APPENDIX A. CITY INFORMATION MODEL TECHNICAL DOCUMENTATION

1. Purpose & Transferability

In building a city information model (CIM) and digital twin for Milwaukee, this accompanying documentation is necessary to describe how the models are built, their structure, the method of online deployment, and the relevant notes for their specifications and operations. The purpose of this documentation is to provide something of a user manual for the models that allows other researchers to review the research integrity, re-run analyses if necessary, and transfer the model to other cities.

While a single set of professional best practices does not yet exist for CIM and digital twins, there are standards that can be referenced. Specifically, these include ISO 37120, the CityGML standard, and the City Induction Research Project. In creating a CIM and digital twin for Milwaukee, aspects of these standards are adapted to the city and the available datasets. Additionally, the use of the standards is meant to facilitate transferability of this Milwaukee model to other cities.

2. Milwaukee Model Framework & Data Structure

2.a. Model Framework

Because this research utilizes a multitude of primary source reports and spreadsheet datasets, the model requires a structured workflow to ensure that the data is digitized, processed, analyzed, and managed correctly. The intent is to have the workflow be transferable from one city to the next, thereby facilitating both desktop analysis by researchers and public use of online tools. This transferability dictates that data is managed in formats that are readable across multiple software platforms.

To complete this research, the model framework creates a flow of data from the digitization process to the construction of the CIM and digital twin. The primary sources utilized for the historic datasets are the repository for the foundation of the CIM. The challenge in using these sources is that none of them were digitized into point, line, polygon, or attribute tables that could be pooled and blended. The first step in this research was to digitize these datasets into shapefiles with joined attribute tables for analysis. These organized datasets then fed the CIM for analysis. To function within the digital twin, select datasets incorporated z values to indicate the height of spatial objects. Thereafter, the data and its accompanying documentation was made available for public use in an open data portal.

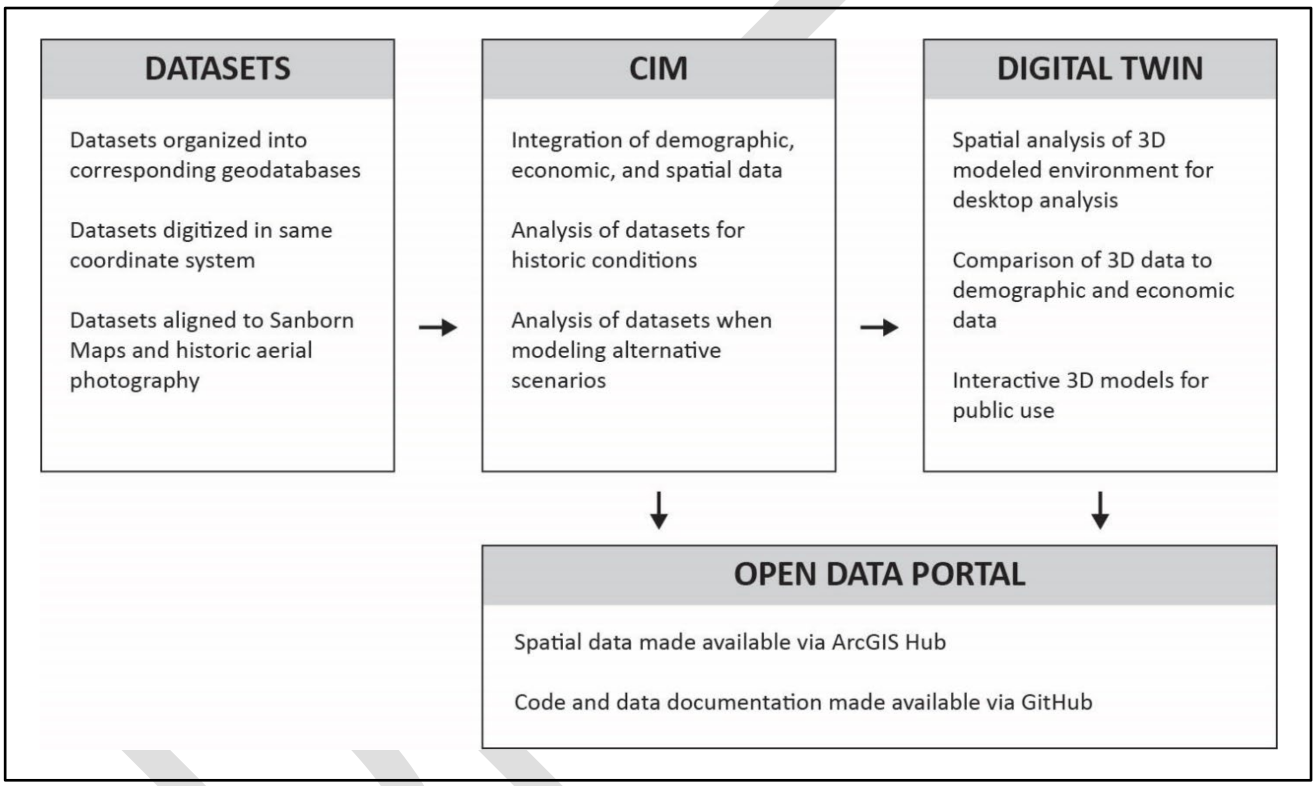


Figure A.1: Model Framework for Data Workflow

2.a.i. Aligning with Professional Standards

At present, there is no all-encompassing professional standard for CIM or digital twins. There are emerging best practices for cities implementing the technology, but there is no handbook or guidebook. This is an emerging field where a significant amount of experimentation is ongoing. However, because these tools are used by government officials and the general public, emphasis is placed on building high-quality models with data integrity and accuracy.

