## Position and Request

AIRA will be seeking supplemental funding from CDC under our current Cooperative Agreement structure to support the community-wide expansion of an address cleansing and geocoding service. The selected service has been proven to be a valuable and cost-effective resource through extensive pilot testing, and supplemental funding would allow AIRA to expand access to this service as a no-cost member benefit to all member IIS throughout the country. Although much detail follows to substantiate these key points, the outcomes can be summarized as follows:

- Six IIS programs representing multiple IIS products (Envision, STC, WIR, and awardee-developed) were successful in connecting with the SmartyStreets service.
- Five of these sites tested batch data validation for data at rest, user-interface API real time standardization/validation, and HL7 real-time validation. All methods were successful.
- On average, data available demonstrated that deliverability increased from 71% to 88%.
- The proportion of addresses determined to be undeliverable, representing unsuccessful mailing costs avoided, averaged 12%.
- The standardization, validation and geocoding processes available through SmartyStreets is expected to improve the accuracy and completeness of incoming and existing data, while also improving core processes such as deduplication.

## Background and Context

In late 2015, the Joint Development (JD) Advisory Workgroup, a group made up of representative entities from across the IIS program and vendor community, held an in-person meeting in Atlanta, Georgia. The goal of the in-person meeting was to validate a project selection process, select an initial JD project, and define the preliminary scope and project plan for the effort. The project review process at the in-person meeting led to the selection of an address cleansing project focused on selecting, validating, and providing access to an address cleansing and geocoding service that could provide data quality benefits to IIS across the community. A consultant was hired to conduct a lengthy evaluation process of several candidate services, considering services offered, technical support options, licensing costs, ease of use, detail included with address metadata, and other customers/clientele that use the product. Results of the evaluation were presented to the JDI Advisory Workgroup who, in conjunction with AIRA staff, selected the SmartyStreets[1] service for pilot testing.

SmartyStreets provides address standardization, validation, and geocoding for addresses submitted through a manual or automated batch process, through a real-time user interface API (Application program interface), or through real-time data exchange. Standardization, validation, and geocoding services are distinct in their functions:

- Address standardization is the process of making an address uniform, or changing addresses to adhere to United States Postal Service (USPS) standards for address formatting.
- Address validation is the process of checking this standardized mailing address against an authoritative database to see if that address is a deliverable address (also called verification).

[1] https://smartystreets.com/

- Geocoding provides the GPS coordinates (or the latitude and longitude) of a physical address. All three services are included in the basic SmartyStreets product.

Note: The SmartyStreets service does not include support for the USPS Address Correction Service (ACS) that manages the movement of individuals by tracking mail holds and forwarding.

## Pilot Design and Participant Selection

Following the selection of the SmartyStreets service, AIRA purchased five million record validations from SmartyStreets to evaluate the service and developed a pilot plan and evaluation criteria to assess the success of address validation both quantitatively and qualitatively. In an effort to ensure applicability for all age groups, results were stratified by three age groups:

- 24-35 months (to mirror AFIX childhood evaluations)
- 13-18 years (to mirror AFIX adolescent evaluations)
- 65-70 years (to focus on an adult age range with significant vaccine activity)

AIRA then reached out to the awardees engaged in the JDI Advisory Workgroup to select pilot participants. Pilot jurisdictions were evaluated to ensure a representative mix of diverse platforms and sizes. Four jurisdictions **(bolded below)** were selected as formal pilot sites to evaluate the batch component, and two additional jurisdictions were added later to test select aspects of connectivity:

- **Washington (using the STC platform) – Batch**
- **Florida (using an awardee-developed platform) - Batch**
- **Delaware (in partnership with the Envision product) – Batch and User Interface API**
- **New York State (using the Wisconsin Immunization Registry [WIR] platform) - Batch**
- Michigan (using an awardee-developed platform) – User Interface API and Real-time HL7
- Envision's test product (set to Kansas geographic state specifications) – User Interface API

## Batch Pilot Efforts and Results

The AIRA team worked with the four batch pilot sites to get them connected with the SmartyStreets service. All four sites were ultimately able to connect with the service. Initial connectivity was achieved through a brief joint call/webinar between the participating states and AIRA Staff. On average, setting up the initial connection reportedly took approximately 30 minutes for Washington, Florida, and Delaware. One jurisdiction, New York State, had considerable difficulty connecting to the SmartyStreets service due to firewall and configuration issues, but was ultimately able to achieve a successful connection. (This State's experience is detailed in the section on "Pilot Limitations" below.)

