

# Ben Bubnick

DATA SCIENCE · DATA CURATION · DATA INTEGRATION

3419 Superior Park Drive, Cleveland Heights, Ohio, 44118, U.S.

☎ (+1) 216-556-1612 | ✉ ben.bubnick@gmail.com | 📷 bubnicbf | 🌐 benbubnick

## Summary

<b>Database Design</b>	ETL/EDW design with SSIS/SQL/T-SQL/PostgreSQL in CI/CD pipeline
<b>SME/Utility Roles</b>	AI & Machine Learning, Data Curation/Wrangling & Integration
<b>Big Data Platform</b>	Python, Pig, Ruby, kafka, and Hive/Impala on Hadoop & Azure
<b>Health Care Analytics</b>	HEDIS Measures & Tailored Analytics from EMR, HL7, 837, CMS, etc.
<b>Demonstrated Success</b>	Lab Services Award, Eminence and Excellence Award, Honey Badger

## Experience

### Merative

#### DATA SCIENTIST

Jul. 2022 - Present

Python, Azure, Postman

Technical lead on the delivery of large & complex AI and Cognitive led transformation programs and works with Merative & Partners to optimize tools and delivery methods.

- Led HIPAA patient deidentification project from 200+ million individual patients for content analytics
- Created testing framework for ETL & implementation processes on new data platform

### IBM

Jun. 2015 - Jul. 2022

Nov. 2021 - Jul. 2022

#### ADVISORY DATA SCIENTIST

Python, GoLang, Azure, Docker, SQL/PostgreSQL, Tableau

Technical lead on delivery of Cognitive & AI led programs using diverse data sources, such as electronic health record data, insurance claims data, scheduling and financial data, social determinants of health data.

- Created machine learning Covid-19 tool that reduced need for hospitalization identification
- Led delivery on tailored analytics and reporting project for client administrative data
- Developed data migration validation analysis on 150+ billion patient records
- Developed surgical smart case card using combined CNN/RNN model

Dec. 2020 - Nov. 2021

#### Promotion SENIOR DATA CURATION ARCHITECT

Gradle, Python, Hive/Impala, Spark, kafka, Kibana, SQL

Senior member of Data Onboarding & Governance for integration of healthcare data and enrichment of metadata. Configuration of data models, analytics, reports, and visualizations.

- Developed clinical data integration of 100+ billion structured & semi-structured records
- Developed topic modelling project on nearly 6 million semi-structured provider notes data
- Curated 25 pages of AI workflow documentation and mentored new hires in AI tuning process
- Created AI Surgical Scheduling Optimization tool currently evolving into new offering

Nov. 2017 - Dec. 2020

#### DATA CURATION ARCHITECT

Ruby, Pig, SQL/SSIS/Oracle, Hive/Impala, kafka, Kibana

Lead member of Data Onboarding & Governance, member of the Watson Foundation for Health Solutions Architecture team.

- Managed multi-org enterprise integration with over 70 interconnected data sources
- Developed new data validation and review process that reduced total rework by 25%
- Reduced mapping rework by 45% with through new knowledge-driven initiatives

## SOFTWARE DEVELOPER

*Sep. 2016 - Nov. 2017*  
*Ruby, Pig, Perl, Jenkins, SQL/MySQL/Oracle, DB2, PL/SQL*

Technical delivery lead for client integrations which has included the ordering, installation, and configuration of required hardware, including VPN devices to support client integrations.

- Reduced time spent on knowledge transfer by ~80% by reorganizing project documentation
- Eliminated need for a \$30,000 annual contract by developing data curation tool
- Reduced cost-per-hire by \$5,210 by re-developing education program

## DATA SCIENTIST

*Nov. 2015 - Sep. 2016*  
*Ruby, Pig, Jenkins, Hadoop/HDFS/CDH, SQL/MySQL/Oracle*

Data Ingestions, pre-analytic enrichments, analytic augmentation, data governance, and implementation of assigned client integrations.

- Managed team of 12 remote data science contractors for 36 integration projects
- Created over 125 new knowledge base articles, making up 30% of the team's online content
- Developed automated data integration tool to cut out 28% unnecessary development time

## DATA SCIENTIST CONSULTANT

*Jun. 2015 - Nov. 2015*  
*Ruby, Pig, React, NodeJS, JavaScript, SQL, Oracle, Jenkins*

Identified clinical, financial, and operational data elements within different health provider systems to develop data transformations and providing meaning to the data.

## AmTrust Financial Services

### SOFTWARE ENGINEER

*Mar. 2014 - Jun. 2015*  
*VB.NET, ASP.NET, C#, CSS, Javascript, HTML5, SQL, PL/SQL*

Created multi-tier MVC components and web forms for commercial insurance policy generating software.