That said, three standards do exist that are substantially related to CIM and digital twins: International Organization for Standardization (ISO) 37120, CityGML Standard (v3.0), and the City Induction Research Project. The ISO standard proposes a set of data criteria to monitor smart cities; CityGML is a data structure and storage framework for 2D mapping and 3D models; and, the City Induction Project is a workflow for urban design analysis partially supported by artificial intelligence. The ISO and CityGML standards, specifically, are intended to facilitate the transferability of a model's framework and data storage from one city to another. The City Induction Project is a theoretical proposal for ongoing operations and the iterative growth and refinement of the CIM. When the three standards are used simultaneously, they provide a more complete approach to build, operate, and manage CIM and digital twins.

The intent of this research is to align the Milwaukee CIM and digital twin as closely to these standards as possible. This will be a good faith effort to develop an original model for Milwaukee as closely aligned as possible to existing standards given what is feasible within the constraints of available datasets.

2.b. Data Structure

For each georeferenced map and digitized dataset, separate file geodatabases hold the corresponding point, line, and polygon files. New geodatabases were created for each map or dataset that represented a different year or time period. For a singular dataset from one data source that had a direct relationship to other datasets from other data sources, those datasets were then organized into a common geodatabase for a defined time period.

Each geodatabase may contain multiple file types, but only represent demographic, economic, or spatial data. These data categories are not mixed in the geodatabases. For example, the digitized files from the Sanborn Maps are held as line and polygon shapefiles in their respective geodatabase, while Census data is digitized based on historic ward maps for the city in a polygon shapefile joined to a .dbf attribute table in its own geodatabase. These datasets are stored separately, but can be overlaid on one another for a demographic and spatial comparison.

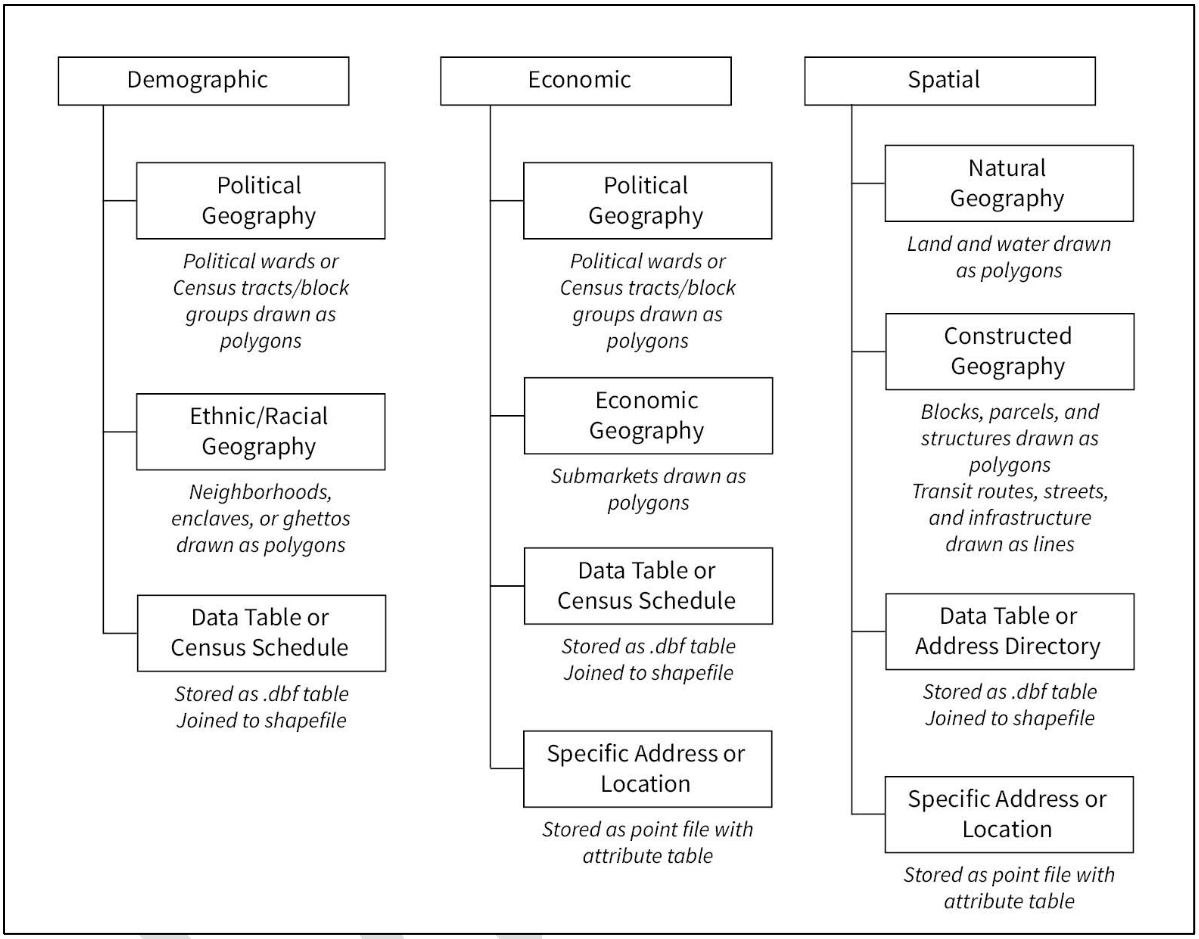


Figure A.2: Data Structure of File Geodatabases

All data used in this research has been projected into NAD 1983 HARN StatePlane Wisconsin South FIPS 4803 Feet. This coordinate system is the same system utilized by the Milwaukee County Land Information Office. Importantly, the georeferenced Sanborn Maps (1910) and historical aerial photography library are all georeferenced in this coordinate system. To follow Milwaukee County's practices, the research data was similarly digitized and stored in this coordinate system.

