

SOFTWARE

Open Access



FHIR PIT: an open software application for spatiotemporal integration of clinical data and environmental exposures data

Hao Xu¹, Steven Cox¹, Lisa Stillwell¹, Emily Pfaff², James Champion², Stanley C. Ahalt^{1,2} and Karamarie Fecho^{1*} 

Abstract

Background: Informatics tools to support the integration and subsequent interrogation of spatiotemporal data such as clinical data and environmental exposures data are lacking. Such tools are needed to support research in environmental health and any biomedical field that is challenged by the need for integrated spatiotemporal data to examine individual-level determinants of health and disease.

Results: We have developed an open-source software application—FHIR PIT (Health Level 7 Fast Healthcare Interoperability Resources Patient data Integration Tool)—to enable studies on the impact of individual-level environmental exposures on health and disease. FHIR PIT was motivated by the need to integrate patient data derived from our institution's clinical warehouse with a variety of public data sources on environmental exposures and then openly expose the data via ICEES (Integrated Clinical and Environmental Exposures Service). FHIR PIT consists of transformation steps or building blocks that can be chained together to form a transformation and integration workflow. Several transformation steps are generic and thus can be reused. As such, new types of data can be incorporated into the modular FHIR PIT pipeline by simply reusing generic steps or adding new ones. We validated FHIR PIT in the context of a driving use case designed to investigate the impact of airborne pollutant exposures on asthma. Specifically, we replicated published findings demonstrating racial disparities in the impact of airborne pollutants on asthma exacerbations.

Conclusions: While FHIR PIT was developed to support our driving use case on asthma, the software can be used to integrate any type and number of spatiotemporal data sources at a level of granularity that enables individual-level study. We expect FHIR PIT to facilitate research in environmental health and numerous other biomedical disciplines.

Keywords: Data integration, Spatiotemporal data, Modular software design

Background

Researchers and healthcare practitioners across fields of biomedicine acknowledge the tremendous impact that environmental exposures have on health and disease. For example, airborne pollutant exposures have been linked to diseases as diverse as asthma [1–6], diabetes [7–9], cardiovascular disease [10], dementia [11], mental health disorders [12], obesity [13], liver disease [14], and premature mortality [15]. Yet, informatics tools to study the interaction between environmental exposures and health outcomes at the level of the individual are largely non-existent. For instance, the fields of epidemiology and environmental

health focus primarily on population-based correlations between trends in spatiotemporal exposures and population-level health outcomes [15]. Longitudinal clinical studies likewise are limited in their ability to collect subject-level data on environmental exposures, typically relying on survey-based self-report [5] or expensive personal monitors [6]. Electronic health record (EHR)-based research also is limited because such records do not contain data on environmental exposures apart from basic demographics.

Herein, we present FHIR PIT (Health Level 7 Fast Healthcare Interoperability Resources Patient data Integration Tool) as an open-source software application designed to overcome challenges in environmental health research and related fields and provide an innovative solution to enable investigation into the impact of individual-level environmental exposures on health and disease.

* Correspondence: kfecho@copperlineprofessionalsolutions.com

¹Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27517, USA

Full list of author information is available at the end of the article



Implementation

This work was conducted under a study protocol that was approved by the Institutional Review Board at the University of North Carolina at Chapel Hill.

Motivation

FHIR PIT is a complex, custom, open-source software application that uses geocodes and time stamps of varying resolution (e.g., hour, day, year) to automatically integrate multiple sources of spatiotemporal data, irrespective of the degree to which the data depend on space and time. FHIR PIT was motivated by our research and development of the Integrated Clinical and Environmental Exposures Service [ICEES [16]]. ICEES was developed as part of the Biomedical Data Translator program in response to a need to openly expose clinical data that have been integrated at the patient and visit level with environmental exposures data [17, 18]. FHIR PIT provides the integrated clinical and environmental exposures data to support ICEES.

Implementation overview and spatiotemporal data sources

For initial research and development of FHIR PIT, clinical data on patients from UNC Health Care System were

integrated with a variety of public data on environmental exposures, including: airborne pollutant exposures from the US Environmental Protection Agency; roadway exposures from the Federal Highway Administration’s Highway Patrol Monitoring System, within the US Department of Transportation; roadway exposures from the US Census Bureau’s Topologically Integrated Geographic Encoding and Referencing system; and socio-environmental exposures from the US Census Bureau’s American Community Survey. (A graphical overview of the FHIR PIT integration pipeline can be found in Fig. 1. A list of currently available feature variables can be found in Supplementary Table 1. This table and additional documentation are maintained and regularly updated on the ICEES OpenAPI.) Importantly, the integration step is conducted within a secure environment and under a protocol that was approved by our institution’s Institutional Review Board because data integration necessitates the use of patient geocodes (i.e., primary home residence), date/time stamps, and patient identifiers—data elements that are considered Protected Health Information under the Health Insurance Portability and Accountability Act (HIPAA).