## PPG Industries

### COLOR SCIENTIST

*Apr. 2012 - Jun. 2013*  
*VB.NET, C#, SQL, C++, FORTRAN*

Developed radiative transfer models, spectroscopic profile analysis, and cross-instrument correlation software.

## University of Cincinnati

### RESEARCH ASSISTANT

*Sep. 2006 - Sep. 2010*  
*FORTRAN, IRAF, IDL, JMP, LabVIEW, Maple, Mathematica, PDL*

Gathered data in the short-wavelength, crossdispersed mode SXD on a 0.8–5.4  $\mu\text{m}$  spectrograph at NASA IRTF.

## Volunteer

---

### CSforAll

#### INSTRUCTOR

*Oct. 2018 - May. 2019*

Our goal was to provide computer science instruction in all of the District's high schools, with a special emphasis on students who do not have access to a computer or a stable Internet connection at home.

### HIT in the CLE

#### COACH

*Mar. 2017 - Mar. 2019*

We were using a data science competition to educate students understand the potential of a career in health information technology and to assist IT companies in Northeast Ohio in developing a talent pipeline.

### Code for America

#### ETL GUILD

*Jan. 2015 - Jun. 2015*

The initiative was brought about to apply data science tools to improve and expand public record access.

### T.C.P. World Academy

#### TUTOR

*Sep. 2008 - May. 2009*

Our goal was to teach students how to understand the value of their learning experience and becoming academically involved by participating in society simulation activities for higher learning.

## Awards

---

2018 - 2022	<b>Gold Champion</b> , For 1,477 classroom hours and earning 43 digital credentials	IBM
2018	<b>Lab Services Award</b> , For coordinating internal teams to deliver on time	IBM
2017	<b>Eminence and Excellence Award</b> , For developing a new data curation tool	IBM
2012	<b>Color Lab Automation Award</b> , For expanding data collection and analysis reports	PPG Industries
2010	<b>Mary J. Hanna Fellowship</b> , For STEM students pursuing research-based Masters	UC

## Talks

---

<b>AI Hackathon 2020: Surgery Scheduling Optimization</b>	Dec. 2020
<i>"An algorithm uses past utilization and current waitlist info to predict surgery duration.</i>	IBM
<b>REACH Initiative: IBM Watson Care Insights</b>	2019
<i>Analyze patient records and real-time data via HL7 feeds, integrate insights into physician's EHR workflow.</i>	IBM
<b>Content &amp; Data Analytics Learning Exchange</b>	2018
<i>Introduction of implementation of machine learning algorithms using real-world health data analytics.</i>	IBM
<b>2018 Innovation Research Interchange Fall Networks Conference</b>	Sep. 2018
<i>Using NLP to extract insight from unstructured data and transform it from a data lake.</i>	IRI
<b>Cross Team Training Series</b>	2016
<i>A 7-part series on using Hadoop for data ingestion that shows advantages of data lakes.</i>	EXPLORYS

## Publications

---

<b>Surgical Scheduling Optimization and Procedure Duration Prediction <sup>a</sup></b>	Dec. 2020
<i>Improvement on traditional fixed ratio methods for total procedure time estimation.</i>	IBM
<b>Process and Requirements for CCD Data Integration</b>	Oct. 2017
<i>Processes and requirements for overcoming the many difficulties with CCD data curation.</i>	IBM
<b>AI Paint Formula Prediction from Spectroscopic Measurements <sup>b</sup></b>	May. 2015
<i>Naive bayesian color and effects matching algorithm of complex automotive paint formula.</i>	SHINYAPPS.IO
<b>Characterization of Surface Coatings using RT Models for Color Matching</b>	Feb. 2013
<i>Approximation of Chandrasekhar RT model for atmospheric scattering applied to color matching.</i>	USPTO
<b>Massive Stellar Clusters in the Disk of the Milky Way Galaxy <sup>c</sup></b>	Dec. 2010
<i>Spectroscopic and Photometric analysis of two open clusters in the galactic plane.</i>	OHIOLINK ETD
<b>Near-Infrared Spectroscopy of Candidates Members of the Galactic Cluster [BDS2003] 107</b>	Feb. 2008
<i>Near-infrared classification spectra for candidate members and comparative 2MASS photometry.</i>	PASP
<b>A Novel Method for Low-Toxic Photo-Etch Printing <sup>d</sup></b>	May. 2003
<i>Application of aquatint etching techniques to photo-etch procedures.</i>	MAPC

## Education

---

2022	<b>Certification</b> , Data Science <sup>a</sup>	IBM
2015	<b>Certification</b> , Data Science <sup>b</sup>	Johns Hopkins University
2010	<b>Master of Science</b> , Astrophysics <sup>c</sup>	University of Cincinnati
2004	<b>Bachelor of Fine Arts</b> , Printmaking <sup>d</sup>	Ohio University