2.c. Public Deployment on Digital Platforms

The analysis workflow allows for data management across multiple digital platforms, which allows data sharing and hosting on multiple pieces of software. Fortunately, this hosting is largely made possible because of the variety of software available from ESRI. However, the data does need to be managed appropriately to ensure that 2D datasets can accurately be fed into the 3D visualizations.

The data workflow is generally broken into two components: 1) desktop analysis of 2D and 3D data by the researcher, and 2) online data sharing with interactive tools for public use. The desktop analysis is composed of numerous tasks that focus on data digitization, data management, quantitative analysis of attribute tables, and spatial analysis of attribute tables and 3D models. The public deployment of the data on the website enables any user to utilize the data dashboards, interactive 2D maps, and digital twin. In addition, the open data portal allows the public to download the data uploaded to these tools for their own analysis.

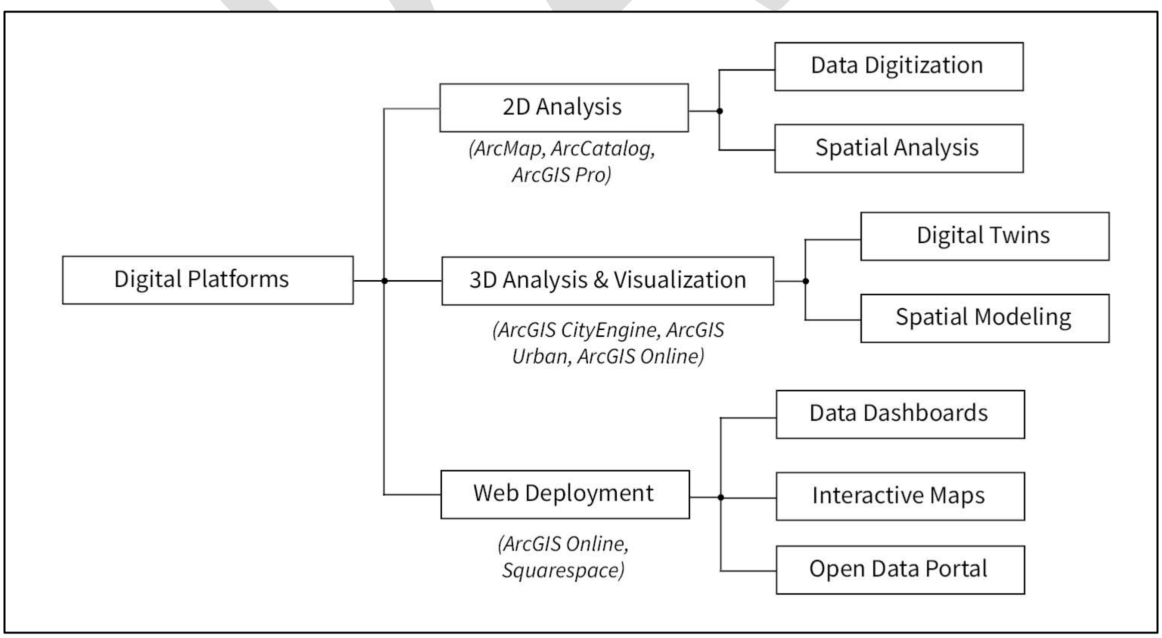


Figure A.3: Workflow & Software Management for Public Deployment of Data

2.d. Processes & Procedures

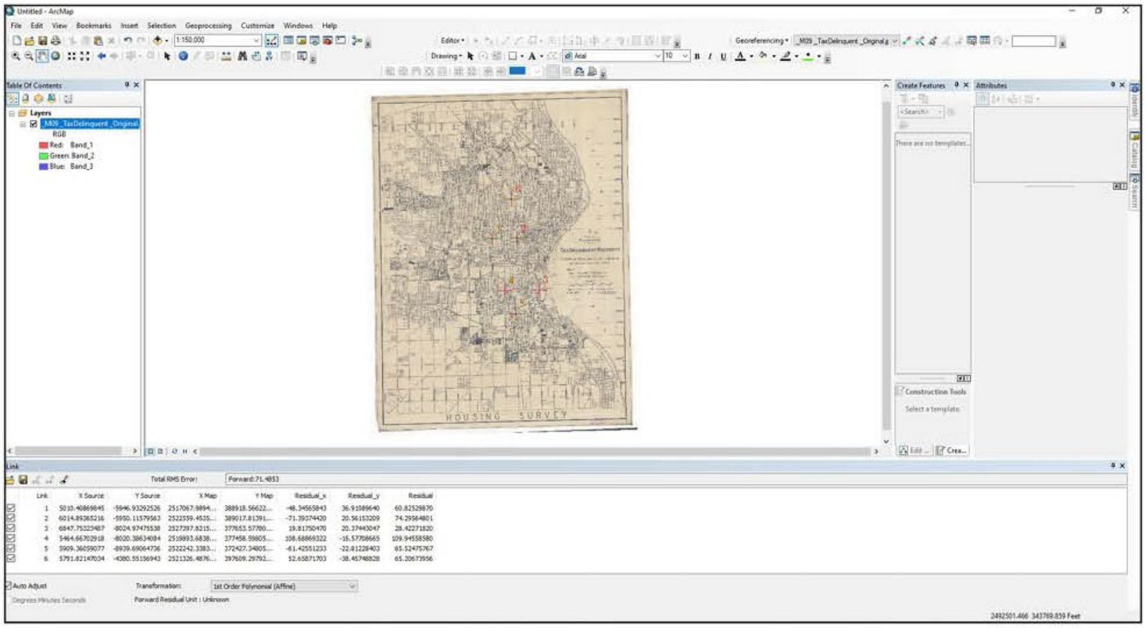
Within specific aspects of the model framework and data workflow, additional processes and procedures were implemented to manage the datasets and their analysis. The focus was on data integrity and ensuring the accuracy of digitized datasets. Because the historic data needed to be converted into modern storage formats, it was imperative that the historic data as it originally appeared be faithfully re-produced for contemporary software analysis. In maintaining the data integrity and accuracy, this thereby provided a greater level of certainty in the operation of the CIM and digital twin.

