**Translator Integrated Clinical and Environmental Exposures Service (ICEES)**

**ICEES** was developed as part of the [NCATS-funded Biomedical Data Translator Program](https://ncats.nih.gov/translator) (also see [Austin et al. 2018](https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.12595); [The Biomedical Data Translator Consortium 2018a](https://ascpt.onlinelibrary.wiley.com/doi/full/10.1111/cts.12591); [The Biomedical Data Translator Consortium 2018b](https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.12592)). “Green Team”, based at the University of North Carolina at Chapel Hill and one of 11 Translator teams, led the research and development of ICEES.

ICEES offers access to observational clinical data on all patients in the Carolina Data Warehouse for Health (CDWH) with an asthma-like phenotype (defined below). The data additionally contain data derived from several public databases on chemical exposures (e.g., airborne pollutants) and sociological exposures (e.g., estimated household income) (**Appendix A**). The exposures data have been integrated with the clinical data at the patient and visit level. The ICEES clinical data were derived from fully identified patient dataset, but the data have been 'binned' or recoded in order to protect patient privacy, while also providing open access to the data via a Translator API and ensuring compliance with §164.514(b) of [HIPAA, 'Safe Harbor' method for patient de-identification of medical records](https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification). Note that while the design of ICEES was driven by a use case on asthma, the general framework and approach is flexible and scalable and can be adapted for virtually any disease or use case.

ICEES is designed to offer four main functionalities and several related functionalities.

1. *Cohort discovery: users define a cohort using any number of defined feature variables as input parameters, and the service returns a sample size.*

*Associate name with cohort identifier: users can name a defined cohort using an ICEES ID and text description*

*Return cohort identifier associated with a name: users can obtain an ICEES ID for a named cohort*

*Feature identifiers: users can obtain a list of identifiers for a given feature variable; identifiers are available from the following sources (where available): NCIT, UMLS, SNOMED, MONDO, HPO, MeSH, ICD9/10, LOINC, ENVO, CHEBI, CHEMBL, PUBCHEM, RxNORM, SMILES, CAS*

*Cohort dictionary: users can obtain a list of all cohorts and their definitions for a given ICEES integrated feature table*

*2. Feature-rich cohort discovery: users select a predefined cohort as the input parameter, and the service returns a profile of that cohort in terms of the available feature variables.*

*3. Hypothesis-driven 2 x 2 feature associations: users select a predefined cohort and two feature variables, and the service returns a 2 x 2 feature table with a corresponding Chi Square statistic and P value.*

*Hypothesis-driven 2 X 2 feature associations with combined binning option: users select a predefined cohort, two feature variables, and bins, which can be combined, and the service returns a 2 x 2 feature table with a corresponding Chi Square statistic and P value.*

*4. Exploratory 1 X N feature associations: users select a predefined cohort and a feature variable of interest, and the service returns a 1 x N feature table with corrected Chi Square statistics and associated P values.*

ICEES can be used for scientific inference and discovery, although important caveats must be considered. The main considerations when working with ICEES are outlined below.

*1. All feature variables have been binned or recoded.*

*2. The integrated feature tables are designed for different 'study' periods (currently defined as calendar years).*

*3. The integrated feature tables are designed to provide access to either patient-level data or visit-level data.*

*4. All inferences must be made with respect to the binning strategy, 'study' design, and type of integrated feature table.*

