CITY COUNCIL
CITY OF NEW YORK

------------------------ X

TRANSCRIPT OF THE MINUTES

         Of the

COMMITTEE ON TECHNOLOGY

------------------------ X

                    September 17, 2018
                    Start: 10:22 a.m.
                    Recess: 11:52 a.m.

HELD AT: 50 Broadway – Committed Rm.
                 14th Fl.

B E F O R E: PETER A. KOO
                  Chairperson

COUNCIL MEMBERS: Robert F. Holden
                  Brad S. Lander
                  Eric A. Ulrich
                  Kalman Yeger

A P P E A R A N C E S (CONTINUED)

Emily Newman, Acting Director, Mayor's Office of Operations, MOO

James Perazzo, Acting Director, Mayor's Office on Data Analytics, MODA

Don Sunderland, Deputy Commissioner of Data Management and Integration, Department of Information Technology and Telecommunications

Pauline Toole, Commissioner, Department of Records and Information Services, DORIS

Alex Camarda, Senior Policy Advisor, Reinvent Albany

Rachel Byrd, Public Utility Law Project

Noel Hidalgo, Beta NYC

Yale Fox, CEO of Rentlogic

1

2   [sound check] [pause] [gavel] [background

3   comments]

4   CHAIRPERSON KOO: Good morning. I'm

5   Council Member Peter Koo, and I am the Chair of the

6   Committee on Technology. I want to welcome you all

7   to our hearing. The hearing will focus on Intro 986,

8   a Local Law in relation to the format of data in

9   agency reports. Intro 1094 sponsored by the Speaker,

10   a Local Law in relationship to oversight access to

11   agency data, and Intro 1098 sponsored by Council

12   Member Kallos, a Local Law in relation to

13   digitization of historic data. The Open Data Law and

14   the resulting Open Data Portal have by almost any

15   measure been a success. An open data census

16   acknowledged New York City as a leading municipality

17   in providing data transparency to the public through

18   the Open Data Portal. This success is credited both

19   to the underlying law itself as well as to the

20   implementation and hard work that has been

21   demonstrated by the Department of Information

22   Technology and Telecommunications, and the Mayor's

23   Office of Data Analysis and Data Analytics. The

24   hearing will focus on how government data can be

25   easily accessible on propriety and machine readable

2 for further use, studies and analysis. We expect to

3 see continued efforts in providing accessible data

4 and data transparency. I look forward from the

5 panels today, and I would like to thank the

6 Technology Committee staff by putting together this

7 hearing. I would also like to recognize the

8 Technology Committee members, which will—they come

9 later. Thank you. So—so we're going to have the

10 first panel from the Mayor's Office of Operations

11 leading by Emily Newman, and oversees the Department

12 of Records Information Services, Pauline, too, and

13 also have the Department of Information Technology

14 and Telecommunications. So, will counsel swear the

15 members in? Yeah. [pause]

16   LEGAL COUNSEL: I need to ask you to

17 raise your right hand. Do you swear to tell the

18 truth, the only truth, and nothing but the truth

19 before our committee in your testimony today?

20   EMILY NEWMAN: Yes.

21   PANEL MEMBERS: [in unison] Yes.

22   LEGAL COUNSEL: Thank you. You can

23 start.

24   EMILY NEWMAN: Great. Thank you. Good

25 morning, Chairman Koo and members of the Committee on

2   Technology. My name is Emily Newman. I'm the Acting

3   Director of the Mayor's Office of Operations. Thank

4   you for the opportunity to testify today on Intro

5   1094, a Local Law in relation to oversight access to

6   agency data, and Intro 986, a Local Law in relation

7   to the format of data in agency reports. I'm joined

8   today by Laura Negron the City's Chief Privacy

9   Officer; James Perazzo, the City's Acting Director of

10   the Mayor's Office on Data Analytics, and Don

11   Sunderland, Deputy Commissioner for Data Management

12   and integration at the Department of Information

13   Technology and Telecommunications all of whom will be

14   available to answer questions. The Mayor's Office of

15   Operations works to make New York City Government

16   more effective and efficient. Operations includes the

17   Mayor's Office of Information and Privacy and the

18   Mayor's Office on Data Analytics. The Office has a

19   key role in meeting the city's work on open data and

20   privacy protection each relevant for today's hearing.

21   The Administration shares the Council's belief in

22   effective data sharing among city entities and

23   supports data focused research. To this end, the

24   administration has implemented various initiatives to

25   help facilitate data access and exchange across city

2 entities while striking a balance with important

3 legal privacy considerations as well. Being nine

4 months pregnant means you run out of breath quickly.

5 So, I apologize. [laugher] A lot of words. A key

6 example of the Administration's commitment to data

7 sharing is the Citywide Data Integration Initiative

8 created in 2015 as part of the Administrations 10-

9 year strategic plan, One NYC to help strengthen the

10 city's capacity for data integration. Developed and

11 managed by Operations, this initiative provides a

12 privacy—a privacy compliant one city approach to data

13 using a centralized technology platform to ensure

14 more effective and efficient use of city resources.

15 This framework requires an approved scope of work and

16 executed legal project agreement for every multi-

17 agency data sharing project involving identifying

18 information in which—in which each and every data

19 element or category requested must be authorized by

20 relevant agency privacy officers. While this is a

21 time intensive process, this approach helps to

22 advance important cross-agency work while at the same

23 time ensuring that the city complies with applicable

24 state, federal and local laws that protect New

25 Yorkers personally identifying information. While

2 this initiative is complex, it helps make key city

3 programs possible including Homestat and Pre-K for

4 all. We'd be happy to provide further information

5 about this initiative and discuss how the Council may

6 participate to help advance its data and research

7 goals. This Administration is also committed to open

8 data, the policy that makes city data available to

9 the pubic wherever possible. Since 2012, the Open

10 Data Portal has grown to include more than 2,000 data

11 sets, and that number is growing. New York City Open

12 data is a world class program with a citywide scope.

13 It's highly visible cross-agency program dedicated to

14 transparency and open governments while also

15 encouraging research and analytical best practices.

16 Open data has been used by New York City residents

17 and City agencies to conduct research and inform

18 important policy decisions. For example, open data

19 includes the city's tree census and information that

20 helps emergency responders formulate the best routes

21 to get to where they need to go. There are many

22 other examples some of which can be viewed in this

23 year's Open Data for All Report. This administration

24 is deeply committed to protecting the privacy of New

25 Yorkers' personal information and advancing privacy

2 best practices. In furtherance of this commitment

3 the Mayor's Office of Information Privacy was

4 established by executive order in 2018. Protecting

5 the privacy of sensitive personal information is

6 critical to ensure compliance with applicable laws

7 and regulations and promote residents' trust in their

8 government. This is particularly relevant for

9 vulnerable people who may be harmed in their—who may

10 be harmed if their information is improperly shared.

11 We know the Council is similarly committed to

12 protecting New Yorkers' identifying information as

13 evident in Local 245 and 247 of 2017, known together

14 as the Identifying Information Law. These laws,

15 which establish the Chief Privacy Officer roles

16 restrict the collection and disclosure of identifying

17 information across more than 175 city agencies and

18 offices. While we share the Council's values around

19 data and privacy, we would like to provide comments

20 on two of the bills on the docket today. As written,

21 we believe Intro 1094 is not a feasible strategy for

22 efficiently accessing city data for three primary

23 reasons. First it empowers the Chief Privacy Officer

24 with the sole authority to approve Council employees'

25 access to any information held in the city's Central

2 Data Platform. Authorizing access to city data for

3 research and analytic purposes requires a fact

4 specific legal review and determination made in

5 collaboration with agency privacy officers. This

6 review is based on the laws, and legal privileges

7 protecting the confidentiality of the information.

8 From our experience, implement the citywide Data

9 Integration Initiative we believe this approach

10 proposed would not expedite access to data. From a

11 feasibility standpoint, it is not possible for the

12 Chief Privacy Officer to assess every data request

13 made under this bill within the proposed timeline.

