Geoffrey Malafsky

gmalafsky@technikinterlytics.com
Fairview, TX

CLEARANCE LEVEL: TS Current

SUMMARY RECENT ACTIVITIES

- Founder, CEO, Chief Scientist of Technik Interlytics developing highly automated expert system for deep inspection, correction, and unification of disparate data into High Fidelity data for AI/ML, Data Science, Compliance, Risk, Financials
- Founder, CEO, Chief Scientist of Colligo Real Data producing High Fidelity broad aggregated data sets for FinTech, PropTech, Supply Chain, eSports
- Chief Data Scientist Pentagon Comptroller financial audit data aggregation and reconciliation of transactional feeder, accounting in Big Data environment with customized business rules and significant data quality and interoperability challenges. Same for DoD CIO analyzing, collecting, merging, and mining disparate data sources for IT systems/assets.
- Chief designer, developer, and implementer of Big Data appliance for 100TB scale secure data aggregation and processing in on-premise secret MIL network
- Chief Data Scientist and Technology Reviewer for ongoing web and live radio broadcasts on new data related data technologies for AI, Big Data, and end use channels (xxxTechs)

EDUCATION

- Ph.D., Chemistry, The Pennsylvania State University, University Park, PA, 1989
- B.A, Chemistry, Cornell University, Ithaca, NY, 1978

CORPORATE HISTORY

- Colligo Real Data LLC, CEO, 2018-present
- Technik Interlytics LLC, CEO, 2018-present
- Phasic Systems Inc., CEO, 2009-2018
- TECHi2, President, 2000-2009
- SAIC, Dir C4ISR Innovation, 1997-2000
- Global Assc, Sr Research Analyst, 1996-1997
- Naval Research Laboratory, Staff Scientist, 1989-1996

TRAINING/CERTIFICATIONS

- Security+ (current)
- Linux (current)

- Systems Engineering
- MEMS

PROFESSIONAL ORGANIZATIONS/AWARDS

- IEEE
- ACM

OVERVIEW

Dr. Malafsky has 25 years of experience leading all phases of complex technical projects. This spans formal research programs as a scientist at the Naval Research Laboratory to designing and developing commercial software products for enterprise data normalization. He has served as technical expert in advanced programs for clients including DoD OSD (CIO, OUSD), DARPA, ONR, Dept of Navy, Dept of Energy, USMC, NGA, and Dept of Transportation. He has started and led two companies: TECHi2, an advanced technology consulting company where he was founder, President, and technical leader on numerous projects; and Phasic Systems Inc., a data science company where he was founder and CEO, supplying the DataStar Data Normalization software system and professional services.

He is regularly featured on web and radio broadcasts for Data Science and Big Data technology as an expert in multiple areas, including: Big Data, Data Science, systems architecture, Data Virtualization, and Data Normalization. He is the designer of the commercial system PSICLOPS which is a low-cost single unit Hadoop environment targeted for parallel processing of computationally intensive operations like Data Normalization, Data Mining, and BI analysis.

Dr. Malafsky has deep experience and expertise in identifying the root causes of large-scale data environment inconsistencies, adjudicating them in open facilitated work with clients, documenting the full business context, semantics, data models, business rules, and analytical logic in a variety of technologies and vendor tools. This specialized capability has been proven many times as solving complex challenges in a fraction of the time and cost as other methods, with substantially higher technical and organization success. This uses key technologies streamlined for rapid, visible, team-oriented success including: semantic models, common data models, NoSQL data integration, services based consolidated data stores, system architecture, and knowledge capture. He leads multi-day workshops on Data Science and Data Normalization that demonstrate how to use the framework and Data Science to normalize enterprise data for greater BI insights, application consolidation, and data accuracy.

He was the lead Data Scientist, Big Data system architect, data aggregation designer, financial data reconciliation rule developer, and senior executive consultant for change management program in DoD OSD to make pentagon financial data auditable. He led all aspects of the initial technical program including supplying the Big Data technology, developing architecture documentation, creating content management procedures, building custom Hadoop data processing Java code that automatically detected data errors and applied ML-based corrections, and worked with financial experts to elucidate and implement entire range of complex business rules for audit reconciliation. Program successfully overcame myriad challenges including severe source data quality and structural incompatibility. Processing grew to hundreds of concurrent users on system Dr. Malafsky managed involving working with Hadoop developer community due to industry leading usage.

He was Data Scientist on the DODCIO's IT Portfolio BI Project that collected, merged, and mined IT related inventory, programmatic, and financial data across the DoD to identify opportunities for cost reduction and consolidation, as well as to support regulatory reporting requirements within the DoD and

to OMB. Dr. Malafsky led team that collected, analyzed, and correlated the data definitions and actual semantics of data within multiple systems, including: DITPR, CADAT, SNAP-IT, FPDS, EDA, FedBizOpps. He identified numerous inconsistencies and unknowns in documentation and determined both root causes of problems as well as actual real data semantics with the DataStar system and Data Normalization methodology. The actual data was collected and put into accurate data structures, including a new common data model made for a normalized view across sources fitting DODCIO needs. This data was provided for analyst access via the DataStar system. Full data model analysis was performed and results documented in DataStar. This information was shared with collaborative offices including the DoD Inspector General and Defense Logistics Agency.

