

Open Data for Open Science

Microsoft Environmental Informatics Framework

Yan Xu
Sr. Research Program Manager
Microsoft Corporation

Microsoft Environmental Informatics



Since 2010

Vision: facilitate seamless access to environmental data and information

Focus: data discoverability, accessibility, and consumability

Objectives:

- advance the technology use in environmental research
- create design wins using Microsoft technologies to
 - Foster innovations in computational environmental research
 - Advance interoperability of data and information sharing
 - Facilitate citizen science for environmental research
- Enhance connections among multiple disciplines and stakeholders

Connect with Environmental Research



[Microsoft Environmental Research Workshop 2010](#)

- 50+ from academia, industry, government, and Microsoft
- Focus:
 - Mind-swap between Environmental scientists and Microsoft technologists
 - Computational challenges in environmental research
 - Mutually beneficial collaborations



Microsoft
Environmental Research Workshop
2010

Initial Findings



Grand Challenge: vast amount of heterogeneous data

- Necessary to use data from different sources
- Not easy (little incentive) to share data
- Need tools to enable producers & consumers to adopt standards

Unique Challenge: it doesn't end with scientific publications

- Compelling presentation of knowledge is critical to
 - Influence policy makers and the general public
 - Enable citizen science to scale the effort.

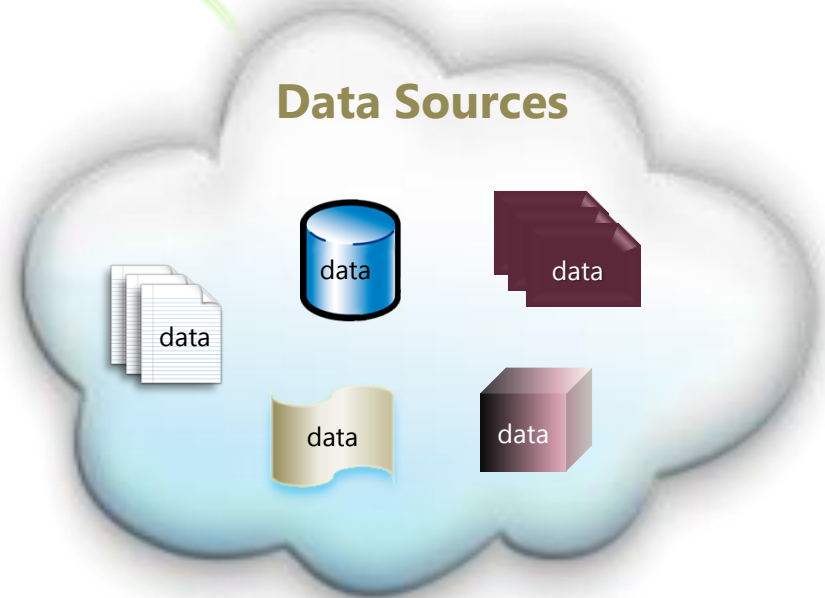
A large, yellow, multi-pointed starburst graphic with a blue outline, containing the text 'Social Impact!'.