14 Nor would it be for any city official or agency to

15 review within days, every relevant record for the

16 potential application of each confidentiality

17 protection or legal privilege before providing access

18 to City Council members and staff. Secondly,

19 personally identifiable information is heavily

20 governed by many federal and state laws that the city

21 cannot overcome by Local Law even if a privacy

22 training course were completely. Lastly, [coughs]

23 based on our conversations with our partners at

24 DOITT, we understand that a clean room is not an

25 industry standard, and would not mitigate the risk of

data being misused or removed from city computers. We

welcome the opportunity to discuss alternative

strategies with the Council for efficiently accessing

city data where permitted by law. We also want to

take this opportunity to highlight the value of open

data and the richness of what the Open Data Law

requires and provides. As mentioned earlier, many

agencies and offices currently use open data to do

extensive internal research, and the tool is useful

to both the public and City employees wishing to

conduct effective, important research. Under the

Open Data Law any regularly maintained data that

appears in a report or can otherwise be made public,

is either already public or in the process of being

made public—I'm sorry, of being made available on

Open Data Portal. In contrast, Data Bridge is a

piece of tech—of technical architecture and not

ultimately a data source itself. We would,

therefore, encourage a continued conversation about

the best way to maximize and fully leverage the

information already available through open data for

the Council's analytic and research purposes. With

regard to Intro 986, Operations similarly understand

and agrees that what seems to be the spirit of this

2 bill—with what seems to be the spirit of this bill.

3 Data on Open Data should not only be available, but

4 also reasonable accessible meaning that it could be

5 used for analytic—for analysis readily. Most data

6 that is currently in publicly available reports is

7 available on the Open Data Portal with few exceptions

8 where it is either not feasible, legal or meaningful.

9 We look forward to discussing this bill further as

10 well, and hope to work with the Council to find a

11 solution. Thank you for the opportunity to testify

12 today. As you know, this Administration is dedicated

13 to using research and data to make informed policy

14 decisions and improve the lives of all New Yorkers.

15 We look forward to collaborating with you to find

16 workable strategies and solutions for accomplishing

17 our shared goals. Thank you. [pause]

18   COMMISSIONER TOOLE: Good morning, Chair

19 Koo and members of the Committee on Technology and

20 staff. I am Pauline Tool the Commissioner of the

21 Department of Records and Information Services,

22 commonly called DORIS. DORIS is responsible for

23 preserving and providing access to the historical and

24 contemporary records of New York City government

25 ensuring that city records are properly maintained

2 following professional archival and record manager—

3 management practices and making city governments

4 records available to diverse communities to the

5 municipal archives and the municipal library.   The

6 department shares the City Council's goals to

7 increase accessibility and usability of the city's

8 historic records. In Fact the Municipal Archives has

9 been overseeing the development of an open sources

10 integrated solution to preserver and provide access

11 to both archival and library records. It will be

12 fully operational in 2020. We are currently storing

13 over 185 terabytes of digital archival information

14 both born digital and newly digitized. We expect

15 that number to grow exponentially as we continue to

16 acquire born digital records, and as we digitize

17 historical records in various formats. This solution

18 will ensure the preservation of historical records

19 over the long term and provide access to that

20 material to the public. Digitization efforts in the

21 past year have included nine million historical vital

22 records, executive orders, liquor licenses, borough

23 president photographs, Department of Finance 1940s

24 tax photographs, Alms House Records, bodies in

25 transit records and more. Our online gallery host of

2   1.1 million photographs and records that are freely

3   available to the public from anywhere. Our end goal

4   is to make the library and archival records available

5   online worldwide. Intro 1098 by Council Member

6   Kallos would require archivists to conduct a search

7   through 246,000 cubic feet and 185 terabytes of

8   historical records, correspondence, maps, drawings,

9   building plans, photos, genealogical records, film,

10   et cetera to locate information in a non-narrative

11   form, assess the value of putting that information

12   into another format and include that newly created

13   document on the Open Data Portal. The value of hard

14   records alone is enormous. The equivalent of three

15   Olympic size swimming pools filled with boxes of

16   paper. It would require decades of dedicated staff

17   time to implement this proposal, and the exercise

18   would yield very little of public value. The value

19   to the public of the archival collections is that

20   they document city government's activities and

21   primary source records. This history can't be

22   tabulated. It's rich and nuanced and requires people

23   to read and draw conclusions, and offer the insights

24   they learned from perusing the collections.

25   Requiring the archivists to extract statistical

2   information and create new records removes the

3   information from its original context. This would

4   up-end decades of practice archivists appraise,

5   preserve and make available entire collections in an

6   unedited format for research purposes, Creating

7   subsets of data drawn from the archival records is

8   the role of researchers not archivists, and I want to

9   just reference a document I could share with you

10 later. It's a copy of the Flushing Lists, and it is

11 a list of possessions taken from a woman's estate

12 when she died. You don't know any—if you just put

13 that information online, you wouldn't know where was

14 she from, what was her relationship to her neighbors

15 or anything else that's relevant for historians who

16 do this kind of research. So, I and the department

17 look forward to working with the City Council to

18 further develop solutions to our shared policy goals

19 of increased—of increasing the accessibility of city

20 government's historical records, and making them

21 available to diverse populations. I thank you for the

22 opportunity to testify, and if you have questions, I

23 will try to answer them.

24   CHAIRPERSON KOO: Thank you. We are

25 joined by Council Member Holden. Yeah. I have a few

2   questions, and the first on Intro 986. Is there any

3   citywide standard on how agencies should provide data

4   in a report? Do we have any standard now? Any—

5   DOITT, anything special? (sic) No?

6   EMILY NEWMAN: I don't believe that we

7   have a citywide standard. Each agency maintains

8   their data, and—and uses it in their reports as is

9   appropriate for the report requirements.

10   CHAIRPERSON KOO: So, what—what data is

11 on the Mayor's Dashboard? Any of you know that?

12   EMILY NEWMAN: On the Mayor's Dashboard?

13   CHAIRPERSON KOO: Yes.

14   EMILY NEWMAN: Which dashboard are you

15 referring to, the--

16   CHAIRPERSON KOO: [interposing] The one

17 that he uses on his phone.

18   EMILY NEWMAN: Oh. The data that we make

19 available comes from the Citywide Performance

20 Reporting Tool, which is available publicly as well.

21   CHAIRPERSON KOO: Yeah. So, if someone

22 sends you—they send—they just send you a report with

23 the data in PDF, would you be able to use this data?

24 I mean if I send you a report in PDF, can you—can

25

2 you—can DOITT read the data? I'm sure someone can,

3 but someone cannot read it.

4   DEPUTY COMMISSIONER SUNDERLAND: I mean

5 it—it depends on the form—the form that the data

6 actually takes in the PDF itself. I mean at this

7 time, there are certainly technologies that can

8 attempt to read text actually making sense of the

9 text and parsing it into data readable format would

10 be a significant piece of work for everyone that you

11 wanted to do this with. It wouldn't be something

12 that could be done generically. As far as the

13 underlying data that might appear in charts or

14 tables, there wouldn't be references within the

15 report necessarily that would allow you to map back

16 to the original data that produced it.

17   CHAIRPERSON KOO: Yeah, that's why the—

18 the main goal of this Introduction 986 is to make

19 agencies use machine readable formats when sending

20 reports to other agencies.

21   EMILY NEWMAN: I just want to add that

22 much of the underlying data where it's relevant for

23 the Open Data Law is available publicly. Again, it's

24 going to depend on the report, and whether there's

25 any personally identifiable information there, but

2   much of the data is already available or is part of

3   the 400 plus data sets that we're working over the

4   next year to add online to open data.

5   CHAIRPERSON KOO: So, so, do you know how

6   many reports still provide data only in the PDF

7   format?

8   EMILY NEWMAN: I do not.

9   CHAIRPERSON KOO: No. Would it be a

10   difficulty, a technical difficulty to provide data in

11   a machine-readable file?

12   EMILY NEWMAN: Again, no. I think it

13   depends on what the data is that's being asked for,

14   and whether there's any personally identify--

15   identifiable information included.

16   CHAIRPERSON KOO: Okay. So, now I go to

17   questions on [pause] 1094. So, what—what data does

18   the Mayor's Office of Operations use to track agency

19   performance besides the CPR?

20   EMILY NEWMAN: I'm very happy to say we

21   have the Annual Mayor's Management Report coming out

22   today. [laughter] So, your timing is great.

