Public Hearing: Government Oversight of Forensic Science Laboratories

Assembly Standing Committee on Codes

Assembly Standing Committee on Judiciary

Assembly Standing Committee on Oversight, Analysis, and Investigation

Wednesday, February 8, 2017

Assembly Hearing Room 1923, 19th Floor
250 Broadway, New York, New York

Testimony by:

Sumana Harihareswara
Founder and Principal
Changeset Consulting
PO Box 6542
Astoria, NY 11106
sh@changeset.nyc

Thank you for inviting my testimony on the government oversight of forensic science in the state of New York. I'm here to discuss steps New York should take to improve the reliability and effectiveness of our practices in forensic science, and I'm grateful to the Assembly Standing Committees on Codes, on Judiciary, and on Oversight, Analysis and Investigation for your time.

Specifically, today I'll be discussing the software we use in our forensic science labs. I am a programmer and manager who has worked in the software industry for over a decade, and I am the founder of Changeset Consulting, where I work on multiple public sector and private sector software projects.

First I'll talk a little about two different types of software, proprietary and open source, and why the distinction is important. Then I'll go into reasons why the open source approach would help with auditability, transparency, cost, and efficiency in our labs. And then I'll make some requests of you, regarding procurement and validation.

Proprietary and Open Source

We use both proprietary and open source software in all parts of industry and government. To explain the difference, and where open source is superior, I need to explain what source code is.

When programmers write software, what we write is called "source code". You can think of it as a recipe for the computer to carry out, like a recipe for baking a loaf of bread, and then the application or system that you use, like Microsoft Windows or Google's GMail, is like the finished loaf of bread.

Proprietary software is also known as “closed source” software. We, the users, can use the end result, but the vendor who makes the software won’t let us see the recipe, so we don’t know what they’re putting into the bread. We have to simply trust our vendors to give us quality software, even though they’ll turn around a month or several years later and tell us to upgrade because the new version fixes a slew of defects.

In our forensic labs, for example, in New York State we pay Porter Lee Corporation to use, for instance, their closed source software Crime Fighter BEAST (Bar Coded Evidence Analysis Statistics and Tracking) and Laboratory Information Management System (LIMS). For service between 2015 and 2019, we’re paying them $749,333.00.[1] Instruments are also affected; for instance, we have paid Applied Biosystems for equipment that comes with closed source software.[2] And because it is proprietary, we cannot inspect the source code to see what mistakes it might be making, and we can’t improve it ourselves.

Open source software, on the other hand, is where we get to see the recipe. Open source software is software that may be inspected, modified and redistributed freely by anyone.[3] When software is open source, there is a specific vendor who takes charge of integrating suggested improvements for everyone’s benefit. If you use an Android phone, or browse the web using Chrome or Firefox, or if your website uses WordPress or Drupal, for example, you’ve used open source software. Your offices, as well as other government agencies

[1] Contract C001714, retrieved via http://wwe2.osc.state.ny.us/transparency/contracts/contractresults.cfm?ID=1379821 . Since early 2012 the state’s payments to Porter Lee Corporation have totalled $365,017.44, per a search on http://wwe2.osc.state.ny.us/transparency/checkbook/ .

[2] The Division of State Police paid $9,233,067.89 to Applied Biosystems/Life Technologies Corporation (owned by Thermo Fisher Scientific) between 2012 and early 2017.

[3] Open Source Initiative’s Frequently Asked Questions https://opensource.org/faq

(including national security and intelligence agencies)[4] frequently choose open source software, because it's more trustworthy, because we can fix defects and share those fixes with other users, and because we don't have to pay license fees. A very popular example is Linux, which we often use to run website servers.

One more distinction: using **open source software** does not mean we have to open any **data**. Whether we're using proprietary or open source software, I'm not proposing any changes today to our data protection rules.

Auditability and Transparency

We need auditability and transparency in the science our labs do, and in their administration, such as chain of custody tracking. I'll speak first about the science.

The 2016 PCAST report[5] focused on:

> (1) the need for clarity about the scientific standards for the validity and reliability of forensic methods and (2) the need to evaluate specific forensic methods to determine whether they have been scientifically established to be valid and reliable.

And software is prone not just to big defects that crash the system so we can easily find them, but to intermittent defects that give us wrong output. The Association for Computing

[4] The New York State Senate, for example, is working to open the source code for all of its software projects, per its webpage https://www.nysenate.gov/coming-soon , and is working to help volunteers and businesses who are writing open source software that can help government work better: "Government tech stakeholders gather at state hackathon", January 22 2013, https://opensource.com/government/13/1/hackathon-hosted-new-york-state-senate .