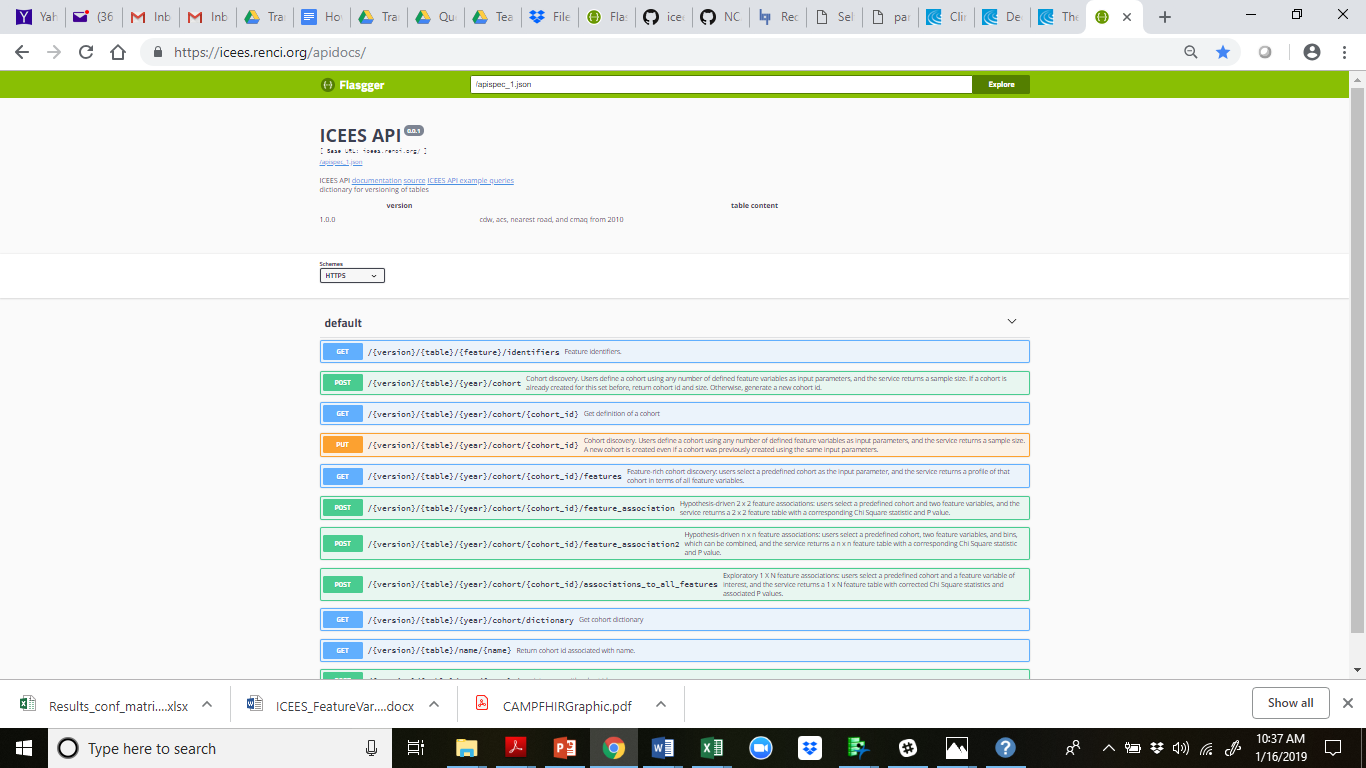
*Access to ICEES is open to anyone and is not subject to regulatory constraints; however, users are encouraged to abide by the following ICEES terms and conditions:*