23   CHAIRPERSON KOO: Yeah.

24   EMILY NEWMAN: We have a whole book with

25   45 agencies included where we track quite a bit of

2 data on each agency. You can see historic trends and

3 so we—we track much of that data monthly. Some of

4 it's annual, but we have a rich amounts of data that

5 we make public that we track for each agency.

6 　　　　　　CHAIRPERSON KOO: So, has the

7 Administration ever shared any of public data if a

8 member of the public using the long disclosure

9 agreement?

10 　　　　　　EMILY NEWMAN: [pause] [off mic] It

11 depends.

12 　　　　　　CHAIRPERSON KOO: No, guess, no?

13 　　　　　　EMILY NEWMAN: Could—could you repeat the

14 question, please?

15 　　　　　　CHAIRPERSON KOO: Yeah, has the

16 Administration ever shared any public data with a

17 member of the public using a long disclosure

18 agreement?

19 　　　　　　EMILY NEWMAN: Are you speaking about the

20 Citywide Data Integration Agreement?

21 　　　　　　CHAIRPERSON KOO: Yeah, yeah, long

22 disclosure.

23 　　　　　　EMILY NEWMAN: I—I guess I don't—I don't

24 fully know the answer. We share data across

25 agencies. We share data with our contractors and

2 vendors using non-disclosure agreements. We try to

3 be very careful and make sure our Chief Privacy

4 Officer can add details, but we try to be very

5 careful with the sharing of data, but we certainly

6 don't just sort of casually share data with the

7 public where it doesn't make sense. Although we, as

8 I've said, make a lot of data available through the

9 Open Data Law and the Open Data Portal.

10   CHAIRPERSON KOO: So, one is specifically

11 to researchers or vendors who need this—need this

12 data set.

13   EMILY NEWMAN: Uh-hm.

14   CHAIRPERSON KOO: And from the agency.

15 So, you have done it?

16   COMMISSIONER TOOLE: If you're speaking

17 about non-disclosure agreements--

18   CHAIRPERSON KOO: [interposing] Yes.

19   COMMISSIONER TOOLE: --it is the vehicle

20 by which the information is shared with a vendor for

21 example or a researcher. Typically, there is either

22 a contract in place with the vendor or the researcher

23 or if there is a multi-agency agreement, an MOU, it

24 would be reviewed and approved by the City's Law

25 Department. There would be a business use case or a

2   research proposal that sometimes has to go through an

3   IRB, Institutional Research Board, and if all of the

4   legal requirements are made, we may add a

5   nondisclosure agreement to just ensure that there's

6   additional protections in place for the information.

7   CHAIRPERSON KOO: So, what make you feel

8   secure disclosing this information with members of

9   the public?

10   COMMISSIONER TOOLE: I can't speak to

11   whether—I mean I haven't personally worked on

12   agreements that have involved disclosure to members

13   of the public, but we have worked with a number or

14   agencies and vendors under agreements that have very

15   strict private and security restrictions.

16   CHAIRPERSON KOO: Okay. So—so, what data

17   or data sets does a data breach have that the Open

18   Data Portal does not have? What's—what's the

19   difference between them?

20   DEPUTY COMMISSIONER SUNDERLAND: I—I

21   don't know that we've done a specific accounting on

22   that. In general, the majority of the data in in Data

23   Bridge is—is also on Open Data.

24   CHAIRPERSON KOO: Do you also use open

25   data?

2   DEPUTY COMMISSIONER SUNDERLAND: Correct.

3   CHAIRPERSON KOO: So, can you or anyone

4 else in the Mayor's Office use Data Bridge for

5 oversight or over agency operations? Do you--?

6   JAMES PERAZZO: Data Bridge is—is less

7 commonly used as an oversight tool, and more often to

8 support the analysis conducted by the programs that

9 own the data and going into that—into the tool.

10   CHAIRPERSON KOO: So how many agencies or

11 people with city agencies have access to Data Bridge?

12   JAMES PERAZZO: I—I don't know the

13 numbers off the top of my head, and my guess is

14 almost every city agency.

15   CHAIRPERSON KOO: Do you have access to

16 it?

17   JAMES PERAZZO: In—in some form or

18 another yeah, the—yeah, of course, the access is

19 limited to, you know, or data appropriate.

20   DEPUTY COMMISSIONER SUNDERLAND: There

21 are a lot of access to Data Bridge to do analysis on

22 their own data. They don't use Data Bridge to access

23 other people's data unless there's been an agreement

24 of some sort that's been created around that.

25

2   CHAIRPERSON KOO: Okay. So—so how is the

3 Administration currently storing the data in

4 historic—historical records? You know this is

5 relating to 1098, you know, Intro.

6   COMMISSIONER TOOLE: Yes, sir. All—all

7 of our data is servers and it's backed up at a site,

8 and we're in the process of moving it to the Cloud to

9 increase public access and-and reduce cost.

10   CHAIRPERSON KOO: So, you're in the

11 process of storage in the Cloud?

12   COMMISSIONER TOOLE: Yes, sir.

13   CHAIRPERSON KOO: How long will it take

14 to—to do the process?

15   COMMISSIONER TOOLE: [laughter] I don't

16 think the process will ever be complete given the

17 volume of historical records that we have, but as we

18 digitize more, we'll be acquiring more Cloud storage

19 so that it's all available as quickly as possible.

20   CHAIRPERSON KOO: When you talk about

21 Cloud, which Cloud are you talking about? Do you

22 mean you hire some outside agencies like Amazon or

23 Google. It's all through Cloud.

24   COMMISSIONER TOOLE: [interposing] The—

25 the city—the city's contract currently is with

2 Amazon, but I know its exploring other options in the

3 Cloud and, you know, our view is we want the most

4 secure and least costly Cloud access possible, and

5 we're not talking about confidential records. We

6 were talking about historical records. We're talking

7 about things that people need to get access to no

8 matter where they live. So, that's—we—we look

9 forward to just buying as much as we can as cheaply

10 as possible and putting all these records up there.

11   CHAIRPERSON KOO: So, when you sign a

12 contract with Amazon, is it for how long is the

13 contract?

14   COMMISSIONER TOOLE: It's a citywide

15 contract. I don't know the terms of it. DOITT

16 negotiated it, but it allows agencies to purchase the

17 data that they need when they need it. So, if-if we

18 have 185 terabytes of data we want to put up, we

19 would buy the amount to put that up, but as we

20 digitize more, we would buy more. It's all dependent

21 on what we have. It's not like we're buying

22 something to use in the Cloud that we're not using.

23   CHAIRPERSON KOO: So, what happens if you

24 discontinue the contract? Does it—do you have to

25 remove all this data to another storage place?

2   COMMISSIONER TOOLE: Well, I'm hoping we

3 don't get there, there, but I think we would have do

4 that, right?

5   CHAIRPERSON KOO: Oh, it is—it would be a

6 complicated process, though, right to remove—to move

7 like such a big data like—

8   COMMISSIONER TOOLE: [interposing] I want

9 to—

10   CHAIRPERSON KOO: --to say from Amazon to

11 Google.

12   COMMISSIONER TOOLE: --I'm going to defer

13 to my friend from DOITT because I don't—I don't know

14 that part of the technology, but I think it's doable.

15   DEPUTY COMMISSIONER SUNDERLAND: Yeah, I

16 mean one of the things about putting it up there is

17 you put it up in form that can be portable by the

18 time it's all done. So, it could be—it could take a

19 while to do it, but you should be able to deport it

20 to another Cloud vendor probably without a lot of

21 changes.

22   CHAIRPERSON KOO: It's not like you will—

23 for something with a U-Haul and you go there and—

24   DEPUTY COMMISSIONER SUNDERLAND: [laughs]

25 It's not like that, but there is—I mean, but there is

2   catalogue—cataloguing software and programs that

3   would have to new—move to the new vendor, too. There

4   would be some effort but it's not—it's not as if—it's

5   not reflective of the massive nature of the data.

6   It's a lot more of—of, you know, just getting it work

7   and under the new vendor.

8   CHAIRPERSON KOO: Uh-hm. Okay. So, has

9   the Department of Records and Information Service

10   worked with any entities that with expert—with

11   expertise in converting books or other paper records

12   to digital formats?