**Social
Impact!**

Common Problems with Data



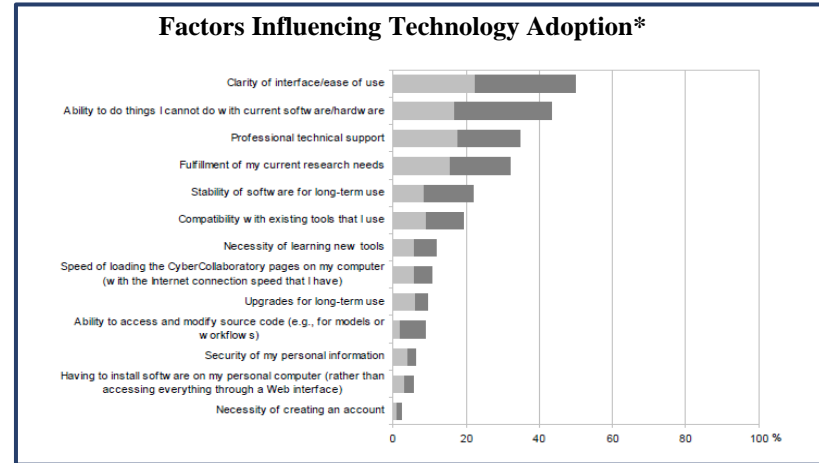
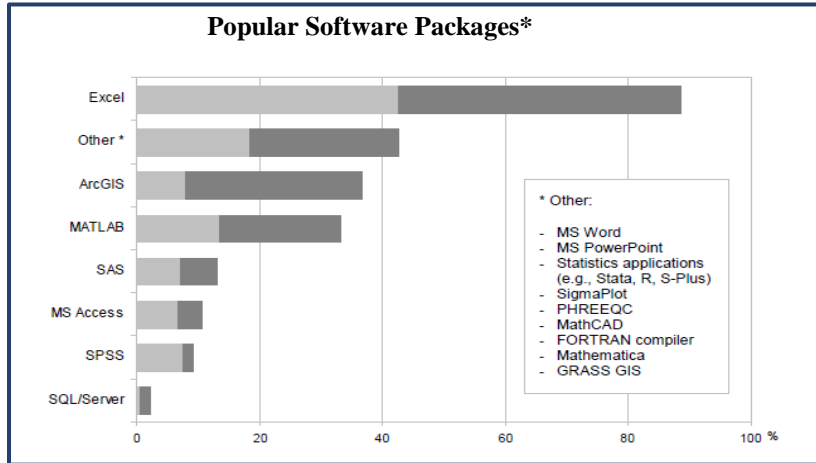
- To use data from different sources
 - Non-standard formats, scales, and units
 - Lack of data quality control
 - Lack of metadata
 - Difficult to repurpose data for different (my) tools
- To share data
 - Lack of incentive
 - Expect no credit
 - Need extra resource and tools
- To properly maintain data, need
 - Versioning
 - Provenance
 - Curation



The Norms of Technology Use



*[Cyberinfrastructure for the waters networks: a Survey of AEEP and CUSHAI Members, K.A. Lawrence et al, May 2006](#)



The Lowest Common Denominator



[HTTP://WWW.](http://www)

Requirements for adopting new technologies:

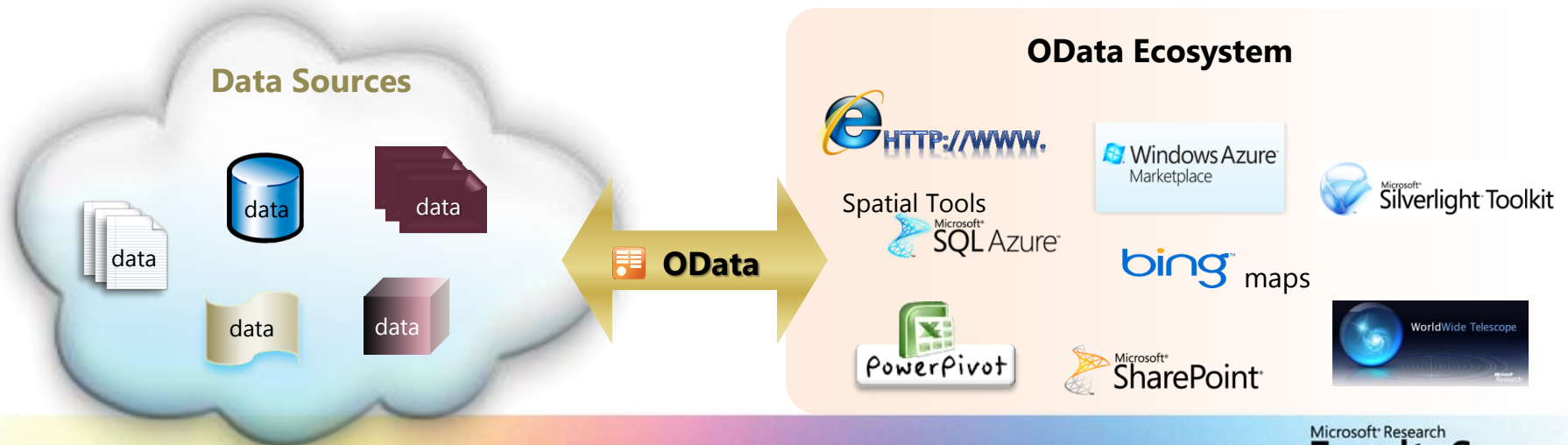
- Easy of use
- Additional value
- Professional technical support