[5] "Report to the President: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods; Executive Office of the President; President's Council of Advisors on Science and Technology", September 2016, https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf

Machinery's US Public Policy Council (USACM) in January 2017 said, in its "Statement on Algorithmic Transparency and Accountability"[6]:

> An algorithm is a self-contained step-by-step set of operations that computers and other 'smart' devices carry out to perform calculation, data processing, and automated reasoning tasks. Increasingly, algorithms implement institutional decision-making based on analytics, which involves the discovery, interpretation, and communication of meaningful patterns in data. Especially valuable in areas rich with recorded information, analytics relies on the simultaneous application of statistics, computer programming, and operations research to quantify performance.
>
> There is also growing evidence that some algorithms and analytics can be opaque, making it impossible to determine when their outputs may be biased or erroneous.

Being able to see the source code is crucial to auditing, accountability, and transparency–to understanding the decisions our forensic software is making, and systematically checking whether those decisions are appropriate.

On the administrative side, we lack proper auditability in our forensic science labs partly because of deficiencies in our software. For instance, I spoke with a lab worker who said that LIMS structures its permissions for who can do data entry in a counterproductive way. In order to give workers the ability to update records of evidence, LIMS administrators also need to give those workers the ability to completely delete a record of evidence at any time. And when we use proprietary software, it's harder for us to get different pieces of equipment to talk to each other, so workers have to manually retype figures or findings from an instrument into a computer, introducing the possibility of human error and reducing auditability.

[6] January 12, 2017, "Statement on Algorithmic Transparency and Accountability", https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

Cost and Efficiency

Now I'll talk about getting value for our taxpayer dollar.

With proprietary software, we don't know exactly what we're paying for. I'll use the example of BEAST/LIMS. I spoke with a lab worker who said it's often unclear whether a new software update for BEAST/LIMS is available, or has gone out to a given lab, and we don't know whether those updates have fixed known issues in the software. In 2013, New York's Inspector General uncovered a problem with the interaction of police and a forensics lab that stemmed from a lack of update notifications generated by BEAST as of 2009.[7] Neither the lab worker nor I were able to confirm whether BEAST has addressed this issue.

Applied Biosystems and Porter Lee Corporation are both vendors with a lock on the market that are, according to the lab workers I've spoken with, unresponsive to defect reports and requests for support or improvements, even when we are paying for support contracts on top of expensive licenses. For example, New York state labs have been asking for some time for a module that lets them link their DNA instruments with BEAST/LIMS, but, according to the lab worker I spoke with, Porter Lee has not provided it. I am mentioning these vendors as examples, not to single them out, but because this is a common dynamic with vendors of proprietary software.

[7] State of New York, Office of the Inspector General, Report of Investigation of the Onondaga County Health Department Center for Forensic Sciences, April 2013, Catherine Leahy Scott. https://www.ig.ny.gov/sites/default/files/pdfs/OnondagaCountyFullReport.pdf "This case was initially reported as a "Shot(s) Fired" case in the BEAST system rather than a homicide...the case status was changed to a homicide in the BEAST system on December 16, 2009. However, when SPD personnel were asked if updating the case type in the system generates a notice of the change to the Crime Lab, they said they did not know. In fact, the Crime Lab advised the Inspector General that an update does not generate such notice."

If we were to contract with a vendor or set of vendors to build or improve open source software for our labs, we would have less vendor lock-in, and it would be easier for our IT staffs to track down defects and integrate different systems together. Once companies and agencies other than the main vendor have the ability to suggest improvements, you can increase the quality of labs' tools faster, which can increase labs' throughput and reduce backlogs. And given how much money we spend on this software, for licenses and support, if we band together with other state governments, an open source alternative would benefit us all.[8]

Procurement

The New York State Office of Information Technology Services already directs all agencies to consider open source software when purchasing;[9] I'd ask you to further require that, in the procurement of software for forensic laboratories, open source software not just be considered but preferred. I'd also ask that we require that vendors be able to import and export data in standardized, open formats, and document and publish the formats they use, to allow labs to more easily chain together instruments and other software.

[8] "What's the Return on Investment for Open?" by Karl Fogel, October 15, 2010, Civic Commons. http://archive.civiccommons.org/2010/10/roi-of-open/ "if your jurisdiction

- plans to use the software for a long time, and
- therefore plans to maintain the software anyway, and
- it's something other jurisdictions might need too

…then there's a good chance that opening it up could be a responsible decision."

[9] New York State Information Technology Policy, IT Policy: Enterprise Plan to Procure Policy NYS-P08-001, https://its.ny.gov/sites/default/files/documents/NYS-P08-001.pdf

Validation

In order to find our software's intermittent defects, especially when the software is performing operations that humans cannot double-check fast enough for police needs (like probabilistic complex DNA genotyping or massive facial recognition operations on hundreds of thousands of faces), isolated validation tests and the attestation of the vendor is not good enough. We need a mechanism of independent verification and validation. We should work with our labs, the Commission on Forensic Science, and other forensics bodies to move towards IEEE[10] software standards for V&V processes, per the recommendations in International Society for Forensic Genetics.[11] And we should facilitate uniform implementation across the state.

Thank you, and I welcome your questions.

[10] Institute of Electrical and Electronics Engineers.

[11] "DNA Commission of the International Society for Forensic Genetics: Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications", September 4, 2016, FSI Genetics.