

Anthony S. Kosky, Ph.D.**CONTACT INFORMATION**

Home Address: 2219 Acton Street
Berkeley, CA 94702
Home Phone: (510) 883-1673
Mobile Phone: (510) 499-1673
E-mail: anthony@anthonykosky.com
Web: <http://www.anthonykosky.com>

SUMMARY

My background is in computer science research and I have more than ten years experience working within the bioinformatics and life sciences communities. I have extensive experience both as a manager of and technical contributor to large-scale commercial and research-oriented software development projects, particularly web-based and client-server applications with major data management and data analysis components. I have made significant contributions, both theoretical and practical, to the fields of computer science and bioinformatics, specializing in database systems and data integration problems, and have had my work published in leading computer science and bioinformatics journals and conferences. I have experience in designing and implementing systems for data management, exploration and analysis of molecular-biology, clinical and genomics data from wide varieties of heterogeneous data sources. I have managed large software engineering teams working on multiple projects under challenging circumstances, and have worked as an individual developer and in smaller development groups. I have a proven track record of bringing large-scale and ambitious software development projects to successful completion, both on-time and under-budget.

EMPLOYMENT

- June 2007 – Present: **Corimbia Inc.**
Consultant.
Consulting and implementation on Laboratory Information Management System (*LIMS*) for Childrens' Hospital Los Angeles. System implemented using Microsoft SQL Server and ASP.NET technologies.
- August 2005 – April 2007: **Axiopie Inc.**
Chief Technical Officer (August 2005 – April 2007)
Consultant (March 2005 – July 2005)
Strategic and technical planning and leadership for Catalyzer, an XML-based data management system for bio-medical and scientific research applications with both web-based and thick-client interfaces. Responsible for working with management team to set technical and business directions, as well as leadership of software developers, development plans and many other tasks as needed. Consulted on SBIR grant proposal for Data Transformation and Integration Tools to extend Axiopie's *Catalyzer* product line. See below for details.
- September 2006 – December 2006: **Haddon Hill Group, Oakland, CA**
Consultant.
Haddon Hill Group is a IT consulting firm working which had been working primarily in the financial sector. Worked with them in order to adapt their techniques and business plan for the Life Sciences and Biotech markets.
- September 1997 – December 2004: **Gene Logic Inc.**
Senior Director, Enterprise Application Systems (November 2003 – January 2005),
Director, Data Management Tools (August 2001- November 2003),

Senior Group Leader (August 1998 – August 2001),
Senior Computer Scientist (September 1997 – August 1998).
 Working in *Software Development* division. See below for details.

- September 1995 – September 1997: **Lawrence Berkeley National Laboratory**
Research Scientist.
 Working in the *Information and Computer Science Division, Data Management Research and Development Group*, on the *Object-Protocol Model (OPM)* toolset.
- November 1988 – August 1989: **STC Technology Ltd., Harlow, England.**
Software Engineer.
 Working in *Formal Methods Dept.* on *RAISE* (Rigorous Approach to Industrial Software Engineering) project – and ESPRIT funded project. Involved with the implementation of a language based editor and tool set for the *RAISE Specification Language*.

PROFESSIONAL ACTIVITIES (RECENT)

- **National Science Foundation**
Panelist for CISE Directorate, Division of Information and Intelligent Systems grants review panel.
- **3rd/4th International Workshop for Data Integration in the Life Sciences (DILS '06/'07)**
Program Committee Member