Advance the Norms



Microsoft Environmental Informatics Framework -

- a solution to engage Microsoft technologies with environmental research
- a strategy to amplify social impact of environmental research
- a common agenda among internal stakeholders



Example – Digital Urban Informatics

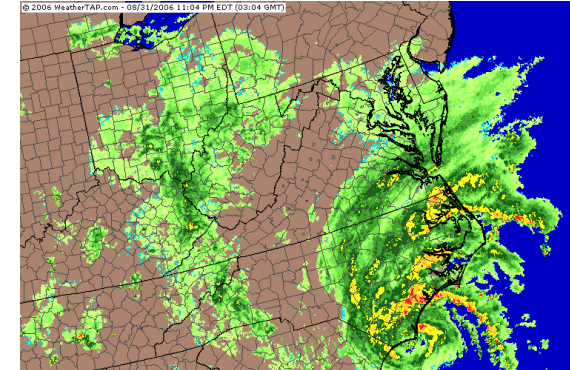


➤ Objectives

- Research - Apply innovative computational thinking to environmental applications by developing a new "Digital Urban Informatics" computational framework addressing both long-term sustainability issues and short-term situational awareness
- Education - Promote citizen science

➤ Key technologies

- Microsoft WWT Earth,
- Azure Cloud Computing,
- Windows Phone 7 app + cloud, s
- OData
- Modeling in Azure, on-demand, in near-real-time
e.g. Modflow parallel ensemble runs for groundwater sustainability



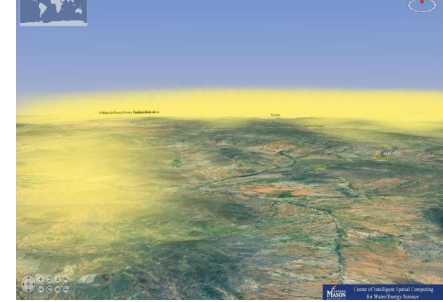
Example – GEOSS Clearinghouse

➤ Objectives

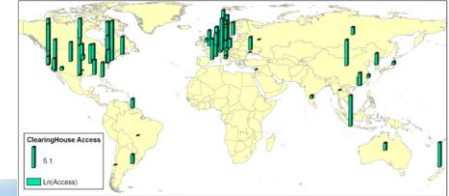
- Share Global Earth Observation Data Among 140+ Countries to Address Global Challenges on Natural Hazards and Emergency Responses
- Support Global End Users to Discover, Access, and Utilize EO Data
- Provide Responses to End Users in Seconds

➤ Key technologies

- Windows Azure - Responding to Massive Concurrent End Users
- SQL Azure and SQL Spatial - Managing Millions to Billions of Metadata Records
- Bing Maps and World Wide Telescope - Visualizing EO Data



THE GLOBAL EARTH OBSERVATION SYSTEM OF SYSTEMS



Example – Forecasting Dust Storms

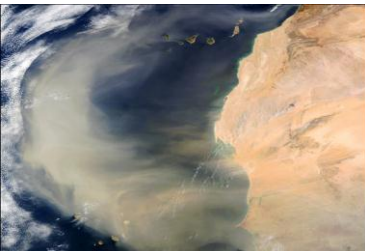
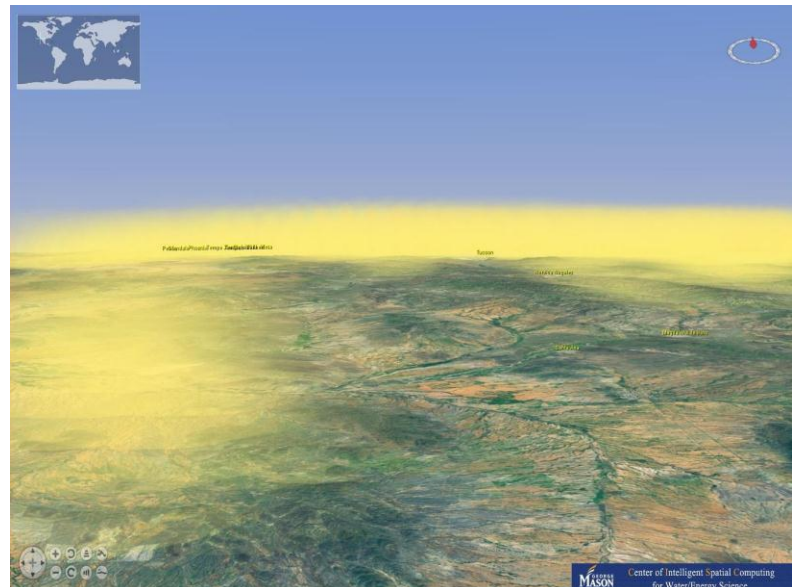


