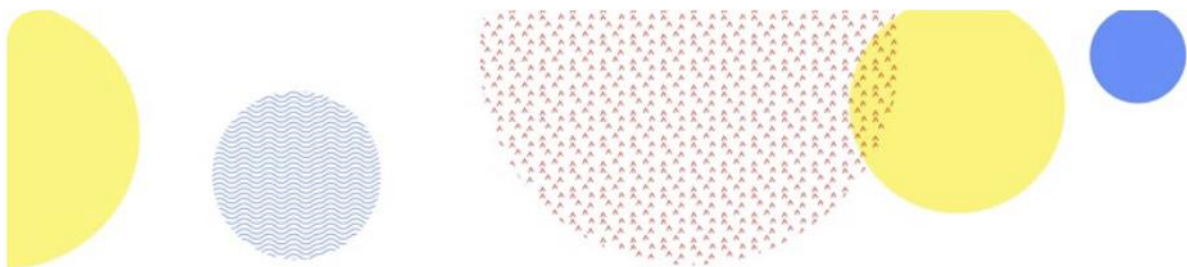


## Module 2: Environmental data

# Enviromental data analysis

- It focuses on collecting and interpreting **data** related to the enviroment
- It includes data about air, water, soil and ice to assess enviromental health
- Key steps in data analysis include collection, inspection, cleaning, trasformation and modeling
- The goal is to understand the impact of human activites on the enviroment