**"The Translator Integrated Clinical and Environmental Exposures Service (ICEES) is providing you with Data that have been de-identified in accordance with 45 C.F.R. §§ 164.514(a) and (b) and that UNC Health Care System (UNCHCS) is permitted to provide under 45 C.F.R. § 164.502(d)(2). Recipient agrees to notify UNCHCS via NC TraCS in the event that Recipient receives any identifiable data in error and to take such measures to return the identifiable data and/or destroy it at the direction of UNCHCS.Restrictions on Recipient’s Use of Data. Recipient further agrees to use the data exclusively for the purposes and functionalities provided by the ICEES: cohort discovery; feature-rich cohort discovery; hypothesis-driven queries; and exploratory queries. Recipient agrees to use appropriate safeguards to protect the Data from misuse and unauthorized access or disclosure. Recipient will report to UNCHCS any unauthorized access, use, or disclosure of the Data not provided for by the Service of which Recipient becomes aware. Recipient will not attempt to identify the individuals whose information is contained in any Data transferred pursuant to this Service Agreement or attempt to contact those individuals. Recipient agrees not to sell the Data to any third party for any purpose. Recipient agrees not to disclose or publish the Data in any manner that would identify the Data as originating from UNCHCS. Finally, Recipient agrees to reasonably limit the number of queries to the Service per IP address within a given time interval, in order to prevent rapid ‘attacks’ on the Service. We kindly request that Translator team members provide proper attribution for any products (e.g., manuscripts, podium presentations, software) derived from work related to Green Team's clinical datasets. Attribution should include acknowledgement of the funder (National Center for Advancing Translational Sciences [NCATS], Biomedical Data Translator Program awards, OT3TR002020 and OT2TR002514), the North Carolina Translational and Clinical Sciences (NC TraCS) Institute (NCATS, Center for Translational Science Award, UL1TR002489), UNC Hospitals and Health Care System, and all Green Team members who contributed to the work.”**

[**ICEES API**](https://icees.renci.org/apidocs/)

**ICEES Preliminary Results**

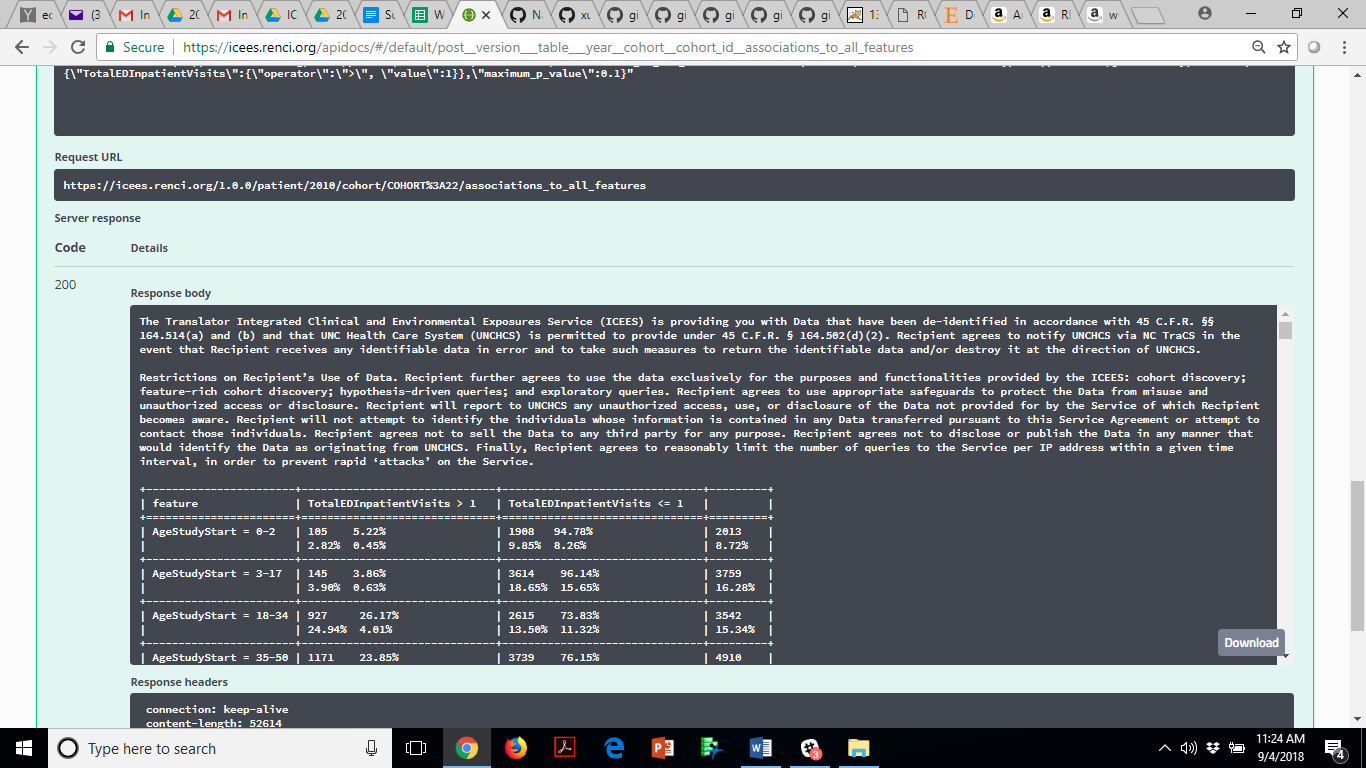
The UI for the ICEES API remains under development (**Figure 1**). While it is functional, it is not very user-friendly. Scientific and technical team members have been working together to improve the design.



