

# Open Data AI – DC Compass

## A GIS-Centric Approach to Open Data

Matt Sokol, Chief Data Officer  
DC Office of the Chief Technology Officer

**GeoGov Summit 2024**

# DC's AI Journey

- **Mayor Bowser establishes order for DC's AI Values and Strategic Benchmarks in February 2024**
  1. Articulates six AI values agencies must consider for a tool
  2. Establishes an AI Advisory Group on AI Values Alignment to conduct regular community and stakeholder engagement
  3. Identifies AI strategic benchmarks in the form of internal governance standards and strategic plans
  4. Creates an AI Taskforce to examine the District's internal posture on governance and to support agency leadership in meeting benchmarks and to advise the Mayor on responsible and effective use of AI

# DC's AI Journey

- **Six AI Values**

1. Clear Benefit to the Residents
2. Safety and Equity
3. Accountability
4. Transparency
5. Sustainability
6. Privacy and Cybersecurity

- **Handbook for AI Values Alignment**

Discusses the AI Values in detail and provides agencies with key concepts, considerations and thoughtful examples for design, development, deployment and maintenance.

# DC Open Data

- Movement widely started in the mid-90's
- Making government data freely available to use or share
- Promotes transparency and empowers the public
- Allows for anyone to explore and make new insights
- DC's first Open Data Portal created in 2014
- Contains over 1,400 datasets and apps, maps and initiatives



[opendata.dc.gov](https://opendata.dc.gov)

# What's the Problem?



**Open Data is fantastic but....**

**There's a lot of data available.**

*I searched "procurement" and got 20 datasets returned, which do I use? Wait, I need to review every datasets metadata?*

**Despite how easy the tools have become, if you're not a data person, it's not that easy.**

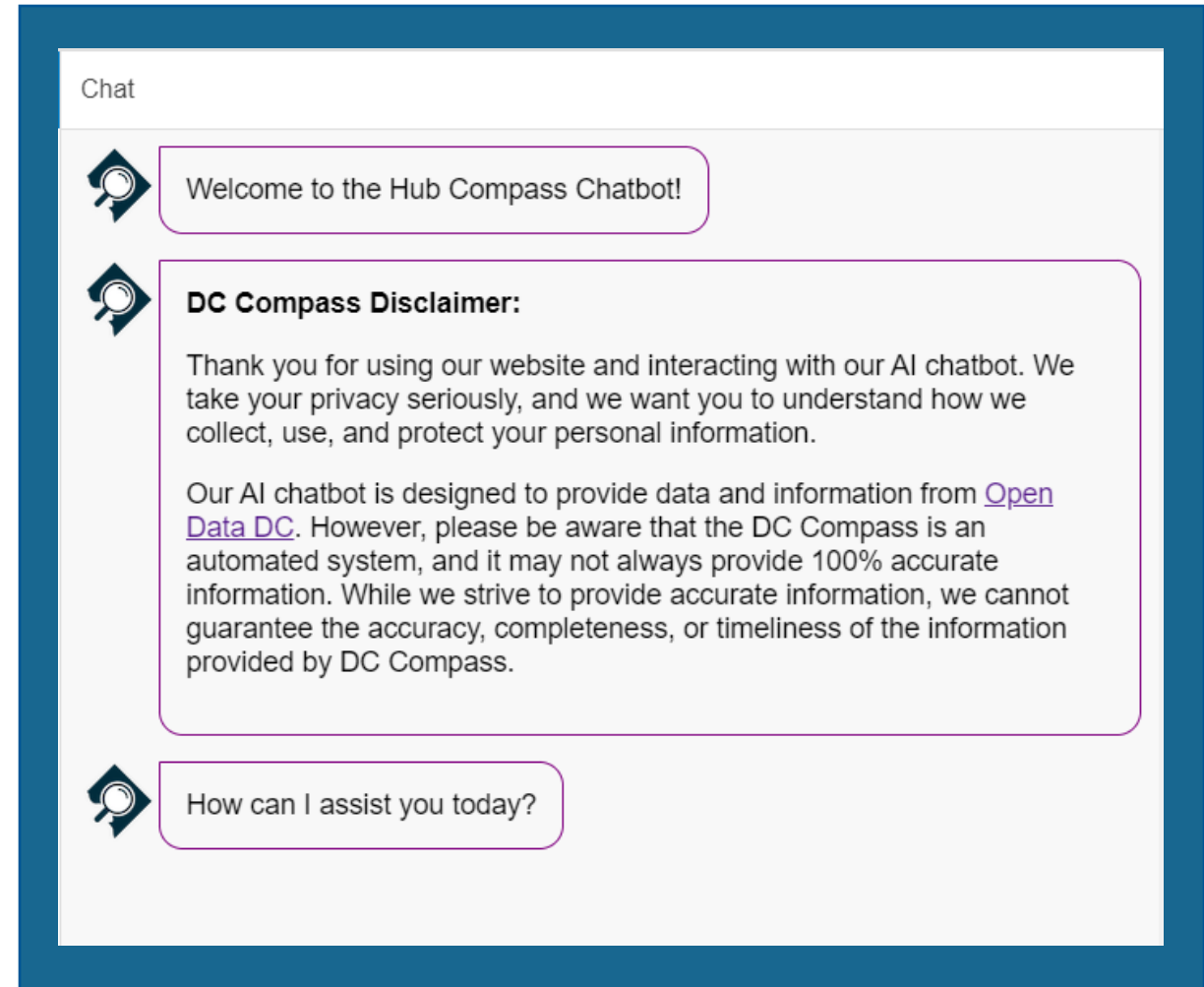
*I need to write a query to get what I need. How do I filter hundreds of thousands of records?*