Multiple integration steps are required to harmonize across these data sources, which vary in spatiotemporal

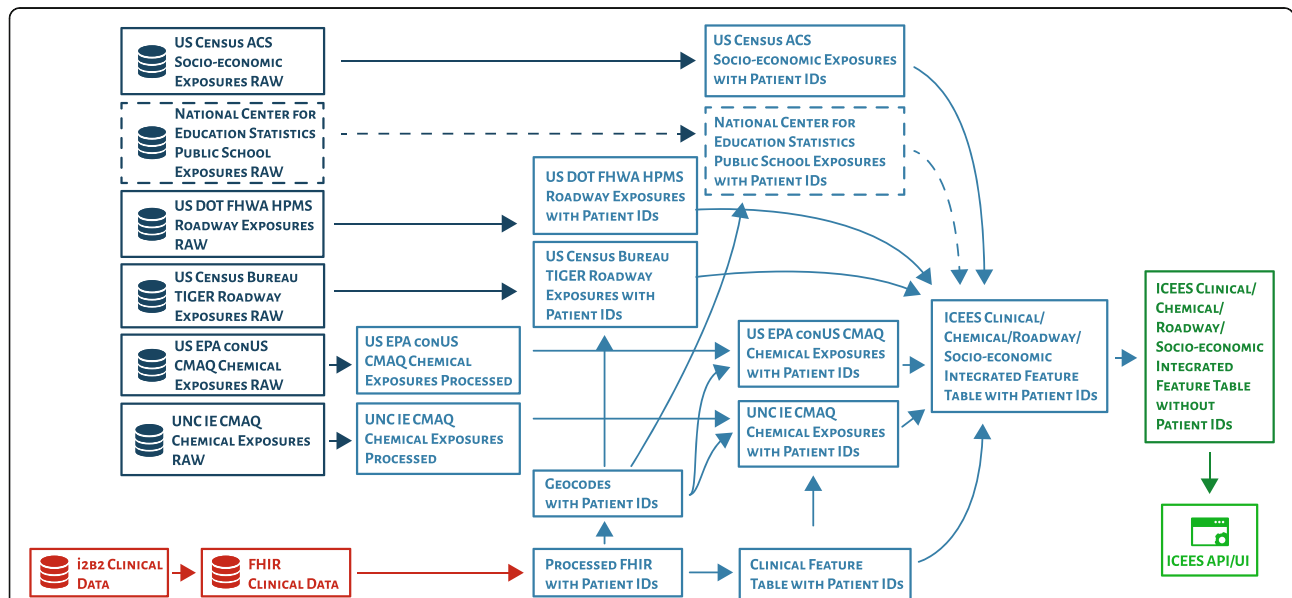


Fig. 1 An overview of the integration steps embedded in the FHIR PIT software application pipeline. API = application programming interface; FHIR = Health Level 7 Fast Healthcare Interoperability Resources; ICEES = Integrated Clinical and Environmental Exposures Service; UI = user interface; US Census ACS = US Census Bureau’s American Community Survey; US Census Bureau TIGER = US Census Bureau’s Topologically Integrated Geographic Encoding and Referencing system; US EPA conUS CMAQ = US Environmental Protection Agency’s conUS Community Multiscale Air Quality modeling data; US DOT FHWA HPMS = US Department of Transportation, Federal Highway Administration, Highway Patrol Monitoring System. Red color = sensitive, fully identified clinical data; dark blue color = public data on environmental exposures; light blue color = secure, firewall- and Institutional Review Board–protected integration steps; dark green color = de-identified, binned integrated feature tables; light green color = ICEES OpenAPI. (Note that data from the National Center for Education Statistics have not yet been integrated using FHIR PIT, but an approach is under development to integrate data on school exposures with home exposures data and clinical data, thereby addressing issues related to patient mobility and differential exposures. A simplified version of the FHIR PIT pipeline was published in JAMIA 2019;26(1):1064–1073 and is reprinted in adapted form here with full permission from the publisher. In contrast to the simplified version of the FHIR PIT pipeline, the version shown here includes and clearly distinguishes all of the data sources and integration steps that are assembled by the current version of the pipeline.)

resolution and format of geocodes and time stamps. For example, patient primary home residence is coded as latitude and longitude in the patient data, whereas the American Community Survey data are provided at the Census block level. Airborne pollutant exposures are available at hourly estimates, daily estimates, or annual averages, depending on the exposure entity and source year. Roadway data are provided as GIS shape files, with latitudes and longitudes in WGS84 decimal format, which is the World Geodetic System for expressing latitude and longitude. Separate software code is required to convert the spatio-temporal representation of the data used by each data source into a common format that allows integration across data sources. In addition, separate mappings are required to link patient identifiers and geocodes with each non-clinical data source, thereby supporting the final integration step that merges the different data sources.

