

Database Design

**Financial Intelligence Center –
Release 2**

Version 8

DOCUMENT CONTROL



Change Record

Date	Author	Version	Change Reference
3/21/2012	Karl Dollard	1.0	Initial version
3/23/2012	Karl Dollard	1.1	Incorporated feedback from Zac Perry
6/28/2012	Karl Dollard	2.0	Incorporated R1 scope
7/10/2012	Karl Dollard	2.1	Updated content, formatting
7/17/2012	Karl Dollard	2.2	Incorporated feedback from design review
7/31/2012	Karl Dollard	2.3	Incorporated feedback from deliverable review
12/14/2012	Suryanadh Majji	2.4	Incorporated R2 scope
1/14/2012	Suryanadh Majji	8	Updated R1 data dictionary based on deliverable review



Sign-Off

Name	Title/Position	Date	Signature

TABLE OF CONTENTS

DOCUMENT CONTROL 2

INTRODUCTION 4

 PURPOSE..... 4

DATA MODEL..... 4

 SHARED DIMENSIONS 4

New Dimensions 4

Existing Dimensions 5

 SUBJECT AREAS/STAR SCHEMAS 5

R0 Scope:..... 5

R1 Scope:..... 5

R2 Scope:..... 7

DATA DICTIONARY..... 9

 DIAGRAMS 9

R0 Scope:..... 9

R1 Scope:..... 15

R2 SCOPE:..... 23

FDNY NYFIRS: 24

 ENTITIES & ATTRIBUTES 34

 EXCLUDED ENTITIES..... 34

INTRODUCTION

PURPOSE

The purpose of this document is to describe the data model that will be used for the R0 release of the Financial Intelligence Center project. This data model will be used to provide data via the Oracle Business Intelligence (CPR), Palantir, and SAS applications.

Approach

The CPR data warehouse serves two primary purposes for the FIC project. First it is the central collection point for the majority of the data that is to be made available to both the Palantir and SAS. Second, the data will be surfaced through the existing CPR application for agency users who wish to do less sophisticated analysis of the data using an application where they already access other data sets.

To support these requirements, we have followed the following principles while modeling the data sources:

- Overall, we have attempted to keep the data as close as possible to the original structure, so as to preserve natural relationships for investigative and predictive modeling.

- In many places the data will be denormalized. This will not only preserve the relationships, but will also reduce the amount of effort for all teams, by reducing the number of distinct data mappings. This will also facilitate modeling the data for deeper analytics, which works more efficiently with denormalized structures.

- For file based data sources (including DEEP and any other MQ based sources) we have created the extraction for all the elements, but have not necessarily modeled every element to the target. In some cases, this serves to add significant complexity to the model, but add very little investigative or analytical value. These sections are listed later in this document for review and validation.

- For data sources to which we have a direct connection we will validate with the business owners the elements to be extracted, as extracting every element adds complexity to the model with little investigative or analytical value in return.

DATA MODEL

SHARED DIMENSIONS

To support cross data source analysis, several new common 'dimensions' - essentially a group of attributes - will be created. Additionally, several existing dimensions within the CPR data warehouse will be referenced, as appropriate.

NEW DIMENSIONS

Person - This will store information about individuals. Exact matches from different sources will share a common record, but no other matching will be attempted. This dimension is designed in such a way that it can be readily integrated with a future MDM or other data enrichment project.

Location - This dimension will be a collection of the various location based datasets. At the base level, this will be defined by the combination of a BIN (Building Identification Number) and BBL (Borough Block Lot). Additional information, such as address and zoning data will be associated with this table and available for querying.

Business - Similar to the Person dimension the business dimension will store business related attributes. Exact matches will share a common record, but no other matching will be attempted. Future MDM implementations or 3rd party data (DnB, Experian, etc) could be used to enrich this dimension.

EXISTING DIMENSIONS

Date - This dimension stores information about dates - the rollup of day to week to month to year, day of week, day of the year, city holidays, and other related information. This simplifies reporting as this information is standardized and pre-calculated.

Time - Similar to the Date dimension, the Time dimension will store information about each second within a 24 hour period.

Community Board - The community board dimension is a higher level view of the location dimension. This provides a common location dimension across nearly every data set within the CPR data warehouse.

Agency - The agency dimension will not only indicate from where the data has been received, but also is commonly used to provide agency based security for datasets that should not be shared with other agencies.

SUBJECT AREAS/STAR SCHEMAS

The following subject areas will be provided for analysis within the CPR data warehouse. This data will be available for Palantir and SAS, though the concept of a subject area may not exist to the same extent in those applications.

R0 SCOPE:

DOB Jobs - This subject area will contain information about DOB Jobs, the items required for the jobs, and the permits that have been applied for. This information will be related to all of the shared dimensions except for the Time dimension.

HPD Building Registrations - This subject area will contain information about the registered contacts for HPD buildings. This data will be related to the Person, Location, Community Board and Agency shared dimensions.

ECB Notice of Violations - This subject area will contain information about violations that have been sent to ECB to be adjudicated. This will also contain information about the individual charges that are a part of the violation as well as the outcome. This information will be related to all the shared dimension except for Time and Business.

HPD Violations - This subject area will contain information about violations levied by HPD. This data will be related to all shared dimensions.

HPD Inspections - This subject area will contain information about property inspections conducted by HPD. This data will be related to the Date, Location, Community Board and Agency shared dimensions.

R1 SCOPE:

DCA Licenses: DCA licenses 55 industries and enforces the laws that govern these industries. This subject area will contain information about the different licenses and the businesses which are licensed. The data in this subject area will be related to all of the shared dimensions with the exception of Time.

DCA Inspections: DCA licensed businesses are subject to intermittent inspections to ensure they are comply with the current laws. This subject area will contain information about the inspection requests, inspection results, and inspection tests. The data in this subject area will be related to all of the shared dimensions with the exception of Time.

DCA Violations: From an inspection or a follow upon a complaint, it may be determined that a business is in violation of a specific law or set of laws. This subject area will contain information about the violation and the outcome of the violation. The data in this subject area will be related to all of the shared dimensions with the exception of Time.

DCA Complaints: DCA captures the complaints placed against business. This subject area will contain information about the complaints placed against both licensed and non-licensed business. The data in this subject area will be related to all of the shared dimensions with the exception of Time.

DOHMH Permits and Licenses: All the permits/ licenses issued at the DCA Licensing center are stored and tracked in CAMIS. The majority of the application materials for Licensing and Permits are available on the NYC Business Express website. This subject area will store the information about the permits/licenses. The data in this subject area will be related to all of the shared dimensions with the exception of Time.

DOF Ownership: DOF captures information related to the property owner in several different applications/files. The information captured includes owner name, property address, BBL, Tax and Building Class, and mailing recipient name, address, and contact information. The data in this subject area will be related to all of the shared dimensions with the exception of Time and Business.

DOF Property Charges: When an entity becomes the owner of property in New York City, they are subject to numerous laws, codes and taxes. The amount of taxes each property owner is charged depends on several factors, including the tax class of the property, size of the property and eligibility of tax exemptions/abatements. This subject will eventually store this information. This data will be related to the Location, Date, Agency and Community Board shared dimensions.

DOF ACRIS – Property Documents: ACRIS is the system used by the Department of Finance (DOF) to record property documents for Manhattan, Brooklyn, Queens and the Bronx. These documents include deeds, new and paid mortgages, property transfers and tax payments made. Additionally, due to integration with Richmond County, information for Staten Island is also available, though with slightly less detail. This data will be related to the Location, Business, Person, Agency, Date and Community Board shared dimensions.

NYCSERV Payments: NYCServ is operated by the Department of Finance and is New York City's primary mechanism for many NYC related payments. The NYCServ application currently processes payments from 3 separate channels: Internet, IVR (phone), and Business Center (in-person). Payments can be made for multiple charges including Parking Violations, ECB Violations, Property Tax, among others. This data will be related to the Date, Time, Person and Agency shared dimensions.

HPD Complaints: HPD receives complaints from tenants regarding issues with apartment buildings. Note that currently CPR also receives HPD complaints, but a different subset of the data, both in terms of volume and scope. This data will be modeled separately, with considerations for later incorporating this into the overall CPR SR model.

DOB Complaints: DOB receives complaints regarding construction and building code violations. Note that currently CPR also receives DOB complaints, but a different subset of the data, both in terms of volume and scope. This data will be modeled separately, with considerations for later incorporating this into the overall CPR Service Request (SR) model.

FDNY Investigations: TBD

R2 SCOPE:

FDNY NYFIRS: The FDNY is the largest combined Fire and EMS provider in the world, and also the largest municipal fire department in the United States. This data set includes information about all Fires or other incidents that the FDNY responded to. The information includes the date and time of the event occurred, the number of civilian deaths, firefighters and civilians injured, and the source and cause of the fire. This data set is related to the Agency, Location, Community Board, Date and Time shared dimensions.

FDNY CIDS: FDNY maintains a database of building information commonly known as CIDS (Critical Information Dispatch System). CIDS provides information on building entryway and design and other atypical or unique features of a specific location that could be useful for firefighters responding to an incident at that location. This data set is related to the Agency, Location, Community Board and Date shared dimensions.

Liquor License: A list of all liquor licenses is available from Sheriff's website. This dataset includes about the licensee as well as the license. This data set is related to the Agency, Location, and Community Board shared dimensions.

Lis Pendens: This data set consists a list of pending foreclosures. The data includes the docket number, property, and plaintiff and defendant details. This data set is related to the Agency, Location, Community Board, Person, Business, and Date shared dimensions.

Section 8 Housing: Section 8 authorizes the payment of rental housing assistance to private landlords on behalf of low-income households, and operates through NYCHA/HPD/DHS. This dataset is used to augment the Location shared dimension with a Section 8 flag.

DOF Tax Liens: A tax warrant is a legal action against an individual or business and creates a lien against their real and personal property. This data set consists of two components – the 90-warning that a tax lien will be filed, and the notification that a lien has been filed. The data set contains information about the property, the property owner, and all the relevant charges. This data set is related to the Agency, Location, Community Board, Person and Business shared dimensions.

EBT Reimbursement: EBT (Electronic Benefit Transfer) Reimbursements is a list of all reimbursements paid to vendors who do not participate in electronic processing of EBT transactions. This data set contains information about the Business and the total amount reimbursed. This data set is related to the Agency, Location, Community Board and Business shared dimensions.

Cigarette Inspection Results: The Sheriff's Department, on behalf of the Department of Finance, conducts inspections to ensure that all requisite taxes have been paid on Cigarettes sold within NYC. This dataset contains the Sheriff's cigarette tax inspection results along with the address details, inspection date, and cigarette inspector, and seizure details. This data set is related to the Agency, Location, Community Board and Date shared dimensions.

BIC Customer Registration: The mission of the Business Integrity Commission (BIC) is to eliminate organized crime and other forms of corruption and criminality from the industries it regulates. BIC responds to inquiries, issues violations for regulatory infractions and also resolves complaints. This data set consists of a list of all customers of private waste haulers. This data set is related to the Agency, Location and Community Board shared dimensions.

DOF Property Assessment: The Department of Finance regularly assesses the value of all properties in NYC. This dataset contains the information about these assessments, along with additional descriptive information about the property. This data set is related to the Agency, Location, Community Board and Date shared dimensions.

Licenses-Permit Superset: A superset is a collection of the data from different subsets where each of these subsets can be logically grouped together. This superset comprises the data from three different data sources/exchanges – DOB Jobs Permits, DCA Licenses, and DOHMH Permits – so all the licenses and permits' data across these three systems will be represented in a combined way. This superset is related to the Agency, Location, Community Board, and Date shared dimensions.

Violation Superset: A superset is a collection of the data from different subsets where each of these subsets can be logically grouped together. This superset comprises the data from three different data sources/exchanges – HPD Violation, DCA Violation, and ECB Violation – so all the data pertain to Violations across these three systems will be represented together. This superset is related to the Agency, Location, Community Board, and Date shared dimensions.

Location Superset: A superset is a collection of the data from different subsets where each of these subsets can be logically grouped together. This superset comprises the data from the following data sources/exchanges so there will be a holistic view of the location information across these 24 systems. This superset is related to the Agency, Community Board, and Date shared dimensions.

- BIC Customer Registry
- DCA Violations
- DCA Complaints
- DCA Licenses
- DCA Inspections
- DOB Complaints
- DOB Jobs
- DOB Permits
- DOF ACRIS
- DOF Cigarette Tax Inspection
- DOF Ownership
- DOF Property Charges
- DOF Tax Liens
- DOHMH Permits
- ECB Notice of Violations
- FDNY CIDS
- FDNY NYC Fires
- HPD Building Registration
- HPD Complaints
- HPD Inspections
- HPD Violations

- HRA EBT Reimbursement
- Liquor License
- Lis Pendens

DEP Complaints: The Department of Environmental Protection receives complaints regarding storm water, water pollution and other pollution related issues. This data set contains information about the complainant and the resolution of that complaint. This data set is related to the Agency, Location, Community Board, Person, and Date shared dimensions.

DEP Water Bills: This data set contains a list of water bill charges and amount paid by customer and location. This data set is related to the Agency, Location, Community Board, Person, Business, and Date shared dimensions.