**How can I easily find out about city initiatives related to datasets?**

*I searched on crashes in the city, tell me about the Vision Zero plan and initiative.*

# Open Data & AI – DC Compass

- Combines the power of generative AI and Open Data
- Assistant for users to ask questions about Open Data
- Uses data summary and GIS functions
- Allows for text or voice interaction and language translation
- Only utilizes Open Data (Level 0 per the [DC Data Policy](#))



# DC Compass - Examples

What species of trees are on the block of 200 I Street SE?



The block located at 200 I Street SE has the following tree species: Kentucky coffeetree (*Gymnocladus dioicus*), Sweetgum (sterile) (*Liquidambar styraciflua* 'Sterile'), Sugar maple (*Acer saccharum*), American linden (*Tilia americana*), and Sweetgum (*Liquidambar styraciflua*). For more information about the Urban Forestry Street Trees dataset, please visit <https://opendata.dc.gov/content/f6c3c04113944f23a7993f2e603abaf2>. The map below shows the locations of the trees in the area.

AI responses may be incorrect. Please provide feedback to help us improve.



What is the the nearest metro station to the John A. Wilson building?



The closest subway station to the John A. Wilson building is L'Enfant Plaza. For more details about this station, please visit [http://www.wmata.com/rail/station\\_detail.cfm?station\\_id=82](http://www.wmata.com/rail/station_detail.cfm?station_id=82).

How many crashes were there in March 2023?



There were 1711 accidents in March 2023 in Washington, DC. You can find more information about this data set at <https://opendata.dc.gov/content/70392a096a8e431381f1f692aaa06afd>.



Add Dataset to Map

Zoom to Location



# DC Compass – Maps and Analysis

Chat



Welcome to the Hub Compass Chatbot!



## DC Compass Disclaimer:

Thank you for using our website and interacting with our AI chatbot. We take your privacy seriously, and we want you to understand how we collect, use, and protect your personal information.

Our AI chatbot is designed to provide data and information from [Open Data DC](#). However, please be aware that the DC Compass is an automated system, and it may not always provide 100% accurate information. While we strive to provide accurate information, we cannot guarantee the accuracy, completeness, or timeliness of the information provided by DC Compass.



How can I assist you today?

Which species of trees are on the block of 200 I Street SE?



