work harmoniously with the end goals of our users.

Ground Station as a Service (GSAS)

AWS Ground Station is ushering in a new era of space-to-cloud services by lowering the ground segment cost of entry for satellite owners and operators.

With an on-demand pay-as-you-go-by-the-minute infrastructure, AWS Ground Station mitigates the heavy upfront costs traditionally associated with ground station setup and operations. By integrating with other AWS services, it streamlines the process of data capture from satellites, storage, analysis, and distribution. While this service undoubtedly democratizes access to space-based data, teams are encouraged to trial AWS Ground Station with caution. Understand its compatibility with your existing satellite configurations (for example, AWS Ground Station currently S and X band operational frequencies), data processing needs, and the specific regulatory environment in which you operate. Its full potential is best realized when aligned with an AWS-service-integrated strategy elsewhere within an organization's technical ecosystem.

Map CoDesign

There is value in co-creating maps with intended audiences particularly when trying to change behavior.

Though this technique is not new, its value is becoming increasingly apparent in terms of conservation, water management, and emergency planning. Co-design is an umbrella term in design, which can encompass a wide range of different ways of engaging users in the creation of an end design. In the geospatial realm, this can apply both to the interface design and importantly to the generation of maps or boundaries that exist within the system. In one of our projects we're creating a tool for mapping the boundaries of water systems, and we need those boundaries to be regularly maintained over the years. By hosting workshops where the users are all adding their boundaries together, with attention from supporting organizations helps ensure that the system becomes usable faster. At this stage, we've had users add their boundaries to a prototype to inform the design, and as we build further, these in-person human-centered design sessions will be key to furthering adherence and delight in using the system. Data from other organizations like Rare.org also supports the effectiveness of this method. We are particularly impressed by their finding that shows a 44% increase in adherence with fishing regulations after using this and other methods. The most important impact of this method is remembering that geospatial data is connected to the lived experience of people on the ground and therefore engaging them in reviewing or contributing data collectively can be impactful in ensuring the data reflects their needs and experiences.

Geospatial ChatGPT interface

As the ecosystem of LLMs expands, we see multiple efforts leveraging their ostensible awareness of major geospatial interfaces to simplify usage.

Recent projects include: labs-gpt-stac, its successor haystac, and sat-gpt. While we are encouraged by the power of natural language abstraction, these underlying LLMs are not perfect, and methodologies of measuring their effectiveness are in their infancies. Thus we are recommending organizations assess these technologies in comparison with known-correct methodologies to get a feel for tradeoffs and efficacy.

On orbit Edge Computing

Edge computing moves computer storage and processing to the "edge" of the network, where it is closest to users and devices and, most critically, as close as possible to data sources.

Satellites usually capture data and transfer it to a ground station in order to be processed further. However, space-to-ground data downlink can be expensive (~$70-100/pass with AWS Ground Station), and very high-resolution satellites may capture more data than they can send to the ground affordably. Having on-orbit data storage and processing enables some processing work to be done in orbit, requiring only a subset of data to be downlinked. There can be a lot of value here; for example, by deploying lightweight machine learning models to the satellite, ML models can be run on-orbit and the results downlinked to the ground. This could result in cost reduction, faster decision-making, better data security, and better data resiliency. If two-way communication is supported, the ML models can be updated ad-hoc, increasing mission flexibility. EOI Space is an example of a constellation that plans to launch with edge computing capabilities.

Analysis Ready Data (ARD)

A widely adopted ARD standard will enable new applications in data analysis, allowing users to better integrate and fuse multiple data sources.

Analysis Ready Data (ARD) is the white whale of geospatial data. Used by data providers all over, but meaning nothing, a true ARD standard has not emerged. The CEOS-ARD (previously CARD4L) specifications have shown technical promise but struggled with wide-spread adoption. The recent OGC ARD Standards Working Group (SWG) aims to use the CEOS-ARD specs as a starting point and in partnership with ISO, develop an ARD standard. A widely adopted ARD standard will enable new applications in data analysis, allowing users to better integrate and fuse multiple data sources. The challenge in generating an ARD specification is to achieve adoption across both commercial and federal sectors.

Standards

Adopt x Standard

COG

Adopting Cloud-Optimized GeoTIFF (COG) takes on added significance as it is now a community standard within the Open Geospatial Consortium (OGC).