911: The 911 systems provide a single number for all emergency calls. Those calls are then routed to the appropriate agency or agencies for action. This dataset is an aggregate of several other datasets and provides, for the first time in City history, a consolidated view of response time from call initiation to arrival on scene. This data set is related to the Date and Time shared dimensions. The individual sets are described below:

Magic: This data set contains telephony information from the 911 call center. It contains information about the time of the call, the telephony related actions associated with that call (transfers, conferences, talking) and the agents associated with those actions. It does not contain any information about the content of the call.

Sprint: This data set contains a textual log of the information recorded in Sprint – the incident management system used by 911 call takers and dispatchers. This data set contains information about the call, the incident, and how the incident was handled. This data set is in a text file that is structured much as if it had been printed, and must be reformatted and parsed before being used.

Starfire: This data set is the output of FDNY's CAD system. It contains information about all incidents that were handled by FDNY and has the FDNY specific disposition of the incident, details about the equipment used and the time of arrival.

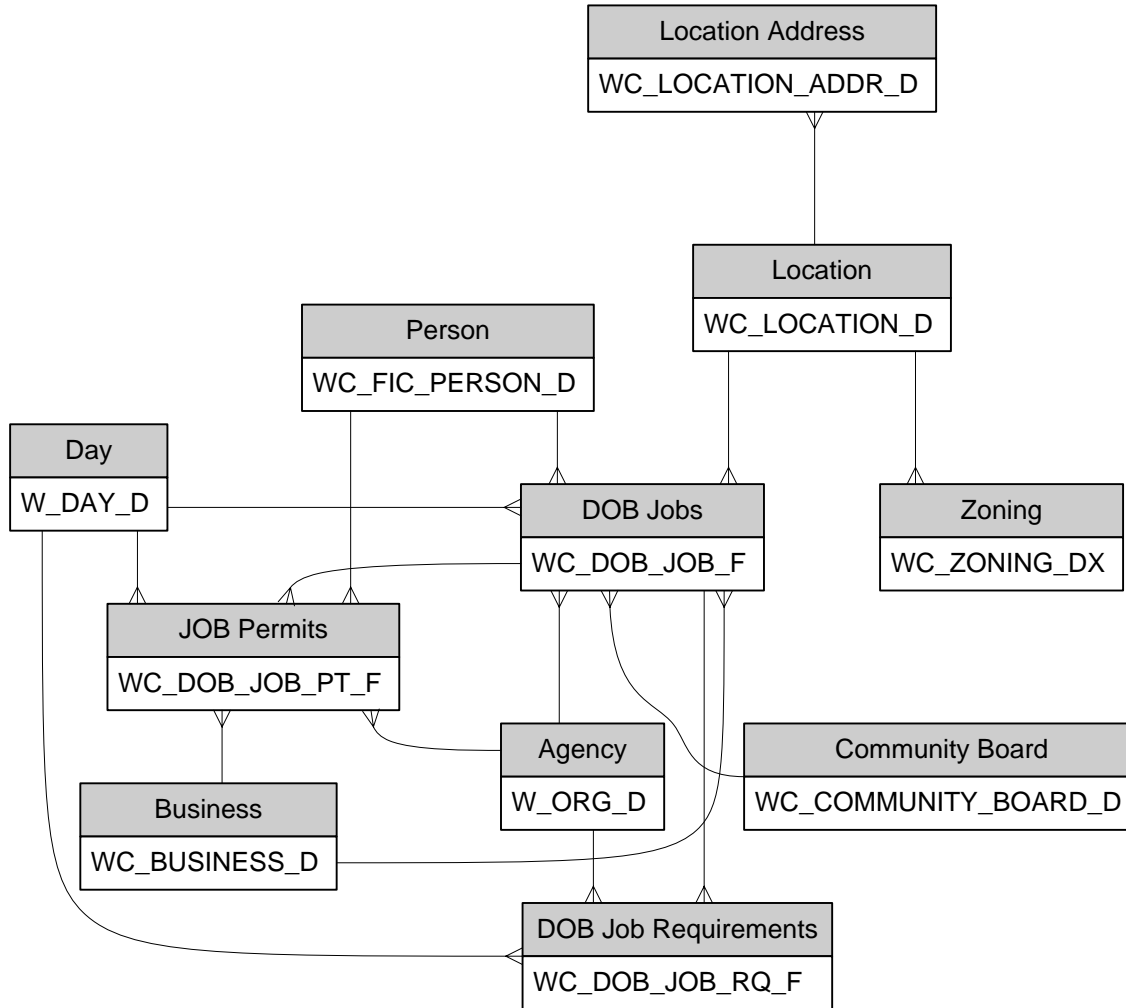
EMS: Like Starfire, EMS has its own CAD system that contains information about all incidents that were handled by EMS. This data set has information about the EMS disposition of the incident, the arrival time and outcomes of the incident.