**Figure 1.** The ICEES API/UI.

ICEES was developed with 10 embedded safeguards in order to ensure patient privacy and abide by HIPAA Safe Harbor. Those safeguards include the fact that usage agreements are returned to users with all queries of the service (**Figure 2**).

1. **Example ICEES API Output**



**B. ICEES Recommended Terms and Conditions of Use**

The Translator Integrated Clinical and Environmental Exposures Service (ICEES) is providing you with Data that have been de-identified in accordance with 45 C.F.R. §§ 164.514(a) and (b) and that UNC Health Care System (UNCHCS) is permitted to provide under 45 C.F.R. § 164.502(d)(2). Recipient agrees to notify UNCHCS via [NC TraCS](https://tracs.unc.edu/index.php/who-we-are/contact) in the event that Recipient receives any identifiable data in error and to take such measures to return the identifiable data and/or destroy it at the direction of UNCHCS.

**Restrictions on Recipient’s Use of Data.**

Recipient further agrees to use the data exclusively for the purposes and functionalities provided by the ICEES: cohort discovery; feature-rich cohort discovery; hypothesis-driven queries; and exploratory queries. Recipient agrees to use appropriate safeguards to protect the Data from misuse and unauthorized access or disclosure. Recipient will report to UNCHCS any unauthorized access, use, or disclosure of the Data not provided for by the Service of which Recipient becomes aware. Recipient will not attempt to identify the individuals whose information is contained in any Data transferred pursuant to this Service Agreement or attempt to contact those individuals. Recipient agrees not to sell the Data to any third party for any purpose. Recipient agrees not to disclose or publish the Data in any manner that would identify the Data as originating from UNCHCS. Finally, Recipient agrees to reasonably limit the number of queries to the Service per IP address within a given time interval, in order to prevent rapid ‘attacks’ on the Service.