The final product of the FHIR PIT software pipeline is a set of “integrated feature tables”, with feature variables binned or recoded and data de-identified according to §164.514(b) of HIPAA for subsequent open access via the ICEES OpenAPI.

Implementation details

FHIR PIT consists of several transformation steps or building blocks that can be chained together to form a transformation and integration workflow. Several of these transformation steps are generic, such that they can take in any data that conform to a certain format. Thus, the incorporation of new types of data amounts to adding new transformation steps or reusing generic steps. FHIR PIT is implemented using Apache Spark. Spark is used to easily parallelize and distribute the data transformation steps. A Python script is used to simplify the application interface to the transformation steps. FHIR PIT supports building containers in both Singularity and Docker. This feature allows the application to run on different machines and platforms with portability.

Table 1 FHIR PIT plugin names and functionalities

Plugin name	Functionality
FHIR	Consolidates different FHIR resources for each patient and extracts geocodes
ToVector	Extracts features from FHIR
EnvData	Preprocesses environmental data source
CSVTable	Converts to ICEES integrated feature table
ACS	Preprocesses US Census Bureau ACS data source
ACS2	Preprocesses US Census Bureau ACS data source, v2; this includes a “ur” field for “urban or rural” residence
NearestRoad	Preprocesses nearest road data source for US Census Bureau TIGER data source
NearestRoad2	Preprocesses nearest road data source for US DOT FHWA HPMS data source
NOOP	No operation

Table 2 FHIR PIT field names and functionality

Field name	Functionality
name	Designates name of given step instance
dependsOn	Defines other step instances that given step instance depends on
skip	Determines whether given step instance should be skipped; if skip is “true”, then this step will not be run; skip function allows for partial re-execution of pipelines that have not been completely executed
step	Defines the given step instance
step.function	Designates the function name for given step instance; this is usually a class name
step.arguments	Delineates specific arguments for given step function; the arguments vary according to the step function

Each block in FHIR PIT is implemented as a plugin consisting of a set of Scala classes that can be plugged into the pipeline. FHIR PIT is configured using a YAML file, and steps can be switched on or off for rapid re-execution of the pipeline. The plugins consist of both generic building blocks such as joining of tables and data set-specific building blocks such as preprocessing of environmental data (Table 1). The input and output of each plugin can be configured so that the output of the previous step in a pipeline configuration can be fed as input for the next step.

One of our goals for implementation of the pipeline is to enable automatic and rapid re-execution. Given the extensible number of input files and parameters, we use the Dhall configuration language to author configuration files and avoid code duplication. Dhall code is converted to a YAML file that is then read by the pipeline. An example YAML configuration of a step in the FHIR PIT pipeline is provided below, with fields defined in Table 2.

```

- name: EnvDataSource2012
  dependsOn:
  - PerPatSeriesToVector2012
  skip: false
  step:
    function: datatrans.step.EnvDataSourceConfig
    arguments:
      environmental_data: /data/env
      fips_data:/data/spatial/LCC.shp
      patgeo_data: /data/processed/2012/geo.csv
      indices: []
      indices2:
      - ozone_daily_8hour_maximum
      - pm25_daily_average
      statistics: []
      start_date: 2012-01-01T00:00:00-05:00
      end_date: 2013-01-01T00:00:00-05:00
      output_file: /data/processed/env/2012/%i

```

Writing the entire FHIR PIT pipeline configuration in YAML would necessitate rewriting the pipeline for every new calendar year and every new data set. With Dhall, we are able to create a function in the configuration that can be instantiated for each new calendar year or data

```
let envDataSourceStep = λ(skip : Bool) → λ(year : Text) → Step.EnvDataSource {
  name = "EnvDataSource${year}",
  dependsOn = [
    "PerPatSeriesToVector${year}"
  ],
  skip = skip,
  step = {
    function = "datatrans.step.EnvDataSourceConfig",
    arguments = {
      environmental_data = "/data/env",
      fips_data = "/data/spatial/LCC.shp",
      patgeo_data = "/data/processed/${year}/geo.csv",
      indices = [] : List Text,
      statistics = [] : List Text,
      indices2 = [
        "ozone_daily_8hour_maximum",
        "pm25_daily_average"
      ],
      start_date = start_year year,
      end_date = end_year year,
      output_file = "/data/processed/env/${year}/%i"
    }
  }
}
```

set. A simplified version of this function to address additional years is shown below.

To instantiate this for calendar year 2012, we simply need to specify the following parameter:

```
envDataSourceStep False "2012"
```

To extend this function for multiple calendar years, we specify an additional parameter:

```
List/map ["2012", "2013", "2014"]
(envDataSourceStep False)
```

Here, the *List/map* function takes a list of terms and a function, applies the function to each element in the list, and returns a list of values.