DATA DICTIONARY

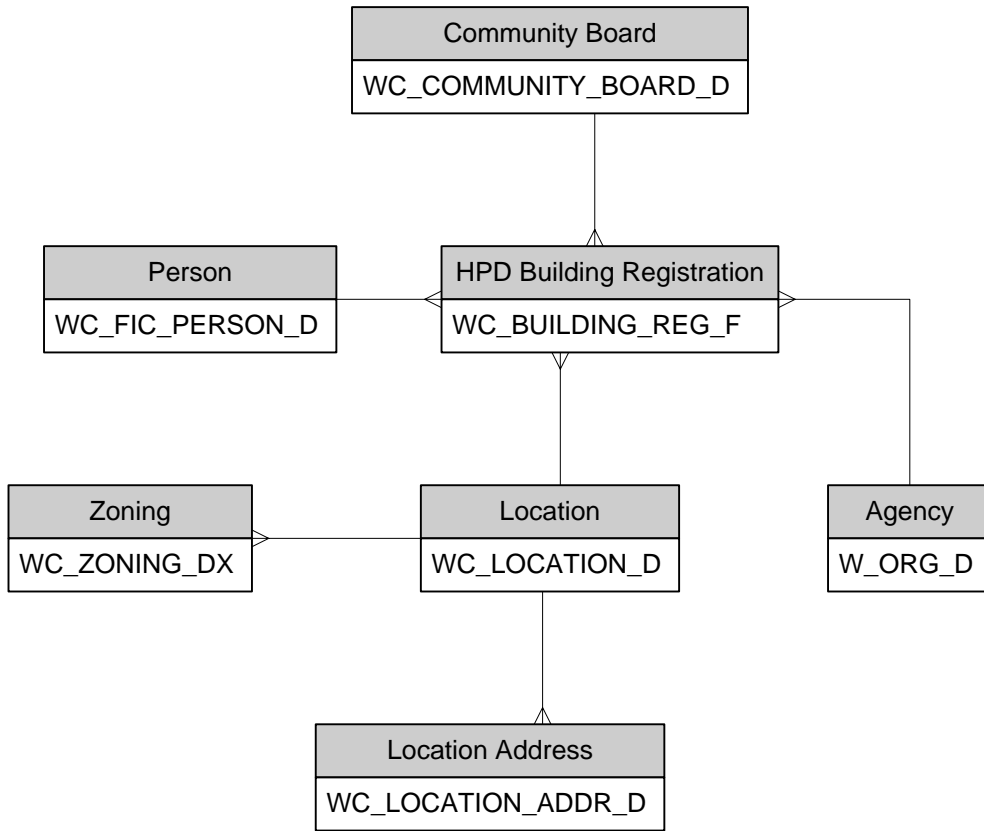
DIAGRAMS

RO SCOPE

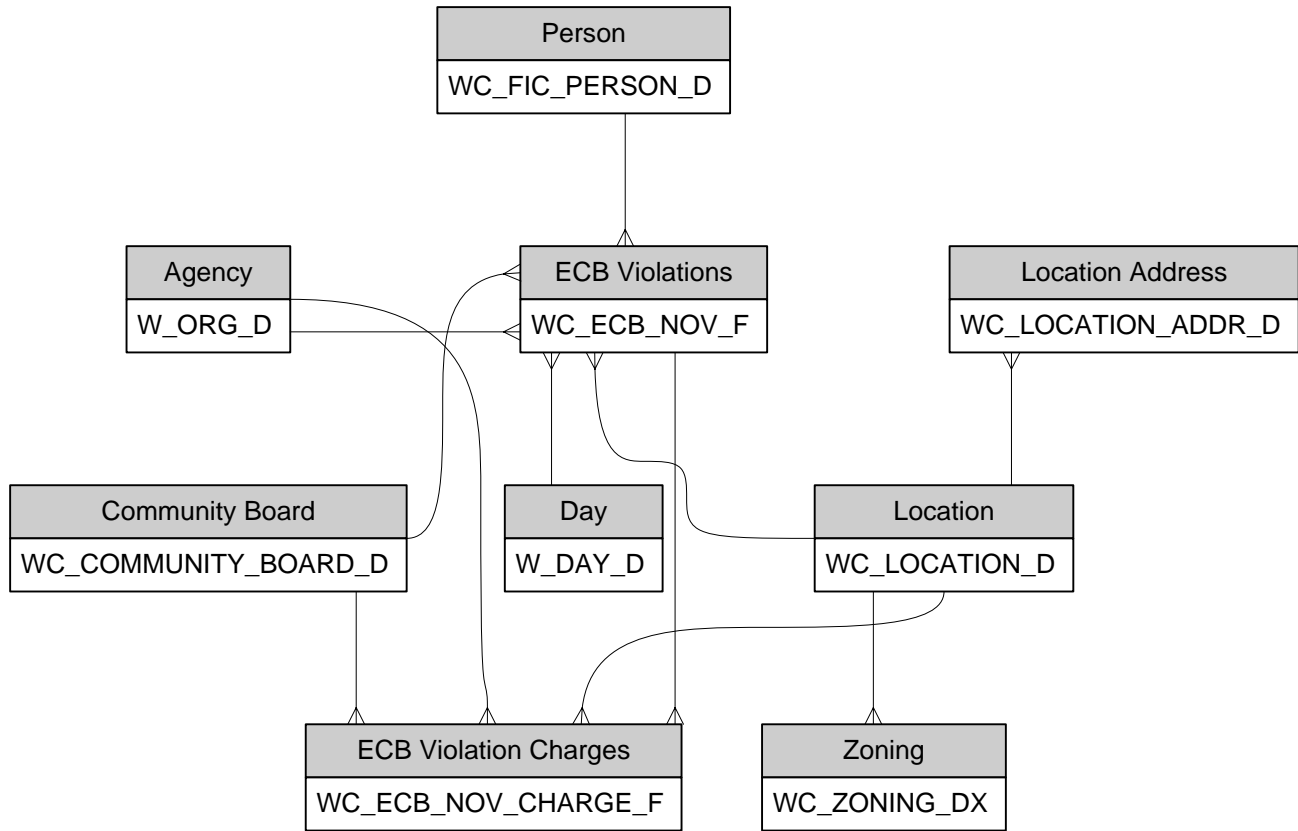
DOB Jobs, Permits and Required Items



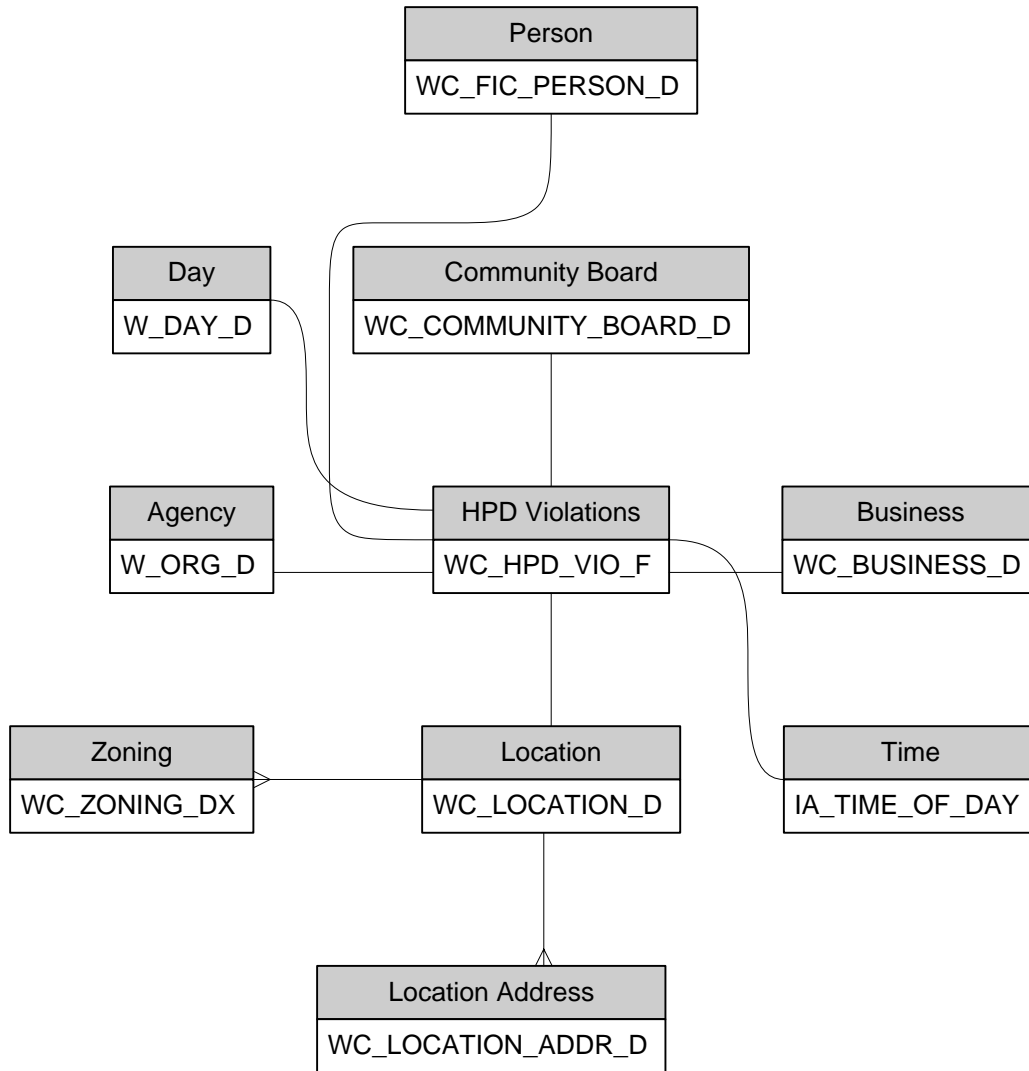
HPD Building Registrations



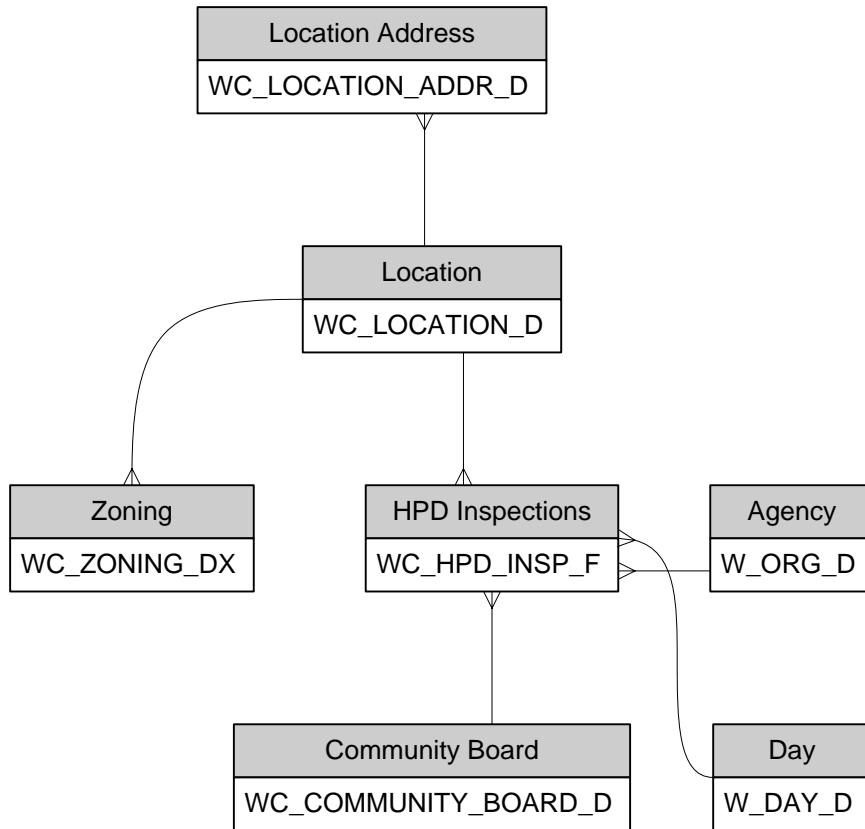
ECB Notice of Violations (NOV)