2.d.i. Georeferencing Maps

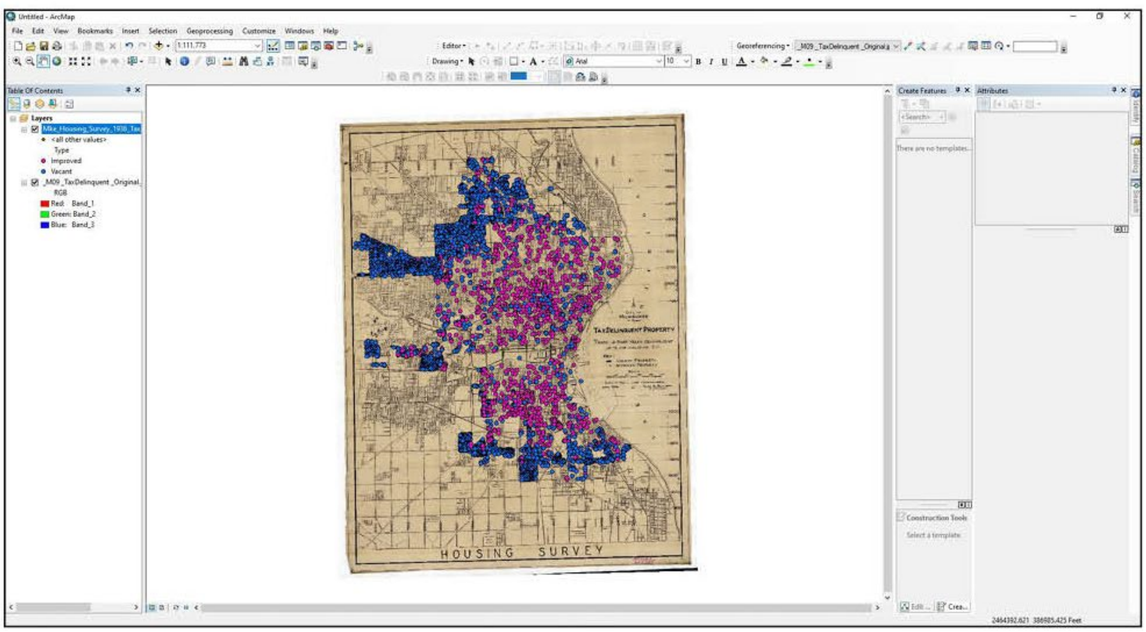
Archival collections of historic maps of Milwaukee provide critical insight into the spatial structure of the city at various points in time. Most importantly – and obviously, digitizing historic ward maps and then joining them to decennial Census data creates the foundation for this research's data analysis. However, due to the volume of maps available, the opportunity existed to create custom databases from scanned maps. To achieve this, the scanned images needed to be georeferenced in ArcGIS. The images of these maps and the accompanying data provided important primary source material to further develop my research conclusions.

A key element in my research's analysis methodology was pooling multiple datasets from different sources to create data overlays and draw conclusions. Throughout the course of the research, it became apparent that historic datasets had not been digitized for contemporary use. As a result, the majority of the data was available in primary source materials in their original format. While significant amounts of narrative and data were available in reports, a substantial amount was available on maps. The maps created two opportunities for analysis: 1) they could be georeferenced and their features digitized into standalone databases specific to that source; and/or, 2) the maps could be georeferenced, their features digitized, and then the data from one map could be related to another map for comparison. This process yielded the creation of numerous databases resulting from primary source maps.

Georeferencing is a digitization process that relates points on a source map to a reference map, thereby accurately placing and scaling the source map in real-world space. To be effective, the process requires high-resolution photos in JPG or TIFF format. In ArcMap, the coordinate system for the map space is set. Control points are placed on the source map and then referenced to a second set of control points on the reference map in real-world space. Once a sufficient number of control points have been placed, the georeferencing data for the image is updated. After the source map has been successfully georeferenced, its features can then be digitized as point, line, or polygon feature classes.



Georeferencing Tools: Control Points & Links Table



Digitized Data on Georeferenced Map

Figure A.4: Georeferencing Process & Data Digitization

2.d.ii. Digitization Process

For historic data sources, many have been scanned as PDF or high-resolution image files; however, the data, images, and maps they present have not been georeferenced or digitized into modern software. Thus, the data sources are valuable for research purposes, but they need to be modernized into contemporary systems to be useful. Each of these sources needs to be digitized into either spreadsheet data, a georeferenced map, or georeferenced point, line, and polygon files.

The digitization process can be categorized as one of two sub-processes for Milwaukee: 1) digitize data from historic narrative or first-person accounts onto an existing georeferenced map (i.e., Sanborn Maps or Milwaukee County historic aerial photography), or 2) georeference a map and then digitize its features into point, line, and polygon files. The georeferenced map is then saved as a TIFF image. The final output of this effort is the storage of the digitized data in file geodatabases organized by either the original data source or a broader theme or time period. This digitization process occurred for every primary source data, image, or map that is cited in the research.