There are four types of trees on the block of 200 I Street SE: Kentucky coffeetree, Sweetgum (sterile), Sugar maple, and American linden. For more information about these trees, please refer to the Urban Forestry Street Trees dataset at <https://opendata.dc.gov/content/f6c3c04113944f23a7993f2e603a>

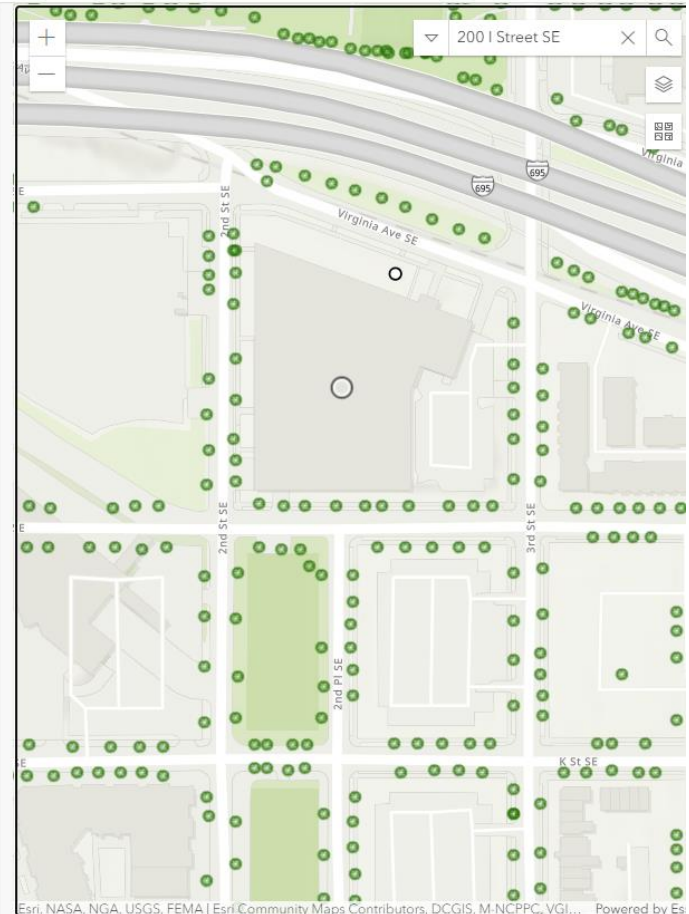


Add Dataset to Map

Zoom to Location

AI responses may be incorrect. Please provide feedback to help us improve.