HPD Violations

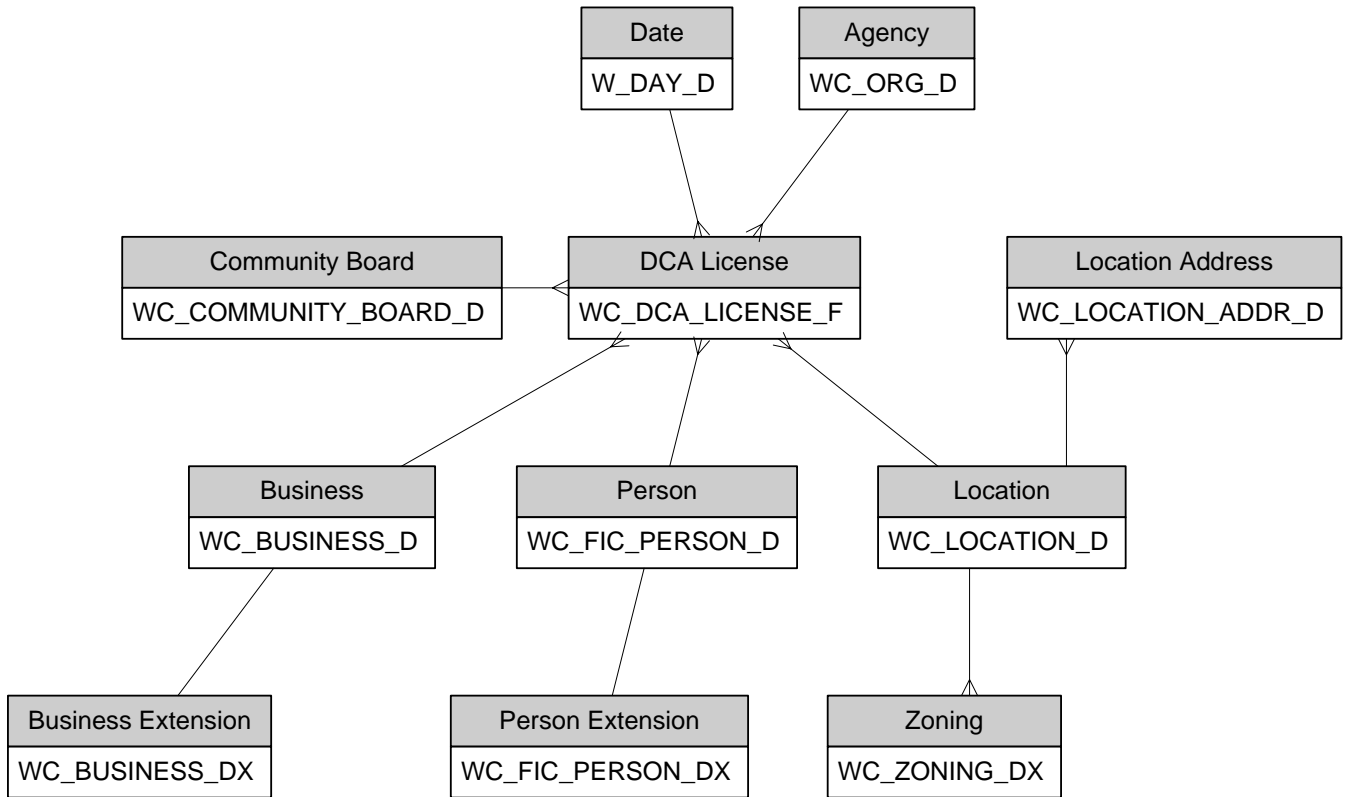


HPD Inspections

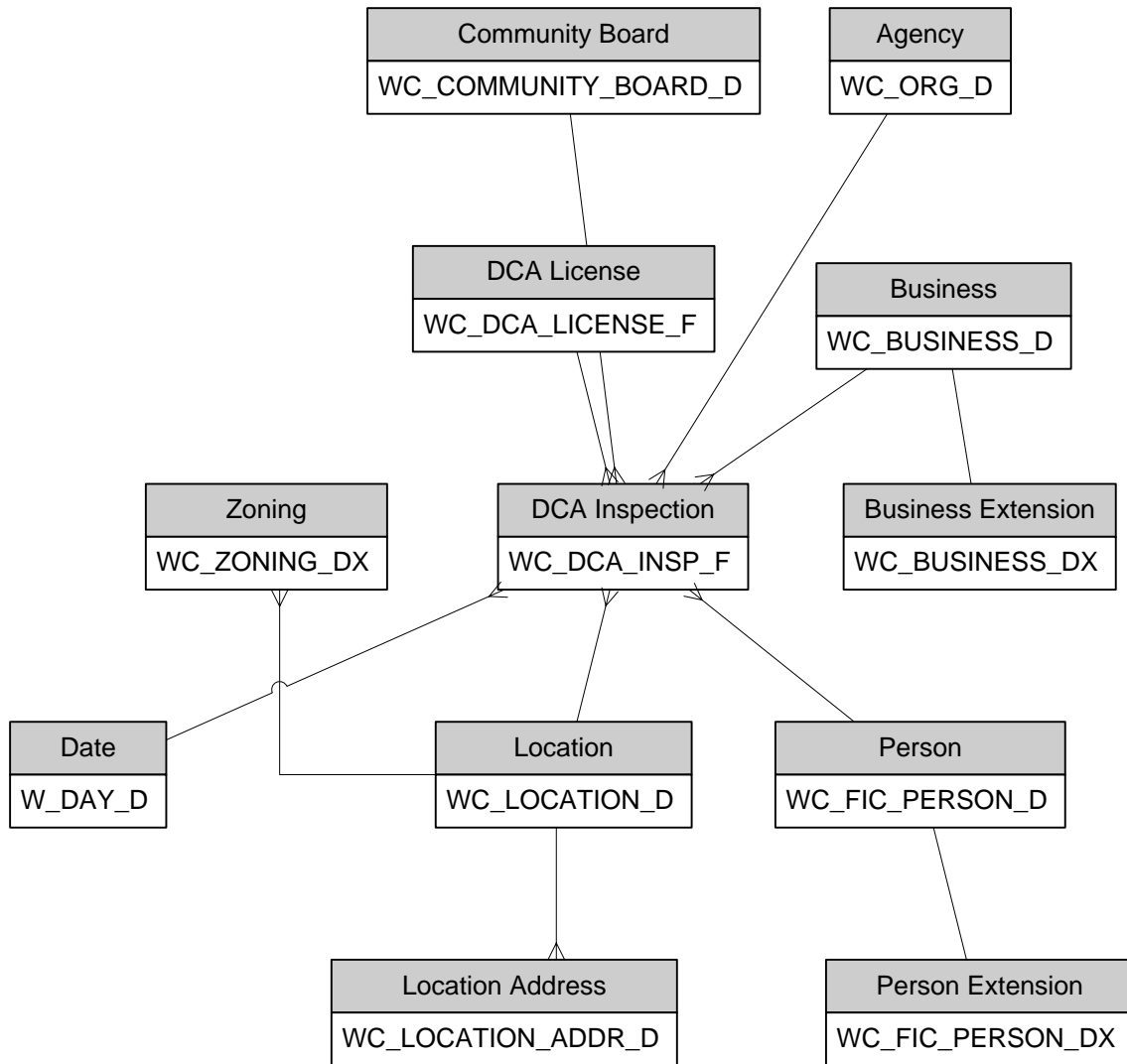


R1 SCOPE

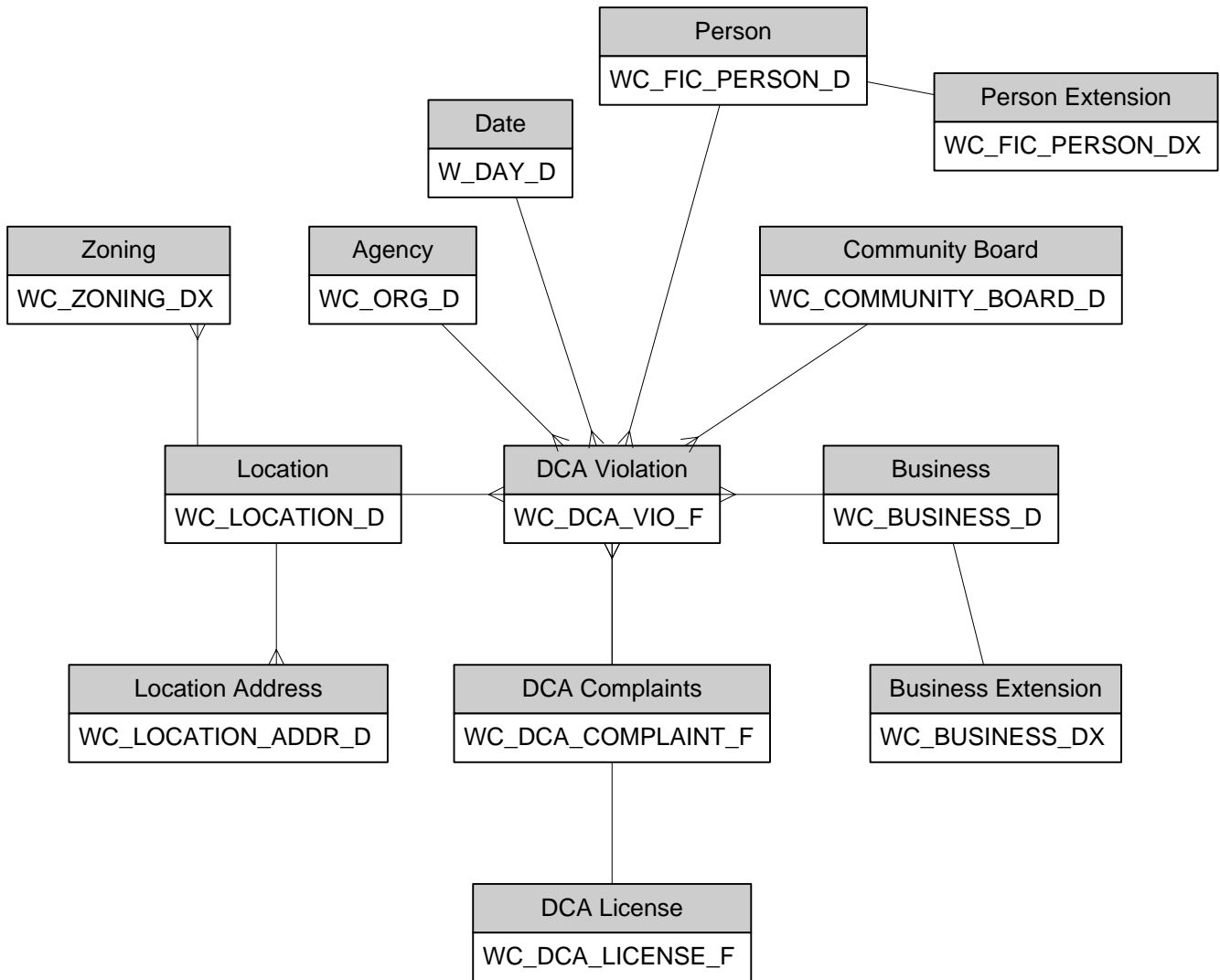
DCA CAMIS – Licenses



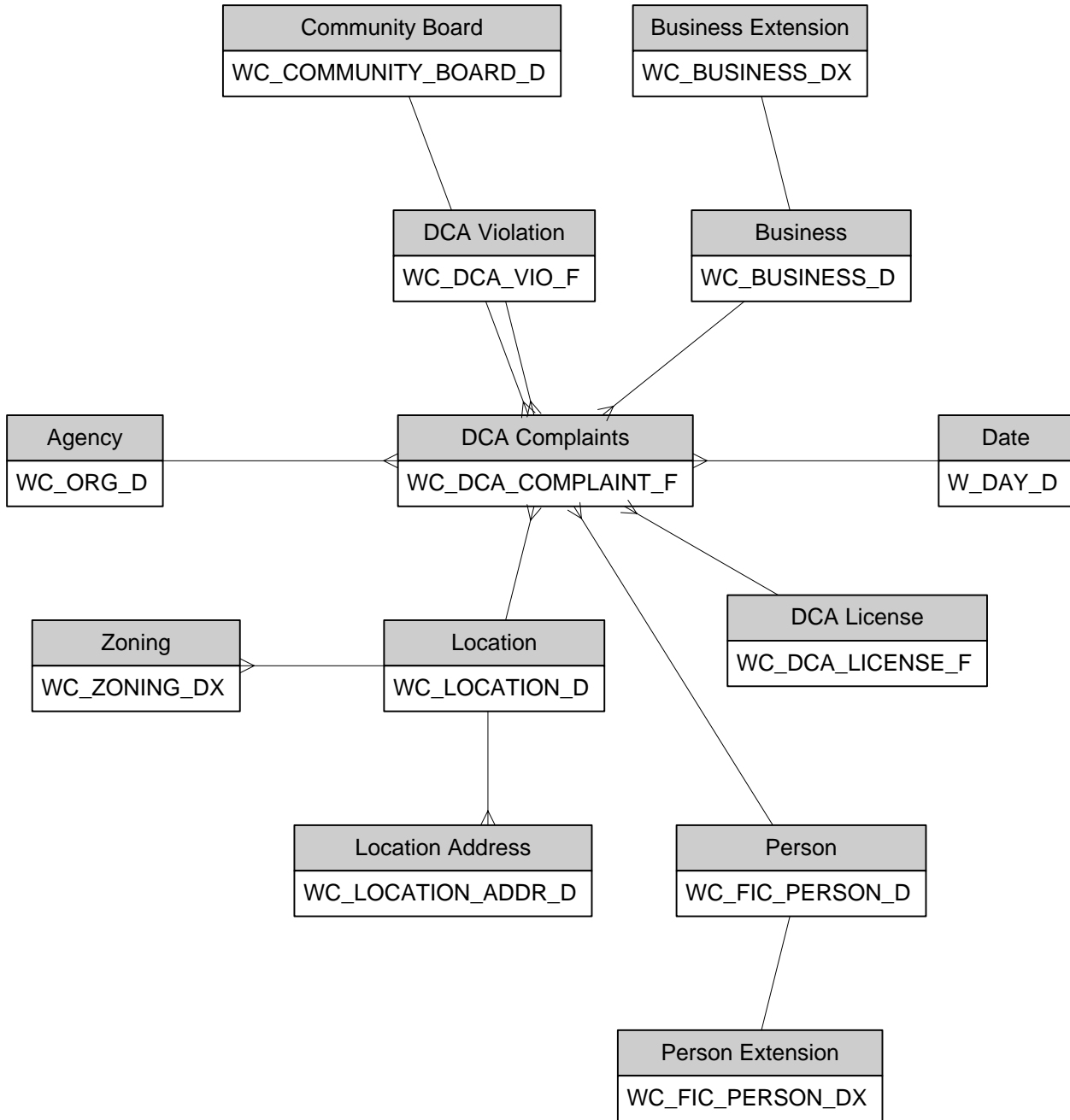
DCA CAMIS – Inspections



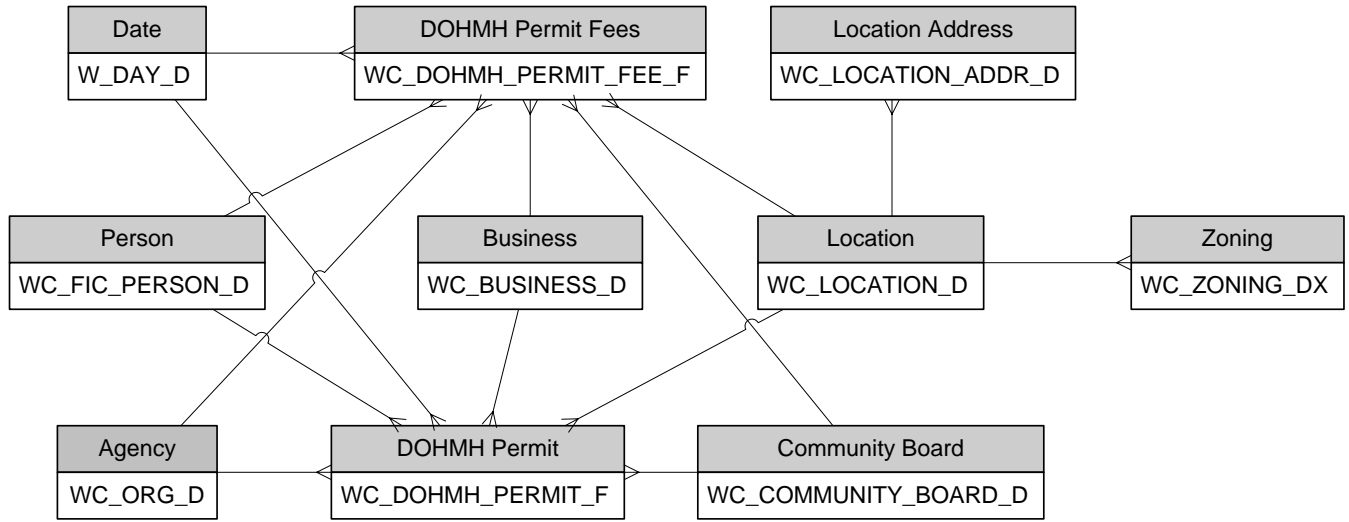
DCA CAMIS – Violations