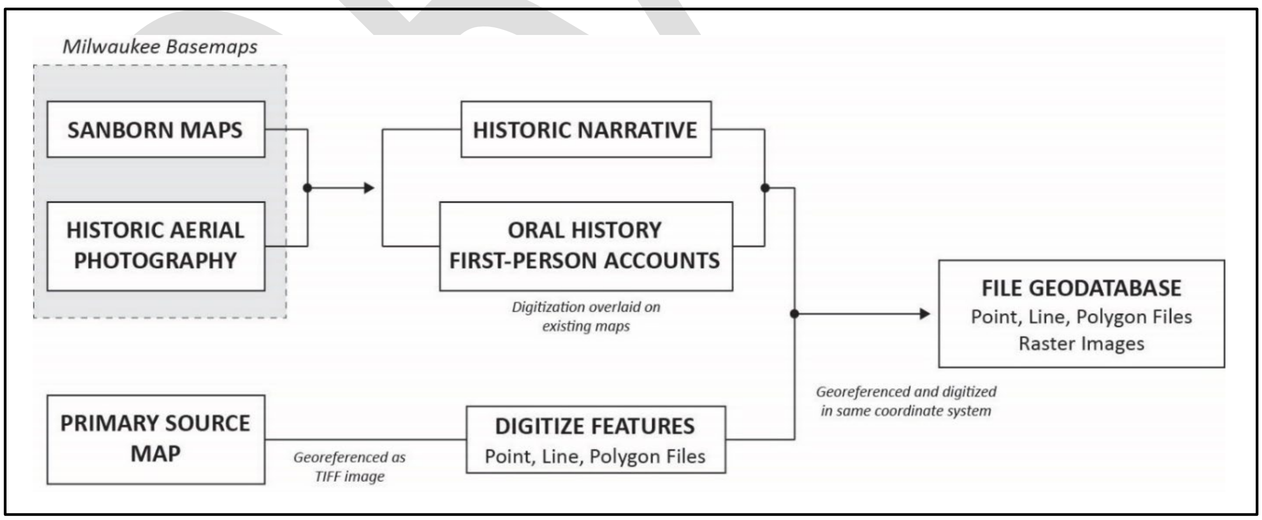


Figure A.5: Digitization Workflow of Primary Source Material

2.d.iii. Data Auditing Process

Following the digitization of a data source, the newly created dataset needs to undergo an audit for quality control. The purpose of the audit is to ensure that the newly-created dataset accurately reflects the data or map presented in the primary source material. This quality control process is required for every primary source that is digitized.

Two types of audits need to occur for each data source: an attribute audit, and/or a feature audit. An attribute audit is used for spreadsheet data when data tables from a source are digitized into .dbf tables that are joined to a shapefile or for standalone spreadsheets that are fed into a data dashboard. An attribute audit is focused on ensuring that the data is transcribed accurately from the original source to the new table. A feature audit is used on spatial data to verify that point, line, and polygon features are digitized accurately from a georeferenced map into a feature class stored in a file geodatabase. This type of audit needs to confirm that all features have been digitized and stored properly. Additionally, for features that need to be defined by attribute data, the corresponding attribute table for each feature class needs to be audited to ensure that the attributes were transcribed accurately. These two types of audits are integral processes in the construction of a CIM to ensure data accuracy, thereby ensuring its integrity. Without regular audits, significant errors can be introduced into the model at its inception, which can cause additional errors during analysis and scenario modeling.

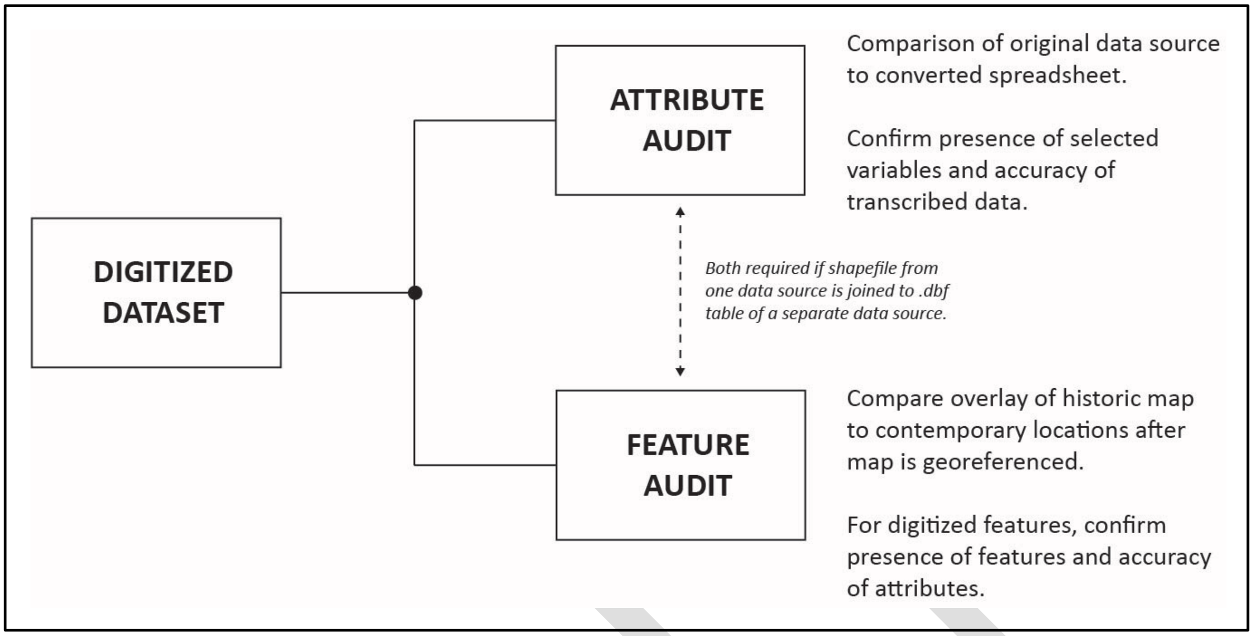


Figure A.6: Audit Process of Digitized Datasets

3. Milwaukee Model Specifications & Notes

The Milwaukee model contains datasets and attributes that are unique to itself. While the CIM's framework and data structure are meant to be transferable to other cities, aspects of the model are Milwaukee-specific because of locally available datasets. To accurately document these unique attributes, the following discussion and notes identify elements and characteristics requiring further clarification.

