OREGONLIVE

The Oregonian

OREGONIAN MEDIA GROUP

April 25, 2017

Dear Chair Williamson and members of the committee:

HB 3399 is about efficiency. These days, government workers routinely use databases to collect and compile information to do their jobs. The information stored in these databases belongs to the public. But as a reporter, I can tell you, it doesn't often feel that way.

In my job as a data reporter at The Oregonian, I frequently request spreadsheets and database tables from government agencies. I often start by asking for documentation of that data. There are many terms for this kind of information – record layouts, database schema, data dictionaries, code books – but essentially, records often exist that describe the indisputably public information stored in these structured formats. These descriptive records should be public. That's the logic of this bill.

As a reporter, I want to know what fields a database contains and how, in the general sense, those fields are organized. That information is important to me because it tells me what I can request efficiently. From the record layout, I can get a general sense of how complicated a given export query might be and which fields I'm willing to give up, without sacrificing accuracy or understanding while saving myself and the agency the expense of legal review. It tells me what questions I should be able to answer with the data and what patterns I might be able to spot. Identifying these patterns and answering these questions, I think you would agree, is in the public interest.

HB 3399 would require public agencies to make sure software vendors can export data in an open format – that is, allow data to be analyzed to identify bottlenecks and breakdowns in government services and tighten the relationship between public goals and public outcomes. The bill would also make it explicit that the record layouts I described earlier are public and not subject to any exemptions.

I came to Oregon from Arizona, where I had few problems getting this type of information. In that state, metadata -- data about data -- is public. Period. So among the first things I did in my job here was ask Portland Police Bureau for an inventory of databases and related documentation. I was astounded by the response. I learned that the company Portland chose for its new records

management system was Versaterm and Versaterm considers record layouts to be trade secrets, at least in Oregon, where it can claim that. To be clear, I was not asking for any computer code or any documents that would describe Versaterm's special sauce -- its interfaces or behind-the-scenes magic. I was asking for the equivalent of a table of contents. And I was denied.

As a result, data requests from Portland can feel like a game of 20 questions. I'm sure you can imagine that's frustrating to the agency as well as to me, the reporter. Let me give you another example. I knew there were questions about who Portland was labeling gang members and whether those designations were appropriate. I read the bureau's directives to learn about that process, and I discovered that Portland kept what it called a "gang affiliates list." I asked for the corresponding data. I knew from an RFP for the Versaterm system that Portland had kept a table in its previous database that contained information about designated gang members, and I knew the solicitation was asking for something similar. I didn't know, because I didn't have a record layout, that the relevant fields were scattered among database tables in the new Versaterm system. This created a far more drawn out and painful negotiation process than was necessary or expedient. The bureau quoted me more than $1,000 to provide a list clearly referenced in their own directives.

I want to emphasize that this bill does not seek to do something revolutionary. Software companies that serve public sector clients should understand that data must be exportable, as the data belongs to the public. And any agency that hopes to evaluate its work based on the data it collects must have a record layout. Professional and reliable analysis requires a shared understanding of meanings.

Some Oregon agencies demonstrate this professionalism. The Judicial Department, for example, provides record layouts for its court data. Those documents describe which tables exist, which fields are in which table and the meanings of the codes used. That facilitates meaningful analysis, such as the Unequal Justice series published by our colleagues at InvestigateWest and Pamplin Media.

We hope that you will see the value of extending this professionalism to other agencies across the state when they seek new database software.

Now I'd like to turn to HB 3361, a bill that seeks to do something different, but complementary. HB 3361 would create a state data portal and appoint a chief data officer, who would set data standards and manage an inventory of state

data sets. I support Oregon adopting a forward-looking attitude toward data, but I have a couple of lingering concerns about this bill.

First, HB 3361 outlines which data should be considered publishable or not publishable on the state's web portal. The bill should be amended simply to state clearly that it does not in any way alter an agency's obligations under Oregon public records law.

Second, HB 3361 incorporates the concept of the "mosaic effect," an idea that comes out of a national security context. It's been around since the 1960s, but was revived post 9/11. This concept relies on hypothetical abstractions, not real risks. It can telescope infinitely -- one piece of information here can be combined with some hypothetically existent piece of information there to reveal something damaging that cannot be described here in this minute.

Let me give you two examples of how this logic has been used by the federal government. In a 1987 case, the feds argued that "anyone possessing the employment histories of DEA agents could piece together a mosaic of the agency's worldwide structure, capabilities, and enforcement activities." In 2004, the court upheld this kind of extreme hypothetical thinking. The ACLU sought the total number of section 215 FISA applications placed by the FBI. It did not seek access to the applications' dates, disposition or content. But the government still denied this request, arguing that if the number the ACLU requested -- a single number -- were combined with the number of FISAs authorized, the number of cases opened and closed each year, "a database could be built with relative ease which would reveal a detailed road map of how the FBI conducts its investigations." I hope it is clear that both of these hypotheticals are nonsensical. This concept has been tailored to stoke fear and expand secrecy. It has no place in Oregon law, if we are to continue to be a state that values transparency and accountability.

Sincerely,

Carli Brosseau
The Oregonian