# Dan Hammer

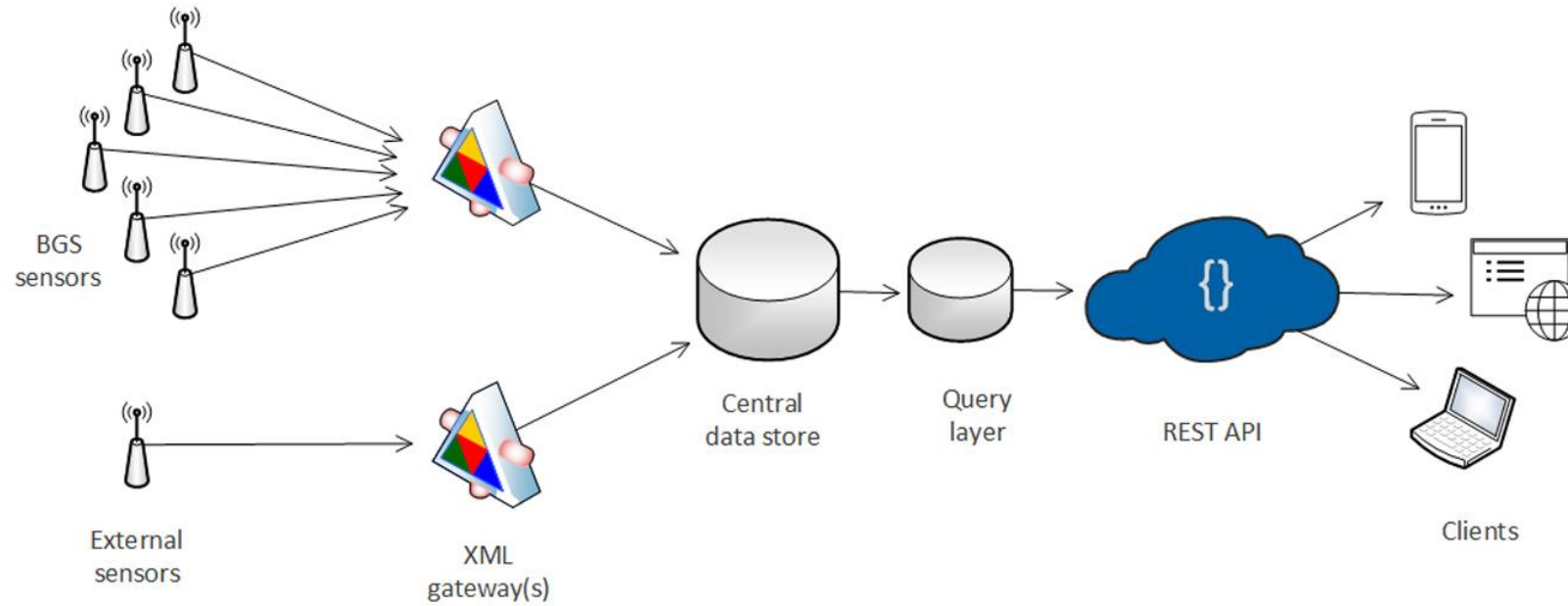
## Data Science for the Environment



# Types of environmental data

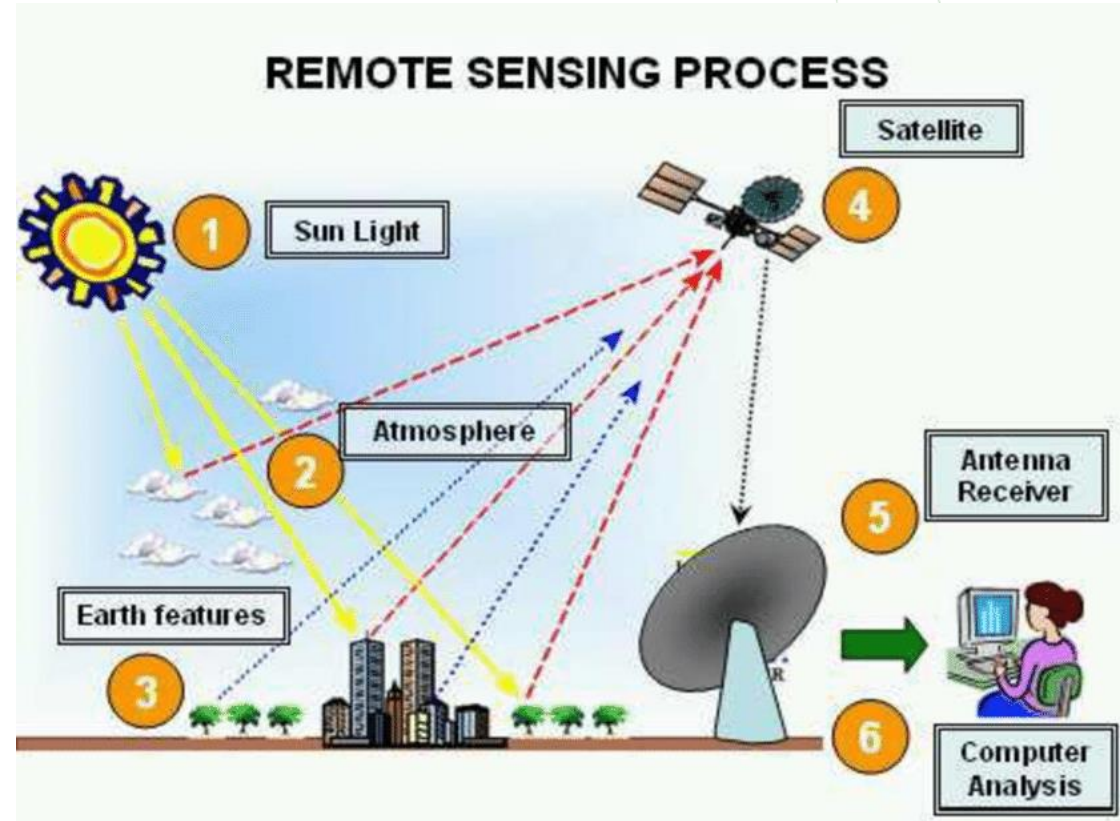
- Sensor data (e.g., air quality, temperature)
- Remote sensing (e.g., satellite imagery)
- Crowdsourced data (e.g., citizen science)
- Historical datasets (e.g., weather archives)
- Model-generated data