13   COMMISSIONER TOOLE: Yes, sir. In fact

14   our digital archivist does really. This is her

15   skillset, and she started a year ago, and is really

16   terrific, and we consult regularly with our

17   colleagues at the National Archives, the State

18   Archives. Our team just presented at the National

19   Association of Archivists Convention and received

20   high praise for their work on digitization. So, we—

21   we take all sorts of input in order to do—create the

22   best product possible.

23   CHAIRPERSON KOO: So, only national

24   archives? Do you use--?

25

2   COMMISSIONER TOOLE: [interposing] No.

3   State Archives. There are archives throughout the

4   country and world and our team is in contact with

5   them.

6   CHAIRPERSON KOO: Alright. So, we are

7   also joined by Council Member Kallos and Council

8   Member Holden has a question.

9   COUNCIL MEMBER HOLDEN: The agencies

10 routinely that we've seen—as a—as a Council member

11 I've seen agencies withhold information, and for

12 instance DHS, Department of Homeless Services, I

13 can't find—and they wont' allow—they won't give us

14 access to for instance homeless. The areas that

15 they're—they're in my district that I have the most

16 homeless I'm trying to find out information so I can

17 help so actually—so we can actually get volunteer

18 organizations to help, but yet routinely we don't get

19 that information on where the homeless are coming

20 from, where are they housed so forth and so on. We

21 hope that Speaker Johnson's bill will—will address

22 that, but can you—so, I—I assume it's up to the

23 individual agencies that they're deciding on the data

24 where it's—what they're going to release?

25

2   EMILY NEWMAN: Yes the agency's privacy

3 officers typically make the determinations as to what

4 federal, state and Local Laws apply to either permit

5 or restrict the disclosure of the requested

6 information.

7   COMMISSIONER TOOLE: Is there any

8 oversight on that, that this—does—is it coming from

9 the Mayor's Office or whose? If—if we had, if we

10 challenged that?

11   EMILY NEWMAN: Well, the—the new

12 legislation that the City Council passed this past

13 year the Identifying Information Law does create a

14 role for the Chief Privacy Officer to be able to

15 review data sharing requests based upon the best

16 interests of the city, and that would as between a

17 city agency with another city agency, but—but as I

18 mentioned in the testimony, we cannot overcome

19 federal and state law that may apply to the specific

20 data.

21   COUNCIL MEMBER HOLDEN: Okay, as

22 regarding historical data, is that just all that goes

23 out, or would be put on the Cloud or is it—is that

24 being also some withheld, some of the data?

25

2   COMMISSIONER TOOLE: It's not withheld.

3 The Records are available to people.

4   COUNCIL MEMBER HOLDEN: Okay,

5   COMMISSIONER TOOLE: A lot of it is in

6 hard copy format so a researcher would ask for the

7 records. We'd bring them. They'd review them.

8 They'd do the research.

9   COUNCIL MEMBER HOLDEN: Okay, so—

10   COMMISSIONER TOOLE: But as we did

11 digitized the material it's just put out for people

12 who draw their conclusions where they find it.

13   COUNCIL MEMBER HOLDEN: That really.

14 Okay, thank you. Thanks much.

15   CHAIRPERSON KOO: [off mic] Council

16 Member Kallos.

17   COUNCIL MEMBER KALLOS: Good morning. I

18 want to thank our Technology Committee Chair Peter

19 Koo for his leadership on these issues, and his

20 commitment to focusing on technologies and

21 partnerships we've already worked on around using the

22 Internet of things to monitor our city's

23 infrastructure and improve it, namely the LinkNYC

24 Kiosks. I also want to thank you for chairing this

25 committee hearing on a package of transparency

2 legislation. Transparency is one of my favorite

3 words. [laughs] So, I want to just thank Pauline

4 Toole for the amazing work you've done at the

5 Department of Records and Information Services that

6 we affectionately refer to as DORIS. You've built

7 the Open Records Portal for folks to send FOIL

8 requests and if anyone is interested in seeing a

9 government recork what is the URL to request one

10 through the Open Records Portal?

11   COMMISSIONER TOOLE: [off mic]

12   COUNCIL MEMBER KALLOS: We need your mic

13 on please.

14   COMMISSIONER TOOLE: My—my apologies.

15 NYC@Gov/openrecords, and since you raised that topic

16 can I just state that our very small development team

17 won a statewide award for that work.

18   COUNCIL MEMBER KALLOS: May I ask how

19 many folks were on that team?

20   COMMISSIONER TOOLE:

21   COUNCIL MEMBER KALLOS:

22   COMMISSIONER TOOLE: Four.

23   COUNCIL MEMBER KALLOS: And that entire

24 system was building in free and open source software?

25   COMMISSIONER TOOLE: Yes, sir it was.

2   COUNCIL MEMBER KALLOS: Great. So, I'm

3 here to speak to introduction 1098, which speaks to

4 digitizing archival records. In your testimony you

5 refer to 185 terabytes. In particular, if I wanted

6 to look at old historical photos, where could I look

7 at those?

8   COMMISSIONER TOOLE: You can look at them

9 on our website nyc.gov/records.

10   COUNCIL MEMBER KALLOS: And that sounds

11 like a lot of photos and—and is—are those

12 proprietary? Does anybody own them? Can I—or can I-

13 -

14   COMMISSIONER TOOLE: We the people of New

15 York City own them, and you can go and browse them

16 and you can download a low resolution copy. It may

17 be watermarked, but they're there for public viewing

18 and then we make them available. If people want to

19 publish a book, we have a small fee that they pay in

20 order to do that.

21   COUNCIL MEMBER KALLOS: Great and so

22 Introduction 1098 hopes to build upon the repository

23 of historical data, and I—I think in—in different

24 conversations we've—we've heard about different

25 documents where this—this famous person's birth

2   record was there or this family had their—signed in

3   at a different location. And just is there—is there a

4   way to prioritize certain documents? For instance

5   land use items, and I guess the idea here is for

6   DORIS to work with different agencies. So, the City

7   Planning Commission has a lot of detailed records,

8   and the Department of City Planning recently updated

9   its portal so you can look at different land use

10 items and click on them and see documents from 80

11 years go—sorry, 20 years ago. However, certain land

12 use items are still missing. I—I think one of my

13 most interesting moments as an elected official was

14 we were looking at subway station on Lexington and

15 86th Street, which has a subway—has stairs inside the

16 building and it was based on I believe a 1906

17 easement, which I forced the MTA after several months

18 or years to pull and give to me. Would there be a

19 way through Introduction 1098, and I will admit that

20 MTA is not a city agency, but to work with agencies

21 to perhaps pull some—for instance prioritizing land

22 use documents so that we can reach into the—the

23 decades and—and centuries of old promises that

24 perhaps have never been kept?

25

2   COMMISSIONER TOOLE: Yes and the—the way

3 we prioritize the digitization of the records is the

4 public demand. If the records are frequently being

5 requested, we will put them on a priority list to

6 digitize, and the other reason is if they are fragile

7 and vulnerable, we want to digitize them and make

8 them available in that format so people aren't

9 handling documents that are hundreds of years old

10 that could—could be harmed. We do have funding from—

11 in this year's budget to digitize the building plans

12 for Lower Manhattan from the inception of the

13 Department of Buildings in 18—I forget the year.

14 Let's say 68, something like that, and those will be

15 readily available, but in order to put something in

16 the queue, we would have to really have demand, which

17 is more than one request, but we do make the

18 materials available upon request and maybe, you know,

19 researchers, people who work in Land Use come to find

20 what was the historical—what are the historical

21 reference for the land that they are developing or

22 living in.

23   COUNCIL MEMBER KALLOS: Are there ways we

24 could improve upon Introduction 1098 sot that it

25 would not be a laborious task of little value, but so

2 that it could proactively work with the agencies to

3 identify projects that might be considered of little

4 or not value now, but because it's inaccessible to

5 the general public and people don't even know it

6 exists to integrating it into larger efforts like the

7 City Planning Commission's efforts or it passed a

8 different law relating to privately owned public

9 spaces, and I know that the Municipal Art Society had

10 been trying to go through those documents. So, just

11 trying to digitize public records that relate. I'm

12 now a Land Use Chair so I'm very focused on the land

13 use areas.

