

# Collaborative Development of an Open Matrix Standard

April 30, 2014



Puget Sound Regional Council



**PARSONS  
BRINCKERHOFF**



<https://sites.google.com/site/openmodeldata>

# Purpose Statement

Create a simple standard matrix format for storing and transferring matrix data that can be easily adopted

For	Anyone managing transportation data with a matrix structure
Who	Wants a portable and open format for storing and exchanging that data
That	Is easy to understand, can be used in custom programs/scripts, and supports commercial products
Unlike	Existing solutions that are either proprietary, cumbersome, slow, and/or too large

# Who We Are

- A diverse group of agency, consultant, and research modelers
- Formed a **working group** to define and implement a the file format for use in modeling
- Similar to the HTML working group for example
- Formed in early 2013, shortly after TRB
- Periodically meet to update the standard/project

# We Believe

- An open matrix file will:
  - make model development and application easier across models and software packages
  - help spur industry research and innovation by having a shared data structure (**think shapefile**)
  - enable researchers, students, agencies, and practitioners to collaborate in ways that are currently difficult, cumbersome and error-prone

# What is an Open Matrix?

- Open matrix is a file format (i.e. skims.omx)
- Based on open-source HDF-5 file storage technology
- Binary data storage with compression
- Can store multiple matrices in one file
- Can include multiple indexes/lookups
- Can contain additional attributes for both matrices and indexes

# Availability

- Specification available on our website:  
<https://sites.google.com/site/openmodeldata>
- Interfaces in Python, R, Java, and now C#
- Supported by PECAS
- INRO and PTV have committed to developing interfaces as well
- Hoping to get broad industry acceptance for it

# OMX

- Review of Existing Solutions
- OMX Specification
- Jumping into OMX
- Using OMX for your project or model
- What's Next



# Review of Existing Solutions

(a.k.a. why create another data format?)



# Text Files

**Pros:**     *Simple*  
                  *Universally readable*

**Cons:**     *Enormous*  
                  *Unwieldy for large datasets*  
                  *Difficult to view by column*  
                  *Hard to include multiple tables*

# SQLite Files

*Pros: Widely understood/readable  
Supports multiple tables*

*Cons: Not well-suited to matrix data  
Slow, on some implementations*

# BSON (“Binary JSON”)

*Pros: Lightweight  
Traversable*

*Cons: Designed as a “wire protocol”  
Primarily used for tabular data  
Viewing is difficult*

# Zip Matrix

*Pros: Just a zipped folder of binaries*

*Cons: Viewing is difficult  
Slow for random access to data*

# Vendor Formats

**Pros:**      *Commercial support*  
*Familiar, fast, feature-rich*

**Cons:**      *Licenses isolate users*  
*Public can't have access*  
*Long-term archival problems*  
*Zero collaboration potential*

# HDF5 – “Hierarchical Data Format”

*Pros: Well-established format in use since 1987 (NCSA, physics, etc)*

*Compact & Fast*

*Any platform, any language*

*Flexible*

*Cons: Unfamiliar to travel modelers  
Extremely flexible; too flexible!*

# More about HDF5

<http://www.hdfgroup.org/HDF5>

- Open source, permissive license
- Every HDF5 file is a hierarchical “container” (like a zipfile)
- The root node can have child nodes & subfolders
- Every node can have key/value attributes attached
- Data can be any shape, size, or structure