➤ Objectives

- Provide timely forecasting of dust storm for public health emergency responses
- Provide an intuitive interface for decision makers

➤ Key technologies

- SQL Azure as an data management to access large volumetric data
- World Wide Telescope(WWT) as an interactive 3D/4D visualization framework to render the data
- Azure as an advanced cloud computing platform to support forecasting.



Example – Forecasting Urban Land Use Changes

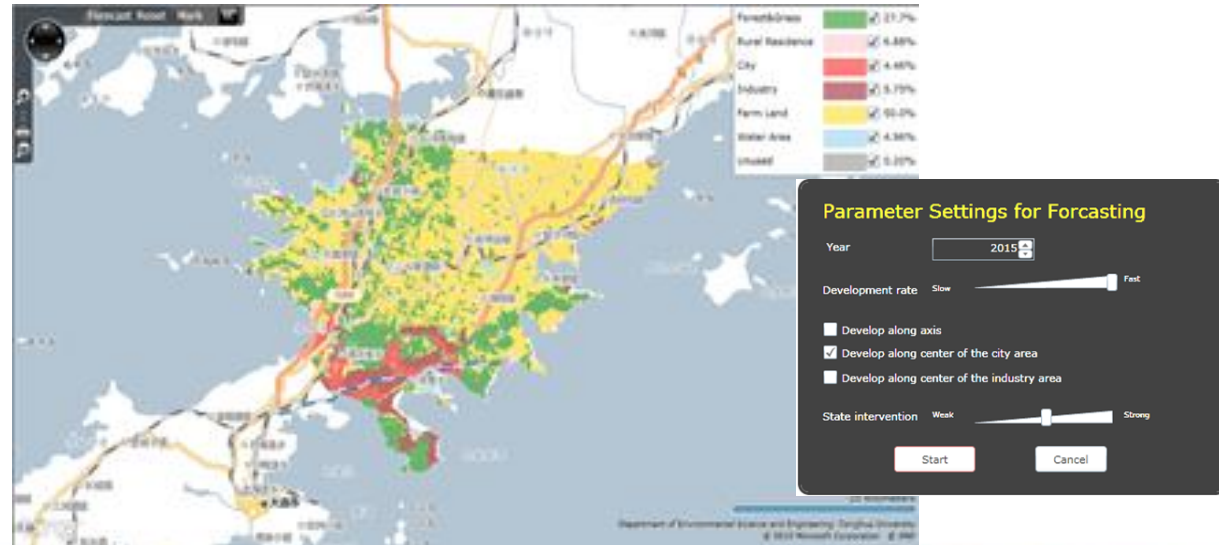


Objectives

- Visually integrated urban land expansion module and environmental forecast module
- Identify potential environmental impacts of various urban land layout over next 10-20 years
- Promote smart decision making

