

# Ben Bubnick

SENIOR AI DATA CURATION ARCHITECT · ANALYTICS TRAINING SPECIALIST

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

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

*"I want to use my 15 years experience in data science to reimagine healthcare by creating, assembling, and improving our business relationships so that all can benefit. I believe that when I inspire people to do the things that inspire them, we can improve the relationship between caregiver and patient together."*

## Summary

<b>Professional Summary</b>	Data Curation, Data Science, Analytics Training, Machine Learning, Scientific Analysis
<b>Big Data Platform</b>	Hadoop, Hive, Impala, Spark, kafka, Oracle, SQL Server, DB2, IMS, PL/SQL
<b>Industries</b>	Health Care, Financial & Insurance, Chemicals Manufacturing, R&D/Scientific Research
<b>Demonstrated Past Success</b>	Subject Matter Expert, Lab Services Award Winner, Eminence and Excellence Award Winner

I manage client resources and drive technical solutions to timely completion. I identify, analyze, and extract key data from heterogeneous systems; and lead technical requirements gathering, data discovery, documentation, and implementation of assigned client integrations. I also lead the technical delivery of assigned client integrations which has included the ordering, installation, and configuration of required hardware, including VPN devices to support client integrations. Configuration and development of software necessary to support client integrations, including ETL and Data Validation. And provide tailored analytics and AI solutions to our customers and internal clients.

## Experience

### Senior Data Architect

IBM

WATSON HEALTH

May, 2020 - PRESENT

Senior member of Data Onboarding & Governance integration of healthcare data and enrichment of metadata. Configuration of data models, analytics, reports, and visualizations. SME to project data architects, delivery, analytics, integration, etc. Directs the configuration of the data onboarding platform and associated infrastructure, and develop improvements to advance stakeholder's business priorities using tailored analytics.

- Led cross-function team as clinical data analyst for 2020/2021 measure updates project
- Created AI Covid-19 tool that reduced need for hospitalization identification time by up to two weeks
- Created machine learning Surgical Scheduling Optimization tool currently evolving into new offering
- Developed large clinical data integration in excess of several 100 billion structured and semi-structured records for life sciences analytics projects
- Led HIPAA patient deidentification project from more than 100 data sources for content analytics and measures
- Placed in the top 6% of all learners at IBM by completing 285 learning hours in 2020, and earning 8 badges in that time

### AI Data Curation Architect

IBM

WATSON HEALTH

Jan. 2018 - May, 2020

Lead member of Data Onboarding & Governance, member of the Watson Foundation for Health Solutions Architecture team. Remediation of data issues during UAT. Builds the data onboarding platform and associated infrastructure, and develop improvements to reducing customer implementation cycle time and data triage activities, and report on the "health of the data" during data onboarding. Tune machine learning classification algorithms for patient matching, and the internal review of similar work by my peers.

- Reduced mapping rework by 45% with through new knowledge-driven initiatives
- Developed new data validation and review process that reduced total rework by 25%
- Managed and developed ambitious multi-org enterprise with over 70 interconnected, project-sized components
- Earned AI certification to peer review tuning algorithm and reviewed over 50% of projects in the last 2 quarters
- Earned 7 badges during IBM Re-skilling Initiative and logged over 200 hours of training in 2018
- Curated 25 pages of AI workflow documentation and mentored new hires in AI tuning process
- Awarded Lab Services Award for coordinating internal project teams to deliver client needs on time
- Awarded GDP bonus for improving revenue per SaaS client by at least 1.34%

## Data Scientist

IBM, Explorys

WATSON HEALTH, EXPLORYS AN IBM COMPANY

Jun. 2015 - May. 2018

Data Ingestions, pre-analytic enrichments, analytic augmentation, and data governance. Identified and extracted key data from multiple heterogeneous systems; and led technical requirements gathering, data discovery, documentation, and implementation of assigned client integrations. Ordered and oversaw installation of & configuration of required hardware, including VPN devices to support client integrations. Configuration and development of software necessary to support client integrations.

- Eliminated need for a \$30,000 annual contract by developing data curation tool
- Principal integration specialist on over 60 projects for 20 different clients
- Chosen as team lead for ~12 remote data science contractors for 36 integration projects throughout 2015 - 2017
- Went above my duties to close project documentation gaps, organizing more than 100 past projects
- Created over 125 new knowledge base articles, making up 30% of the team's online content
- Reduced time spent on knowledge transfer by ~80% by reorganizing project documentation
- Awarded Eminence and Excellence Award in 2017 for analytics project
- Awarded GDP bonus for improving revenue per SaaS client by at least 1.17%

## Data Integration Training Specialist

IBM

WATSON HEALTH

Sep. 2016 - Jun. 2018

Project Management for technical training, providing both basic and advanced software training, along with consulting through cross-team and other training/presentation settings. Designed and managed the internal training, an effective onboarding program, and cross-team development training. Focus on HDFS technologies: Pig, Hive & Impala, Ruby & Java.

- Eliminated average of 3 weeks from employee onboarding process
- Reduced cost-per-hire by \$5,210 by re-developing education program
- Spearheaded new-hire mentoring program that led to ~4,500 increase in total billable hours
- Trained 60% of current data curation teams, including multiple managers

## Software Engineer

AmTrust Financial Services

COMMERCIAL PACKAGE POLICIES

Mar. 2014 - Jun. 2015

Developed web forms for insurance industry using ISO standards and created multi-tier MVC components, Web APIs, and UIs using VB, ASP, Javascript, & HTML5

- Development on two software architecture projects in 2015 and principal developer on offering product in 2014
- Developed over 65 ISO rating based web forms for insurance industry
- Collaborated with developers and business analysts to prioritize needs
- Steered development group into a unit testing project on multi-tier MVC components in 2014

## Color Scientist

PPG Industries

AUTOMOTIVE REFINISH

Apr. 2012 - Jun. 2013

Developed software and radiative transfer models on complex paint mixtures to determine formulation methods and for spectroscopic profile analysis on color match samples.