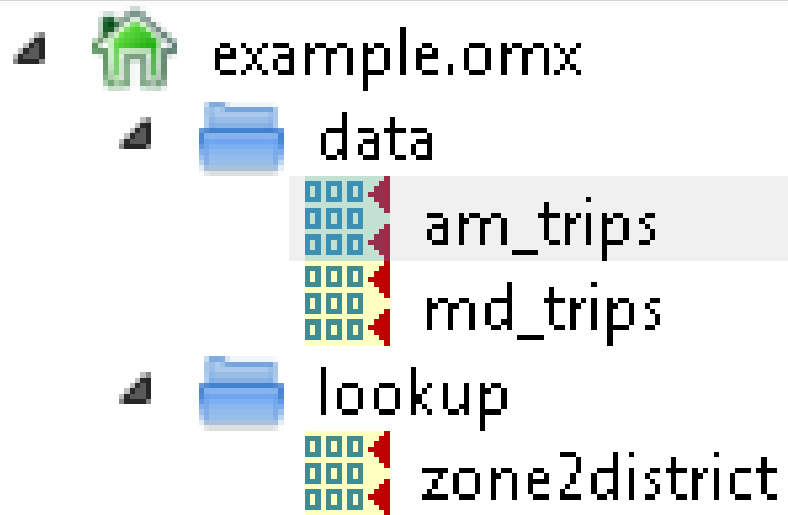
# The OMX Specification

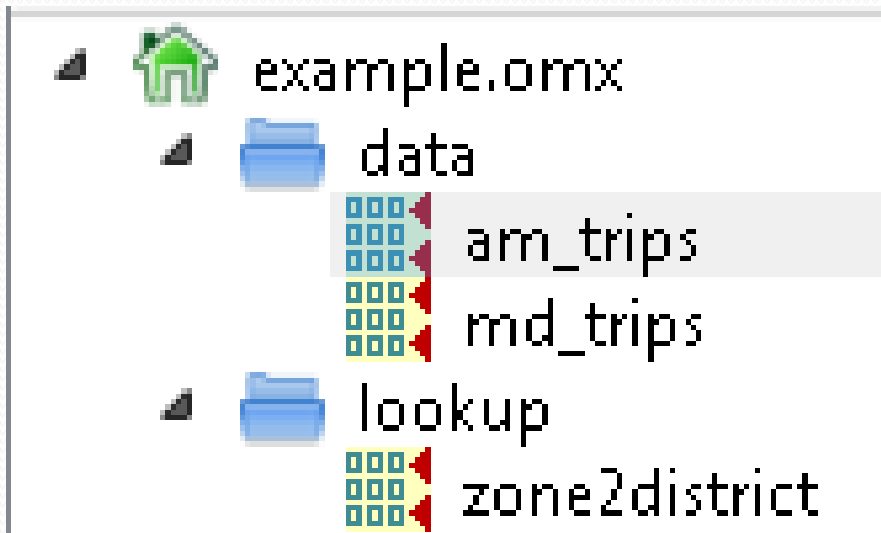
(or, how to rein in the complexity of HDF5)



# OMX Specification

*An OMX matrix file is a structured collection of two-dimensional array objects and associated metadata.*