# Sensor data



*For example: air quality, temperature, humidity, noise levels*

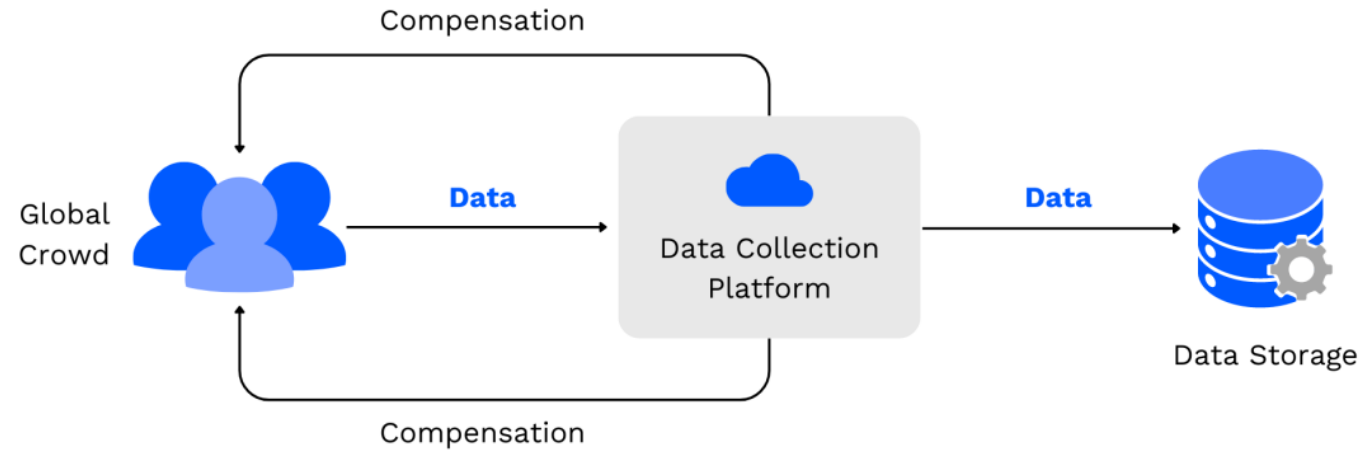
# Remote sensing data



*For example: satellite images, aerial photos, radar scans*

# Crowdsourced data

## Data Collection through Crowdsourcing



AIMultiple

*For example: citizen science contributions, mobile apps, social media reporting*

# Historical dataset



U.S. Department of Commerce  
National Oceanic & Atmospheric Administration  
National Environmental Satellite, Data, and Information Service  
Current Location: Elev: 1518 ft. Lat: 33.4191° N Lon: -111.8444° W  
Station: EAST MESA, AZ US USC0022782

**Record of Climatological Observations**  
These data are quality controlled and may not be identical to the original observations.  
Generated on 05/03/2021

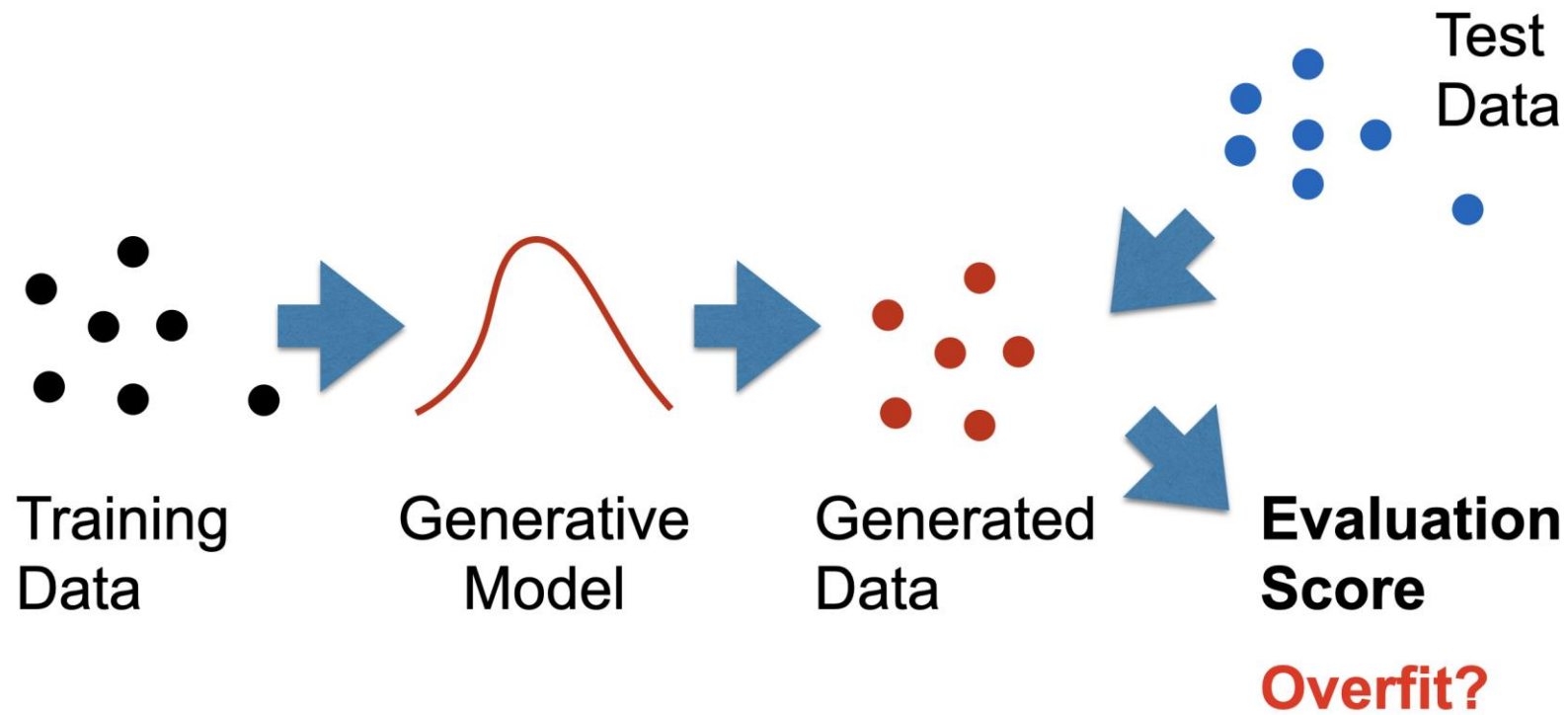
National Centers for Environmental Information  
151 Patton Avenue  
Asheville, North Carolina 28801