The pilot states met regularly throughout the pilot to share experiences/lessons learned that would inform guidance for future implementations. They contacted SmartyStreets technical assistance team as needed to interpret/validate results and/or to troubleshoot connectivity issues. The three states who were able to connect and standardize/validate data all used the service for batch processing, recognizing that cleaning existing data (data at rest) would provide the most immediate value for a very moderate level of effort.

## Quantitative Analysis

Throughout the five week pilot, 835,539 addresses were standardized, validated, and geocoded using the service. Florida and Delaware were able to extract data in a batch format that met our previously mentioned age bands. Washington was restricted by their existing health promotion mailing age cohorts, so grouped their address validations into 24-35 months and 0-6 years. New York State was not able to provide data due to their delayed connectivity issues. Due to the small number of pilots and slightly varied approaches, qualitative results will be shared in a case study format to best highlight results, with some limited data combined across Delaware and Florida.

Florida: 48% of Florida addresses were modified through the SmartyStreets standardization and validation process through fixed abbreviations, corrected spelling, modified zip codes, etc., resulting in an increase of deliverability from 70% to 89%. This increase was greatest among records of 24-35 month olds, in which the deliverability increased from 73% to 92% on average. These corrections have the potential to improve the cost-effectiveness of mailings for reminder recall, improve the completeness and accuracy of IIS data, improve ability to perform geographic-based coverage assessments, and also have broader program benefits through improvements to data quality via better match/deduplication rates within the IIS. Also important, an average of 12% of addresses were undeliverable following SmartyStreets standardization. These undeliverable addresses represent cost savings in unsuccessful mailing costs that are avoided, and ultimately increase the proportion of known good addresses in the IIS database. Had these data not been standardized and validated, mailings would have taken place to an additional 106,041 patients out of 500,000, wasting resources and creating inefficiencies within and beyond Public Health. With address corrections, over 48,000 of those records were corrected to reach their intended targets, while another 57,000 were known to be bad addresses and not mailed.

| Florida Results (update) | Pre-SmartyStreets Deliverability | Post-SmartyStreets Deliverability |
|---|---|---|
| **Yes** | 70% | 89% |
| **Maybe** | 9% | 0% |
| **No** | 21% | 11% |

Delaware: In terms of pre- and post-deliverability, Delaware's results were very consistent with Florida's (see table below). However, significantly fewer of Delaware's addresses were modified through SmartyStreets; only 16% were ultimately changed. This may be an indication of comparing a large state with a smaller state geographically, or with a smaller state population overall, resulting in fewer address issues. Although there were fewer modified records, the end result regarding deliverability was fairly stable in Delaware as compared to Florida, with deliverability increasing from 75% to 87%. In Delaware's data, approximately 21,184 records were undeliverable pre-SmartyStreets corrections; following standardization and validation, 16,218 were determined undeliverable, resulting in cost savings in unsuccessful mailings avoided, while nearly 5000 of the full mailing of almost 126,000 were now deliverable following cleansing.

| Delaware Results (update) | Pre-SmartyStreets Deliverability | Post-SmartyStreets Deliverability |
|---|---|---|
| **Yes** | 75% | 87% |
| **Maybe** | 8% | 0% |
| **No** | 17% | 13% |

Delaware and Florida Combined: When combined, Delaware and Florida's results highlight what specific issues rise to the surface in the validation and standardization process. The following table contains an excerpted list of the most-cited corrections that were made across their combined records:

| Abbreviation | Footnote | Description | Count |
|---|---|---|---|
| **N#** | Fixed abbreviations | The delivery address was standardized. For example, if STREET was in the delivery address, SmartyStreets will return ST as its standard spelling. | 214,126 |
| **F#** | Address not found | The address, exactly as submitted, could not be found in the city, state, or ZIP Code provided. | 42,089 |
| **H#** | Missing secondary number | ZIP+4 information indicates that this address is a building. The address as submitted does not contain a secondary (apartment, suite, etc.) number. | 31,722 |
| **L#** | Changed address component | An address component (i.e., directional or suffix only) was added, changed, or deleted in order to achieve a match. | 26,803 |
| **M#** | Fixed street spelling | The spelling of the street name was changed in order to achieve a match. | 26,130 |
| **A#** | Corrected ZIP Code | The address was found to have a different 5-digit ZIP Code than given in the submitted list. The correct ZIP Code is shown in the ZIP Code field. | 22,095 |
| **V#** | Unverifiable city/state | The city and state in the submitted address could not be verified as corresponding to the given 5-digit ZIP Code. | 7,634 |
| **B#** | Fixed city/state spelling | The spelling of the city name and/or state abbreviation in the submitted address was found to be different than the standard spelling. The standard spelling of the city name and state abbreviation is shown in the City and State fields. | 6,610 |
| **U#** | Unofficial post office name | The city or post office name in the submitted address is not recognized by the United States Postal Service as an official last line name (preferred city name), and is not acceptable as an alternate name. | 4,584 |
| **I#** | Insufficient/in-correct address data | More than one ZIP+4 Code was found to satisfy the address as submitted. | 4,376 |