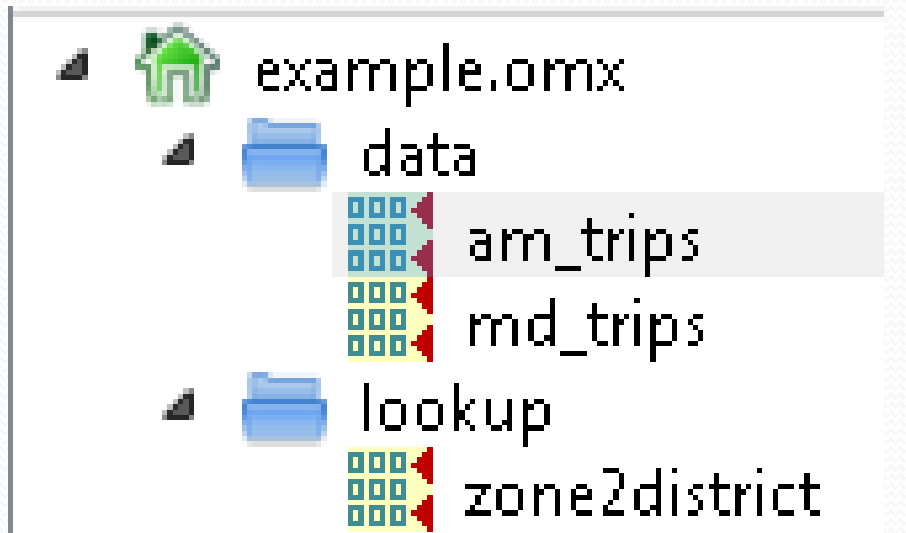




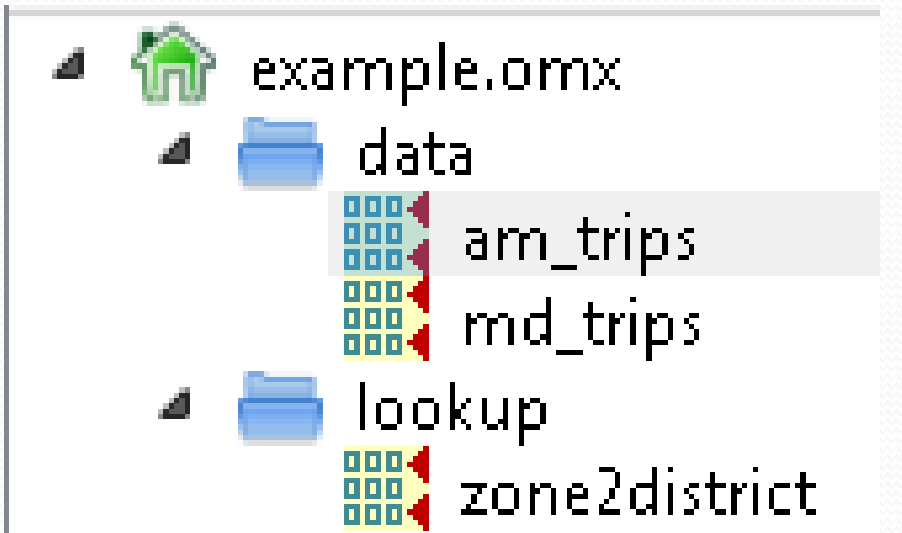
**ROOT node**  
*contains:*

**SHAPE**  
(rows x cols)

**OMX\_VERSION**  
(0.2)



**DATA folder**  
*contains all your  
matrices*

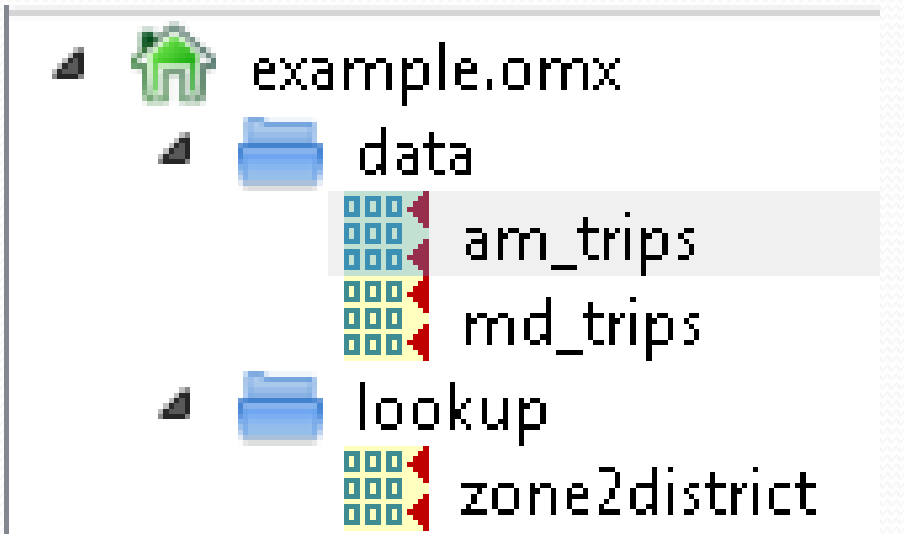


**MATRICES**

*Are named*

*Are all the  
same size*

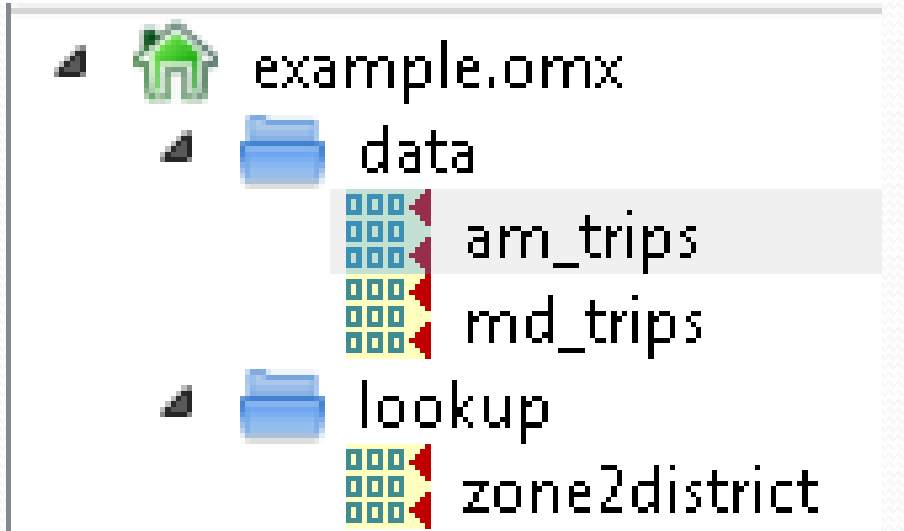
*Are all  
two dimensions*



## ← MATRICES

Can have unlimited key/value attributes:

**TIMEPER:** "AM"  
**MODE:** "transit"  
**VOT:** 11.69



**LOOKUP folder**  
*contains any  
“mappings” of  
row numbers*

*e.g.,  
zone-to-district*

OMX Viewer 1.0.0

File Node Dataset Settings Window Help



Tree of databases

- example.omx
  - data
    - am\_trips
    - md\_trips
  - lookup
    - zone2district
- Query results

	am_trips		md_trips		
	1.0	1.0	1.0	2.0	2.0
1.0	2.0	12.0	2.0	4.0	0.0
1.0	4.0	14.0	2.0	4.0	6.0
1.0	6.0	16.0	0.0	10.0	10.0
2.0	8.0	18.0	2.0	4.0	6.0
2.0	10.0	0.0	2.0	4.0	-19998.0





# Jumping into OMX: Python example



Using OMX for your project or model

# open model data



[Home](#)

▼ [APIs](#)

[C# API](#)

[Java API](#)

[Python API](#)

[R API](#)

▼ [Background Information](#)

[Case Studies](#)

[Data Structure 0.2](#)

[Existing Solutions](#)

[Open matrix requirements](#)

[Overview of the Specification](#)

[Use Cases](#)

[What's A Matrix?](#)

▼ [Files](#)

[OMX Viewer](#)

[APIs](#) >

## Python API

The latest APIs are available on github.

## Python API User's Guide

### Installation

**Prerequisites:** You'll need Python 2.6+, PyTables, and NumPy. On Windows the easiest way to get these is

### Content

[1 The l](#)