14   COMMISSIONER TOOLE: Good. Yes, I think

15 we are—are welcome the opportunity to digitize more

16 and more and more records and work collaboratively

17 with all the institutions you named and other and we

18 do, do that. I think the thing that—the biggest

19 issue for the archive—archivists was the requirement

20 to created a new record based on historic data can—

21 that exists in the original records because it is not

22 part of their practice. It would run afoul of their

23 Code of Ethics. So, it's that creation of records

24 rather than the digitization piece. So, we want to

25 digitize as much as we can, and make it available.

2 COUNCIL MEMBER KALLOS: Thank you.

3 COMMISSIONER TOOLE: Sure.

4 CHAIRPERSON KOO: Alright. Okay, thank

5 you Council Member Kallos. So, what along public

6 data sets do you use for your agency's oversight?

7 [pause] No? There's no answer?

8 COUNCIL MEMBER KALLOS: No. [background

9 comments, pause]

10 EMILY NEWMAN: We can certainly think

11 through it and get back to you. I think as I

12 mentioned earlier, most of what we use for oversight

13 is, in fact, made public. We've go the MMR that's

14 coming out today, the datasets that come through open

15 data, and so most of that information—most of the

16 information that we use for oversight is made public

17 there.

18 CHAIRPERSON KOO: And also I have

19 question on it. You mentioned that the clean room is

20 not a good place—thing. So, what are the alternative

21 strategies to a clean room? [pause]

22 DEPUTY COMMISSIONER SUNDERLAND: Yeah.

23 Okay, thanks. [coughs] You know, I think the idea of

24 creating, you know, outer trails around the data and

25 responsibilities for the parties who use the data,

2 but a clean room in and of itself doesn't have—

3 there's no—no specific terminology in the realm of

4 data that—that refers to clean room and—and defines

5 what that is. It kind of implies more a physical

6 limitation of the data than the actual electronic

7 limitation that will be required. So, in order to—to

8 come up with a—with a solution, and we'd really have

9 to sit down and work through the actual requirements

10 about the data access, and—and-and—and use going

11 forward. There's—there's no general principle that

12 would govern that.

13   CHAIRPERSON KOO: Uh-hm. So, so, going

14 back to the previous questions, so who—who are the

15 vendors, and—and is the data the same on City Bridge

16 as available to the public? Can you name some of the

17 vendors? Would you say--

18   DEPUTY COMMISSIONER SUNDERLAND: Sorry,

19 which—which vendors?

20   CHAIRPERSON KOO: You know the—the

21 vendors who—who—who want to do research right or—or--

22 ?

23   DEPUTY COMMISSIONER SUNDERLAND: I—I

24 don't have any specific examples of—of vendors who—

25 who we share data with.

2   COMMISSIONER TOOLE: There is one vendor

3 I can think of that was part of a citywide data

4 integration Project agreement. I believe they were

5 federally qualified health centers that were doing

6 some work concerning identifiable information and

7 access to healthcare. I should also, if I may, speak

8 to your question raised earlier concerning the use of

9 non-public data at least with respect to the Citywide

10 Data Integration projects that are advancing research

11 and best practices. There is work underway at

12 agencies and agencies working collaboratively to

13 identify the needs of for example high engagers of

14 city services to better understand those populations

15 and to improve coordination of services to those

16 individuals. So--

17   CHAIRPERSON KOO: So, is the data on Data

18 Bridge the same as on Open Portal?

19   JAMES PERAZZO: The same? I mean if—it

20 would depend on—on which data set.

21   CHAIRPERSON KOO: Oh, so you're not--

22   JAMES PERAZZO: [interposing] Since data

23 is—it varies.

24   CHAIRPERSON KOO: --they're not the same?

25

2   JAMES PERAZZO: Well, they're—they're two

3   separate streams. So the determination for what

4   should be eligible for open data is, as we previously

5   said, made by the—the agency owners and their privacy

6   officers and general counsels. Data is—is put into

7   Data Bridge to support analysis of that particular

8   program, and typically most of the data there is

9   public just by the fact that most of the data is

10 public in the first place, but it is from the stream

11 from the agency's source and there isn't really a

12 cross accounting of which things are public, and

13 which aren't. Typically it's—it's filtered down by

14 the time it gets to Data Bridge because it's for a

15 particular purpose.

16   CHAIRPERSON KOO: Council Member Holden,

17 do you have questions?

18   COUNCIL MEMBER HOLDEN: Yes, could I get

19 back to the historical photos for a second. As a

20 researcher and somebody that's worked on the history

21 of the neighborhoods, I was always annoyed at the

22 watermarks, which were all at the time diagonally all

23 across the photo. Do you need to do that? I mean

24 can't we just put a copyright in the corner—

25

2 COMMISSIONER TOOLE: There are--I man the

3 city has a copyright to the records. A printed

4 copyright in the corner is not going to—is not going

5 to assist us in selling the photos and making our

6 revenue targets for the city.

7 COUNCIL MEMBER HOLDEN: How much is—how

8 much do you generate in that for you?

9 COMMISSIONER TOOLE: So altogether

10 between Vital Records and historical photos we

11 generate around a million dollars of revenue.

12 COUNCIL MEMBER HOLDEN: A million

13 dollars, but—so, if the photos are owned by the

14 people so we're selling them to the people. These

15 deals--

16 COMMISSIONER TOOLE: [interposing] People

17 from all over the world, sir.

18 COUNCIL MEMBER HOLDEN: So, I mean so we—

19 if it's all over the world, why not down—I m mean

20 they're low rise and you generate only a million

21 dollars. That's not a lot of money.

22 COMMISSIONER TOOLE: Please tell that to

23 OMB.

24 COUNCIL MEMBER HOLDEN: OMB okay.

25 [laughs] Okay, but I just feel that the watermarks

2   diagonally across the entire photo. Actually, it

3   kind of destroyed the—the—you're looking at the—at

4   the historic lineage of New York City, let's say

5   Manhattan and you—you kind of—you-you spoil the

6   experience. I just want to throw that out.

7   COMMISSIONER TOOLE: I'm sorry you feel

8   that way. I can get back to you on how we might do

9   things differently but, you know, we get a lot of

10   people who look at those historical photos. I'd say

11   most of them don't think they had a spoiled

12   experience.

13   COUNCIL MEMBER HOLDEN: And I can also

14   get people who will tell you that they—they're

15   annoyed with the diagonal water marks.

16   COMMISSIONER TOOLE: Okay.

17   CHAIRPERSON KOO: Council Member Kallos.

18   COUNCIL MEMBER KALLOS: Just to follow

19   along with my colleague's point, even in the private

20   sector at this point companies like Photo Stock and

21   what have you, they—they have very minimal watermarks

22   and—and I would just—I would be happy—I would just

23   say that a million dollars is—is—is almost a rounding

24   error in our budget. So, if we were able to ask OMB

25   that whether or not this—that if we could reduce the

2 fees as it were, whether they could live without it,

3 would you be open to removing the watermark from low

4 resolution non-print publication quality photos?

5   COMMISSIONER TOOLE: I'd have to look

6 into it.

7   COUNCIL MEMBER KALLOS: Oh, okay. I

8 wanted to follow up on the Data Bridge Legislation.

9 Where does that—where does Data Bridge overlap with

10 HHH—sorry, HHS Connect? And so, HHS connect is a

11 service that is offered and owned by the Mayor's

12 Office of Operation. They call themselves Ops. I

13 call them MOO, but it allows agencies to share

14 information after signing a memorandum of

15 understanding. The reason I'm so interested in HHS

16 Connect is because of legislation we passed last year

17 call Automatic Benefits where I would like any New

18 Yorker receiving any benefit for which they are

19 screened and actually any New Yorker who's paying

20 taxes whether they know it or not because it's being

21 deducted from their pay checks, to—for the government

22 to use the information we already noted just mail

23 them a SNAP card in the mail or any other member of

24 40 government benefits that we currently offer. So,

25

2   can you tell me a little bit about the difference

3   between the Bridge and the Connect?

4   EMILY NEWMAN: I can speak about the

5   particular tool you're speaking about, which is the

6   Worker Connect Program.

7   COUNCIL MEMBER KALLOS: Uh-hm.

8   EMILY NEWMAN: You are correct. There is

9   a data exchange interagency agreement that's been in

10 place since I believe 2010. There are five agency

11 data providers for that tool, and it is centered on

12 a—an algorithm based matching tool that can have

13 agency information from the five—any one of the five

14 data providers matched with information that is from

15 another agency that may sign onto the agreement.