This underscores the growing consensus within the geospatial community about the importance and value of COG. COGs are optimized for efficient data retrieval, enabling rapid access to specific regions of interest within large datasets, leading to cost savings by minimizing the amount of data transferred and processed. Additionally, COGs are highly compatible with a wide range of geospatial tools and libraries, facilitating seamless integration into existing workflows. With the growing use of Cloud-Native Geospatial, adopting COGs ensures that your data is not only readily accessible and cost-effective but also future-proofed for evolving cloud technologies and practices.

HDF5

The HDF data structure stores multiple data sets in a hierarchical structure, making it easy to add, remove, and update constituent datasets.

HDF5 has been around for almost 30 years, and, as such, it would not normally qualify for the Radar. However, it has received some significant recent interest and remains under active development. HDF stands for "hierarchical data format". The HDF data structure stores multiple data sets in a hierarchical structure, making it easy to add, remove, and update constituent datasets. HDF5 supports high precision data, compression, and very large amounts of data. It's an open, well documented format, and the core libraries support several programming languages. Originally, HDF5 was an outgrowth of the AEHOO project at the U.S. National Center for Supercomputing Applications and had key early support from NASA and the National Science Foundation. Today, it is now supported by a nonprofit, The HDF Group, whose mission is to ensure continued development of the format. HDF5 is also at the core of the NetCDF format, and it is widely used for publication of scientific data sets. It is receiving renewed interest because tools like kerchunk are making it newly useful and relevant for machine learning. Further, emerging hyperspectral satellite operators are considering HDF5 for storing the hyperspectral data from forthcoming satellites. In addition, some of the recently proposed changes in Zarr are causing it to look more like HDF5. Or, perhaps more accurately, HDF5 is a very complicated data structure with decades of barnacles and crenulations, and if you stripped it down to its basics as a modern, cloud-optimized HDF5, it would look a lot like Zarr.

Adopt x Standard

STAC

Adopting SpatioTemporal Asset Catalog (STAC) is a forward-looking choice, particularly given the increasing adoption of STAC by entities such as NASA and the USGS for remote sensing missions, as well as the growing commercial satellite community.

The core STAC specs are stable, with an increasing number of high-value public datasets made available on AWS and Azure. Perhaps most significant is the rapid evolution of an open-source ecosystem of tools designed to create and effectively work with STAC data at scale. A stable specification for metadata combined with foundational open-source tooling that can be built upon greatly lowers the barrier to entry for working with and distributing geospatial data. The most widely used libraries for working with STAC are probably the stac-utils tools, and we use this broadly in our work.

GeoParquet

GeoParquet builds off the Apache Parquet data format that has been rapidly gaining popularity.

Parquet stores tabular data in a columnar orientation and has built-in compression. This means that it is both smaller to store and quicker to read than conventional data formats (such as csv or GeoJSON). This is especially impactful for cloud-native computing since you don't have to download the whole file to access the data of interest. Despite being fairly new (they just released 1.0.0), GeoParquet is already well supported because it simply extends parquet with a specification for how to represent geographic data. We expect that GeoParquet will be especially interesting to people with a large volume of relatively static vector data who don't want to stand up a whole database just to get easy access to subsets of the data. An example of an existing GeoParquet dataset is Microsoft Building Footprints on Planetary Computer. In addition, Planet Labs has released GPQ, a utility for working with GeoParquet.

PMTiles

A cloud-optimized vector tile format, PMTiles draws parallels with Cloud-Optimized GeoTIFFs (COGs) in the raster domain.

PMTiles stores tiled vector data organized to support spatial operations, along with tile overviews in a single binary format file. This file can be stored in the cloud and provides a serverless cloud-friendly approach to serving map tiles. For applications where users need to visualize large vector datasets over broad regions, PMTiles is worth a trial.

Trial x Standard

Zarr

Zarr is a data storage specification that is optimized for storing groups of multidimensional arrays in the cloud.

It came out of biology (specifically malaria research) but has been widely adopted by the Python geoscience community. Pangeo - a community of scientists dedicated to open, scalable, and reproducible geoscience - specifically recommends adopting Zarr. Since Zarr is mostly interested in specifying how you access data, you can use Zarr even if you don't have Zarr datasets. With tools like kerchunk you can make more traditional data formats that allow byte-range requests (like NetCDF or plain HDF5) to act like Zarr files. This is very appealing because you can use Zarr readers without having to re-write and then store a duplicate of your data.