Washington: Washington State opted to integrate the SmartyStreets process with their Child Profile mailing process. Based on this decision, they ran a number of smaller weekly batches of data, rolled those up into summary results, and conducted their own analysis and evaluation on the end resulting data. On the whole, Washington's childhood addresses appeared to be significantly cleaner than Florida's or Delaware's addresses. That being said, an additional 4% of Washington addresses were determined to be undeliverable in the 24-35 month age band, and this proportion increased slightly to 4.3% when looking at all 0-6 year old addresses. This again represents cost savings and improved data quality in avoiding mailings to undeliverable addresses. Based on their internal process, AIRA was not able to determine exactly how many addresses were standardized and/or corrected in the Washington testing process.

### Qualitative Analysis

The three pilot sites that were able to run data and receive results were asked to estimate the time, resources, and level of satisfaction of working with the SmartyStreets service (note that New York State was not included in the analysis due to timing issues). Although initial set-up of the service was consistently very short (~30 minutes per site), person hours required to fully implement the service (including planning for extracting the data, storing returned results, etc.) ranged from 1-9 hours for batch sites (the variation may be an indicator of larger teams and more lengthy planning processes). The Delaware site reported that the process of development for setting up the real-time API took approximately 60 hours for configuring the use of the API connection for real-time cleansing via the user interface. Person hours providing information back to AIRA for evaluation and development of guidance ranged from 1-4 hours across pilot sites. This consistently low number suggests that reporting use metrics back to AIRA should represent a relatively light lift for sites following community-wide roll-out. All respondents said that implementing the service was easy or very easy, and only one of the sites reported needing additional support from SmartyStreets or AIRA to implement the service ("minimal Q&A via chat" on the SmartyStreets site).

Level of confidence in the data returned from Smarty Streets was rated as high by all of the pilots. Washington State compared the SmartyStreets results against those provided by their mailhouse vendor and investigated differences between how the two services handled specific addresses. These addresses were reviewed with SmartyStreets who further validated the addresses against the USPS Zip+4 database and determined that SmartyStreets handled the addresses correctly. As a result, Washington is working with their mailhouse vendor to determine why the results provided are not consistent with the SmartyStreets and USPS results.

All pilot participants said they would recommend this service to other IIS.

## Pilot Limitations

It is important to note that all of these pilot connections took place using only limited technical resources and information; when this service is rolled out to the full community, it will be accompanied by Implementation Guidance to assist IIS in fully implementing the service.

One site, New York State (NYS), experienced significant internal issues getting connected with SmartyStreets. NYS's system sits behind a commerce site that introduces additional firewalls and other

security measures into the data flow. NYSIIS's technical lead spent over a month working with AIRA staff, SmartyStreets technicians and NYS central IT resources to resolve firewall issues that appeared to hinder their connection to the service. Initially, the NYSIIS technical lead was unable to find the right internal resources with the right knowledge to resolve the connectivity issues, and was forced to abandon the effort to focus on other priorities. In the meantime, the NYSIIS central IT team was able to identify and resolve the issues on their own through configuration changes internally. NYS was then able to connect successfully with the SmartyStreets service in a matter of minutes. It is challenging to know the extent to which this issue might be replicated with other sites, but a key strategy will likely be close collaboration with internal IT resources.

Due to timing limitations, none of the pilot sites have yet imported the validated and standardized data back into their databases. This will require a greater level of effort and the establishment of business rules that was not possible within the time allotted for the pilot. That being said, all pilot sites recognize this is the next step in the process if/when the SmartyStreets service is launched as an ongoing and stable service; at that point, sites are prepared to develop business rules and modify workflows to realize the full value and benefits of the service.

Finally, the pilot focused on validating only patient addresses, but did not explore the benefits of validating and geocoding other provider and stakeholder addresses contained in the IIS database. There is likely significant business benefit to geocoding both patient and provider addresses in terms of mapping and data visualization opportunities. One could envision significant data visualization opportunities through mapping the relationship between provider location and up to date rates, provider type and AFIX results, participation by size of provider practice, and an unlimited number of other factors. All of these opportunities can be explored with the approval of an unlimited license with SmartyStreets.