EDUCATION

- September 1989 – August 1995:
Dept. of Computer and Information Science, University of Pennsylvania.
 PhD. in *Computer Science*. Thesis on “Transforming Databases with Recursive Data Structures”, supervised by Prof. Peter Buneman and Prof. Susan Davidson.
- October 1987 – September 1988:
Dept. of Computing, Imperial College of Science and Technology.
 MSc./DIC (Distinction) in *Foundations of Advanced Information Technology*.
 Subjects Studied include semantics of programming languages, domain theory, theory of functions, functional programming technology, logic programming, models of concurrent computation.
 Dissertation on “Semantics of Object Oriented Programming Languages”, supervised by Prof. Samson Abramsky.
- October 1984 – June 1987:
University of Kent at Canterbury.
 BSc. (First Class Honors) in *Mathematics*.
 Winner of 1987 Rotary and I.M.A. prizes.
 Subjects studied include real analysis, complex analysis, numerical and computational mathematics, topology, discrete mathematics, relativity theory, fractal geometry and complex analytic dynamics.

TECHNICAL INTERESTS AND EXPERIENCE

- *Technical Interests:* Databases and data-management systems; data integration and transformation; schema evolution; object-oriented databases and complex data structures; database programming and query languages; biological and other scientific databases; web-based software applications; bioinformatics and gene expression data analysis.

- *Programming experience:* Extensive experience from programming large-scale, robust systems in C++. Experience with Java, UNIX scripting languages, etc. Experience with various CORBA products, C++ Standard Template Library (STL), XML, source control tools (SVN, CVS, RCS, etc.), Purify, etc.

ROLES AND RESPONSIBILITIES (AXIOPE INC.)

Chief Technical Officer: Responsibilities include working with CEO, operations and marketing teams, and potential investors, to set technical and business directions; project management for all software development; management of a multi-national software development group; writing functional requirements specifications, software documentation and white papers; writing SBIR grant proposals; interaction with customers, potential investors and consultants, and suppliers of third party equipment; software design and testing, and occasionally even a little bit of coding.

Catalyzer is a general purpose data management system which has been used for Electronic Lab Notebook (ELN), Electronic Data Capture (EDC), Image Management, Inventory Management and Specimen Tracking applications. It provides object-oriented data modeling capabilities, a graphical user interface allowing non-technical users to rapidly create and evolve databases, and interfaces to various image and document file formats, scientific equipment such as confocal microscopes, bar-code scanners and printers. An extensible client-server architecture makes it easy to develop new plug-ins and data importers, and to tailor Catalyzer to new applications areas. As CTO, I have been responsible for designing and overseeing the development of new versions and extensions to the Catalyzer product line to handle the data management needs of projects meeting regulatory requirements, such as HIPAA and 21 CFR part 11, to add support for new application areas, and to improve scalability and performance.

PREVIOUS ROLES AND RESPONSIBILITIES (GENE LOGIC)

Management Responsibilities

- *Site Manager:* Responsible for Berkeley satellite office of Gene Logic. Site has about 20-25 employees, including software developers, Software QA testers, project managers, technical writers, managers and administrative assistants. Responsibilities included overall organization, administration, personnel issues and morale, and representing the site to head office in Gaithersburg, MD.
- *Client-Server Development Group:* Responsible for development and maintenance of Gene Logic's Gene Express client/server product and related products. This included the GX Explorer client-side application and UI, the Analysis Engine and various other servers and middleware components, the Run Time Engine (RTE) interfaces, APIs and loading scripts, and so on.
- *Data Integration/Enterprise Systems Group:* Responsible for development and maintenance of GX Connect component of Genesis Enterprise System, allowing integration and management of customer and Gene Logic gene expression and related data. Also responsible for various custom data integration projects.
- *Systems Infrastructure and Integration Group:* Responsible for architecture, system and software requirements for Gene Logic's Gene Express products. Also responsible for deployment and configuration methods and tools, systems management and maintenance tools, update tools and procedures, etc.
- *Systems Engineering Group:* Responsible for requirements management and for specifying and managing the interfaces between software and other components of Gene Logic's information projects. Responsible for working with customer support, marketing, customers and other parts of the company to clarify and specify requirements, develop use-cases, specify training, analyze and verify issues, etc.