[2 Pytho](#)

[2.1 In](#)

[2.2 Q](#)

[3 Pytho](#)

[3.1 G](#)

[3.2 Fi](#)

[3.2.](#)

[3.2.](#)

[3.2.](#)

[3.3 M](#)

OMX API is already available in:

**Python - Java - C# - R**

OMX API is coming soon in:

**Fortran - C++**

OMX native support coming soon in:

**Emme - PTV - DTA lite/Nexta - PECAS**



# Next Steps



Next Steps:

Converters for existing matrices

We will soon have converter programs for existing Cube and Emme matrices

*(running converter requires valid license)*



## Next Steps:

# More and better installers and docs

The website is “feature complete” but we could use some volunteers to help with testing and documentation.

Platform-specific installers would also help get new users up to speed quickly.

# Conclusions

- OMX is a simple standard matrix format for storing and transferring matrix data
- APIs are available for the language you're using
- GUI Viewer is available for Win, Mac, Linux
- Join us, let's modernize the industry together! – Publish and share your existing trip tables in OMX right now.
- Check out our website and email list for more information



# Thanks!

<https://sites.google.com/site/openmodeldata>

<https://github.com/osPlanning/omx>

- Ben Stabler                      stabler@pbworld.com
- Billy Charlton                    bcharlton@psrc.org
- Brian Gregor                    gregorbj@gmail.com
- Chris Frazier                    frazierC@pbworld.com
- Elizabeth Sall                    elizabeth@sfcta.org
- Jeremy Raw                      jeremy.raw@dot.gov
- John Abraham                    jea@hbaspecto.com
- Suzanne Childress              sChildress@psrc.org