SmartFabric Release Notes

Version 20241015

Table of Contents

Overview

The purpose of this document is to provide additional information regarding the 20241015 version of LightBox's SmartFabric data. This release contains data content updates, data coverage expansions and bug fixes. If any questions arise, please contact support by emailing support-dmp@lightboxre.com with any questions.

We provide following documentation along with this release note:

1) DataDictionary – For understanding the definition of our data fields.
2) ERD – For understanding the entity relationship between all data layers. This is a helpful tool for enabling connectivity between different tables.
3) SmartFabric_FillRateReport_CountyLevel – For understanding counts and population rate of each field for each of the data layers, grouped by County.
4) SmartFabric_RecordCount – For understanding total records for all our data layers, grouped by States and Counties.
5) SmartFabric_Statistics – For important high-level statistics on this quarter data release.
6) SmartFabric_VintageReport – For understanding latest refresh dates of our data.

New Data and Data Updates

The focus for this release was on content update and new delivery mechanisms. The extensive updates ensure exceptional data currency and coverage across the U.S., while also enhancing attribution accuracy and location precision. Parquet file format is added as a new option for bulk data delivery format. Additionally, SmartFabric is now available in Snowflake marketplace for easy access.

The spreadsheet SmartFabric_Statistics_20241015.xlsx contains important statistics details of this release. This spreadsheet include:

- Overview – a description of each tab like the below, for ease of use.
- Summary Statistics – a tool to generate summary statistics for national, and any state level or county level geographies.
- County Statistics – contains raw data by county that can be used for analysis and filtering.
- Update Counties – a listing of counties with significant updates in this release. ***Please see below on how to use this list for important information.**

How to Use the Update Counties Sheet

This new addition to our SmartFabric Statistics sheet provides county-level information detailing which spatial layers have undergone significant updates in this release. Each is noted with "Yes" or "No",

flag. 'Parcel Update' column also contains a "New" category that represents counties which LightBox did not previously have parcel coverage for and in this release, we do.
There are instances where a layer for one county is not flagged as updated in the Updated Counties sheet, but you still see changes in the data. This is because all data layers in SmartFabric are connected. Changes in one layer can trigger changes in connected layers. Please refer to our methodology for more details.
There may be instances where a county has been updated, but significant changes within the data are not seen. This can occur in counties where the previous version of source data that LightBox had was already recent, and the update ensures we have the most up to date data.

High Level Updates

- Parcels - We refreshed parcel data for 1126 counties in this release.
- Assessments - We refreshed assessor data for 766 counties in this release.
- Ownership data – We updated ownership data in assessments for 2503 counties with the latest sales transactions.
- Addresses – We made county-level local source updates to counties in the states of CA, CO, FL, ID, KY, MD, MN, NM, NV, OH, PA, WA, WY, MA, NH, and TX.
- Building Footprints – We refreshed footprints for buildings in the states of CA, FL, IA, MO, NC, NE, NH, SC, TN, TX, VA, and WI.

Enhancements

- Addition of PARCEL_LID to the Assessment Table

  We added a new field, PARCEL_LID, to the assessment table for SmartFabric Professional and Core packages. This field will enable a direct join from the parcel table to the assessment table, facilitating more efficient data connections. PARCEL_LID will be added to the end of the assessment table's attribution.

- Parquet file format

  We added parquet file format as a new bulk data format option starting from this release. Parquet files are available at the national and state level. We use geoparquet specifications in the creation of the parquet files (Reference: https://geoparquet.org/). Create Table Statements are provided for Snowflake, and DDL-formatted strings are provided for Amazon Athena and PySpark to help you get started with SmartFabric data.

- For address, we removed inconsistencies in house number reporting and converted decimal (where existed) to simple numeric. For example, 5.000 Main St were converted to 5 Main St. This impacted 240 counties.

- SmartFabric and Zoning data are now available in Snowflake. The data listings in Snowflake have the same content as our bulk data offering. Free trial listing can be found at: https://app.snowflake.com/marketplace/listing/GZTSZTDMS2/lightbox-parcels-assessments-buildings-and-addresses-of-us-properties

- SmartFabric and Zoning data are now available as LightBox Feature Service.

Known Issues

- Street Parcels – Our current product design is to keep street parcels which are delivered to us from our sources as is. There is a roadmap item to add a flag field to the parcel data which will help client to identify Right-Of-Way (ROW) parcels with ease. Please see 'Important Upcoming Changes' section for the upcoming schema change.

- Parcel record count drop – NOTTOWAY county, VA has a 32% decrease in parcel record count. This is because that in previous release we used a ID field that kept many sub-property level parcel in the data. This release, we identified the correct ID field for real properties which leads to better match between parcel and assessment records. Thus, the decrease is a result of quality enhancement.

- Assessment record count drop – YAKUTAT county, AK has a significant decrease in assessment record count. (Dropped from 786 records in last release to 81 records in this release.) This is because that the previous assessment records were fabricated from deed by the assessors and were not reliable. Last year the county assessor created the real assessment database. Thus, the decrease is a result of higher quality data that is being provided.

- Address record count drop – For 18 Counties listed below, address point count dropped by 10% as we updated our sources, improved address standardization process and deleted inaccurate data points.