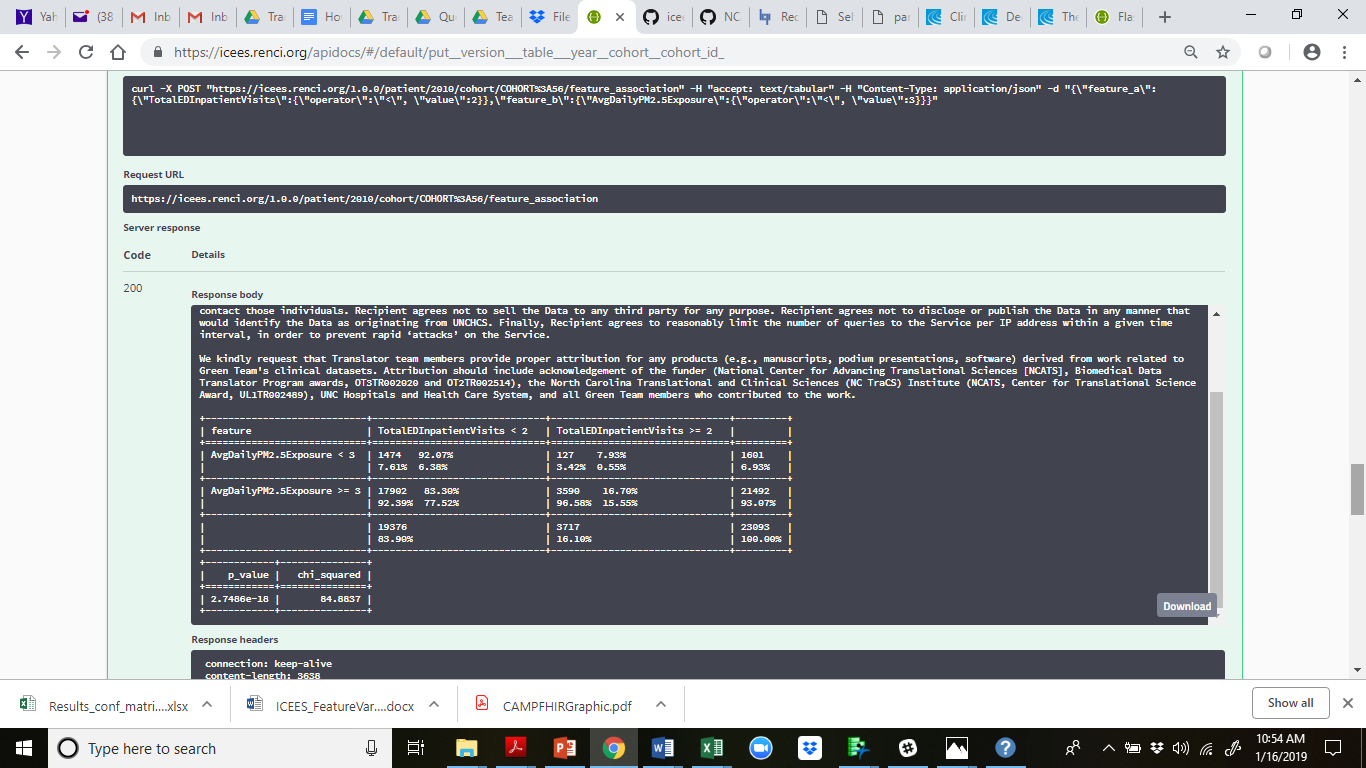
**Figure 2. A.** Screenshot of output from the ICEES API. **B.** A more user-friendly version. Terms and conditions of use. As a safeguard, a set of usage agreements is returned to users as the initial response to all queries of the service.

While we are still analyzing the integrated feature tables that are behind the ICEES API and refining the binning strategy for feature variables, the preliminary results look promising.