16 That is distinct from a privacy perspective.

17 Actually, it is the model, which is—which is based in

18 a use case process, a business use case process that

19 looks as I have spoken earlier about—about how to

20 share data across agencies when there's identifiable

21 information. That model was used to develop the

22 citywide data integration framework, which exists

23 today, and yeah, that's from a privacy perspective.

24 I can speak to that. From a technology perspective,

25 I defer to my colleagues to talk about how Data

2 Bridge fits into that, but it is a separate tool that

3 is used under a specific agreement. I should also

4 mention that the Citywide Data Integration Agreement

5 is now signed. It's a master framework agreement

6 signed by 47 city agencies, and I'm happy to have a

7 further conversation about the—with the Council if

8 you're interested in—in taking advantage of that

9 framework.

10   COUNCIL MEMBER KALLOS: Yes, please.

11   JAMES PERAZZO: So, on the technology

12 side and Don can correct me if I'm wrong, but the

13 Workers Connect product is that fundamentally is one

14 of the same. It's a massive integrated (sic)

15 management solution that's about matching data and

16 servicing the results to the appropriate people. The

17 Data Bridge is a data warehouse, which allows

18 programs to pull data in from various different

19 sources, and combine them in ways that are

20 advantageous to their particular analysis calls, and

21 digalize it through various different front-end tools

22 including some that we make public on nyc.gov.

23   COUNCIL MEMBER KALLOS: When the city,

24 which published this in a journal, which is the only

25 reason I know about it I guess. They just switched

2   so you could be a little bit more transparent with

3   some of the great work we do. When the city

4   purchased data from Experian to identify the four-

5   year-olds in the city, was that through City Bridge

6   or is that data now residing in City Bridge or was

7   that done separately?

8   EMILY NEWMAN: I believe—if I may, I

9   believe you're talking about the Pre-K Outreach

10 Program.

11   COUNCIL MEMBER KALLOS: Yes.

12   EMILY NEWMAN: That team was centered at

13 the—at a combination of City Hall and Department of

14 Education--

15   COUNCIL MEMBER KALLOS: [interposing]

16 MODA.

17   EMILY NEWMAN: --and—okay. There was an—

18 an agreement by which—I can't speak to the Experian

19 Agreement. I can only speak to the city agency data

20 exchange for that agreement, but I can—I don't know.

21 Maybe I'll defer to my colleagues on the other

22 details.

23   DEPUTY COMMISSIONER SUNDERLAND: Yes, I—I

24 wasn't familiar with the Experian agreement per se,

25 but this is the way analytical projects like this

2 come about. So, this was an ad hoc project very near

3 term that was required to pull together data from a

4 number of disparate—disparate sources both externally

5 and internally to try to identify where the four-

6 year-olds were, and then we used a variety of city

7 tools to be able to do that. I think we used that

8 analytics platform. We made these data bridges,

9 while I don't recall specifically, but that's that

10 way—that's the thing to remember about Data Bridge is

11 that it's not in—it's not a piece of infrastructure

12 with tentacles that extend throughout the city

13 through which we can tap anything that we want. It's

14 a tool that we can implement on ad hoc basis to solve

15 specific problems, and that's really—that's really

16 the way that it's been working, but the need to solve

17 analytical problems often times will involve

18 agreement with—with vendors for external data as well

19 as—as leveraging internal data, and that's really a

20 lot of what we do with Data Bridge.

21   CHAIRPERSON KOO: [off mic] Any

22 questions?

23   COUNCIL MEMBER KALLOS: So, I guess I'm

24 not—so is HHS Connect or Worker Connect just city

25 agency data and Data Bridge is external data? How

2 would—what is the difference between the two

3 products? Are they—

4   DEPUTY COMMISSIONER SUNDERLAND: They're

5 separate systems.

6   COUNCIL MEMBER KALLOS: Do they—do they

7 have overlapping data?

8   DEPUTY COMMISSIONER SUNDERLAND: It's

9 possible. I don't know.

10   COUNCIL MEMBER KALLOS: I guess what's

11 the difference?

12   DEPUTY COMMISSIONER SUNDERLAND: Well, as

13 I said, it's—so the—in the Worker Connect example the

14 data is sourced from a number of agencies. It's run

15 through a master data management solution, and the

16 end products of that matching is surfaced to the—the

17 user who logs in with their credentials.

18   COUNCIL MEMBER KALLOS: Okay. So,

19 anything that is in Worker Connect should be in Data

20 Bridge because Data Bridge is—is larger and doesn't

21 have restrictions on who can do it. You're—it's just

22 the—the warehousing and the--?

23   DEPUTY COMMISSIONER SUNDERLAND: Data

24 Bridge does contain more data, but it is—the two

25 programs are—are separate from each other. So, it

2   would be a matter of coincidence if some of the data

3   used to do the matching for Worker Connect were

4   brought into Data Bridge for—in support of a

5   different program. Yeah.

6   COUNCIL MEMBER KALLOS: And—and forgive

7   me. I was a little late. So, and—and your—you

8   support the—the idea of having the clean room for

9   examining of the Data Bridge?

10   DEPUTY COMMISSIONER SUNDERLAND: We

11   don't. I mean we're not familiar with the—with the—

12   with the use of the term clean room vis-à-vis data

13   per se except in the cases where—I mean the other

14   reference we could find was in the financial industry

15   where during mergers and acquisitions they created an

16   isolated area. They have documents to due diligence.

17   So, it's not really something that exists within—to

18   my knowledge within the realm of electronic data.

19   COUNCIL MEMBER KALLOS: Great, but do you

20   support the legislation?

21   DEPUTY COMMISSIONER SUNDERLAND: Yeah,

22   not as it's written. It—it would take a

23   conversation. We're interested in having

24   conversation around the intention of the legislation

25   but there's a—there's a great deal that would have to

2 be detailed and for us—in order for us to understand

3 how to respond appropriately.

4   COUNCIL MEMBER KALLOS: So, I guess one

5 thing I wanted to share is just this concept of if

6 you're at an agency you might be asking for well

7 maybe I need this point of data and that point of

8 data, and so you get to see—you get to see—I'm going

9 to go with the metaphor of the three blind people and

10 the elephant, and so one person grabs the tail and

11 it's a snake, one person grabs the leg and it's a

12 tree. By being able to see the enormity of the full

13 dataset, and all the different pieces that's where

14 somebody who is an information architect like I am

15 can say, okay, so we don't actually know how many

16 four-year-olds we have because we don't have a list

17 of four-year-olds, but we have a list for DOHMH of

18 all the live births and then we have meta data over

19 here from the vaccination records. So, now we have

20 an accurate count of the number of people who

21 received their three-year vaccination and therefore

22 if we match that to other things like a—a voter file

23 or city ID or what have you, we—we can now take eight

24 different things that gave us different pieces of the

25 elephant and identify all the four-year elephants in

2   the city of New York to get them Pre-K [laughter] and

3   to not wait for them to apply, but to actually just

4   send them a letter or even an email so we have that

5   information to say hey, just wanted to let you know

6   yes there was an application deadline, but we didn't

7   assume that you needed to apply for it. So, FYI, if

8   your four-year-old wants to start Pre-K, here is a

9   seat waiting for you. Wouldn't you like to take it?

10   EMILY NEWMAN: I can't speak to the

11   policy considerations or the legal considerations

12   about reaching out in that manner, but with the Pre-K

13   effort that has been underway since 2014, we have I

14   believe it's about 7 or 8 different agreements in

15   place that we're working to consolidate across

16   agencies, and we have done a significant amount of

17   data managing to be able to identify those—those

18   four-year-olds. It's not exactly where it needs to

19   be only because of the privacy laws that, you know,

20   we've had to cobble together a patchwork solution.

21   We do work very closely with agencies and the

22   Administration and our partners to try to find

23   workable solutions, but I think you know, we've been—

24   accomplished quite a bit in the years, I believe

25   through data integration work in this area.