State	FIPS Code	County Name
CT	9003	Hartford
CT	9001	Fairfield
CT	9009	New Haven
CT	9007	Middlesex
CT	9011	New London

CT	9013	Tolland
CT	9015	Windham
CT	9005	Litchfield
KY	21149	McLean
KY	21055	Caldwell
MN	27111	Otter Tail
MS	28129	Smith
NE	31073	Gosper
OH	39005	Ashland
OH	39075	Holmes
TX	48435	Sutton
TX	48287	Lee
WV	54097	Upshur

- Building footprint gaps – We have developed a system to be able to target specific neighborhoods where we have signals that building footprints exist but are not currently in our database. Each quarter we are targeting these specific areas and making a significant addition to our building footprints data. Counties that these gaps have been identified and updated will be found in the Updated Counties tab of the SmartFabric Statistics sheet.

- Building Footprint Height Accuracy – We are constantly refining our process for accurately modeling buildings heights. For this release, we have updated building heights for 109 counties.

- Overlapping parcels were reported in San Francisco, CA. We are currently working with the county's GIS department for a solution to identify the correct parcels and remove the redundant ones.

Updating and Making Use of the 20241015 Release

For existing clients of LightBox's SmartFabric there is the question of how you best make use of the new data. In any update of the data there are numerous things that can change, below are a few examples:

- Parcel/Building polygon changes - new parcels/buildings, deleted parcels/buildings, reshaped parcels/buildings.
- Building grouping changes – the relationship between buildings can change, for example what once was a 2-polygon region object with one address may become 2 single polygon regions each with its own address.
- Primary Address changes – an address may see a significant change as we improve the addressing (from no address to an address, or 10 Smith St to 12 Smith St), or may see an improvement in the hygiene of the address (from a ZIP5 to a full ZIP+4).

- Secondary Address changes – the numbers and details of secondary address and secondary unit information may change (through addition, deletion and modification).
- Unique Identifiers – all unique IDs (both LID and ObjectID fields) are subject to change.
- Removal of PO Boxes coming from the USPS ZIP+4 data lead to a decrease in secondary address numbers in certain locations but an increase in data quality.

There are several strategies that you can use. In each, an understanding of LightBox's LID structure and relationships is key. Please see our LID FAQ and Relationship FAQ for more information on each of these.

Comprehensive Use of New Data

Upload the new data into your production or operational environment and discontinue using the previous version of data. This is a good strategy if you do not perform any value add or improvements on the data (i.e. you have no work that you want to preserve).

Our general recommendation is to always make use of the entire new release of data, or at least an entire state. For example, if Kansas' coverage is significantly improved, drop the prior version's Kansas data, and upload the new data for Kansas.

You may wish to assess the significance of an update versus the work involved. For example, new coverage for any given state may consist of mostly rural areas. If these counties are of less importance for your business, you may want to skip utilizing new data for various states for the current release. It is also important to check the updated counties list, though, as updates to counties with high population counts may tilt your decision in favor of updating an entire state.

Uploading Just "The Deltas"

We hear and use the term "The Deltas" to refer to any change or additions in coverage. We do not recommend blending coverage between releases or to only update partial counties. It is important to note that there will be changes in a county even though that are not listed as such the Updated Counties sheet. This occurs specifically in the addresses and relationship files as in addition to the local sourcing completed and noted in our SmartFabric Statistics sheet, each quarter we bring in nationwide address sources and run all addresses through our address standardization. We also re-establish all relationships in our 6 relation files to reflect updates and enhancements to our matching logic.

It is possible to use the "Updated Counties" tab in the "SmartFabric_Statistics" sheet to identify new counties of coverage. Extract only the features with the County FIPS code for a county with new coverage. Then insert only those records into your tables from the prior release. For example, if you

want to add coverage for New Castle, Delaware you can add in the entire DE*.* files. If you want to add in coverage for the city of Fort Smith in AR, then you then you can select records with FIPS 05137 and insert those records into your prior release.

Next and Future Releases

Our next release will be the January 2025 vintage data. In this next release we will continue to make advances on our parcel and building footprint coverage. We are also planning on important changes mentioned below:

- **Planned FIPS Code Change in Connecticut**
  The U.S. Census has defined a new geographic "Planning Region" FIPS for the State of Connecticut and will use these new codes as County FIPS in future reporting. LightBox has reviewed this change and determined that we should update to the new Planning Region FIPS for Connecticut to benefit our clients.
  For more information on the Connecticut FIPS changes, please review the following links:
  Final Changes to County Equivalents in Connecticut (census.gov)
  Federal Register: Change to County-Equivalents in the State of Connecticut

- **Planned Schema Changes - Introduction of ROW_FLAG to the Parcel Table**
  We are introducing a new field named ROW_FLAG (subject to change) in the parcel table of our database schema. This field will serve as a flag to indicate if a parcel is a non-real property, carrying a category code for types like Right-of-Way. The purpose of this field is to help clients easily remove non-real properties from their processes, particularly for those who consider non-real properties as noise in their use cases.

These changes may impact your current ETL process. Please reach out to your Customer Success Manager or our support team at support-dmp@lightboxre.com with any questions or concerns.