

**UC Tech Awards 2023 Candidate**

**Category: DESIGN (Also applied in the INNOVATION category)**
**Name:** Robert Krumm, Programmer IV, UROLOGY, UCSF Staff (1)
**Number of people:** (1)
**Location:** UCSF

1. **Person submitting the application/nomination**
	1. Robert Krumm, Programmer IV, UROLOGY, UCSF Staff
	2. **Email address:robert.krumm@ucsf.edu**
	3. **UCSF**
2. **Award category** Design, Innovation
3. **Name of person, name of the team, or name of the project to receive the award** Robert Krumm
4. **All project team members -** Albright, Jonathan <Jonathan.Albright@ucsf.edu>; Cowan, Janet <Janet.Cowan@ucsf.edu>; Lopez, Karen <Karen.Lopez@ucsf.edu>; Tenggara, Imelda <Imelda.Tenggara@ucsf.edu>; Odisho, Anobel <Anobel.Odisho@ucsf.edu>; Cooperberg, Matthew <Matthew.Cooperberg@ucsf.edu>; Mane, Sanika <Sanika.Mane@ucsf.edu>
5. **Which location was affected by the work?** Urology Oncology

**Summary** Despite huge advances in both open-source(R-studio) and commercially available analytics (Tableau) their use in medical research is impeded by the shortage of “analyzable” data. ERM systems like EPIC model the patient “chart” model which collects all available data into a folder that requires inspection to determine the course of treatment, lab results, pathology reports, and physician notes as created. The UODB project is designed to extract, transform and connect unstructured source into a fully relational data lake with a suite of web-based apps and compatible with any analytics suites.

**Narrative**

In recent years, the tech world has provided a variety of fabulous data analytical tools from open-source environment like R studio to commercial products like Tableau which allow professionals like the tech literate faculty of UCSF to directly perform sophisticated statistical analysis eliminating layers of friction in the research process. But all of this requires “analyzable” data, that is, highly structured, fulling integrated database that can produce the X and Y variables is the prerequisite for any sophisticated bioinformatic research. The UODB project is a standard SQL relational database with a single page style web UI interface that provides an end-to-end research environment that services all the member of the community to contribute to and extract from a single data source all of the data needed to support a new level of bioinformatics. The UODB consumes data from a wide variety of sources, then converts, transforms and extracts highly structured relational data. For example, the backend apps parse using various tools structured data from notes, smart data elements, full pathology report, Redcap databases and personal spreadsheets into highly analyzable formats such as the OMOP style defined by the Athena project. It allows direct connection between bioinformatics and biobank specimens so that given a set of research criteria it can determine what slides or images are available for the study. The level of integration allows for application that have not been possible before. One area is “rule” based concepts that allow for basic statistics, e.g., “how many prostate cancer patients have we treated in the last decade?” that current medical records cannot answer. The potential is unlimited.

**UC Tech 2023 Awards Program Background Information**

**Eligibility**

1. **Robert Krumm, Staff**
2. Those working closely **with tech and/or IT teams:**
	1. Imelda Tenggara,
	2. Janet Cowan,
	3. Anobel Odisho,
	4. Karen Lopez,
	5. Matthew Cooperberg,
	6. Jonathan Albright
3. Email addresses: Albright, Jonathan <Jonathan.Albright@ucsf.edu>; Cowan, Janet <Janet.Cowan@ucsf.edu>; Lopez, Karen <Karen.Lopez@ucsf.edu>; Tenggara, Imelda <Imelda.Tenggara@ucsf.edu>; Odisho, Anobel <Anobel.Odisho@ucsf.edu>; Cooperberg, Matthew <Matthew.Cooperberg@ucsf.edu>; Mane, Sanika <Sanika.Mane@ucsf.edu>
4. The project must be operational and have been implemented within the last **3 years**
May 2020-April 2023 continuing
5. Those working in **cross-location** and **cross-functional** teams are encouraged to participate No

**Criteria:**

1. **Complexity:** ERM model of patient centered, procedural event sequence was replaced with a fully relational object model where the ERM model structure are converted to attributes of data concept objects. ERM concept specific table/list model is replaced with a simple object model: object => Date/Time att, Patient att, Event att, Concept Att, Measurement Att. SQL/relation principle used to virtualize tables/List for backward compatibility.
2. **Impact:** The evolving apps now cover all of the domains of the research model that previously used disconnected data sources.
3. **Mission Alignment:** The project succeeded in using a single data source for all the tasks that require data support. Once created tasks that were not possible prior to its creation are now the basis of key research work. The extensible property of the data model allowed it you take on new, novel tasks not envisioned at the time of design. Example, the Biobank structure was extended from physical specimens to images of specimens created with Super HD scanners still totally integrated with all conceptually related data elements – e.g., provide the surgeon who performed the procedure from which the slide was created that was used to create the image in the inventory.