Execution of the FHIR PIT pipeline generates a report of skipped tasks, succeeded tasks, failed tasks, and errors from failed tasks.

airborne pollutant exposures on asthma. The validation data set consisted of ~ 160,000 patients with “asthma-like” conditions from UNC Health Care System and the environmental data sources depicted in Fig. 1, focusing initially on data from calendar year 2010 [19–21]. FHIR PIT was used to integrate the clinical and environmental data and then de-identify the data and bin feature variables before openly exposing the integrated data using ICEES. ICEES was queried using the following input parameters:

```
Feature variables: {
  "feature_a": {
    "AvgDailyPM2.5Exposure": {
      "operator": "<",
      "value": 3
    }
  },
  "feature_b": {
    "TotalEDInpatientVisits": {
      "operator": "<",
      "value": 2
    }
  }
}
Version of data: 1.0.0
Table: patient
Year: 2010
Cohort ID: COHORT:60
Cohort ID: COHORT:72
```

Results

We validated FHIR PIT in the context of our driving use case for research and development of ICEES: impact of

ICEES returned the following JSON output, which is also displayed in graphical form in Fig. 2.

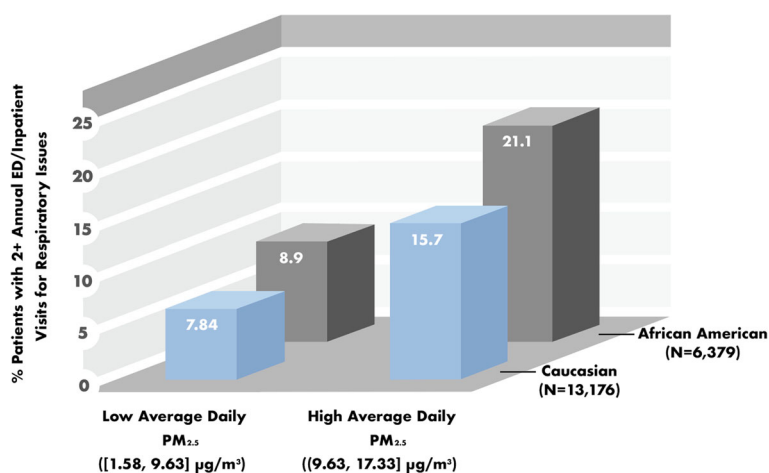


Fig. 2 Racial disparities in the impact of airborne pollutant exposures on asthma exacerbations. Sample sizes are: N = 6379 African American patients; and N = 13,176 Caucasian patients. PM_{2.5} = particulate matter < 2.5-µm in diameter. Levels of PM_{2.5} exposure were binned in FHIR PIT using pandas qcut and expressed as ranges. $\chi^2 = 28.2841, P < 0.0001$ for African Americans; $\chi^2 = 47.0133, P < 0.0001$ for Caucasians

```

"version": "1.0.0",
"return value": {
  "total": 6379,
  "feature_matrix": {
    {
      "frequency": 297,
      "total_percentage": 0.04659921799249254,
      "row_percentage": 0.05854923953252515,
      "column_percentage": 0.911842944788276
    },
    {
      "frequency": 4776,
      "total_percentage": 0.74787066938391598,
      "row_percentage": 0.9414547604967475,
      "column_percentage": 0.7893302329424228
    }
  ],
  "chisquared": 0.00895703321472233
},
{
  "frequency": 29,
  "total_percentage": 0.004546167110832419,
  "row_percentage": 0.0222922067381317,
  "column_percentage": 0.00895703321472233
},
{
  "frequency": 1279,
  "total_percentage": 0.20018811725979857,
  "row_percentage": 0.29778479320161602,
  "column_percentage": 0.210946976705765735
}
},
"feature_b": {
  "feature_name": "TotalEDInpatientVisits",
  "feature_qualifiers": [
    {
      "operator": "<",
      "value": 2
    },
    {
      "operator": ">=",
      "value": 2
    }
  ]
},
"feature_a": {
  "feature_name": "AvgDailyPM2.5Exposure",
  "feature_qualifiers": [
    {
      "operator": "<",
      "value": 3
    },
    {
      "operator": ">=",
      "value": 3
    }
  ]
},
"p_value": 1.0475066725063578e-7,
"columns": [
  {
    "frequency": 326,
    "percentage": 0.05110518890108168
  },
  {
    "frequency": 4053,
    "percentage": 0.9488948110989184
  }
],
"rows": [
  {
    "frequency": 5073,
    "percentage": 0.795265715629409
  },
  {
    "frequency": 1306,
    "percentage": 0.204734284370591
  }
],
"chisquared": 28.284141190389228
},
"version": "1.0.0",
"return value": {
  "total": 13176,
  "feature_matrix": {
    {
      "frequency": 976,
      "total_percentage": 0.07407407407407407,
      "row_percentage": 0.08721293986881423,
      "column_percentage": 0.9216241737488197
    },
    {
      "frequency": 10213,
      "total_percentage": 0.7752732240437158,
      "row_percentage": 0.91278781610311858,
      "column_percentage": 0.4833045305246462
    }
  ],
  "chisquared": 0.00629933211900425,
  "total_percentage": 0.04181360201511335,
  "column_percentage": 0.07837582621180336
},
{
  "frequency": 1902,
  "total_percentage": 0.14435336976320584,
  "row_percentage": 0.25818263979848689,
  "column_percentage": 0.15896954691355383
}
},
"feature_b": {
  "feature_name": "TotalEDInpatientVisits",
  "feature_qualifiers": [
    {
      "operator": "<",
      "value": 2
    },
    {
      "operator": ">=",
      "value": 2
    }
  ]
},
"feature_a": {
  "feature_name": "AvgDailyPM2.5Exposure",
  "feature_qualifiers": [
    {
      "operator": "<",
      "value": 3
    },
    {
      "operator": ">=",
      "value": 3
    }
  ]
},
"p_value": 7.05098065486281e-12,
"columns": [
  {
    "frequency": 1059,
    "percentage": 0.08037340419307833
  },
  {
    "frequency": 12117,
    "percentage": 0.9196265988069217
  }
],
"rows": [
  {
    "frequency": 11191,
    "percentage": 0.84934729811779
  },
  {
    "frequency": 1985,
    "percentage": 0.1506527018221008
  }
],
"chisquared": 47.0133228473862
}