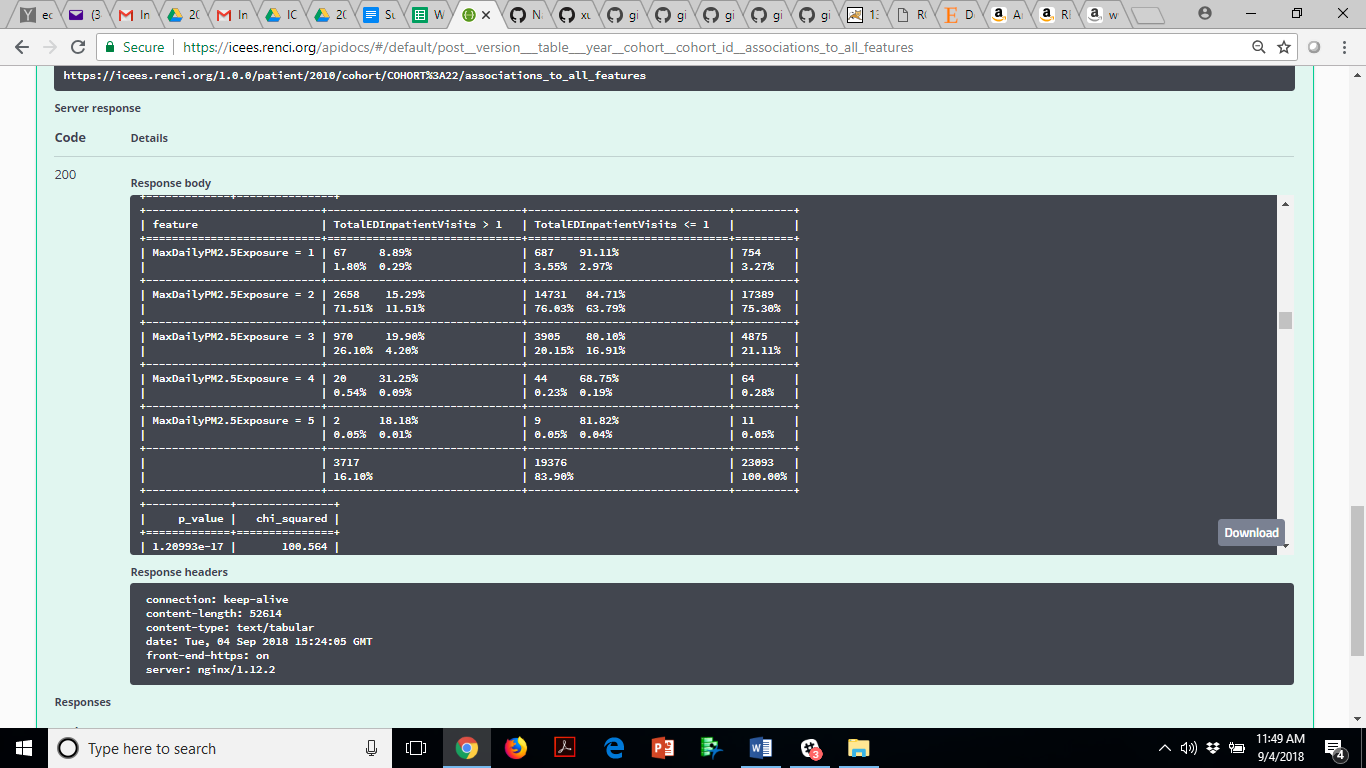
For instance, as part of our driving use case on asthma, Green Team developed a list of “competency questions” (CQs). CQ1 states: Among pediatric patients with an 'asthma-like phenotype', is exposure to particulate matter <=2.5 micrometers in diameter (PM2.5) and ozone associated with responsiveness to treatment? In other words, are exposures higher in patients who are non-responsive to treatment than in patients who are responsive to treatment?

CQ1 was developed to validate the prototype Translator system, our data sources, and our overall approach; i.e., a wealth of literature demonstrates an association between PM2.5 and ozone exposure and responsiveness to treatment (or asthma exacerbations), so we should be able to replicate this finding. Indeed, the preliminary results are clearly in the intended direction (**Figure 3**).

1. **Average Daily PM2.5 Exposure and ED/inpatient visits for respiratory issues**



1. **Maximum Daily PM2.5 Exposure and ED/inpatient visits for respiratory issues**



**Figure 3.** Screenshot of output from the ICEES API. The results demonstrate a significant association between patient-level exposure to PM2.5 (**A**: average daily PM2.5 exposures; **B**: maximum daily PM2.5 exposure) and number of ED or inpatient visits for respiratory issues over a one-year ‘study’ period.

**Green Team's Asthma-like Cohort**

**Asthma-like cohort**: At present, ICEES is restricted to patients with an ‘asthma-like’ phenotype. However, we are expanding ICEES to include additional patient cohorts (e.g., pain, obesity, diabetes, drug-induced liver injury).