Map



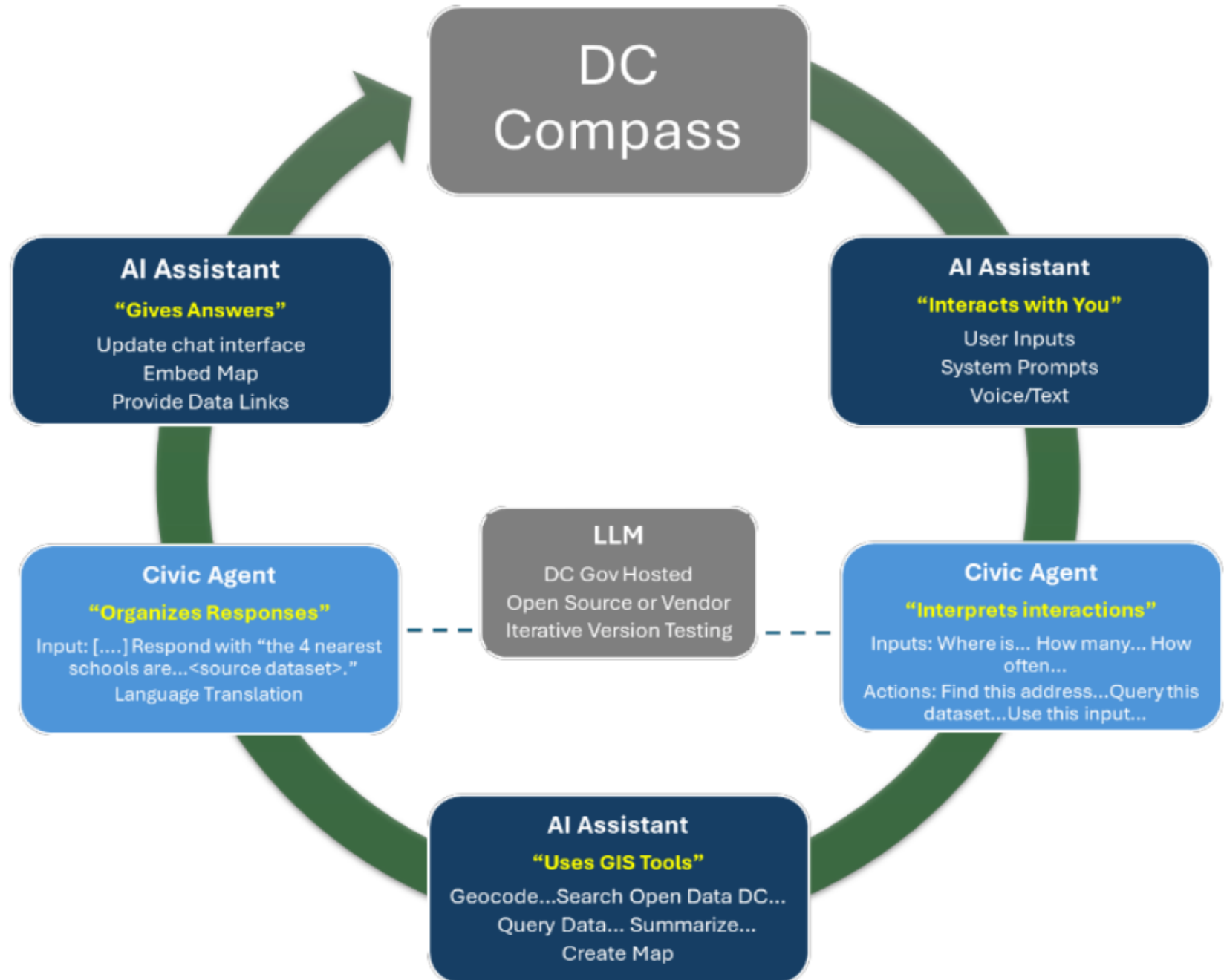
Table

Urban Forestry Street Trees - UFA Street Trees (Total: 208,878 | Selection: 0)

	Scientific Name	Common Name	Genus Name
<input type="checkbox"/>	Gymnocladus dioica	Kentucky coffeetree	Gymnocladus
<input type="checkbox"/>	Nyssa sylvatica	Blackgum	Nyssa
<input type="checkbox"/>	Liquidambar styraciflua 'Sterile'	Sweetgum (sterile)	Liquidambar
<input type="checkbox"/>	Ulmus americana 'Princeton'	Princeton elm	Ulmus
<input type="checkbox"/>	Quercus phellos	Willow oak	Quercus
<input type="checkbox"/>	Acer saccharum	Sugar maple	Acer
<input type="checkbox"/>	Tilia americana	American linden	Tilia



# DC Compass - Design



# Other Considerations

## Verification & Validation

Provide links back to the data source to allow users to validate the answers

Human-in-the-loop

## Values Alignment

How does it align with the six values?

What can/can't it do?

What data and systems does it have access to?

How does it use your information?

Sustainability of costs

## Feedback

Built in Survey123 form to capture experience, good & bad

Discover issues and opportunities for improvement

## Challenges

Alignment drift

Only as good as the underlying (meta)data quality

Continued understanding of how LLM models operate to reduce chances of incorrect information

# What's Next?



- Currently in public beta
- Continued improvement of existing functionality, model processes and more efficient use of tokenization
- On-going model evaluation: GPT 3.5 vs 4o vs Llama
- Additional visualizations and GIS functions
- More conversational data workflows