```

These results indicate that the proportion of patients with two or more annual emergency department or

inpatient visits for respiratory issues was higher among patients exposed to relatively high average daily levels of particulate matter <2.5-µm in diameter (PM_{2.5}) than among those exposed to relatively low average daily levels of PM_{2.5}. Moreover, asthma exacerbations, as defined by two or more annual emergency department or inpatient visits for respiratory issues, were more common among African Americans than among Caucasians.

We then examined prednisone use in relation to asthma exacerbations among African Americans and Caucasians. The ICEES query is shown below.

```

Feature variables: {
  "feature_a": {
    "TotalEDInpatientVisits": {
      "operator": "<",
      "value": 2
    }
  },
  "feature_b": {
    "Prednisone": {
      "operator": "=",
      "value": 0
    }
  }
}
Version of data: 1.0.0
Table: patient
Year: 2010
Cohort ID: COHORT:60
Cohort ID: COHORT:72

```

ICEES returned the following results, which are shown in tabular form in Table 3.

Table 3 Relationship between prednisone use and asthma exacerbations, defined as two or more annual ED or inpatient visits for respiratory issues, among African Americans and Caucasians

	Patients with < 2 annual ED/inpatient visits for respiratory issues N (%)	Patients with ≥ 2 annual ED/inpatient visits for respiratory issues N (%)	Chi square, P value
African Americans (N = 6379)			
<i>Prednisone</i>			
No	4536 (89.41%)	1078 (82.54%)	$\chi^2 = 46.4781$,
Yes	537 (10.59%)	228 (17.46%)	$P < 0.0001$
Caucasians (N = 13,176)			
<i>Prednisone</i>			
No	10,071 (89.99%)	1675 (84.38%)	$\chi^2 = 54.8241$,
Yes	1120 (10.01%)	310 (15.62%)	$P < 0.0001$