Key technologies

- Bing Maps and SilverLight
- HPC computing

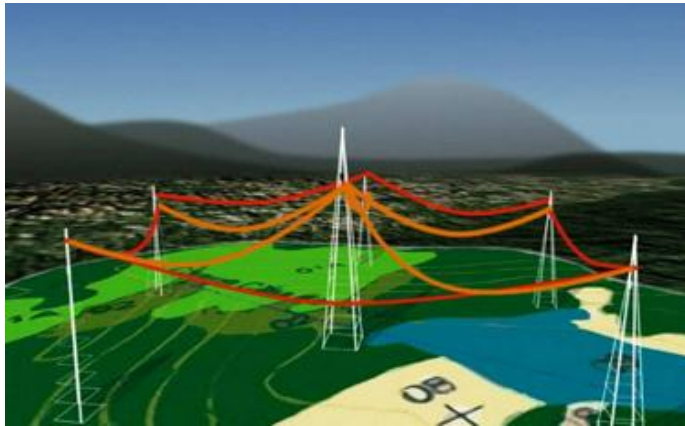


Example – Wireless Sensor Networks in Brazilian Forest



➤ Hundreds of sensors to measure the vital signs of the rainforest

- Temperature
- Water vapor
- Solar radiation

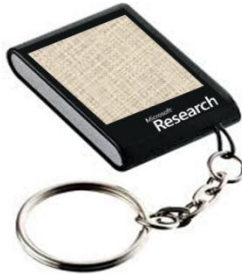
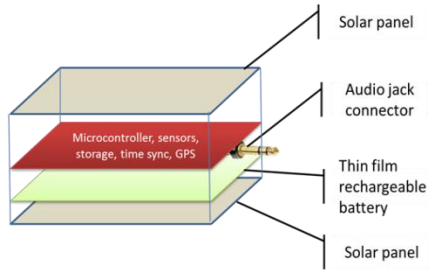


Deployment of towers with sensors (*in red*) placed on cables (*in orange*) above the rainforest canopy.



One of six towers equipped with sensors to measure temperature, water vapor, and solar radiation

Example – Participatory Environmental Monitoring Toolkit



Mobile sensors



Personal gateway



Cloud Storage & Compute



Decision support

➤ Objectives

- Facilitate socially inclusive environmental observation
 - Time & GPS location
 - Temperature & Humidity
 - CO2
 - H2S
- Leverage existing Microsoft technologies and user communities
- Deliver a HW+SW toolkit in open source form

➤ Key technologies

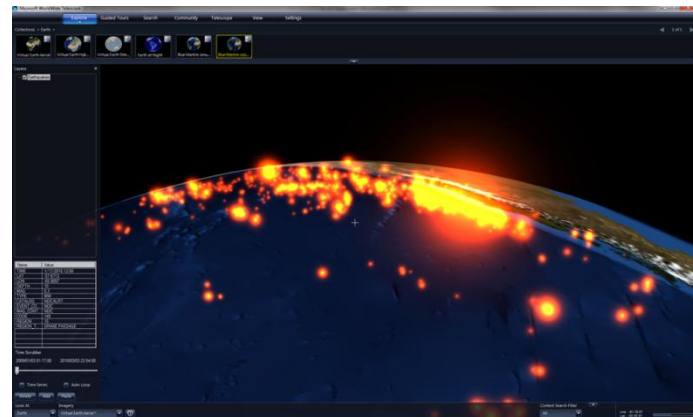
- Microsoft Research low energy GPS location sensing and mobile data collection services
- OData
- World Wide Telescope(WWT)
- Windows Azure

Demo – WWT and Geo-Data Visualization



WorldWide Telescope (WWT)

- A visualization software environment
 - Enables a computer to function as a virtual telescope (astronomers call it “the best VO (virtual observatory) implementation”)
 - Visualizes geo-data in 4D (space + time)
 - Integrated with Excel
 - Allows data sharing with controlled access – WWT Community
 - Empowers high-quality, intuitive, and interactive visual presentation via “WWT tour”
- Datasets under consideration
 - Seismic event distribution against sbuductionslab slab models (USGS NEIC)
 - Standardized-format datasets (OGC, WxS, NetCDF, Shapefile, CSV, HDF, ...)
 - Dataset and model output concept: plugging data generators directly into WWT
 - Draped raster, e.g. MODIS ocean, land and atmospheric products
 - Alternatice topography, e.g. ice sheet thickness and bathymetry
 - Climate change thematic datasets, e.g. monthly sea ice extent from NSIDC
- Free for research and education use