3.a. Data Dictionaries

3.a.i. Sanborn Fire Insurance Map (1910)

The fire insurance maps from the Sanborn Map Company remain as arguably the most authoritative catalogue of America's built environment in the late 19th century and the first half of the 20th century. While historical research about Milwaukee benefits from the numerous mapping and cataloguing efforts that date to the city's founding, there is no better resource than the Sanborn Maps to understand the city's historical spatial condition. Thus, the maps should be utilized as a trusted resource, but researchers must understand the historical purpose of the maps contrasted with their current use.

The Sanborn Map Company was originally founded by D.A. Sanborn as the D.A. Sanborn National Insurance Diagram Bureau in 1866. The company's explicit purpose – and, by extension, the purpose of its maps – was to provide catalogues of maps to insurers to assist in risk management efforts related to fire protection in American municipalities. The company represented an early entrant into the burgeoning business of national insurance map publishing.[1] In an essay written for the Library of Congress' extensive collection of Sanborn Maps, the purpose of the maps is summarized well:

> The maps were designed to assist fire insurance agents in determining the degree of hazard associated with a particular property and therefore show the size, shape, and construction of dwellings, commercial buildings, and factories as well as fire walls, locations of windows and doors, sprinkler systems, and types of roofs. The maps also indicate widths and names of streets, property boundaries, building use, and house and block numbers.[2]

At the time of their publishing, Sanborn Maps were not used by the general public. The maps were of specific use for the fire insurance industry as reference material when underwriting a policy. It was not until years later that the maps became of interest to professionals in other industries and members of the public.

Professional interest in Sanborn Maps has grown significantly by various groups in the engineering, surveying, urban planning, and historical research professions. In their contemporary use, the maps' value is well recognized because of their accuracy and detail. The maps provide a unique perspective for researchers because of their status as a near-complete catalogue of the spatial structure and building unit inventories of numerous American cities. Today, the maps provide important details about the platting structure of cities (parcels, blocks), urban transit networks (streets, alleys, railroads, waterways, ports/docking), fire suppression infrastructure (water mains, fire hydrants, fire alarms, building sprinkler systems), and characteristics of buildings (fire protection, height, windows, doors, building materials,

---

[1] Indiana University Libraries, "Sanborn Fire Insurance Maps History" (undated).
[2] Walter Ristow, "Introduction to the Collection: Fire Insurance Maps in the Library of Congress" (undated).

land use). Just as important, however, is what the maps lack; there is no discussion or data about land topography, property values, and land use regulations (i.e., zoning). In select instances, building attributes refer to a specific ethnic or racial community, but more detailed demographic information about residents is absent. Thus, the maps are a valuable tool to research urban built environments; but, they need to be used in conjunction with other, historical data sources to develop more comprehensive demographic and economic portraits of cities.

The digitization of Milwaukee's Sanborn Maps serves as a central element in the CIM's framework. Whereas contemporary Milwaukee already relies on multiple, detailed datasets to describe the built environment, historical data sources are disparate with different maps presenting a variety of spatial attributes and conditions in the city. Importantly, the Sanborn Maps accurately outline the city's spatial structure. The spatial structure is used in two ways: 1) to provide a dataset of attributes of the city's blocks, parcels, and structures for the CIM, and 2) to provide shapefile polygon data to extrude the blocks and structures for the digital twin. This thus provides spreadsheet and spatial data to create the CIM's 2D maps for spatial analysis and the digital twin's 3D renderings.

Table A.1 provides the attribute table for the structure characteristics of Milwaukee's Sanborn Maps as displayed in the CIM and digital twin. In addition to the written descriptions on the maps, multiple other legends were consulted to properly code the structure attributes. These primarily include details made publicly available by Environmental Data Resources, Inc. (EDR), California State University-Northridge (CSUN), and the Library of Congress.

Table A.1: Attributes of Structures in Sanborn Map Feature Class

Name	*Type*	*Field Length (number of characters)*	*Values*	*Description*
Company Name	Text	200	Unique	If identified, name of company located in the structure.
Building Name	Text	200	Unique	If identified, the type of building on the parcel.
Land Use Code*	Text	5	R – Residential RT – Residential-Transient C – Commercial W – Warehouse M – Manufacturing P – Public or Institutional U – Utility T – Transportation	The abbreviation used to identify the land use type of a structure.
Land Use Type	Text	50	Educational Entertainment Fire Police Streetcar Hotel Religious Railroad Residential Residential – Duplex Residential – Flats Residential – Tenement Residential – Boarding Residential – Secondary Residential – Mixed Commercial Commercial - Secondary Industrial	Identification of land use of a structure.

			Industrial - Elevator Warehouse Municipal	
Personal Transit Type	Text	20	Stable Auto	For buildings housing either horses or cars for transportation.
Bldg_Height	Short Integer	- - -	Unique	Height of a building in stories.
Bldg_Height_Feet	Short Integer	- - -	Unique	Height of a building in feet.
Abbreviations*	Text	25	*Refer to EDR abbreviation guide.	The abbreviation used to identify the use or business located in a structure.
Neighborhood	Text	100	Walker’s Point Bayview Third Ward Lindsay Heights Halyard Park Brewer’s Hill Harambee	Identification of a structure in select neighborhoods from Milwaukee’s Neighborhood Identification Project and contemporary planning efforts.
Floor_Plate_Area	Double	- - -	Unique	Geometry area calculation in square feet of first floor of building footprint.
No_DU	Short Integer	- - -	Unique	Number of dwelling units per residential structure, as indicated by the number of addresses listed. Tenement buildings show one dwelling unit if the total number of dwelling units are not otherwise indicated on the map.
No_Bus	Short Integer	- - -	Unique	Number of businesses per structure. Determined by either counting the number of street addresses per building or counting the number of businesses identified in the building. If the two counts conflict, the lower number is indicated in the database.