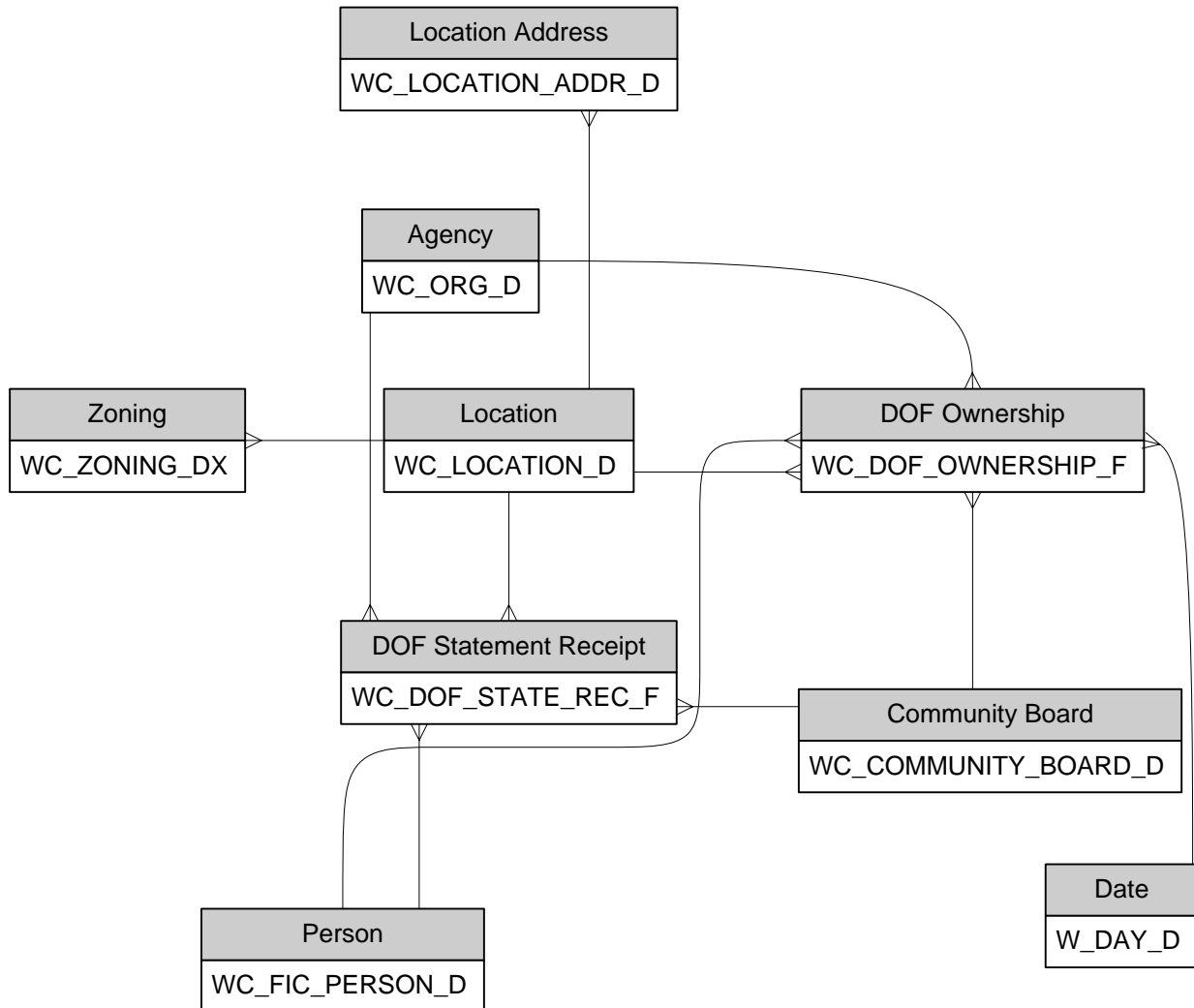
DCA CAMIS – Complaints



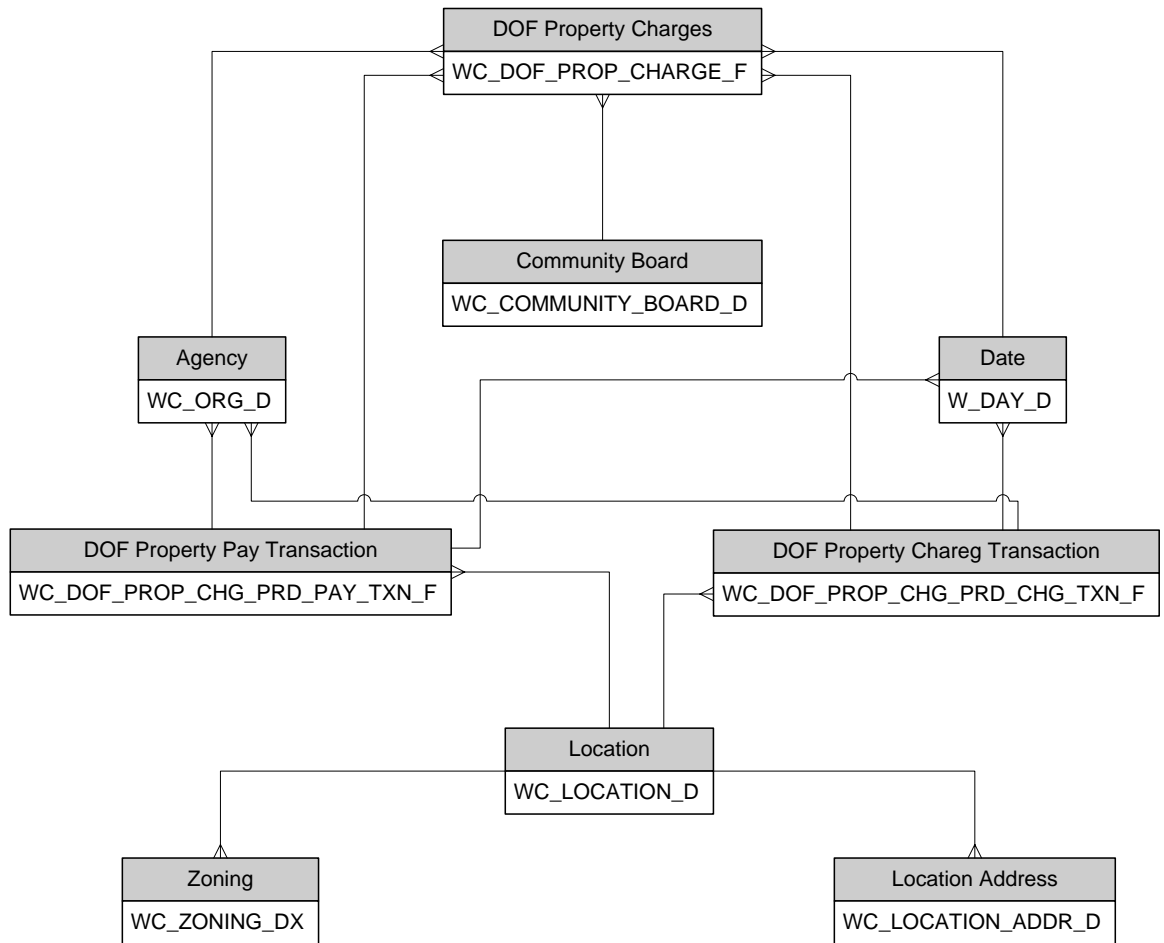
DOHMH – Permits/Licenses



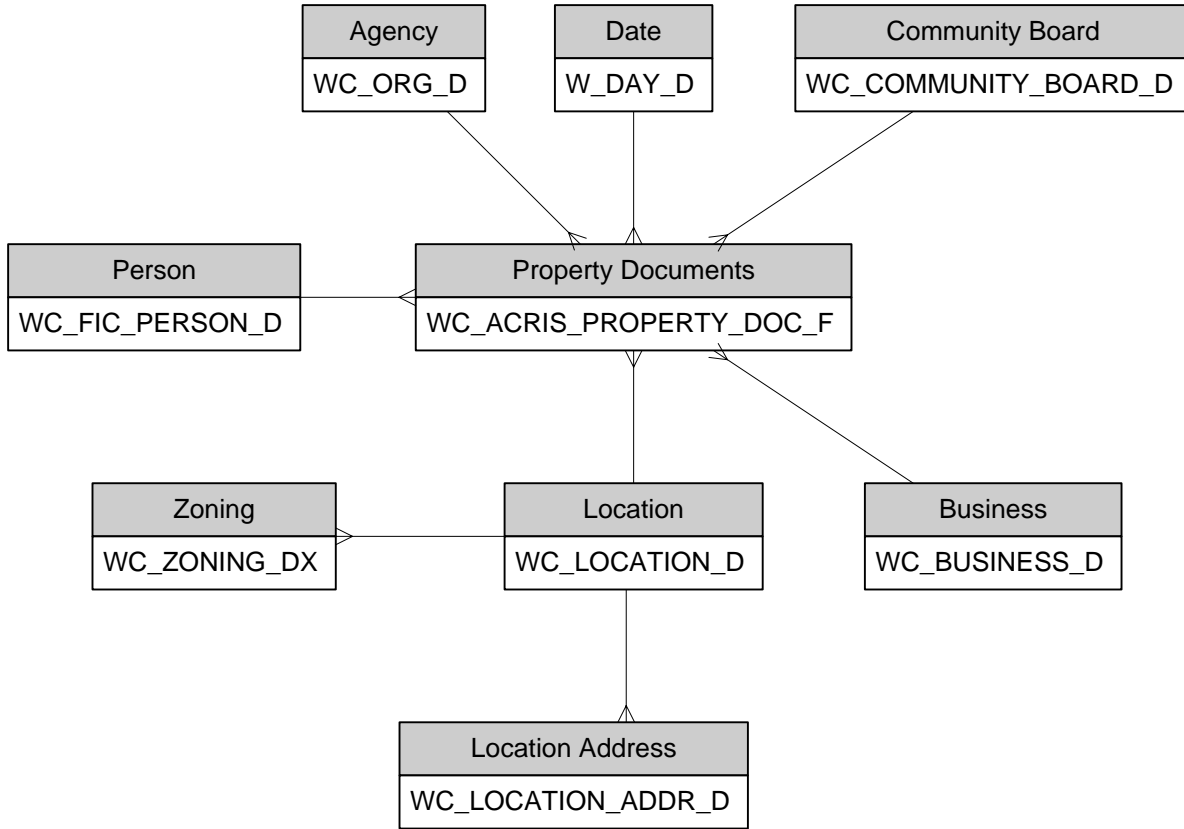
DOF -- Ownership



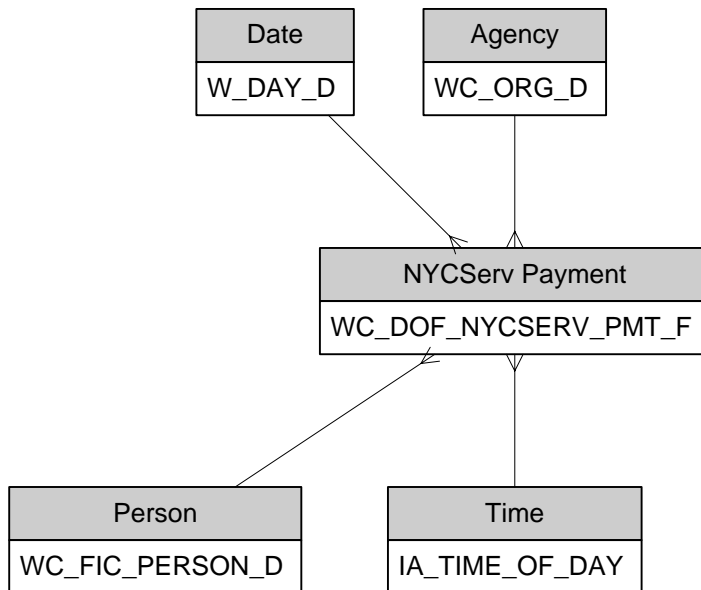
DOF -- Property Charges



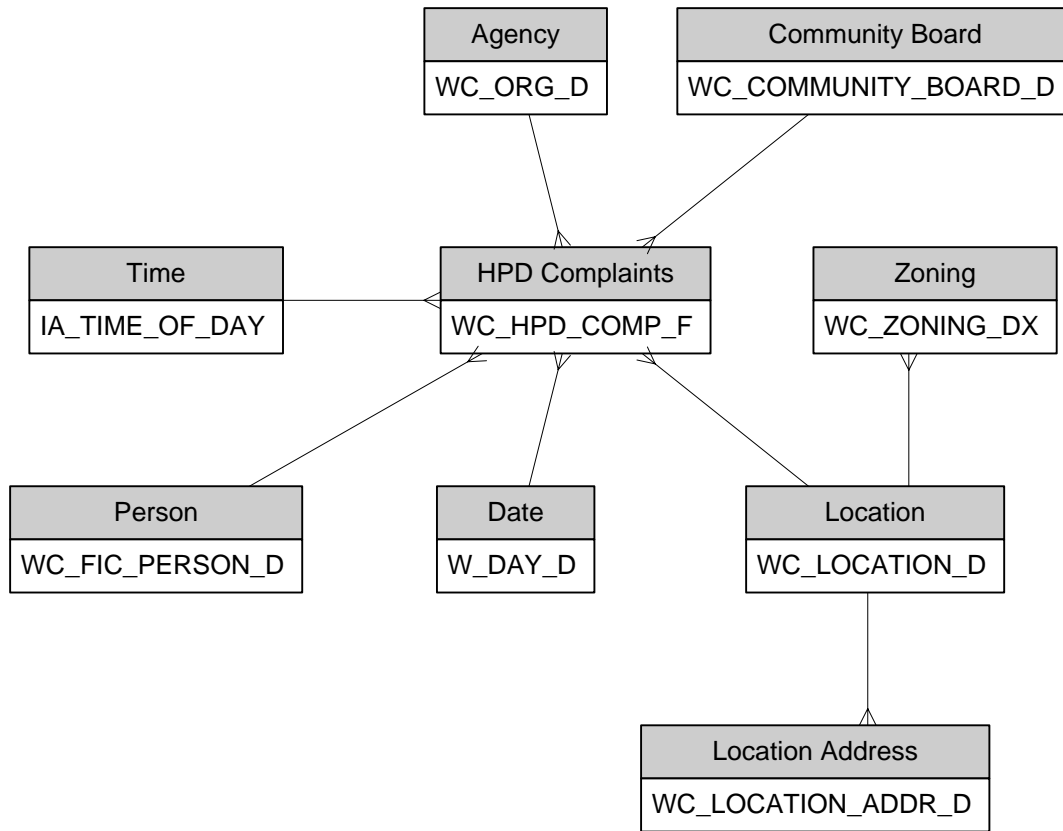
DOF ACRIS – Property Documents