Abbreviations: ED, emergency department

```

{
  "version": "1.0.0",
  "return_value": {
    "total": 4379,
    "feature_matrix": [
      {
        "frequency": 4536,
        "total_percentage": 0.7110832418874432,
        "row_percentage": 0.8079800498753117,
        "column_percentage": 0.8941454760496748
      },
      {
        "frequency": 1078,
        "total_percentage": 0.16999200501646025,
        "row_percentage": 0.19201995012468828,
        "column_percentage": 0.8254211332312404
      },
      {
        "frequency": 537,
        "total_percentage": 0.08418247374196583,
        "row_percentage": 0.7019607843137254,
        "column_percentage": 0.10585452395032525
      },
      {
        "frequency": 228,
        "total_percentage": 0.03574227935413074,
        "row_percentage": 0.2980392156862745,
        "column_percentage": 0.17457866674875958
      }
    ],
    "feature_b": {
      "feature_name": "Prednisone",
      "feature_qualifiers": [
        {
          "operator": "=",
          "value": 0
        },
        {
          "operator": "<=",
          "value": 0
        }
      ]
    },
    "feature_a": {
      "feature_name": "TotalEDInpatientVisits",
      "feature_qualifiers": [
        {
          "operator": "<=",
          "value": 2
        },
        {
          "operator": ">=",
          "value": 2
        }
      ]
    },
    "n_value": 9.26500631308178e-12,
    "columns": [
      {
        "frequency": 5073,
        "percentage": 0.795265715629409
      },
      {
        "frequency": 1306,
        "percentage": 0.204734284370591
      }
    ],
    "rows": [
      {
        "frequency": 8614,
        "percentage": 0.4900752469039035
      },
      {
        "frequency": 765,
        "percentage": 0.11992475309609657
      }
    ],
    "chi_squared": 46.47805966021564
  }
},
{
  "version": "1.0.0",
  "return_value": {
    "total": 13376,
    "feature_matrix": [
      {
        "frequency": 10071,
        "total_percentage": 0.764344022208082,
        "row_percentage": 0.857398263285493,
        "column_percentage": 0.8999195782325082
      },
      {
        "frequency": 1675,
        "total_percentage": 0.1271250758955677,
        "row_percentage": 0.14260173676145071,
        "column_percentage": 0.8438287153652393
      },
      {
        "frequency": 1120,
        "total_percentage": 0.0850030382270795,
        "row_percentage": 0.7832167832167832,
        "column_percentage": 0.10008042176749174
      },
      {
        "frequency": 310,
        "total_percentage": 0.0235276259664236,
        "row_percentage": 0.21678321678321677,
        "column_percentage": 0.1561712846347607
      }
    ],
    "feature_b": {
      "feature_name": "Prednisone",
      "feature_qualifiers": [
        {
          "operator": "=",
          "value": 0
        },
        {
          "operator": "<=",
          "value": 0
        }
      ]
    },
    "feature_a": {
      "feature_name": "TotalEDInpatientVisits",
      "feature_qualifiers": [
        {
          "operator": "<=",
          "value": 2
        },
        {
          "operator": ">=",
          "value": 2
        }
      ]
    },
    "n_value": 1.3181707519332912e-13,
    "columns": [
      {
        "frequency": 11191,
        "percentage": 0.84934729811779
      },
      {
        "frequency": 1985,
        "percentage": 0.15065270188221008
      }
    ],
    "rows": [
      {
        "frequency": 11746,
        "percentage": 0.8914693181906497
      },
      {
        "frequency": 1430,
        "percentage": 0.10853066180935034
      }
    ],
    "chi_squared": 54.82406086877697
  }
}

```

These results indicate that prednisone use was more common among patients with asthma exacerbations than among those without asthma exacerbations, as expected given that prednisone is generally reserved for patients with severe disease [22]. While this finding was true for both African Americans and Caucasian, the effect was more pronounced among African Americans than among Caucasians.

In sum, we successfully applied FHIR PIT to integrate clinical and environmental data and then openly expose the data for interrogation via ICEES, thereby replicating and extending published literature demonstrating the impact of exposure to airborne particulate matter on asthma (e.g., 4) and the existence of racial disparities in asthma exacerbations [23].

Conclusion

We developed FHIR PIT as an open-source spatiotemporal data integration tool. We are currently using FHIR PIT to generate integrated clinical and environmental data for open exposure and interrogation via ICEES. While FHIR PIT was developed and validated in the context of a driving use case designed to evaluate the impact of airborne pollutant exposures on asthma, the software application has broad applicability in any use case that requires integrated spatiotemporal data for individual-level analysis. Indeed, we are currently extending FHIR PIT to support investigations into the impact of environmental exposures on primary ciliary dyskinesia, drug-induced liver injury, and several additional conditions. We believe that FHIR PIT will facilitate research in environmental health and many other biomedical disciplines.

FHIR PIT is under active development, with new data types and sources planned for the use cases noted above and others. The modular design of FHIR PIT will allow us to rapidly adapt the pipeline for these new data types and sources and automatically execute the pipeline to generate new ICEES integrated feature tables, thus providing flexibility and extensibility. These features will facilitate the adoption and adaptation of FHIR PIT for use in other applications and at other institutions.

Availability and requirements

- Project name: FHIR PIT
- Project home page: Software code and instructions for downloading FHIR PIT can be found at: <https://github.com/NCATS-Tangerine/FHIR-PIT>
- Operating system(s): Linux
- Programming language: Scala, Python
- Other requirements: Java 8 or higher
- License: MIT
- Any restrictions to use by non-academics: none

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12911-020-1056-9>.