Build_Mat	Text	100	Wood Wood, metal clad Wood, brick clad Wood, brick veneer Wood, brick veneer, iron clad Wood, iron clad Brick Cement block Concrete Concrete block Reinforced concrete Metal clad Concrete and brick	Type of construction materials used for a structure.
Alley_House	Text	100	Y – Yes N – No	Identification of a residential structure as an alley house. Alley and rear houses are interpreted as synonymous in this dataset.
S_by_S	Text	50	Y – Yes N – No	Identification of residential structures as side-by-side housing sharing a common wall.

*Notes: *Indicates field documented by Environmental Data Resources, Inc. (EDR).*

4. Digital Twin Development in ArcGIS CityEngine

A hallmark of my dissertation research is the ability to blend and pool large datasets to draw conclusions from a variety of sources. This advantage is most clearly seen in maps, diagrams, tables, and interactive 2D tools on my research website. An added strength and emerging technology is the 3D rendering of these datasets in ArcGIS CityEngine. The data workflow for this relies on procedural modeling of 3D features that allows for dynamic data analysis and visualization of the built environment.

The data workflow for 3D renderings is reliant on the organization and quality of 2D datasets. As a result, the datasets need to be managed in a cohesive workflow that facilitates an efficient analysis of spatial and spreadsheet data in 2D space, which is subsequently transformed into 3D models. The central element of this workflow is data attribute fields in the individual database feature classes. The organized management of the data attribute fields operationalizes all subsequent analyses in this research. To manage the digital twin development for my dissertation research, key aspects of the data attribute fields were organized to specifically fit within the data workflow.

4.a. Feature Class Management

Within the digital twin, each “shape” – or 3D feature – relies on data attributes to be called in the procedural modeling. This operation of “calling” visualizes the shape in 3D space. The shape not only visualizes itself based on its attributes, it also retains additional attributes that allow for further refinement in the modeling. However, this refinement is only made possible if the feature classes are properly designed with visualization and interactivity in mind. Because the digital twin can only visualize a certain number of attributes before the features become illegible, it must be decided which features will be visualized, how the attributes of those features are stored in the feature class, and how the attributes will be used in the procedural modeling to build the 3D environment.

During the course of my dissertation research, it became apparent that data attribute consolidation was necessary to create efficient lines of code in the procedural modeling. The challenge in this approach was that my dissertation research was based on deductive analysis, whereas the procedural modeling is based on inductive thinking. Because of this dissonance, the procedural modeling was an iterative process requiring deliberate and judicious management of the data attributes.

To illustrate this inductive process, an example of data attribute consolidation is relevant. In the structures feature class digitized from the Sanborn Maps (1910), each structure has two data fields for land use: “Land_Use_Type_General” and “Land_Use_Type.” “Land_Use_Type_General” is a consolidated data field generated from the specific attributes of “Land_Use_Type.” The field “Land_Use_Type_General” was designed specifically for the development of the digital twin to code the shapes of the buildings into one of six general land use types. This consolidation of data fields allowed for two processes to occur simultaneously: 1) the visualization of the land use of individual structures in the digital twin, and 2) the storage of specific land use types in the data attribute table for all structures. Table A.2 details the consolidation of these data fields.

Table A.2: Consolidation of “Land_Use_Type” into “Land_Use_Type_General” Fields

Land_Use_Type		***Land_Use_Type_General***
Residential Residential – Boarding Residential – Female Boarding Residential – Duplex Residential – Flats Residential – Tenement Residential – Mixed Residential – Secondary	→	Residential
Commercial Commercial – Secondary Hotel Office	→	Commercial
Industrial Industrial – Elevator Industrial – Secondary Railroad	→	Industrial
Warehouse Warehouse – Secondary	→	Warehouse
Boat House Net House Pier	→	Maritime
Community Educational Entertainment Hospital Municipal Religious	→	Community

4.b. CGA Rule Files

Procedural modeling in ArcGIS CityEngine is built on the programming language of computer generated architecture (CGA). The basic structure of CGA is to extrude and manipulate shapes with functions that operate on data attributes. The simplicity of CGA lends itself to a variety of applications in 3D modeling. Basic models extrude features to display massing, while others provide additional detail in schematic models or rendered models that display façade features and materials. To extrude features, a CGA rule file needs to be written for each data layer. These rule files specify the rendered display attributes of the features.

For the digital twins in my dissertation research, schematic modeling was used to provide a reasonable level of detail about the built environment. The CGA rules were written to be easily adapted to the display of a variety of feature attributes. As a result, the code was designed to be streamlined and efficient. Additional detail was only introduced when specific attributes were meant to be displayed. For example, land use models were developed for neighborhood analysis with the data field “Land_Use_Type_General” visualized with six colors representing the different attributes. Figure A.7 shows code excerpts from two different CGA rule files for various features.