Observation Time Temperature: 1700 Observation Time Precipitation: 1700

Year	Month	Day	Temperature (F)			Precipitation					Evaporation		Soil Temperature (F)						
			24 Hrs. Ending at Observation Time		At Obs.	24 Hour Amounts Ending at Observation Time			At Obs. Time		24 Hour Wind Movement (mi)	Amount of Evap. (in)	4 in. Depth		8 in. Depth				
			Max.	Min.		Rain, Melted Snow, Etc. (in)	F i a g	Snow, Ice Pellets, Hail (in)	F i a g	Snow, Ice Pellets, Hail, Ice on Ground (in)			Ground Cover (see °)	Max.	Min.	Ground Cover (see °)	Max.	Min.	
2018	01	01	65	30	60	0.00													
2018	01	02	71	47	62	0.00													
2018	01	03	73	51	64	0.00													
2018	01	04	64	50	57	0.13													
2018	01	05	59	48	54	0.28													
2018	01	06	58	47	55	0.71													
2018	01	07	55	45	48	0.67													
2018	01	08	52	40	50	0.19													
2018	01	09	57	34	53	0.00													
2018	01	10	55	35	52	0.00													
2018	01	11	60	35	54	0.00													
2018	01	12	65	31	60	0.00													
2018	01	13	67	33	60	0.00													
2018	01	14	63	33	59	0.00													
2018	01	15	60	37	55	0.00													
2018	01	16	61	33	56	0.00													
2018	01	17	66	35	60	0.00													
2018	01	18	70	37	62	0.00													
2018	01	19	67	35	64	0.00													
2018	01	20	68	39	65	0.00													
2018	01	21	70	39	65	0.00													
2018	01	22	75	43	70	0.00													
2018	01	23	73	39	65	0.00													
2018	01	24	65	40	62	0.00													
2018	01	25	66	42	63	0.00													
2018	01	26	67	44	65	0.00													
2018	01	27	70	47	66	0.00													

*For example: past weather records, environmental archives, long-term monitoring*