Additional file 1: Supplementary Table 1. ICEES integrated feature variable tables (v1.0.0, v2.0.0): variable names, descriptions, and binning strategy.*

Abbreviations

API: Application programming interface; ED: Emergency department; FHIR: HL7 Health Level 7 Healthcare Interoperability Resources; FHIR PIT: Health Level 7 Fast Healthcare Interoperability Resources Patient data Integration Tool; ICEES: Integrated Clinical and Environmental Exposures Service; PM_{2.5}: particulate matter < 2.5- μ m in diameter; UI: User interface; US Census ACS: US Census Bureau's American Community Survey; US Census Bureau TIGER: US Census Bureau's Topologically Integrated Geographic Encoding and Referencing system; US DOT FHWA HPMS: US Department of Transportation, Federal Highway Administration, Highway Patrol Monitoring System; US EPA conUS CMAQ: US Environmental Protection Agency's conUS Community Multiscale Air Quality modeling data

Acknowledgements

The authors acknowledge Sarav Arunachalam for provisioning the airborne pollutant exposures data, Stephen Appold for provisioning the socio-environmental exposures data, and Alejandro Valencia Aria for assistance with the roadway exposures data. The authors also acknowledge the clinical expertise provided by David B. Peden. The authors wish to thank Marian Mersmann for proofreading and the Carolina Data Warehouse for Health Oversight Committee at the University of North Carolina at Chapel Hill for their guidance and support of the work described in this manuscript.

Authors' contributions

HX is the lead software designer, developer, and implementer of FHIR PIT; SC, LS, and KF contributed to the design of FHIR PIT; LS contributed expertise on GIS files; EP and JC provisioned the patient data and contributed to the design of FHIR PIT; KF led the validation and analysis of FHIR PIT; SCA provided overall leadership and direction for research and development of FHIR PIT; HX and KF led development of the manuscript and prepared the first draft. All authors reviewed the manuscript, provided critical input, and approved the final submission.

Funding

This work has been supported by the National Center for Advancing Translational Sciences, National Institutes of Health [grant numbers OT3TR002020, UL1TR002489]. The funder had no role in the design of the study; the collection, analysis, and interpretation of data; or the writing of the manuscript.

Availability of data and materials

The data that were used to validate FHIR PIT are openly available via ICEES at <https://icees.renci.org/apidocs>.

Ethics approval and consent to participate

The research described in this manuscript was conducted under a study protocol that was approved by the Institutional Review Board at the University of North Carolina at Chapel Hill with a waiver of informed consent (protocol #16-2978, title "Carolina Biomedical Data Translator", principal investigator SCA, created November 21, 2016, most recent annual approval received in September 20, 2019).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27517, USA. ²North Carolina Translational and

Clinical Sciences Institute, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA.