2   COUNCIL MEMBER KALLOS: I would just note

3   for the sponsors and—and the staff please make sure

4   that Council Members are allowed into this clean

5   room, and I—please add me as a sponsor, and—

6   EMILY NEWMAN: Yes, if privacy laws

7   permit. [laughter]

8   COUNCIL MEMBER KALLOS: I—I—it's—with

9   regards to privacy laws I—I understand it's brought

10   up quite often. I—I've submitted to—I believe your

11   agency has in hand a 30-page memo reviewing--

12   EMILY NEWMAN: We've been looking at it.

13   COUNCIL MEMBER KALLOS: So, it's federal,

14   state, and city privacy laws and privacy amongst

15   employees is more protected than privacy between an

16   agency and the public, and I think the thing that I

17   find most interesting is where you want the most

18   privacy in life is between you and law enforcement.

19   You—you don't want to get in trouble because you—you

20   did something that you thought wasn't a big deal like

21   for instance I don't know a New Yorker who hasn't

22   jaywalked, and so you don't want police to be able to

23   say we—we know that you are a serial jaywalker and

24   you have jaywalked a thousand times this month alone

25   and what have you, but there a broad exemption of

2   federal, state, and city and in-private sector any

3   law enforcement activity. I find it difficult to

4   believe that the—and—and based on my own legal

5   research that the federal, state and city laws

6   prevent using data to give people free food from SNAP

7   benefits, free assistance on their rent through

8   Senior Citizen Rent Increase and Disabled Rent

9   Increase Exemption, and that all of those taken

10   together prevent us from just instead of making some

11   body apply for Pre-K just saying here's your seat.

12   EMILY NEWMAN: Again, I think, you know,

13   we've done a lot of work in the area of solutioning

14   around—within the privacy framework that exists

15   today, and by no means do privacy laws preclude the

16   solutioning around providing important programs and

17   services to—to our—our residents, but it's very much

18   a fact as you know, if you, you know, from your

19   research I'm sure fact based determination based upon

20   the laws that apply to particular data elements and

21   data categories that have to—there has to be a legal

22   pathway to allow it. But not to say that there

23   aren't any, it just—it's a fact specific

24   determination under the various typical laws.

25

2 COUNCIL MEMBER KALLOS: And just one last

3 question for Department of Records Information

4 Services. So some of the city's original records are

5 not maintained by DORIS as far as I understand. I'm

6 looking. I—I know that the CUNY School of Law—sorry—

7 the New York Law School has many of our city's

8 decisions particularly BSA decisions. I do a lot of

9 work in the BSA, and that is—they somehow ended up

10 owning our BSA decisions. I think they may also have

11 COIB decisions. They have an entire portal for

12 searching a lot of government information that I

13 think should also reside with you. Similarly, I have

14 a friend his name is Dennis Harlow. We went to law

15 school together, we—and he may be one of the smartest

16 attorneys I've ever met in my life, and for fun he

17 goes to John Jay College of Criminal Justice in the

18 CUNY system where he goes through the trial

19 transcripts of the County of New York from 1883 to

20 1927 where he's reading through the Court of General

21 Sessions, which is the predecessor to our local

22 courts as well as the tribunals and the police

23 courts, which really went through some of the moral

24 crimes and what have you that in this day and age

25 might seem strange or at least captivates the

2   imagination. You mentioned request for information,

3   but is there just for the sake of this friendship, my

4   friendship with him would you consider digitizing

5   some of these records, and is there also an op—are

6   you tracking people who show up in person and pull

7   records or who show up at partner sites that house

8   additional pieces of—I guess the sensible question is

9   are you tracking how many people checked the book out

10 from the library and the other shoe I found with the

11 NYPDL is they don't actually track how many people

12 take a book off the shelf and read it in the library.

13 So some of the most popular books in any library no

14 one knows that they're popular because they aren't

15 checked out.

16   COMMISSIONER TOOLE: Well—well, there

17 were several pieces to that. So, let me see if I can

18 get them all together starting with the last one, we

19 are not a lending library, but our reference service

20 desk does not the topics and the materials that

21 people are interested in—in perusing, and they have

22 to obtain the for them, right? They people cannot

23 come in and browse the stacks in the Municipal

24 Library because it's a research library not a lending

25 library. We also track when people come in and they

2 ask for a variety of archival records what those are,

3 and that's what's you see what a person is

4 researching, but what kind of records are in demand

5 so then we prioritize digitizing those. We do have a

6 number of the records that you mentioned that are at

7 John Jay. I would have to look into the agreement

8 with the New York Law School who publishes the BSA

9 and other—other records or transcripts. Okay. So,

10 tell your friend to come. We've got it—we've got it

11 all.

12   CHAIRPERSON KOO: Thank you. Yeah, so we

13 have to move on because of time limitations and so

14 thank you all for coming.

15   EMILY NEWMAN: Thank you, Council Member.

16 [background comments, pause]

17   CHAIRPERSON KOO: Oh, Council Member

18 Yeger. Yeah

19   COUNCIL MEMBER YEGER: [off mic]

20   CHAIRPERSON KOO: Oh, no, that's okay.

21   COUNCIL MEMBER YEGER: [off mic] It is

22 not easy. It's what you want all of us to do I

23 think.

24   COUNCIL MEMBER KALLOS: I was—I was

25 stalling for my colleague.

2   COUNCIL MEMBER YEGER: I gave my

3 questions to you and colleagues over there.

4   CHAIRPERSON KOO: We have panel number—

5 number 2 from the public. [background comments] Now

6 we have Gail Fox. [background comments] Rachel?

7 Rachel Berg and Noel Hidalgo and Alex Camarda (sp?)

8 (background comments, pause) So, let's start now.

9 Please identify yourself and then just begin and

10 start. It doesn't matter who, but--

11   ALEX CAMARDA: Are—are we on a clock or--

12 ? Okay. Hi, good morning. My name is Alex Camarda.

13 I'm the Senior Policy Advisor for Reinvent Albany. I

14 won't read through my testimony. I think I'm

15 actually going to comment on what was said by the

16 different agency officials in relation to the bills.

17 So, regarding Intro 986 sponsored by I think Council

18 Member Koo, the chair of the committee, we are

19 generally supportive of the City Council receiving

20 data that's in agency reports. We also believe that

21 this data should be placed in the Open Data Portal.

22 I think you heard from the head of the Mayor's office

23 of Operations that they believe that that's typically

24 done when the—when the data is regularly maintained,

25 which is what the law requires. I can tell you from

2 our experience we've certainly seen tabular data sets

3 in agency reports that are not in the Open Data

4 Portal. I think that's actually more often the case

5 than is the case, and so, we would like to see agency

6 reports that have tabular data regularly maintained

7 or not made available in the Open Data Portal, and we

8 suggested an amendment to the bill to that effect.

9 But I think really what's most important here is that

10 the regularly maintained data that's currently in

11 reports there should be a more aggressive effort to

12 put those data sets in the portal. With regard to

13 the Kallos bill, you heard from the DORIS

14 Commissioner that it's very burdensome to be able to

15 go through all of these historical records to

16 determine which datasets have public value, and which

17 should be placed in the Open Data Portal. Generally

18 speaking, you know, we—we believe the intent of this

19 bill is—is something that's worthwhile. I think that

20 their current framework for processing digitization

21 could be applied to putting data sets in the portal.

22 We heard from the Commissioner that as demanded by

23 the public or by archivists or others that that's how

24 they determine the order or the sequence for

25 digitizing data. So, we think that a similar process

2 could be put in place for putting datasets in the

3 portal. I don't know. It wasn't commented on by the

4 DORIS Commissioner, but when they're digitizing

5 documents we would hope that they are doing so in a

6 manner that allows for extraction of—of the data sets

7 via OCR. I mean sometimes we see data sets—I'm

8 sorry—archived information and it's just PDF'd and

9 then you can't search it for the input for the data

10 in the PDF, and you can't extract the dataset. We

11 would hope that they're doing so in their—in—as they

12 go through and digitize their documents. We think

13 that should be a requirement, and we don't think it

14 would be too burdensome for them to put some data

15 sets in the portal as they're going through and

16 digitizing the priority documents. On the last bill

17 that Johnson Bill, I think what's missing from the—

18 from the previous testimony was there's really an

19 imbalance between the mayoralty and the City Council

20 when it comes to access to data. My understanding is

21 the City Council has to ask agency officials for data

22 whenever they're providing their oversight functions.