# Model-generated data

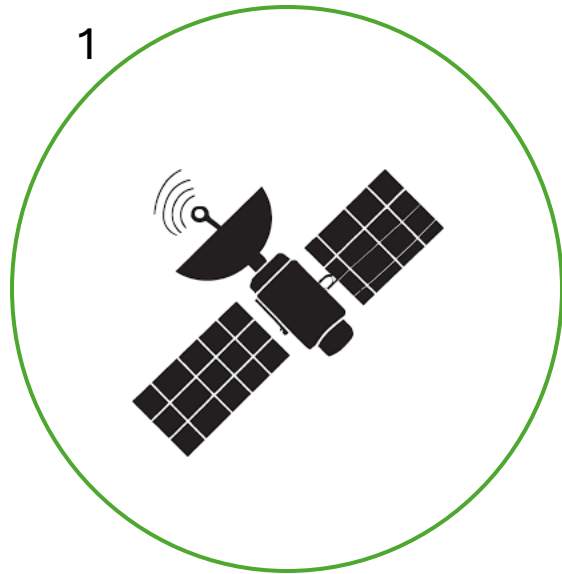


*For example: climate models, pollution dispersion simulations, weather forecasts*

# Environmental data

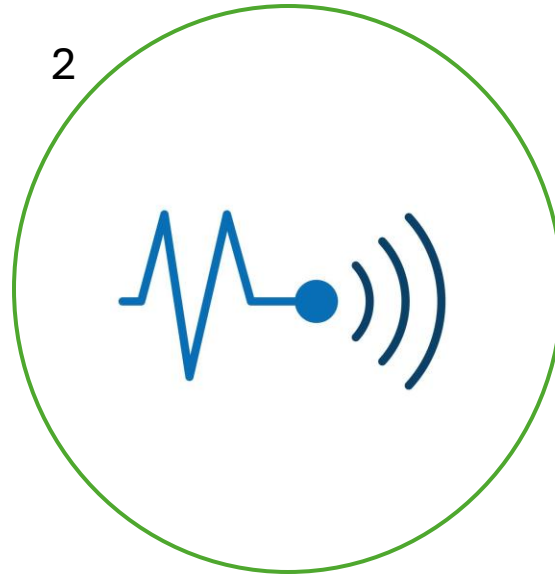
Data Type	Source / Origin	Typical Format	Strengths	Limitations
<b>Sensor Data</b>	Ground-based sensors (e.g., temperature, air quality)	Structured (e.g., CSV, time series)	High frequency; real-time; precise local readings	Limited spatial coverage; sensor maintenance needed
<b>Remote Sensing Data</b>	Satellites, drones, aircraft	Semi-structured (e.g., raster images, GeoTIFF)	Wide spatial coverage; frequent global snapshots	Requires specialized processing; may have cloud/noise interference
<b>Crowdsourced Data</b>	Citizens, apps, social media	Unstructured or semi-structured (e.g., text, mobile reports)	Covers gaps in official data; real-time public input	Varies in quality and consistency; can be biased or incomplete
<b>Historical Datasets</b>	Archives, weather stations, long-term monitoring	Structured (e.g., databases, spreadsheets)	Long-term trends; well-documented	Possible gaps or inconsistencies in older data; standard changes over time
<b>Model-Generated Data</b>	Computational simulations and forecasts	Structured (e.g., NetCDF, gridded formats)	Enables prediction; explores hypothetical scenarios	Depends on model assumptions; sensitive to input data quality

# Environmental data sources



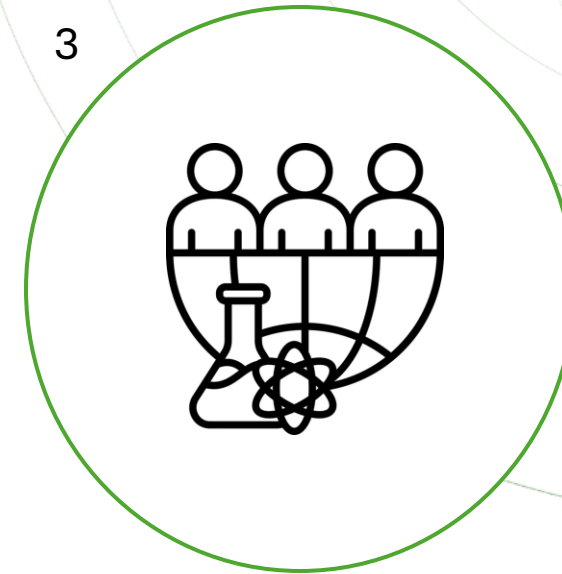
Example: <https://modis.gsfc.nasa.gov/>

To deepen: <https://almanacco.cnr.it/articolo/13371/il-ruolo-dei-satelliti-nella-lotta-al-cambiamento-climatico>