Normalized data was also compiled into custom data sets for BI presentation in a variety of tools which were installed and operated by Dr. Malafsky including: QlikView, Microsoft Excel 2013 SharePoint, PentaHo, and IBM Cognos. In addition, he is working with commercial applications to provide them with normalized system architectures developed and maintained in DataStar to provide flexibility and extensibility not available in the commercial tools while using the built-in analytics and reporting of the information model-based tools, including: Troux, and Mega. It is widely recognized that this work is providing DODCIO with data sets and understanding unavailable previously, and not available in other programs in other agencies.

The normalization processing was done in Hadoop environment for rapid parallel processing. This important capability allows frequent changes to the desired analysis results because the larger DODCIO team can collaboratively assess and document their business rules relative to the actual source data in DataStar system, which Is then used automatically in PSIKLOPS processing. This environment includes the main Hadoop components as well as additional open source tools providing: parallel processing, the Data Lake storage of structured and unstructured data, NoSQL data bases (HBase, Cassandra), relational databases (MySQL), predictive analytics (DataStax R), dashboard BI (PentaHo), and text analytics (SOLR).

He was also the technical lead and Data Scientist for a large client in the Financial and Risk Management field. They analyze risk, broker agreements, underwrite large international agreements. They needed common enterprise data for accuracy, consistency, global view for executive decision making and applications. They have many systems, both commercial and in-house, and multiple data sources (company, financial, history) that grew as project-based systems. In this environment, they tried the typical range of commercial products and consultants without success. They achieved rapid success using Dr. Malafsky's professional services as shown by a special meeting of their senior executive steering committee after only thirty days of contract where unanimous agreement making this work the center of their corporate data Governance, design, integration, and metadata repository. Continued work produced additional key successes including: ingesting from a variety of disconnected documents the architecture and specifications of primary corporate data marts, warehouses, and external partner interchange (Word, PDF, Visio, Excel, architecture tool output) which were automatically correlated into a full architecture of data models, glossaries, code dictionaries, and business system models in DataStar; identification of root causes of data inconsistencies in various groups use of main corporate data warehouse on Company data. The latter was done with the DataStar analytical features that showed specific issues that were then discussed and adjudicated in live meetings with the client facilitated by Dr. Malafsky. The result was merged into a new Company data model during the live meeting in DataStar and accepted for production use. Importantly, this live analysis and decision making also identified errors in the new system-system XML schema interchange specification which were corrected within the same day using the new information in DataStar.

Dr. Geoffrey P Malafsky, Data Scientist, External Technical Expert, Big Data Aggregation

He was the lead technical expert reviewer and corrective action architect for R&D programs in DARPA, ONR, and DOE. These included: peer-peer automated data fusion for urban warfare; MEMS X-band radar for low cost cruise missiles; cognitive technology for advanced command centers; ensemble acoustic phasing of MEMS sonar; data processing and MEMS sensors for condition-based maintenance; commercial transition roadmaps of advanced R&D technologies.

Additional sample experience on large projects as technical team leader and Chief Scientist includes:

Department of Navy SeaWarrior project: Led architecture analysis and design in SeaWarrior Systems Engineering and Architecture for functional requirements, Analysis of Alternatives, system data/hardware/software architecture, detail product descriptions; data element tables; technical standards and conventions; system security and contingency plans; test and evaluation of system components. Produced system architecture detailing activities, systems functions, and data requirements in the Department of Defense Architecture Framework (DoDAF) including: Shared Data Environment; Services Oriented Architecture (SOA); Consolidated data reporting and analysis system; SCORM Metadata Architecture; Integrated Business Rules Repository; Scalable Content Repository; Single architecture for afloat and ashore; Defense Readiness and Reporting System Navy (DRRSN).

National GeoSpatial-Intelligence Agency (NGA): In Knowledge Discovery and Dissemination program, analyzed various technologies developed by NGA to assess how well they would solve operations research challenges. Accessed tools from vendors, DoD and other government programs as well as researched and developed effective technology architectures and operations research systems to support implementation projects. Researched, analyzed and defined a modular federated architecture of distributed metadata registries, databases, and information sources coupled with logical metadata architecture for semantic-enabled knowledge collection, annotation and dissemination. Designed and developed prototype system (MyGEOINT) using semantic model for automated relevancy ranking of intelligence information.

Department of Navy Enterprise Data Environment: Led all phases of analysis, architecture development, system development, data model analysis and design, and data integration, correction, and normalization for a consolidated environment of Human Resources data from many conflicting and disparate data sources. The project covered data lifecycle processes involved in developing and operating an Enterprise Data Environment in a SOA environment. The goal was to integrate the varied functional processes for defining data requirements, building a flexible data model, defining rules for data validity and quality control, cleaning data, sorting large amounts of data efficiently, providing data efficiently in web-based services, and managing the continuous operations.