Patients with an asthma-like phenotype were defined as follows:[[1]](#footnote-1)

*1. Patients with a diagnostic code of ‘asthma’ and prescribed or administered medications that are typically used to treat asthma;*

*2. Patients with a diagnostic code for a respiratory condition other than asthma and prescribed or administered medications that are typically used to treat asthma;*

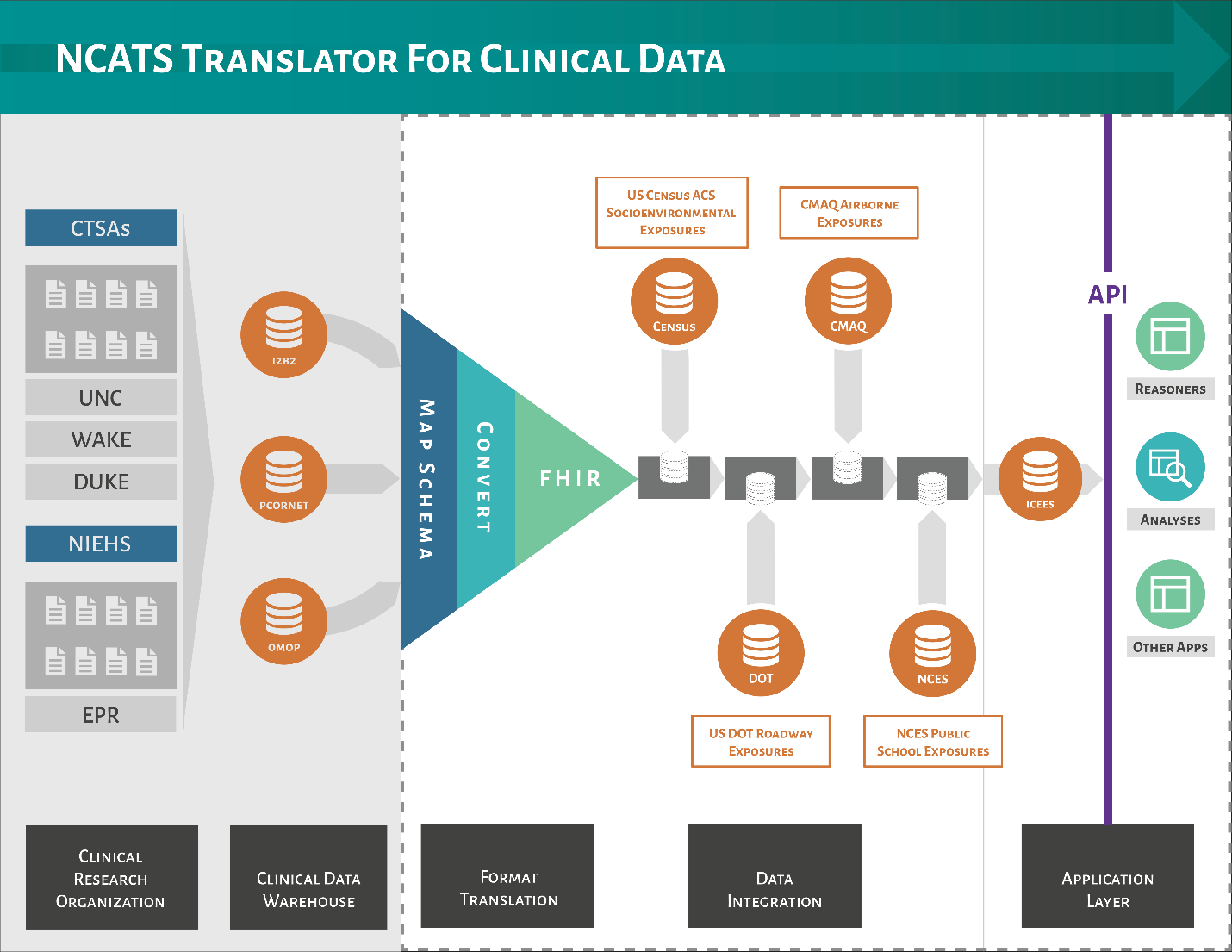
*3. Patients with a diagnostic code for a respiratory condition other than asthma and prescribed tests or procedures that are typically used to diagnosis asthma; or*

*4. Patients with a diagnostic code for a respiratory condition other than asthma and frequent ED visits with albuterol nebulizer administered.*

**A Note about Acknowledgements**

**We kindly request that ICEES users provide proper attribution for any products (e.g., manuscripts, podium presentations, software) derived from work related to ICEES. Attribution should include acknowledgement of the funder (National Center for Advancing Translational Sciences [NCATS], Biomedical Data Translator Program awards, OT3TR002020 and OT2TR002514), the North Carolina Translational and Clinical Sciences (NC TraCS) Institute (NCATS, Center for Translational Science Award, UL1TR002489), UNC Hospitals and Health Care System, and all Green Team members who contributed to the work.**

**Appendix A.** A high-level overview of the ICEES integration pipeline and its relationship to common clinical data models (i.e., i2b2, PCORNET, and OMOP).