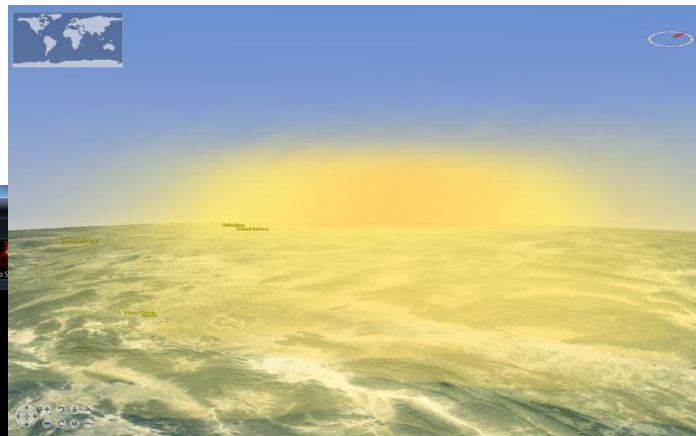


Demo – WWT and Dust Storm Simulation



A mutually beneficial casestudy

- Mind-swap, e.g. at [Open Data for Open Science Developers Training](#)
- Improve science modeling
- Improve computer engineering
- ...



Microsoft Excel

Book2.xlsx

	A	B	C	D	E	F	G
1	Lat	Long	Color	Depth	Date	Mag	
2	26.6667	-112.167	21.915085	-1	2007/07/0	4.146589	
3	31.6667	-109.333	12.977047	-1	2007/07/0	4.146589	
4	31.6667	-109.167	10.850816	-2.61278	2007/07/0	4.146589	
5	31.8333	-109.167	22.231303	-1	2007/07/0	4.146589	
6	31.8333	-109.167	12.357295	-2.61278	2007/07/0	4.146589	
7	32.5	-110.667	15.501434	-1	2007/07/0	4.146589	
8	32.5	-110.667	10.096395	-2.61278	2007/07/0	4.146589	
9	32.5	-109.667	11.507545	-1	2007/07/0	4.146589	
10	32.5	-109.667	10.605312	-2.61278	2007/07/0	4.146589	
11	32.6667	-109.833	28.509170	-1	2007/07/0	4.146589	
12	32.6667	-109.833	15.045910	-2.61278	2007/07/0	4.146589	
13	32.6667	-109.667	14.665606	-2.61278	2007/07/0	4.146589	
14	32.6667	-109.667	13.590615	-2.90849	2007/07/0	4.146589	
15	32.8333	-109.833	13.185535	-1	2007/07/0	4.146589	
16	32.8333	-109.833	11.415855	-2.61278	2007/07/0	4.146589	
17	33.1667	-108.667	11.71204	-1	2007/07/0	4.146589	
18	34	-108.333	10.98493	-1	2007/07/0	4.146589	
19	26.6667	-112.167	13.418444	-1	2007/07/0	4.146589	
20	26.8333	-112	13.149444	-2.61278	2007/07/0	4.146589	
21	26.8333	-112	10.445504	-2.90849	2007/07/0	4.146589	
22	31.5	-109.333	11.749627	-2.90849	2007/07/0	4.146589	
23	31.5	-109.333	12.488841	-3.08279	2007/07/0	4.146589	

Microsoft WorldWide Telescope

Explore Guided Tours Search Community Telescope View Settings

My Collections Combinations Solar System (Sky) All-Sky Surveys Spitzer Studies Chandra Studies Hubble Studies Astrophotography Radio S

Layers >

- ☑ Sky
- ☑ Atmos
- ☑ Clouds
- ☑ Moon
- ☑ Mars
- ☑ Jupiter
- ☑ Saturn
- ☑ Uranus
- ☑ Neptune
- ☑ Pluto
- ☑ Sky

Time Scroller: 2007/07/01 03:00:00 2007/07/02 18:00:00

Time Series Add Auto Loop

Map Columns: Layer Marker

Data In The First Row: WWT Label

Lat: Latitude

Long: Longitude

Color: Color

Depth: Depth

Date: Start Date

Mag: Magnitude

To map additional columns, include the column in your selection range and click the "Update Range"

Distance Units: Kilometers

View in WWT

Time Series: No Results

Look At: Imagery Virtual Earth Asset

Control Search Filter: All 1 of 2

2002 km

Lat: -108.3338
Lon: -109.3338