## Ongoing Pilot Activities

Work continues across the community to test other forms of connectivity with SmartyStreets. Envision continues to work with their Kansas implementation to integrate the real-time user interface standardization and validation API (application program interface) component into their development system. Michigan has integrated the real-time data exchange API into their development system for the Michigan Care Improvement Registry (MICR). They are in the process of testing the standardization and validation of HL7 incoming real-time messages. In both instances, the jurisdictions have validated that the real-time services are functional, beneficial, and do not adversely affect performance time in the development environments. STC has shared the success of their pilot activities with their iWeb Consortium, in the hopes of beginning a joint development process to integrate returned cleansed data into their database. As a late-connecting pilot site, New York State is continuing to test the functionality of SmartyStreets via the batch process, and will share its results with other WIR participants.

Enthusiasm for this service is high among IIS; AIRA has already been contacted by a number of sites that are hoping to integrate the service into their workflows as soon as possible.

To quote one pilot site in forming their long-term plans with this service:

*"We would suggest adding SmartyStreets as part of the day-to-day system processes versus just using it when a mailing comes up. If it's part of the system on the front end, addresses would be standardized on a regular basis and would help with consistency and possibly deduplications. It would also be beneficial to add SmartyStreets to our birth certificate process prior to loading that data. This could also prevent duplicate records and be a way to clean up data on the front end."*

## Project Budget

To fund the pilot, AIRA purchased a finite number of record validations (5 million). To expand the project, AIRA has negotiated with SmartyStreets for an off-label license of SmartyStreets that will offer unlimited address validations to the full IIS community. The cost of the service is $50,000 per year, and the license would be available to all IIS across the US and US Territories to standardize, validate, and geocode their data. The AIRA staff, in conjunction with the pilot sites, believes that this is a cost-effective service that leverages economies of scale, and will positively impact the quality of IIS data substantially.

This request for CDC supplemental funding includes the cost of the unlimited license for three years (note the discount for purchasing a multi-year license):

| | Year 1 | Year 2 | Year 3 | Total |
|---|---|---|---|---|
| **SmartyStreets Service License with Unlimited Validations/Geocodes** | $50,000 | $50,000 | $50,000 | $150,000 |
| **15% discount for purchasing service for multiple years** | | | | -$22,500 |
| **Actual purchase amount** | | | | $127,500 |

Administrative management of the service, including dispensing and tracking authorization keys for the use of the service, and support for implementation would be provided by AIRA as funded under our existing Cooperative Agreements with CDC.

## Project Timeline

Post-pilot activities will continue for the coming months, as will continued development of the Implementation Guidance that will support community roll-out of the SmartyStreets service. The Implementation Guidance is planned to be circulated for community review in April, with anticipated completion by June, 2017.

The community roll-out would ostensibly take place following CDC's approval of this supplemental request. AIRA is hoping for supplemental funding from CDC to ensure widespread access to and success for this project. The success of the pilot has generated significant interest in leveraging the service across the IIS community, so funding would allow IIS to begin benefiting from the SmartyStreets service as soon as mid-2017.

## Conclusion

In the words of another pilot site:

> *"The way we incorporated SmartyStreets into our workflow was quick and easy. It required a small amount of additional time each week, but nothing unmanageable."*

As stated at the start of this document, after thorough evaluation, the AIRA team feels confident that the SmartyStreets product is a low-cost, high-value service that will allow most IIS to standardize, validate, and geocode their data. The following take-home points stand out as a result of thorough evaluation:

- Six IIS programs representing multiple IIS products (Envision, STC, WIR, and awardee-developed) were successful in connecting with SmartyStreets.
- Five of these sites tested batch data validation for data at rest, user-interface API real time standardization/validation, and HL7 real-time validation. All methods were successful.
- On average, data available demonstrated that deliverability increased from 71% to 88%.
- The proportion of addresses determined to be undeliverable, representing unsuccessful mailing costs avoided, averaged 12%.
- The standardization, validation and geocoding processes available through SmartyStreets is expected to improve the accuracy and completeness of incoming and existing data, while also improving core processes such as deduplication.

An unlimited license for the IIS community would provide substantial economies of scale as we envision rolling this out beyond the pilot sites. This service will improve the quality of data in IIS today through standardized, validated addresses to improve functions such as reminder recall. Geocodes will allow for sophisticated data visualization opportunities using IIS data. Cleaner valid addresses will positively impact multiple IIS processes such as merging and deduplication for years to come. Collectively, these efforts will help programs to meet the many IIS Functional Standards focused on data quality through complete, accurate demographics capture. We are excited to offer this service to the full IIS community with CDC's assistance, and we appreciate your consideration of this important request.

Please contact Mary Beth Kurilo, AIRA Policy and Planning Director, at mbkurilo@immregistries.org with any comments or questions related to this supplemental request.