COPC

Cloud-Optimized Point Clouds (COPC) are the point cloud world's answer to the Cloud-Optimized GeoTIFF (COG).

Developed by Howard Butler at Hobu and built on top of the de-facto standard laz format, COPC is a backward-compatible extension of the format that stores indexing data alongside the data themselves, enabling efficient range-based access, which is critical in our network-aware world. COPC is relatively new (2021) and is gaining "market" share; it is currently used to store data from the 3D Elevation Program (3DEP) on the Planetary Computer, planetary data from NASA and the USGS on AWS's Public Dataset Program, and much more. We categorize COPC as Assess due to its youth and lack of deep market penetration, but expect to upgrade it soon as it gains traction. COPC was released with a reference implementation built into the Point Data Abstraction Library (PDAL), which should speed up adoption.

Flatgeobuff

Flatgeobuf is a vector data format tailored for cloud-native applications.

It's designed to manage vast volumes of static data, offering both streaming and random access capabilities, and outperforms many traditional formats with no size restrictions. Flatgeobuf's spatial index allows rapid filtering of features, although the lack of overviews makes it less suitable for multi-level viewing over large regions. Users looking to serve large volumes of vectors should compare Flatgeobuf to GeoParquet and Vector Tiles to see which is best suited to their use case.

SOZip

The concept of cloud-optimizing legacy formats, as seen with SOZip, may not be groundbreaking, but it holds pragmatic value, especially where legacy formats remain dominant in data distribution.

Many industries and systems are deeply entrenched in older data formats due to historical reasons, infrastructure limitations, or the sheer scale of adoption, making a wholesale shift to newer formats challenging and costly. In such scenarios, the ability to enhance the functionality of these legacy formats for cloud environments, as SOZip does for classic ZIPs, can be invaluable. By enabling range reads on compressed archives, SOZip offers a bridge between the traditional and the modern, ensuring that organizations can continue to leverage their existing data archives while also benefiting from the efficiencies of cloud-native operations.

Web Sustainability Guidelines (WSG) 1.0

As the Web Sustainability Guidelines (WSG) explain, "the digital industry is now responsible for between 2-5% of global emissions."

This translates to added pressure on designers, software engineers, and infrastructure teams to implement solutions that minimize the negative impacts on our industry. The WSG collects best practices for building sustainable systems in the areas of User Experience Design, Web Development, Infrastructure, and Business Practices. The guidelines go for breadth and, accordingly, we found some of the recommendations lack specifics regarding how to achieve some of their goals. They make up for this lack of depth by providing links to relevant resources with different approaches to achieving the goals. For example, the guideline "Undertake Systemic Impacts Mapping" identifies the need for understanding variables impacting sustainability on a website but doesn't say how to go about doing that. However, the resources section links to sites, like Digital Eco Design, which describe how to understand the impact of your design. There are literally hundreds of variables that impact the sustainability of a site. The Web Sustainability Guidelines, in draft as of October 2023, are an important step in educating the industry and pushing for greater adoption of sustainability approaches.

STAT

The development of a tasking specification for the geospatial domain has many challenges, but offers unparalleled value for users.

Element 84 has been spearheading community sprints (#1, #2) to help advance such a specification. A unified protocol would streamline the process of requesting and receiving satellite imagery, empowering users to choose data that best suits their needs from the ever-growing list of providers. This highly collaborative initiative has focused on fostering a community, with an impressive array of contributors from across the satellite sector. Such broad-based collaboration signifies not only the industry-wide recognition of the need for such a standard but also ensures that diverse perspectives and requirements are considered. We recommend watching this effort for providers and users of satellite data, as its success promises to drastically simplify workflows and enhance data accessibility.

This was a team effort

Many thanks to all who contributed to make this a reality

Content Contributors:
Robert Cheetham, Catherine Oldershaw, Philipp Schmitt, Matt Hanson, Pete Gadomski, James McClain, Trevor Skaggs, Andrew Pawloski, Jason Gilman, Julia Signell, Matt Bialas, Brad Andrick, Adeel Hassan, Nathan Zimmerman, Justin Polchlopek, Jarrett Keifer, Hector Machin, and Sara Mack

Design
Catherine Oldershaw, Philipp Schmitt

Development
Philipp Schmitt, Terence Tuhinanshu