Example: <https://www.arpae.it/it>

To deepen: <https://openaq.org/>



Example: <https://eu-citizen.science/>

To deepen: <https://www.museonaturalemaremma.it/csi/>  
<https://scistarter.org/>  
<https://www.adventurescientists.org/>

# Environmental data sources

Source	Description	Data Type	Strengths	Limitations
<b>Satellites</b>	Remote sensing from orbiting platforms (e.g., Sentinel, Landsat, MODIS)	Raster data, images, multispectral	Large-scale coverage; consistent over time; global and repeated observations	Requires complex processing; limited resolution for some applications
<b>Sensors</b>	Ground-based instruments measuring physical parameters (e.g., temperature, air quality)	Time series, structured data	High accuracy; real-time measurements; detailed local information	Limited spatial coverage; maintenance and calibration needed
<b>Citizen Science</b>	Data collected by individuals or communities (e.g., observations, mobile apps)	Text, geolocated points, mixed	Cost-effective; covers data gaps; promotes engagement	Varying quality; subjective reporting; non-standardized formats
<b>Legacy Models / Historical Data</b>	Archived model outputs or old measurement records (e.g., past climate models, weather logs)	Structured datasets, reports	Long time spans; useful for trend analysis and calibration	May use outdated methods; limited metadata; format inconsistencies

## Case study: CSI Piedmont

**CSI Piemonte** specifically focuses on the **collection, production, and dissemination of territorial data**

🌐 These data consist of three key components:

1. **A geometric component**
2. **An alphanumeric component**
3. **A relational component**

## Quality criteria for territorial and environmental data


CSI considers the **definition and application of a methodology for data quality control** to be of fundamental importance for ensuring the quality of information



**Quality is often defined as “fitness for users”**



**"Quality, in an objective sense, is defined as the set of properties and characteristics of a product or service that give it the ability to meet expressed or implied needs."**

-  **Quality control is based on assessing the fitness of data for the specific application for which it was generated, bearing in mind that:**
- The methodology **must be consistent with the logical model** of the database and the objectives of the information system.
  - The **technical specifications must incorporate quality control requirements**, not treat them as an afterthought.
  - **Automated control procedures should be extensively used**, in order to quickly detect and correct potential errors.

# Quality criteria for territorial and environmental data



**GLOBAL QUALITY**



**LOCAL QUALITY**

# Quality criteria for territorial and environmental data



- **Exhaustiveness**
- **Currency**
- **Genealogy**



- **Metric accuracy**
- **Resolution**
- **Semantic accuracy**
- **Logical/topological consistency**

## Quality verification by CSI

### Geometric data




- Utmost care and precision during acquisition (technique and equipment),
- accuracy in acquiring control points.

### Associated data

- type and structure of the data clearly defined (record layout),
- decoding table of values in case of classified associated information,
- legend associating graphic symbol – thematic meaning for information represented graphically. .

# The procedures for verifying the quality of territorial data





## Procedures for verifying **overall quality**

-  **For verifying completeness:** the completeness of the data is checked both for the geometric part and for the data part.
-  **For verifying currency:** the indication of the temporal dimension on the data is verified. Data sources may not be directly comparable because the data were collected at different times.
-  **For verifying lineage:** the origin of the territorial data is checked. This is derived from the metadata that must accompany the data.



# The procedures for verifying the quality of territorial data

## Procedures for verifying **local quality**

-  **For verifying metric precision (positional accuracy):** the difference between the position of a point represented on the map and its actual position in the geographic reference system is measured.
-  **For verifying resolution:** the dimensions of the smallest geographic detail represented are identified.
-  **For verifying semantic accuracy:** the correspondence between the qualitative attribute associated with an acquired entity and its actual characterization in the original data is checked.
-  **For verifying logical consistency (or congruence):** any inconsistencies in the data are identified, both from the geometric perspective and in the associated data.





BREAK