

Michael Della Bitta

Providence, RI | 917-477-7906 | mdellabitta@gmail.com

* Summary

Hands-on engineering leader and developer with over 20 years of data, search, analytics, and multi-tier applications development experience. Extensive experience with software, data and cloud architecture.

* Experience

Digital Public Library of America | Director of Technology

March, 2024 – Present | Remote

- Single-handedly ran a large-scale data ingestion and transformation process that included over 40 direct partners, as well as thousands of indirect sources.
- Operated a media pipeline that made DPLA the largest contributor to Wikimedia Commons.
- Prepared systems, software, and processes for acquisition by a national-level partner.
- Oversaw technical criteria for acquisition selection process.
- Led a project to create a novel archive of social media posts in the runup to the U.S. election.
- Performed SRE functions, including defending against large-scale botnets.

Harvard University Library Innovation Lab | Engineering Manager

March, 2023 – March, 2024 | Boston, MA / Hybrid

- Released a public dataset of over 8M U.S. judicial opinions targeted at AI use cases.
- Led a project to retrieve and release over a million openly-licensed volumes from Harvard's contribution to Google Books, which became the founding dataset of the Institutional Data Initiative.
- Led decision-making on headless CMSes, incident response processes, and software architecture.
- Implemented an agile process for scoping, planning and executing on work.

Digital Public Library of America | Director of Technology

September, 2016 – March, 2023 | Boston, MA / Remote

- Primary Investigator of a Sloan-funded grant that established a pipeline between DPLA and Wikimedia Commons, which made the DPLA network the single biggest contributor of assets at over 3.1 million.
- Primary Investigator of a series of grants from the Sloan Foundation, Unfinished, and Netgain Partnership to develop a role for libraries to support research into the impacts of social media on society.
- Invested in tooling and skills and removed roadblocks that led to a year-over-year growth of 40% of items in DPLA's collections, surpassing the 47 million items mark.
- Re-imagined the Tech Team's data practice using modern data tools and architecture allowing for significant growth of capabilities.
- Directed development of Ingestion3, an Apache Spark-based data ingest, transformation, and publishing tool which performed orders of magnitude faster than previous tools in use at DPLA.
- Served as Primary Investigator for "Fostering a New National Library Network through a Community-Based, Connected Repository System," a multi-year, multi-million dollar grant

awarded by the Institute of Museum and Library Services in collaboration with DuraSpace and Stanford University.

ScribbleLive / Appinions | Engineering Manager (Previously: Senior Software Engineer)

2012 - 2016 | New York, NY

Team leader, big data and search expert at a flourishing NYC content marketing/analytics startup.

- Lead a team of NLP, data, platform, and front-end engineers.
- Served as Scrum Master and agile coach. Ran Scrum ritual meetings and review presentations.
- Designed, built and managed a 1 billion document / 2.5 TB Apache Solr installation.
- Drove adoption of continuous integration, automated testing, configuration management, static analysis, code reviews, security, Scrum, and Kanban.
- Planned and implemented cost-saving initiatives that saved ten thousand dollars off the company's monthly infrastructure costs.
- Discovered and fixed $O(N^2)$ complexity in core algorithm that reduced worst case performance from hours to single-digit minutes.
- Designed and led implementation of a ground-up rewrite of Appinions' core analytics engine which provided increased durability, discovery, and delivered vastly richer results.
- Company acquired by ScribbleLive in 2015, formerly Appinions, Inc.

New York Public Library | Senior Applications Developer | 2005 – 2012 | New York, NY

Primary developer of NYPL's Digital Repository implementation which securely stored and digitally-preserved +8M digital cultural heritage items and metadata.

- Streamlined the export and search indexing of NYPL's "Digital Gallery," reducing processing time from weeks to hours, and scaled concurrent request processing by a factor of eleven.
- Created an Apache Hadoop-driven service that performed format verification, digital signing and data migration of 100 TB of digital assets in two weeks.
- Pioneered search at NYPL via a Solr cluster which powers search across several products including the Library's website.

Columbia University | Analyst/Developer | 2001 – 2005 | New York, NY

Client- and server-side developer for Columbia University's initial foray into distance learning.

- Designed and built the content management system used to run Columbia Interactive, Columbia University's award-winning online education clearinghouse Web site.
- Co-designed and built a Java Web application for constructing XML ISO Topic Maps.
- Developed multimedia-rich educational websites for Columbia University's Digital Knowledge Ventures group.

*** Education**

Bates College, Lewiston, ME – Bachelor of Arts – Major: Philosophy

*** Technologies**

Python, Javascript, Typescript, Scala, Java, Ruby, Solr, Elasticsearch, Postgresql, Spark, Node, React, Next.js, EMR, ECS/Fargate, EC2, S3, SES, SNS, SQS, IAM, RDS, DynamoDB, Route 53