Projects

- *Genesis/GeneExpress (Genesis 2.0.1 – 2.0.3 maintenance releases and 2.5, 2.6 major releases):* Overall project lead for software side of Gene Logic's *Genesis* series of products, including *Gene Express*, *BioExpress*, and *Genesis Enterprise System*. This included coordinating the various teams producing the different software components, advising and working with marketing and higher management to plan strategies and monitor development, anticipating, assessing and dealing with technical issues, variances in plans, changes to requirements, and so on. The role requires an extensive knowledge of the entire software architecture, database requirements, underlying gene expression technologies, data models, and other concerns and requirements, both from internal groups and customers. Established quarterly cycle and mechanism for maintenance releases, used for 2.0.1 – 2.0.3, which featured various minor enhancements, bug fixes, and so on, and also allowed more major enhancements to be synchronized with these releases. Genesis 2.5/2.6 was a more major development effort, involving significant improvements in functionality and usability of many components of the system.
- *ASCENTA 1.0 Software:* Overall responsibility for software development for initial release of ASCENTA product. ASCENTA is Gene Logic's "entry-level" gene-expression data product. It provides access to aggregate data for sample sets classified by tissue type, disease, morphology, and other clinical attributes. This project involved working with product managers, scientists and developers who were not familiar with product development, and guiding them through the process to establish user requirements, expected usage patterns, etc., and to working to schedules and plans.
- *Genesis 2.0 Infrastructure and Overall Coordination:* Technical lead and overall responsibility for systems infrastructure for version 2.0 Gene Logic's *Genesis* series of products. This included determining system and software specifications and configurations for GeneExpress and Genesis (enterprise) systems; determining QA and acceptance testing plans and schedules; determining deployment and update procedures and methods; development of configuration and maintenance tools; developing upgrade tools and strategies, including data migration tools for Workspace File System; benchmarking and determining criteria for specifying systems based on number of users, data set and performance requirements; developing systems monitoring and performance tuning tools; etc. Also responsible for overall coordination and synchronization of schedules for GeneExpress and GX Connect projects, and for software components, such as *GX Launcher* that were common to various Genesis components.
- *Probe Intensity Analysis:* Project lead for Probe Intensity Analysis R&D project. This involves developing tools for rapid access of Affymetrix GeneChip probe-level data (i.e. CEL files) from various statistical and programming languages (R, Splus, C++, Perl), and various experimental efforts to make use of this data, such as trying to detect anomalous probe sets, improved summary data and so on.
- *Various Data Integration and Customization Projects:* Responsible for a variety of customization and professional services projects, including one data integration project that involved loading gene expression, gene annotation and sample annotation data from a custom LIMS system, with support for custom chips and incremental updates. Also responsible for certain customer interactions and providing technical support for custom projects.
- *GeneExpress 1.2 and 1.3:* Responsible for overall project management and technical lead for 1.2 and 1.3 releases of GeneExpress system, and some minor releases (1.3.1, etc.). Prior to 1.2 release, this role was not formalized, so responsibilities included putting in place standards and practices for project planning and management.
- *OPM Multidatabase Query System:* Responsible for development of Multidatabase Query System based on the Object Protocol Model and associated database development and management tools. This project included design, development and maintenance of the multidatabase query engine, development of query servers for accessing Oracle and Sybase databases, XML/SQL hybrid databases, BLAST and other bioinformatics tools. Also modeling various molecular biology databases in OPM (Genbank, Medline, Swissprot, MGD, etc.), and mapping certain flat file databases (Genbank, Medline) to hybrid XML/SQL databases. The system was deployed at Smith-Kline Beecham and a

number of academic and research institutions. The project also involved managing customer interactions and requests.

REFERENCES

Available upon request.

SELECTED PUBLICATIONS

A full list and links to all my publications are available at <http://www.anthonykosky.com/bib.html>.