Welcome | Massing_Blocks.cga

```
/**
 * File:    Massing_Blocks.cga
 * Created: 14 Jul 2023 18:44:59 GMT
 * Author:  Kristian Vaughn, UW-Milwaukee SARUP
 */

version "2022.1"

@StartRule
Mass -->
    extrude(0.25)

# Extrude block at standard height of approximately 0.5 feet for all blocks. No distinction between blocks.
```

Massing of Blocks

Welcome | Schematic_Massing_Structures.cga

```
/**
 * File:    Schematic_Massing_Structures.cga
 * Created: July 2023
 * Author:  Kristian Vaughn, UW-Milwaukee SARUP
 * Adapted from a CGA rule developed by ESRI R&D Center and Devin Lavigne, Houseal Lavigne Associates.
 */

version "2022.1"

###   CONSTANTS
const unitScale = 0.3048          # Convert feet to meters
const areaScale = 10.7639         # Convert square meters to square feet

const Floor_Height = 3                      # Floor heights are 3 meters (approximately 10 feet)


###   ATTRIBUTES


@Group("Building Attributes",2)
@Description("Height of the building in stories; land use type; roof feature")
@Order(1)
@Handle(shape=SplitFloors, align=left)
@Range(min=0,max=200,stepsize=1)
attr Bldg_Height = 0
attr Land_Use_Type_General = ""

@StartRule
Generate-->
    color(.8,.8,.8)
    extrude(Bldg_Height * Floor_Height)
    comp(f){top:BuildRoof | all:x.}

BuildRoof-->
    case Land_Use_Type_General == "Residential":
        roofHip(45)
    else:
        x.
```

Massing of Structures

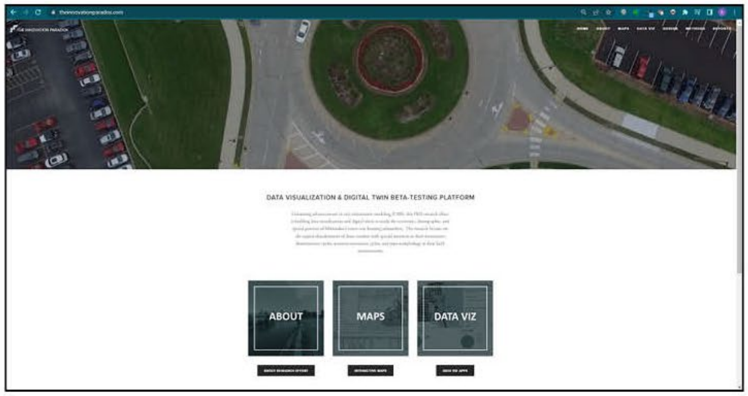
Figure A.7: Excerpts of CGA Rule Files for Schematic Modeling in ArcGIS CityEngine

5. Operational Website & Public Deployment of Digital Tools

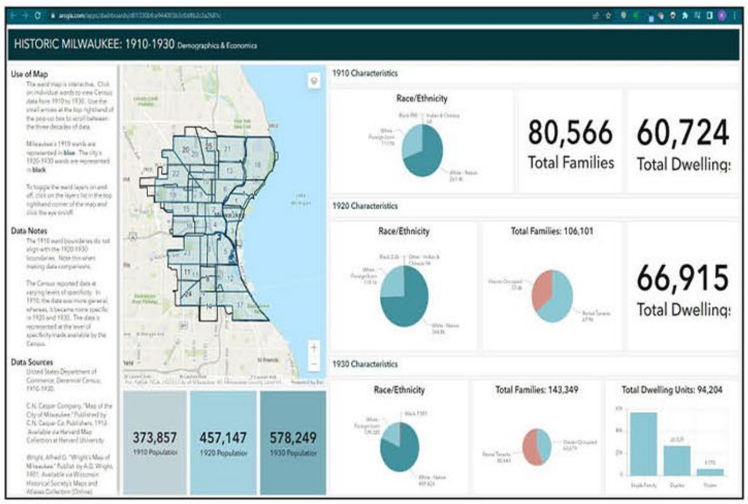
In pursuing research investigating CIM and digital twins, it became apparent during the dissertation process that public access to the tools was necessary. It was not sufficient to simply develop the tools on a private server, excerpt graphics into the dissertation research, and then delete the developed technologies. With this realization, the public deployment of the tools on a website was developed under the auspices of three general ideas:

1. To establish the technologies as viable proofs of concept, they needed to withstand use by and criticisms from the public.
2. The public access would facilitate public education about pressing urban issues, raise public consciousness about historical events, and assist in community engagement.
3. By achieving Ideas 1 and 2, the website would create an accountable and transparent research endeavor that made data publicly available and democratized the information.

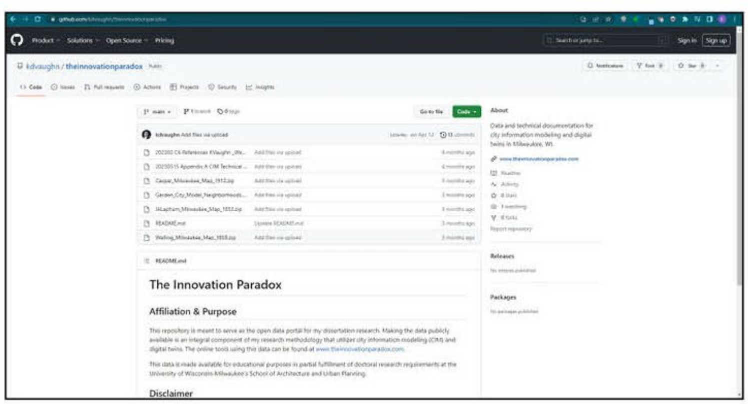
As a result, the website was launched in an initial beta-testing capacity at the onset of the research at the URL www.theinnovationparadox.com. As the research progressed, the website's content became more detailed and nuanced with maturing digital tools. The website presented users with research content in multiple formats: 2D interactive dashboards and maps, 3D digital twins, reports in PDF format, and an open data portal via GitHub. The purpose in providing this diversity of information in these formats was to present research conclusions, demonstrate the methods and sources utilized to arrive at those conclusions, and make datasets available for peer review. While the website did not qualify as an acceptable deliverable of the dissertation in its final format, it was a key element in the research process proving the capability and efficacy of CIM and digital twins.



Website Home Page



Data Dashboard



GitHub Open Data Portal

Figure A.8: Snapshot of Publicly-Available Digital Tools on Website