Received: 19 December 2019 Accepted: 17 February 2020

Published online: 11 March 2020

References

- Delfino RJ, Coate BD, Zeiger RS, Seltzer JM, Street DH, Koutrakis P. Daily asthma severity in relation to personal ozone exposure and outdoor fungal spores. *Am J Respir Crit Care Med*. 1996;154(3 Pt 1):633–41. Available from: <https://doi.org/10.1164/ajrccm.154.3.8810598>.
- Saha C, Riner ME, Liu G. Individual and neighborhood-level factors in predicting asthma. *Arch Pediatr Adolesc Med [Internet]*. 2005;159(8):759–63. Available from: <https://doi.org/10.1001/archpedi.159.8.759>.
- Schildcrout JS, Sheppard L, Lumley T, Slaughter JC, Koenig JQ, Shapiro GG. Ambient air pollution and asthma exacerbations in children: an eight-city analysis. *Am J Epidemiol*. 2006;164(6):505–17. Available from: <https://doi.org/10.1093/aje/kwj225>.
- Mirabelli MC, Vaidyanathan A, Flanders WD, Qin X, Garbe P. Outdoor PM_{2.5}, Ambient air temperature, and asthma symptoms in the past 14 days among adults with active asthma. *Environ Health Perspect*. 2016;124(12):1882–90. Available from: <https://doi.org/10.1289/EHP92>.
- Schurman SH, Bravo MA, Innes CL, Jackson WB 2nd, McGrath JA, Miranda ML, et al. Toll-like receptor 4 pathway polymorphisms interact with pollution to influence asthma diagnosis and severity. *Sci Rep*. 2018;8(1):12713. Available from: <https://doi.org/10.1038/s41598-018-30865-0>.
- Rabinovitch N, Adams CD, Strand M, Koehler K, Volkens J. Within-microenvironment 711 exposure to particulate matter and health effects in children with asthma: a pilot study 712 utilizing real-time personal monitoring with GPS interface. *Environ Health*. 2016;15(1):96. Available from: <https://doi.org/10.1186/s12940-016-0181-5>.
- Krämer U, Herder C, Sugiri D, Strassburger K, Schikowski T, Ranft U, et al. Traffic-related air pollution and incident type 2 diabetes: results from the SALIA cohort study. *Environ Health Perspect*. 2010;118(9):1273–9. Available from: <https://doi.org/10.1289/ehp.0901689>.
- Weinmayr G, Hennig F, Fuks K, Nonnemacher M, Jakobs H, Möhlenkamp S, et al. Long-term exposure to fine particulate matter and incidence of type 2 diabetes mellitus in a cohort study: effects of total and traffic-specific air pollution. *Environ Health*. 2015;19:14–53. Available from: <https://doi.org/10.1186/s12940-015-0031-x>.
- Jagai JS, Krajewski AK, Shaikh S, Lobdell DT, Sargis RM. Association between environmental quality and diabetes in the USA. *J Diabetes Investig*. 2019; Available from: <https://doi.org/10.1111/jdi.13152>.
- Hennig F, Fuks K, Moebus S, Weinmayr G, Memmesheimer M, Jakobs H, et al. Association between source-specific particulate matter air pollution and hs-CRP: local traffic and industrial emissions. *Environ Health Perspect*. 2014;122(7):703–10. Available from: <https://doi.org/10.1289/ehp.1307081>.
- Chen H, Kwong JC, Copes R, Tu K, Villeneuve PJ, van Donkelaar A, et al. Living near major roads and the incidence of dementia, Parkinson's disease, and multiple sclerosis: a population-based cohort study. *Lancet*. 2017;389(10070):718–26. Available from: [https://doi.org/10.1016/S0140-6736\(16\)32399-6](https://doi.org/10.1016/S0140-6736(16)32399-6).
- Bernardini F, Attademo L, Trezzi R, Gobbicchi C, Balducci PM, Del Bello V, et al. Air pollutants and daily number of admissions to psychiatric emergency services: evidence for detrimental mental health effects of ozone. *Epidemiol Psychiatr Sci*. 2019;6:1–7. Available from: <https://doi.org/10.1017/S2045796019000623>.
- McConnell R, Gilliland FD, Goran M, Allayee H, Hricko A, Mittelman S. Does near-roadway air pollution contribute to childhood obesity? *Pediatr Obes*. 2016;11:1–3. Available from: <https://doi.org/10.1111/jjpo.12016>.
- Kim JW, Park S, Lim CW, Lee K, Kim B. The role of air pollutants in initiating liver disease. *Toxicol Res*. 2014;30:65–70. Available from: <https://doi.org/10.5487/tr.2014.30.2.065>.
- Chang SY, Vizuete W, Serre M, Vennam LP, Omary M, Isakov V, et al. Finely resolved on-road PM 2.5 and estimated premature mortality in central north carolina. *Risk Anal*. 2017;37(12):2420–34. Available from: <https://doi.org/10.1111/risa.12775>.
- Fecho K, Pfaff E, Xu H, Champion J, Cox S, Stillwell L, et al. A novel approach for exposing and sharing clinical data: the translator integrated clinical and environmental exposures service. *J Am Med Inform Assoc*. 2019;26(10):1064–73. Available from: <https://doi.org/10.1093/jamia/ocz042>.

17. Biomedical Data Translator Consortium. The biomedical data translator program: conception, culture, and community. *Clin Transl Sci* 2019;12(2):91–94. Available from: <https://doi.org/10.1111/cts.12592>.
18. Biomedical Data Translator Consortium. Toward a universal biomedical data translator. *Clin Transl Sci* 2019;12(2):86–90. Available from: <https://doi.org/10.1111/cts.12591>.
19. Ahalt SC, Chute CG, Fecho K, Glusman G, Hadlock J, Taylor CO, et al. Clinical data: sources and types, regulatory constraints, applications. *Clin Transl Sci* 2019;12(4):329–33. Available from: <https://doi.org/10.1111/cts.12638>.
20. Fecho K, Ahalt SC, Arunachalam S, Champion J, Chute CG, Davis S, et al. Sex, obesity, diabetes, and exposure to particulate matter among patients with severe asthma: Scientific insights from a comparative analysis of open clinical data sources during a five-day hackathon. *J Biomed Inform*. 2019; 100:103325. Available from: <https://doi.org/10.1016/j.jbi.2019.103325>.
21. Pfaff ER, Champion J, Bradford RL, Clark M, Xu H, Fecho K, et al. Fast healthcare interoperability resources (FHIR) as a meta model to integrate common data models: development of a tool and quantitative validation study. *JMIR Med Inform*. 2019;7(4):e15199. Available from: <https://doi.org/10.2196/15199>.
22. WebMD Medical Reference. L. Chang, Reviewer. Prednisone and Asthma. July 5, 2018. Available from: <https://www.webmd.com/asthma/guide/prednisone-asthma#1>.
23. Keet CA, McCormack MC, Pollack CE, Peng RD, McGowan E, Matsui EC. Neighborhood poverty, urban residence, race/ethnicity, and asthma: rethinking the inner-city asthma epidemic. *J Allergy Clin Immunol*. 2015; 135(3):655–62. Available from: <https://doi.org/10.1016/j.jaci.2014.11.022>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