- “Declarative Languages for Advanced Information Technology”, *Journal of Information Technology*, Vol. 3, No. 2, June 1988.
- “A Formal Model for Databases with Applications to Schema Merging”, in *Specification of Database Systems*, Glasgow 1991, Harper and Norrie (eds.).
- “A Basis for Interactive Schema Merging”, (with P. Buneman, S. Davidson and M. VanInwegen), in *Proc. Hawaii International Conference on Systems Sciences*, 1992.
- “Theoretical Aspects of Schema Merging”, (with P. Buneman and S. Davidson), in *Proc. Extending Database Technology (EDBT)*, Vienna, 1992.
- “Facilitating Transformation in a Human Genome Project Database”, (with S. Davidson and B. Eckman), in *Proc. Third International Conference on Information and Knowledge Management (CIKM)*, Gaithersburg, MD, 1994.
- “Observational Distinguishability of Databases with Object Identity”, in *Proc. 5th International Workshop of Database Programming Languages (DBPL5)*, Gubbio, Italy, 1995.
- “Exploring Heterogeneous Molecular Biology Databases in the Context of the Object-Protocol Model”, (with V. M. Markowitz and I. A. Chen), in *Theoretical and Computational Methods in Genome Research*, Suhai, S. (Ed), Plenum Press, 1997.
- “Facilities for Exploring Molecular Biology Databases on the Web: A Comparative Study”, (with V. M. Markowitz, I. A. Chen and E. Szeto), in *Proc. of the Pacific Symposium on Biocomputing*, Hawaii, January 1997.
- “WOL: A Language for Database Transformations and Constraints”, (with S. Davidson), in *Proc. 13th International Conference on Data Engineering (ICDE)*, Birmingham, United Kingdom, 1997.
- “Semantics of Database Transformations”, (with S. Davidson and P. Buneman), in *Semantics in Databases*, Springer Lecture Notes in Computer Science, 1998, Thalheim and Libkin (eds.).
- “Constructing and Maintaining Scientific Database Views” (with I. A. Chen, V. M. Markowitz and E. Szeto), in *Proc. of the 9th International Conference on Scientific and Statistical Database Management*, Olympia, WA, 1997, Hansen and Ioannidis (eds.).
- “Exploring Heterogeneous Biological Databases: Tools and Applications” (with I. A. Chen, V. M. Markowitz, and E. Szeto), in *Proc. of the 6th International Conference on Extending Database Technology (EDBT’98)*, Valencia, Spain, 1998.

- “Advanced Query Mechanisms for Biological Databases” (with I. A. Chen, V. M Markowitz, E. Szeto and T. Topaloglou), in *Proc. of the 6th International Conference on Intelligent Systems for Molecular Biology (ISMB'98)*, June 1998.
- “Object-Protocol Model Data Management Tools '97”. (with Markowitz, V.M., Chen, I.A., and Szeto), in *Bioinformatics Databases and Systems*, Stan Letovsky (ed), Kluwer Academic Publishers, 1999, pp. 187-199.
- “Seamless Integration of Biological Applications within a Database Framework”. (with Topaloglou, T. and Markowitz, V.M.), in *Seventh International Conference on Intelligent Systems for Molecular Biology (ISMB'99)*. Heidelberg, Germany, June 1999.
- “Extending traditional query-based integration approaches for functional characterization of post-genomic data”, (with B. Eckman and L. Laroco), in *Journal of Bioinformatics*, Vol. 17, No. 7, 2001, pp. 587-601.
- “Gene Expression Data Management: A Case Study”, (with V.M. Markowitz and I.A. Chen), in *Proc. Eighth International Conference on Extending Database Technology*”, Prague, 2002.
- “Integration Challenges in Gene Expression Data Management”. (with Markowitz V.M., Campbell, J., Chen, I.A., Kosky, A., Palaniappan, K., and Topaloglou, T.), to appear as a chapter in *Bioinformatics: Managing Scientific Data*, Morgan Kauffman / Elsevier Science, May 2003.