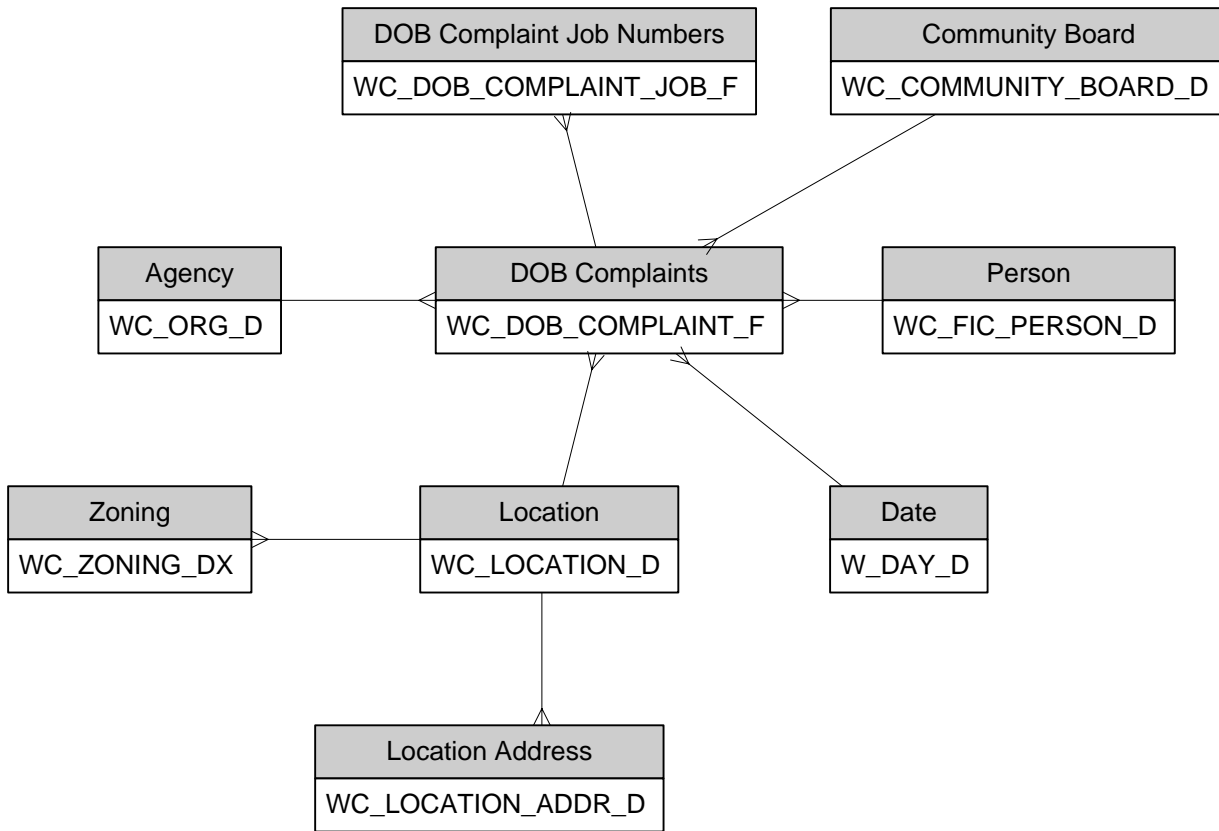
DOF NYCServ – Payments



HPD Complaints



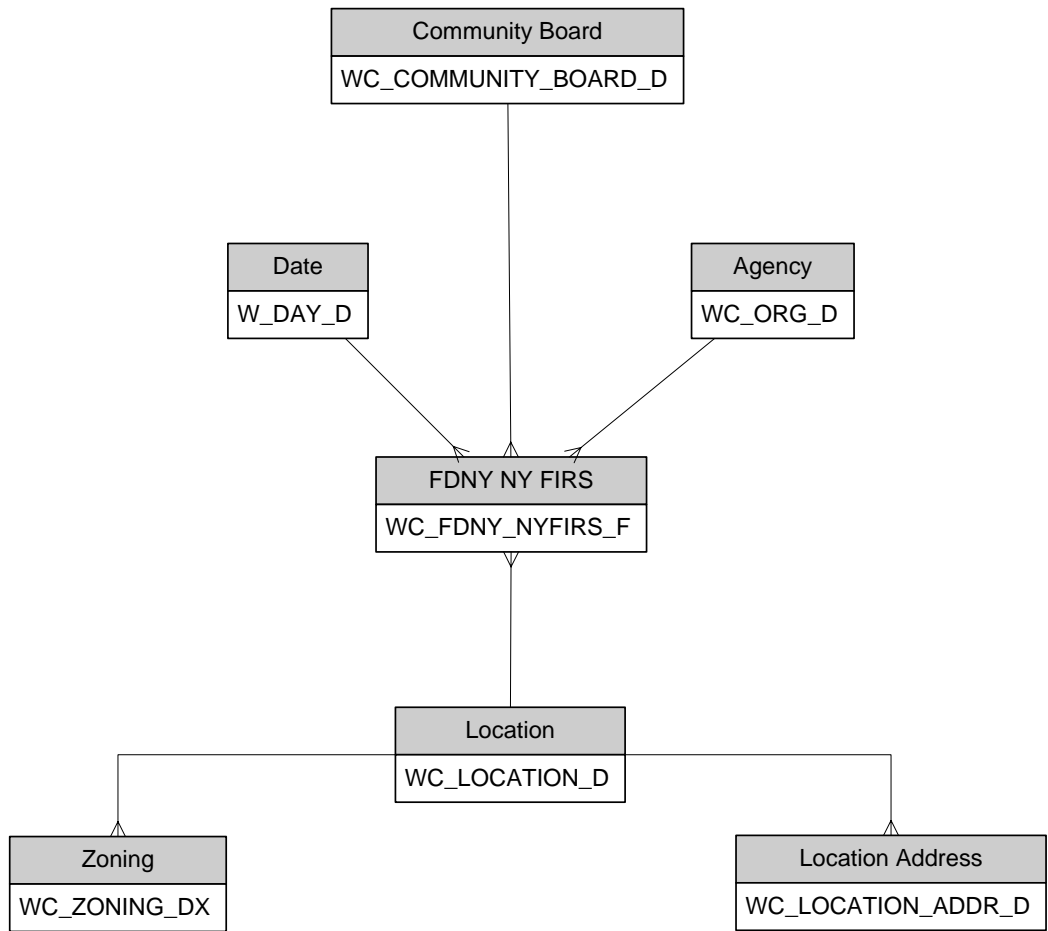
DOB Complaints



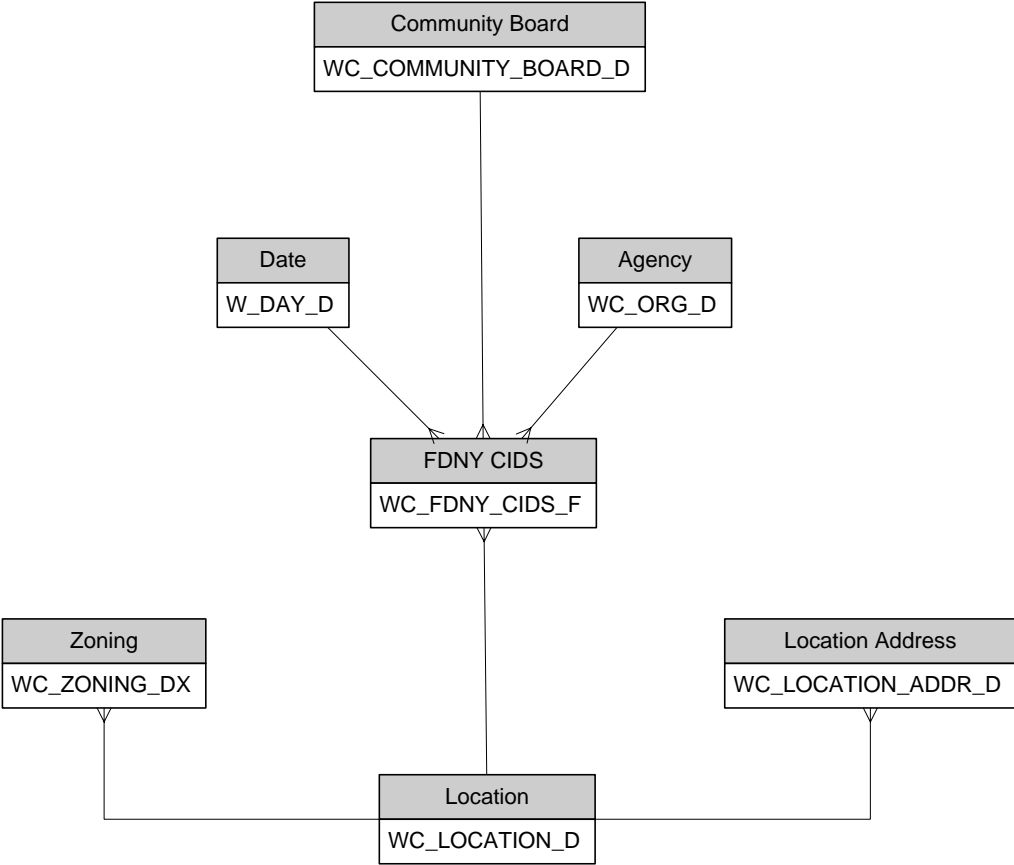
FDNY Investigations: - TBD

R2 SCOPE:

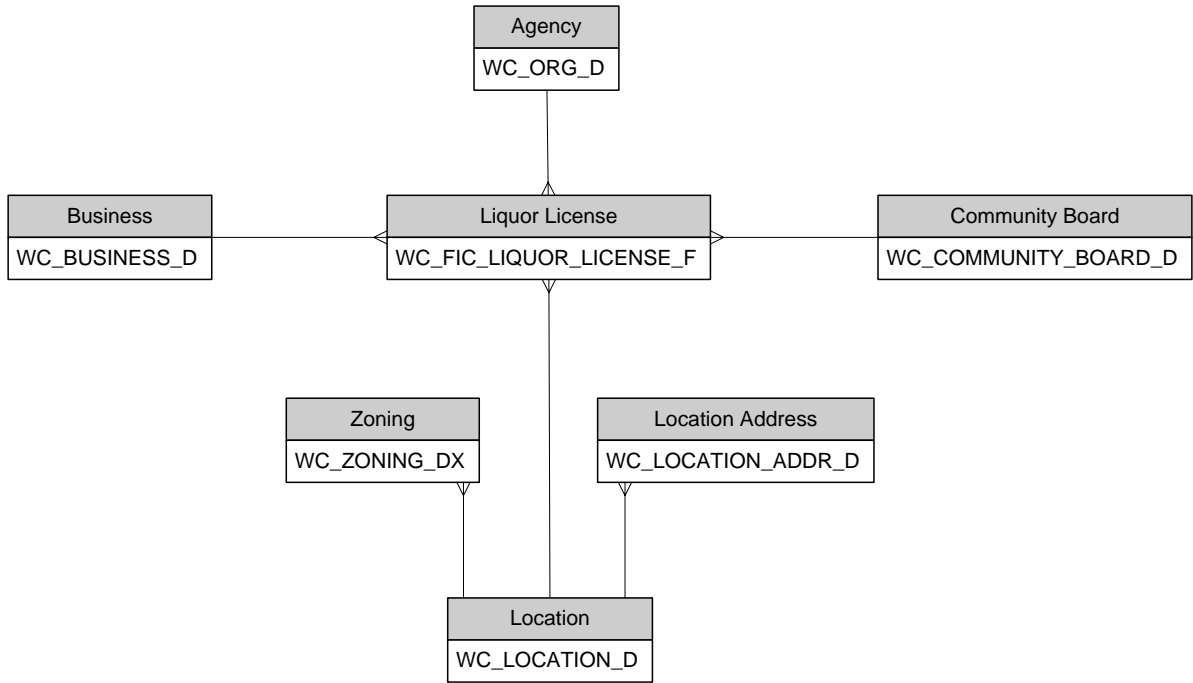
FDNY NYFIRS:



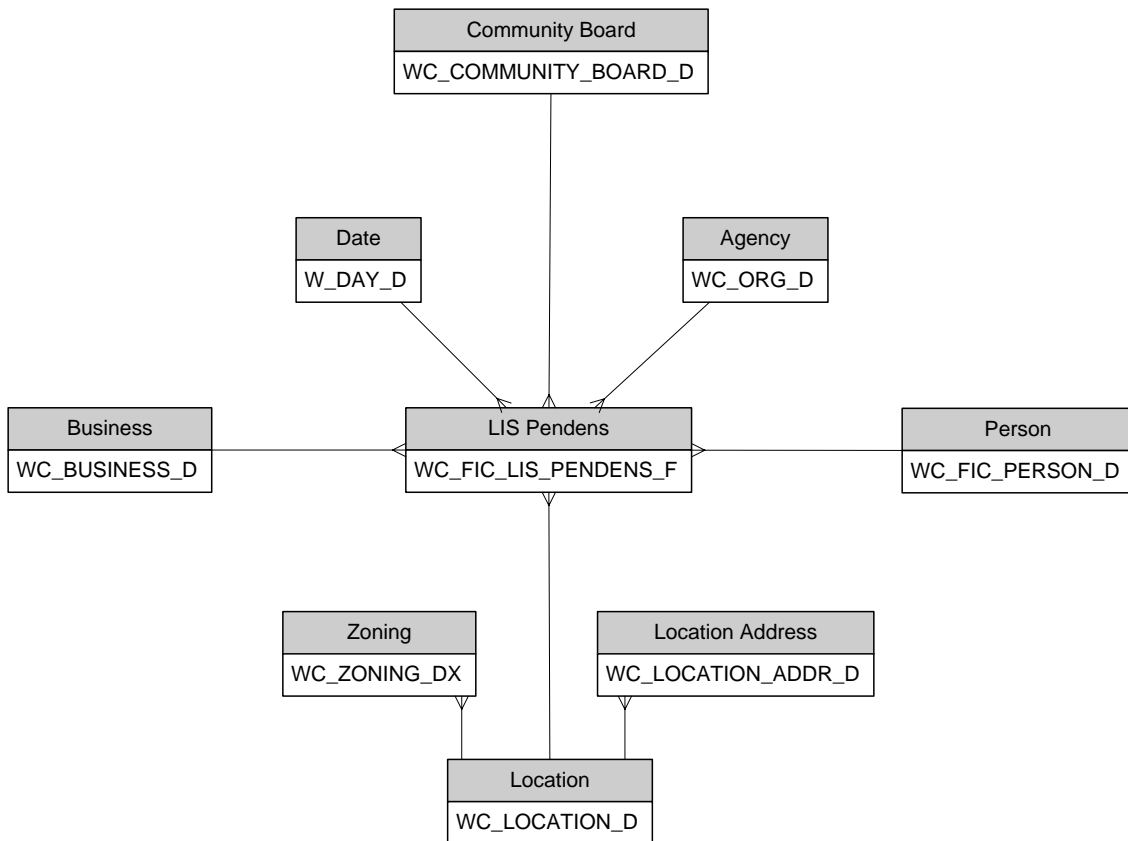
FDNY CIDS:



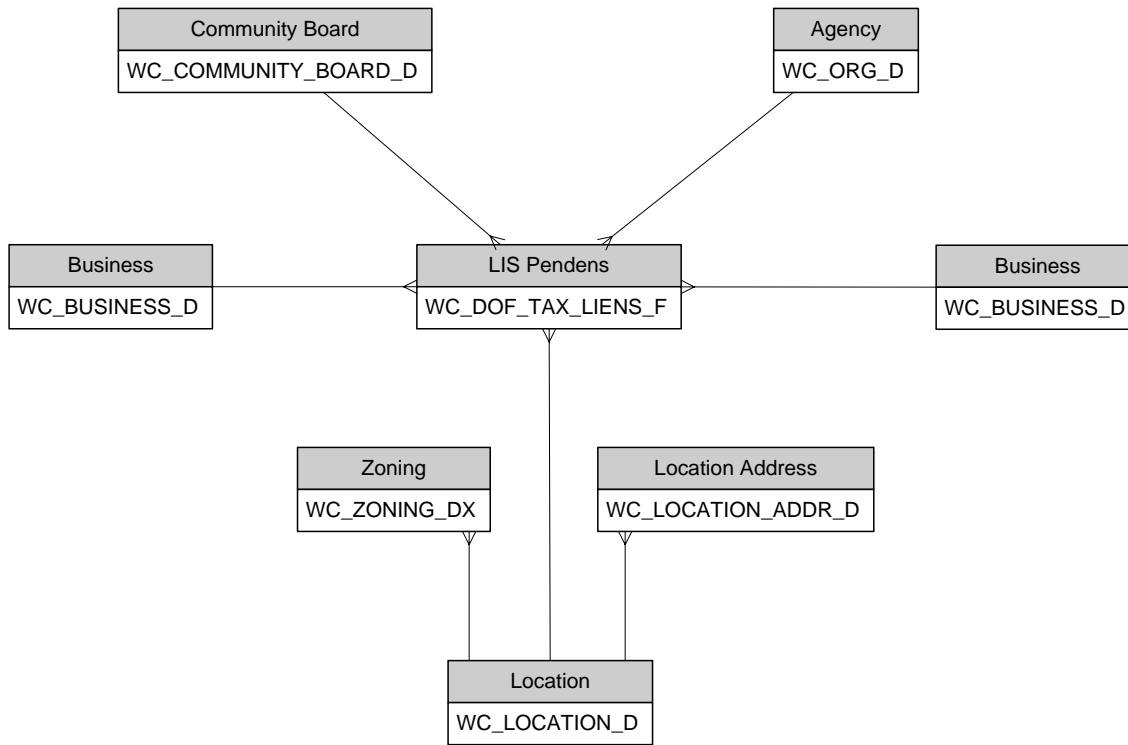
Liquor License



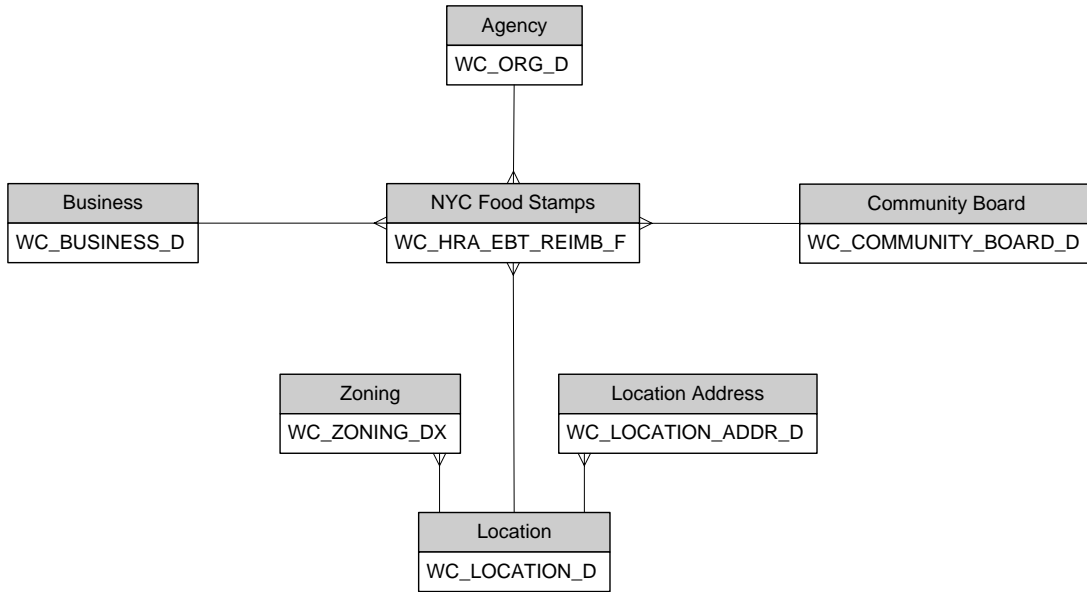
Lis Pendens



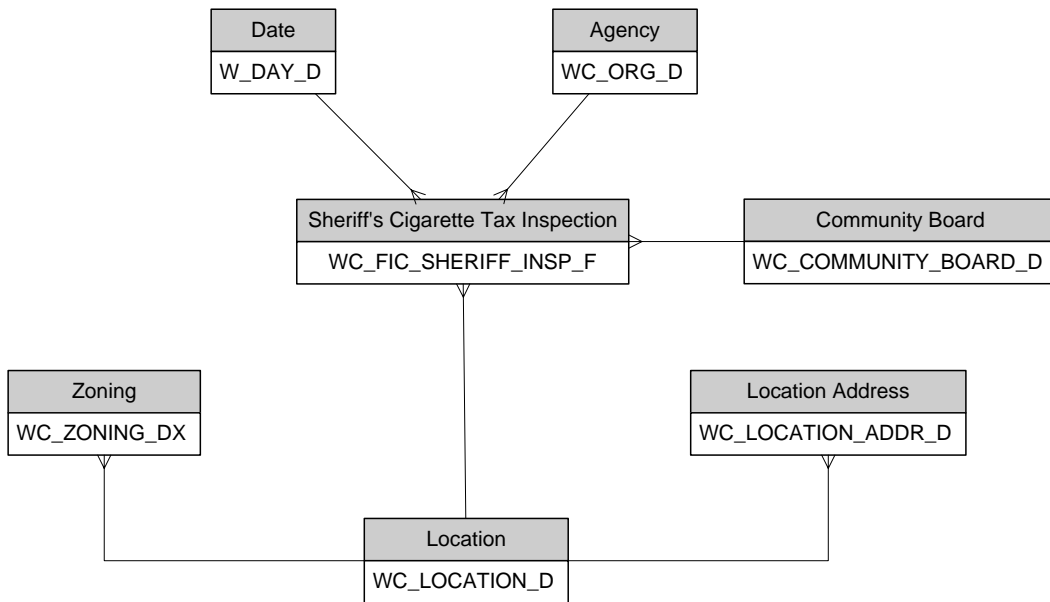
DOF Tax Liens:



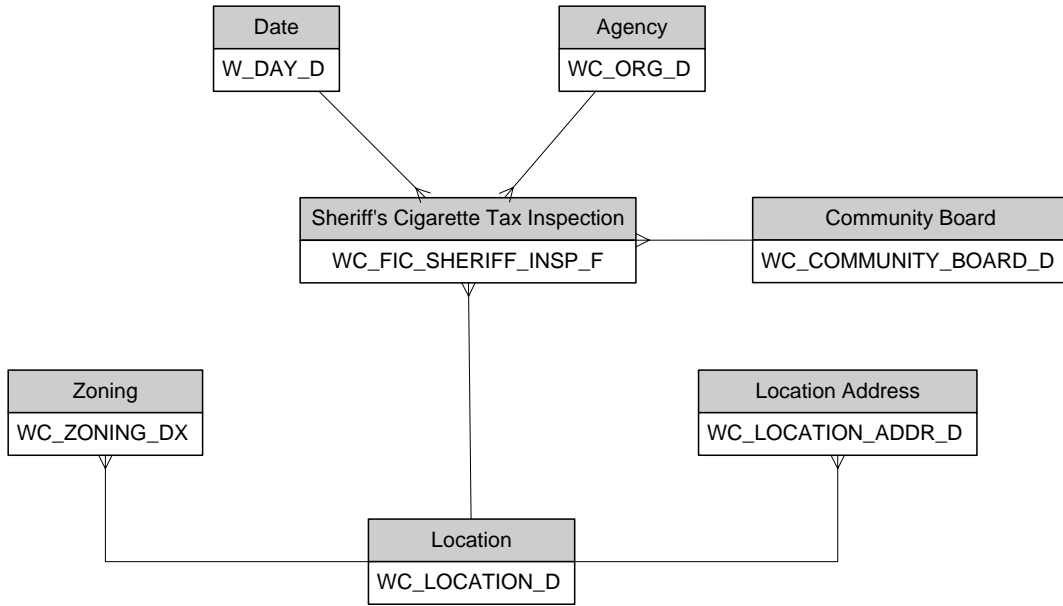
EBT Reimbursement:



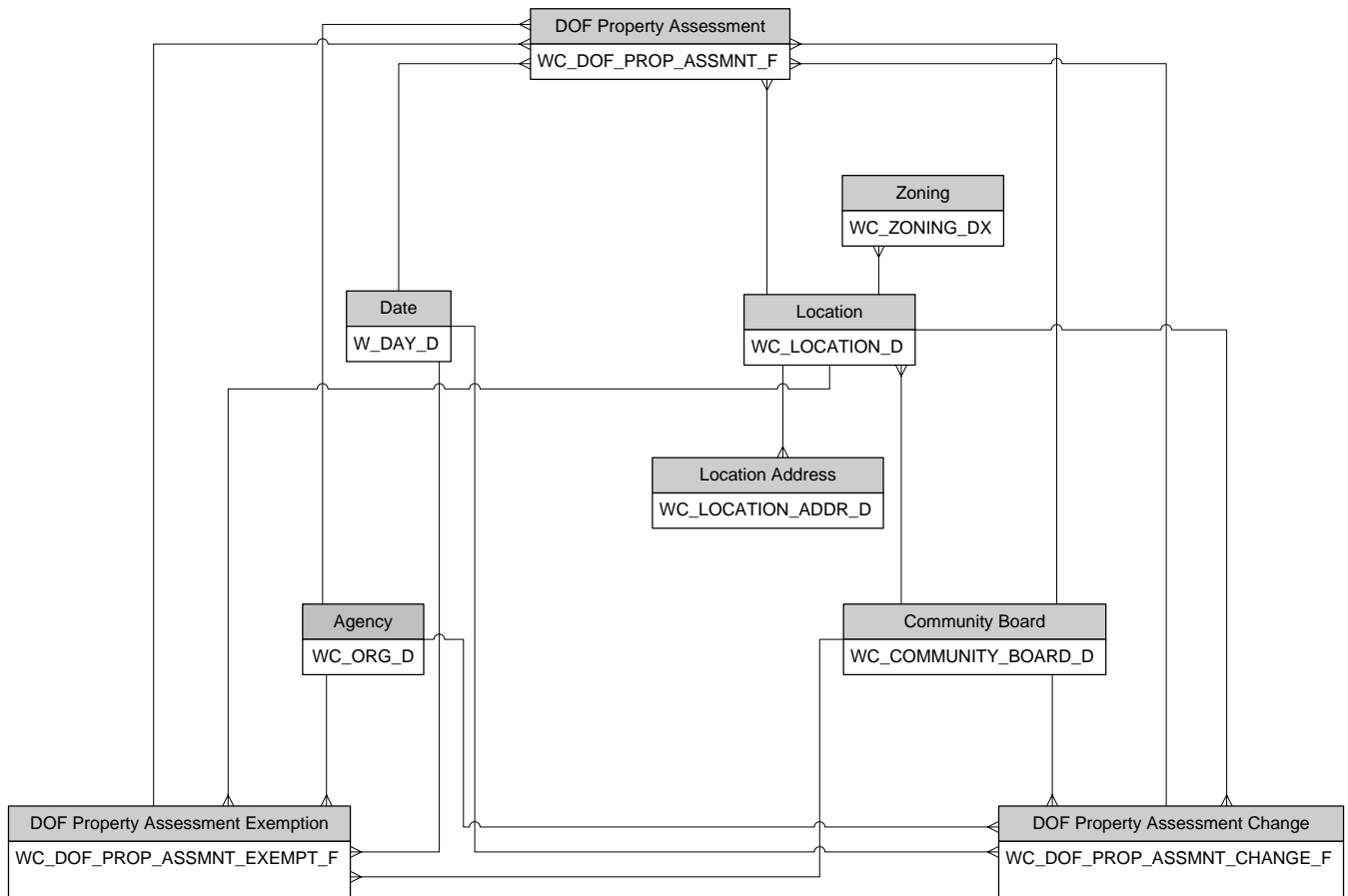
Cigarette Inspection Results:



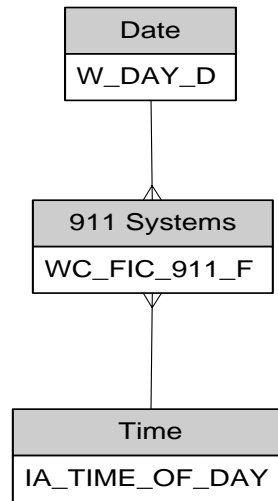
BIC Customer Registration:



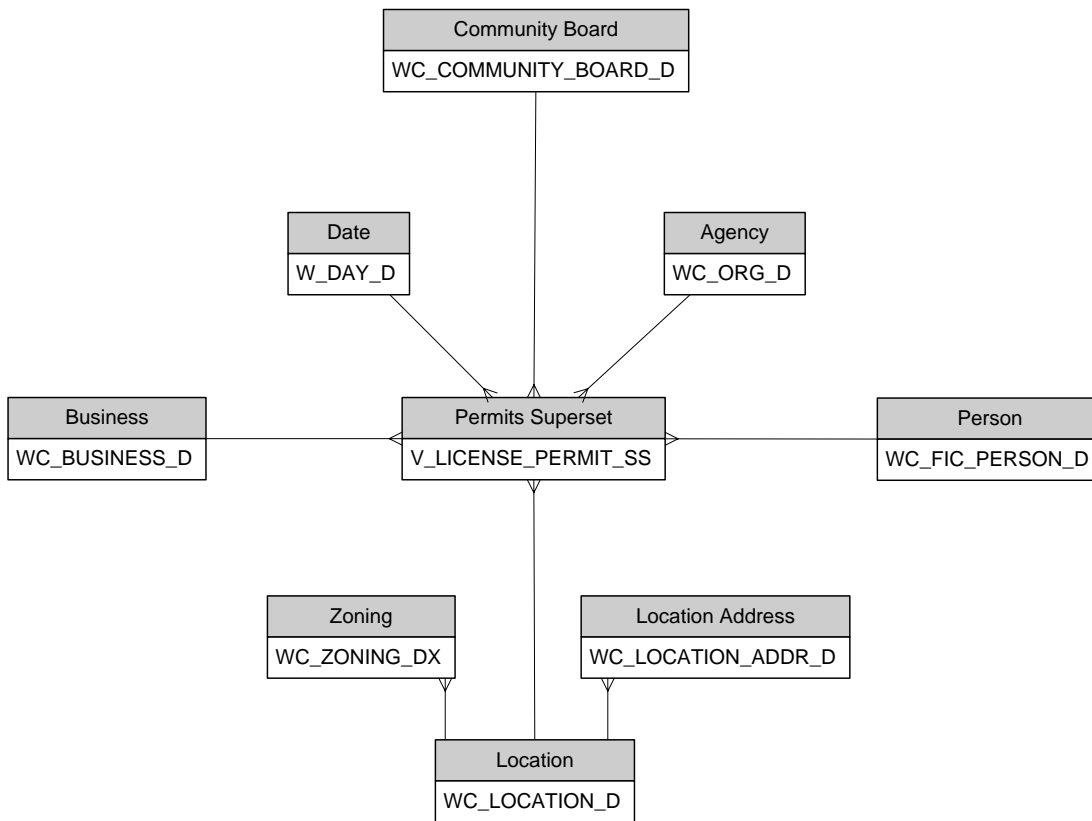
DOF Property Assessment



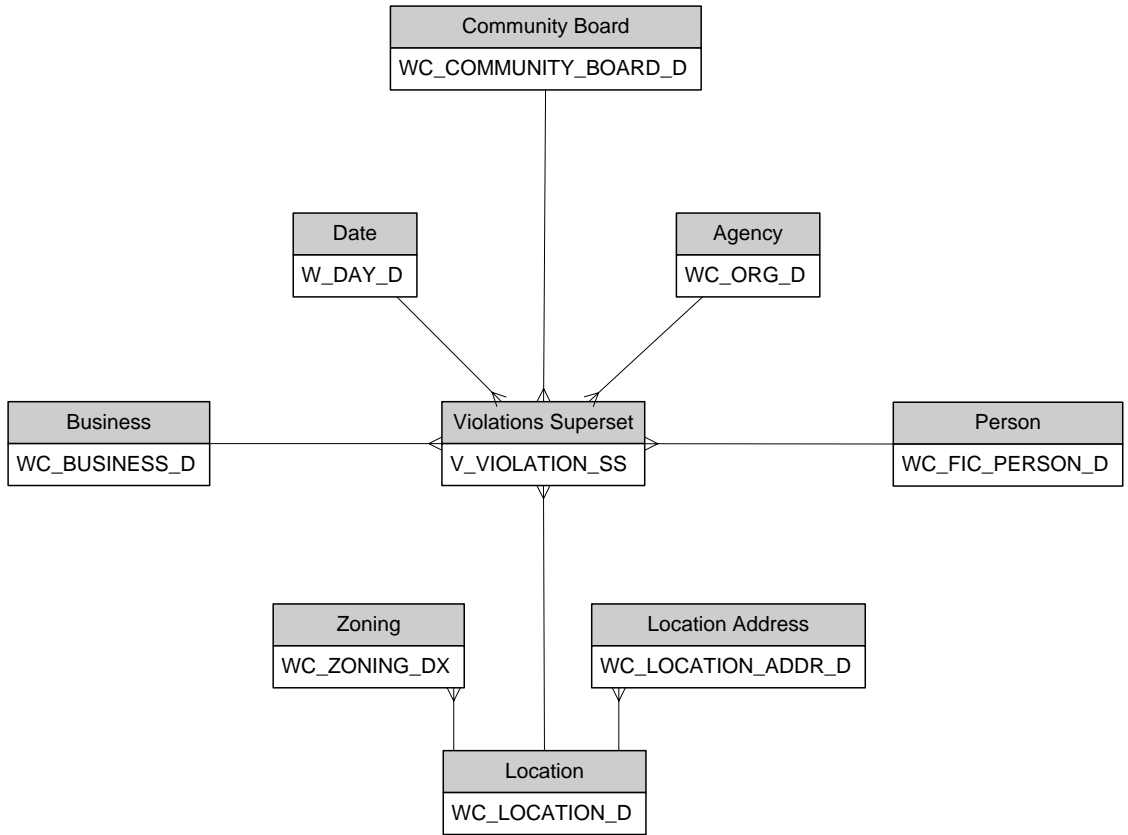
911:

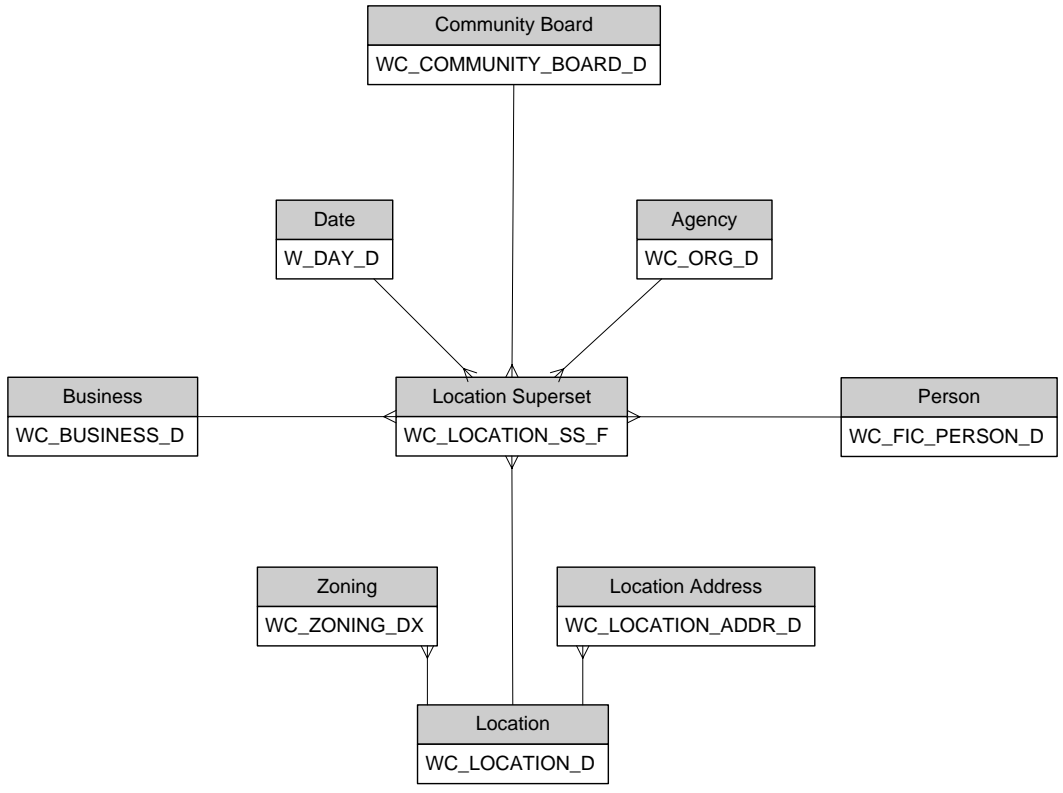


Permit Superset:

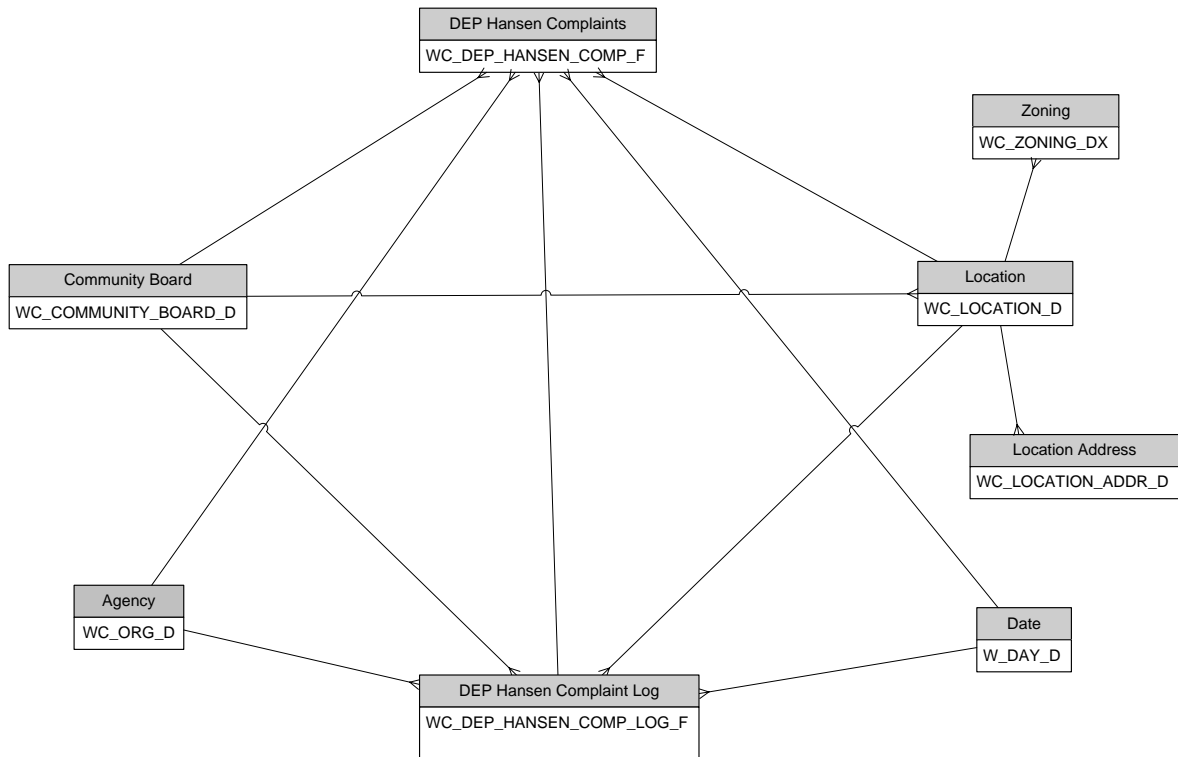


Violation Superset:

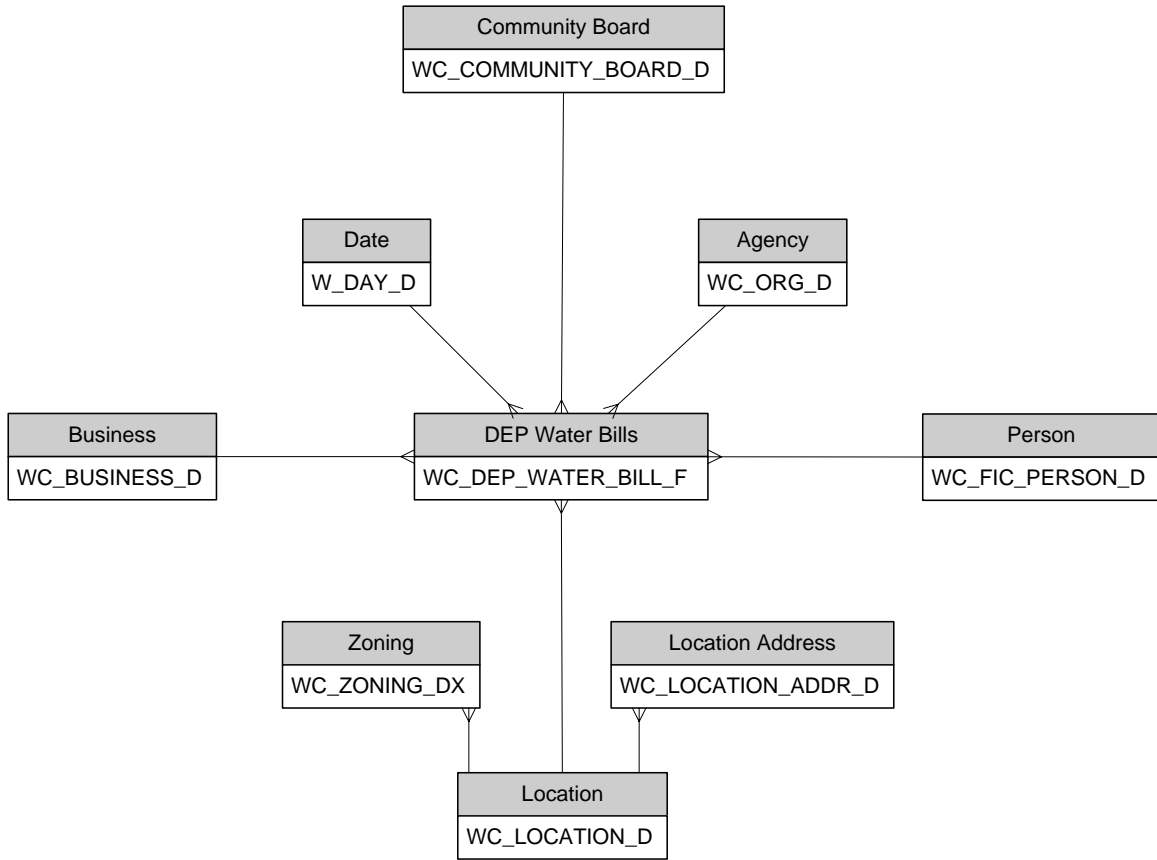




DEP Complaints:



DEP Water Bills:



ENTITIES & ATTRIBUTES

This portion has been withheld under Public Officers Law Section (87)(2)(g)(iii) because it is not a final agency policy or determination.

EXCLUDED ENTITIES

This portion has been withheld under Public Officers Law Section 87(2)(g)(iii) because it is not a final agency policy or determination.

This portion has been withheld under Public Officers Law Section 87(2)(g)(iii) because it is not a final agency policy or determination.