- Improved color matching classification algorithm performance by 40% through new numerical methods
- Wrote and filed application for patent on the application of the Radiative Transfer Equation (Chandrasekhar) to automotive coatings classification models
- Developed cross-instrument correlation software for risk analysis project
- Awarded Color Lab Automation Award in 2012 for color matching laboratory opening

## Adjunct Professor

various colleges

LORAIN COUNTY COMMUNITY COLLEGE, MIAMI UNIVERSITY, CINCINNATI STATE TECHNICAL AND COMMUNITY COLLEGE

Aug. 2010 - May. 2019

Developed curricula in introductory physics & astronomy courses, and technical presentations of complex physical concepts for major, non-major, and remedial students. Wrote concise laboratory instruction manuals, trained in precise measurement techniques & data analysis using DataStudio, and coordinated analysis documentation.

- Exceeded feedback evaluation goals by an average of 16% in 2017 and by 17% in 2018
- Tracked a 28% reduction in student confusion by reorganizing online education materials
- Uploaded over 700 hours of lecture material for convenient consumption

## Scientific Research

University of Cincinnati

DEPARTMENT OF PHYSICS

Sep. 2006 - Sep. 2010

Gathered data using SpeX, a 0.8–5.4  $\mu\text{m}$  spectrograph, at the NASA Infrared Telescope Facility. Observations in the short-wavelength, crossdispersed mode SXD. Multiple analysis projects comparing 2MASS with measured data, statistical classification without independent photometry for cluster properties, and photometric analysis bootstrapped to independent measurements.

- Volunteered 120 hours of Quality Control for the Cambridge Astronomical Survey Unit Variable Stars database
- Published 3 scholarly articles on near-infrared astrophysics over a 24-month period
- Awarded Hanna Research Fellowship in 2010

## Writing

---

### Surgical scheduling optimization and procedure duration prediction

IBM

TECHNICAL DEVELOPMENT TEAM - TOTAL PROCEDURE TIME ESTIMATION

Dec. 2020

- Improvement on traditional fixed ratio methods for total procedure time estimation, using linear regression models based on relevant variables.

### Process and Requirements for CCD Data Integration

IBM

LEAD INTEGRATION DEVELOPER

Oct. 2017

- Technical paper on processes and requirements, overcoming the many difficulties with flat file data curation of CCDs.

### Characterization of Surface Coatings using Radiative Transfer Models for Color Matching

PPG

RADIATIVE TRANSFER PHYSICS

Dec. 2010

- Approximation of Chandrasekhar radiative transfer model for atmospheric scattering is applied to color matching. (U.S Patent Application)

### Massive Stellar Clusters in the Disk of the Milky Way Galaxy

OhioLINK ETD

GRADUATE RESEARCHER

Dec. 2010

- Using spectroscopic and photometric analysis of two open clusters to determine the structure of the plane of the Milky Way Galaxy.

### Near-Infrared Spectroscopy of Candidates Members of the Galactic Cluster [BDS2003] 107

PASP

UNDERGRADUATE RESEARCH ASSISTANT

Feb. 2008

- New near-infrared classification spectra for nine candidate members and comparative 2MASS photometry for several cluster members.

## Talks

---

### IBM Watson Care Insights Reach Initiative

Watson Health

TRAINER

Apr. - Sep. 2019

- Watson Care Insights aims to analyze longitudinal patient records and real-time data points through HL7 feeds to provide practitioners insights into the patient's data, and address these challenges by integrating within the physician's workflow in the electronic health record system.

### Content & Data Analytics Learning Exchange

Watson Health

PEER PRESENTER

Apr. & May 2018

- Introduction of implementation of machine learning algorithms and their applications using real-world health data analytics examples.

### 2018 IRI Fall Networks Conference

Innovation Research Interchange

PRESENTER FOR THE INFORMATION SERVICES/INFORMATION TECHNOLOGY NETWORK TALKS

Sep. 2018

- Cognitive computing can use NLP to extract insights from the unstructured data that holds most of the relevant information, but also it can also perform the necessary transformation accurately at the earliest stages of the data integration pipeline.

### Cross Team Training Series

Explorys

LEAD PRESENTER FOR THE METL GROUP

Sep. 2016

- Seven part series in data ingestion approach using the Hadoop software framework. Data can be aggregated much more quickly and cost-effectively using a data lake framework than in traditional data warehouses, because of the speed and low-cost of massively parallel computing.

## Volunteer

---

2018/19 **Instructor**, Computer Science supplemental high school course hosted by BioEnterprise

CSforCLE/CMSD

2018/19 **Volunteer Coach**, Data Science High School Competition hosted by Case Western Reserve University

HIT in the CLE

2015 **ETL Volunteer**, RTAHeatMap project for Open Cleveland, a Code for America Brigade

Open Cleveland

2008/09 **Tutor**, On-site after school tutoring for T.C.P. World Academy

Community School

## Education

---

### Master of Science in Physics

University of Cincinnati

McMICKEN COLLEGE OF ARTS AND SCIENCES

2008 - 2010

- Thesis - Massive Stellar Clusters in the Disk of the Milky Way Galaxy

### Bachelor of Fine Arts in Printmaking

Ohio University

COLLEGE OF FINE ARTS

2000 - 2004

- Thesis - A Novel Method for Low-Toxic Photo-Etch Printing