*Graphic was designed by Kelsey Urgo.*

We recently developed an extensible clinical data conversion pipeline that invokes a custom software application, CAMP FHIR (Clinical Asset Mapping Program for FHIR), to transform clinical data from common data models (e.g., i2b2, PCORnet, OMOP) into HL7 Fast Healthcare Interoperability Resource (FHIR) files.[[2]](#footnote-2) A subsequent custom software application, FHIR PIT (Patient data Integration Tool), then integrates the clinical data with environmental exposures data from multiple sources before stripping the data of PHI and binning feature variables to create ICEES tables. Of note, FHIR PIT is modular and extensible and can be adapted for virtually any type of data that is of interest to clinical researchers and requires geocodes, dates, and identifiers for integration with EHR data. We expect CAMP FHIR and FHIR PIT to greatly improve the scalability of ICEES and automate the creation of integrated feature tables. CAMP FHIR and FHIR PIT are currently undergoing unit testing and will be released under an open GitHub license.

1. The following codes and parameters were used to identify patients with an ‘asthma-like’ phenotype:

   **Diagnostic codes for asthma and asthma-like conditions** ICD9 493.% asthma ICD10 J45.% asthma ICD9 464.% croup ICD10 J05.% croup ICD9 496.% reactive airway ICD10 J44.% reactive airway ICD10 J66.% reactive airway ICD9 786.% cough ICD10 R05.% cough ICD9 481.% pneumonia ICD9 482.% pneumonia ICD9 483.% pneumonia ICD9 484.% pneumonia ICD9 485.% pneumonia ICD9 486.% pneumonia ICD10 J12.% pneumonia ICD10 J13.% pneumonia ICD10 J14.% pneumonia ICD10 J15.% pneumonia ICD10 J16.% pneumonia ICD10 J17.% pneumonia ICD10 J18.% pneumonia

   **Tests and procedures for asthma and asthma-like conditions** CPT 94010 spirometry CPT 94070 multiple spirometry CPT 95070 methacholine challenge test CPT 94620 simple exercise stress test CPT 94621 complex exercise stress test CPT 31624 bronchoscopy CPT 94375 flow-volume loop CPT 94060 spirometry (pre/post bronchodilator test) CPT 94070 bronchospasm provocation CPT 95070 inhalation bronchial challenge CPT 94664 bronchodilator administration CPT 94620 pulmonary stress test CPT 95027 airborne allergen panel

   **Medications prescribed for patients with asthma-like phenotype** MEDCTN prednisone MEDCTN fluticasone MEDCTN mometasone MEDCTN budesonide MEDCTN beclomethasone MEDCTN ciclesonide MEDCTN flunisolide MEDCTN albuterol MEDCTN metaproterenol MEDCTN diphenydramine MEDCTN fexofenadine MEDCTN cetirizine MEDCTN ipratropium MEDCTN salmeterol MEDCTN arformoterol MEDCTN formoterol MEDCTN indacaterol MEDCTN theophylline MEDCTN omalizumab MEDCTN mepolizumab

   [↑](#footnote-ref-1)
2. Pfaff ER, Champion J, Cox S, Xu H, Fecho K, Krishnamurthy A, Chute CG, Overby Taylor C, Ahalt S. All roads lead to FHIR: an extensible clinical data conversion pipeline. *AMIA Conference Paper*, accepted for publication (October 2018). [↑](#footnote-ref-2)