23 That's how they obtain data about the agencies. That

24 is fundamentally the imbalance and I think this

25 bill's intent is to try to correct that imbalance so

2   that the City Council has access to at least some raw

3   data and we support that effort. As far as the

4   particular as to whether there should be a room—a

5   physical room or some kind of digital access, we're

6   not particular on, but we do support the idea of the

7   City Council accessing data in its rawest form, and

8   creating a more formalized structured manner to do

9   so. I—I don't know to what extent agency officials

10   when they provide data to the City Council is vetted

11   for all the privacy restrictions. I assume they do

12   so through their counsels when they—before they

13   provide the data to the City Council, but the point

14   being any data the Council I believe currently

15   receives is data that has been contextualized, has

16   been further segmented by the agencies themselves,

17   and we'd like to see the Council have access to as

18   much raw data as possible. Thank you and I'll stop

19   there.

20   CHAIRPERSON KOO: Thank you. Yeah.

21   Next. [background comments, pause] Can you use the

22   mic? Yeah.

23   RACHEL BYRD: Hi. Thank you. Is this

24   on? Good morning, Chairman Koo, and members of the

25   committee. Thank you for inviting the--

2 CHAIRPERSON KOO: [interposing] Please

3 identify yourself.

4 RACHEL BYRD: Oh, sorry. Apologies. My

5 name is Rachel Byrd. I'm with the Public Utility Law

6 Project. We are 37-year-old public interest law firm

7 as well as an advocacy and education firm that or

8 organization that's sole mission is to advocate,

9 educate and litigate on behalf of low-income utility

10 rate payers. We highly support the idea and the need

11 for as much data availability as possible. That's

12 been a critical tool in PULP's work. We have for

13 example in our rate cases when the utilities must go

14 to the Public Service Commission and apply it to

15 their rate increases, we at—we access their data, and

16 spend an enormous amount of time crunching it in

17 order to make the case for them getting lower

18 increases and for providing better discounts and

19 consumer services. We also use it—actually, I'm

20 completely off—off my testimony, but it's buried in

21 here somewhere. We've also used it as well—we've

22 done FOIL requests for two years worth of Public

23 Service Commission complaint data about energy

24 service companies. We found 10,000 complaints, and

25 of those we were able to analyze them and find out

2   where and under what circumstances and what

3   populations they were targeting. So, we found that

4   the ESTOS, as we called them, targeted low-income

5   communities, people of color, low English or limited

6   English speaking communities, and seniors.

7   Basically, the people who are most vulnerable to a

8   pitch for lower cost energy. We have done—let's see.

9   I have more in there because it's actually specific.

10   We did and we recently we've been getting more

11   involved as the City has gotten more involved in the

12   question of water bills and liens. We accessed the

13   Open Data Portal, and were able to find, and this was

14   last spring when one of the first sales of water

15   liens came up, we were able to find exactly which

16   boroughs and how many liens were potentially going

17   to be sold in May. It was something—it was I believe

18   7,698. We were able to determine the highest

19   concentration in which boroughs, and we were able to

20   actually help people who did not know they actually

21   have rights to defer payment agreements and other

22   payment plans and other negotiating processes that we

23   encourage and filter down we're available to people.

24   So, having that kind of data available was eye

25   opening. We did not know at the time that that

2 information was available. Oh, I hate these things.

3 I can't find—now I end up looking for it. Accept my

4 apologies somehow.

5   CHAIRPERSON KOO: Okay.

6   RACHEL BYRD: Somehow my computer has

7 decided that I can't look at this. Apologies.

8   CHAIRPERSON KOO: You want to come back

9 later?

10   RACHEL BYRD: Perhaps. Do you want me to

11 continue so I can-?

12   CHAIRPERSON KOO: Yes

13   RACHEL BYRD: Let me see. I just have to

14 decide how I can get home because I just--

15   NOEL HIDALGO: Well, my name is Noel

16 Hidalgo from Beta NYC. I'm appreciative of the three

17 bills that are being introduced today. I have some

18 general concerns that I'll start off. One is that

19 the Mayor's Office of Data Analytics is still missing

20 the Chief Analytics Officer, and we're still missing

21 a Chief Technology Officer. Both of these roles in

22 addition to the suite of agency representatives that

23 you saw in front of you are critically important to

24 implement the City's IT Technology and Data Policy,

25 and I'm concerned that both of these programs, MODA

2 and the CTO's Office will be underfunded through the

3 next budget cycle as we continue to have a lack of

4 very clear leadership in both of those agencies.

5 I'm—similar to Alex I won't exactly read all of my

6 notes because I thought that the testimony from the

7 Administration was quite illuminating in regards to

8 Data Bridge. It's kind of absurd that you have to

9 write up legislation to require access to the city's

10 data. To me, this is almost a testament of a lack of

11 leadership from the Administration. They should be

12 willing to share data and provide opportunities for

13 Council Members and their staff to secure data or to

14 get access to secured data because the NYU Center for

15 Urban Pedagogy actually has such a secure, clean data

16 room. That's actually one of the reasons why they

17 were founded was to help provide a safe, secure,

18 clean, data analytics capacity between governments

19 and private parties, and so it's kind of crazy to

20 hear the agency to say that they don't have examples

21 when there are National Institute of Standards and

22 Technology examples that there's federal information

23 processing standards and publications. Like the

24 federal government is doing this. I have worked in

25 secure data centers. So, it's kind of crazy to hear

2 that there is no examples or there's no best

3 practices when there clearly are, and it's really

4 absurd that they refuse to figure out a way to work

5 with Council to make sure that Council Member Kallos'

6 concerns in regards to children in Pre-K should have

7 access to school. So in regards to the data agencies

8 and report, similar to Reinvent Albany we continue to

9 find examples where reports are published from the

10 Mayor's Office that are released in a PDF that don't

11 have structured data, and that aren't machine

12 readable, and so we support this particular bill.

13 Our community actually asked a few additional

14 questions in regards to making sure that these

15 reports are legislated to be put up on the agency

16 website or at least shared in a—in a place that has

17 access. I know that DORIS is in charge of receiving

18 these types of reports, but there has been some issue

19 with those types of reports being publicly accessible

20 in the past. Second to that is that we would love to

21 know that when an agency discloses what type of data

22 they are—the—the system. How do I say this? When

23 they release data that they describe, also the system

24 that's collecting that data this is critical to a

25 charter revision commission that was created Copic

2   where there is the city's data dictionary or not data

3   dictionary, data catalogue, which was the first data

4   catalogue produced. It specifies exactly what are

5   the fields that are FOILable, and we think that every

6   report that the—an agency hands off to the

7   administration should also very clearly indicate what

8   information is public accessible either through the

9   Freedom of Information Law so that way we can bet

10 around some of the privacy concerns that were

11 expressed by the Administration. And then in regards

12 to 1098, Council Member Kallos' bill in regards to

13 the digitization of historical data. Over the past

14 summer we've been working with community boards to

15 understand their tack and data needs, and what we've

16 discovered is that there are an internal desire to

17 digitize archived agendas, and permits that have some

18 type of stipulations around them. For example,

19 liquor licenses currently are essentially a

20 structured document that's held at the state, but my

21 communities board provide explicit stipulations on

22 how, when, and where alcohol can be served within a

23 liquor establishment, and none of that is structured

24 data. So, we hope that through this bill that there

25 is a conversation and value that is placed upon

2   getting community boards to digitize their historical

3   data, and then in our written testimony we have some

4   other concerns in regards to how did community groups

5   and community institutions are consulted within that

6   particular bill? What exactly and how is the

7   methodology of public valued defined, and then

8   ultimately, you know, how do we use open source tools

9   that essentially build a framework for best practices

10 across all agencies to go through their historical

11 pieces of data that can then be shared publicly and

12 available in a machine readable format.

13   CHAIRPERSON KOO: Thank you. We are also

14 joined by Council Member Lander. Next one.

15   YALE FOX: Okay. Hi, my name is Yale

16 Fox and I'm the CEO of Rentlogic. We are a tech

17 powered standards association that grades every

18 apartment building in New York City as a A, B, C or F

19 based on health and safety standards. Very similar

20 to how the city grades restaurants. Our software

21 uses Open Data from HPD, DOB, and ECB to evaluate how

22 well a building complies with the city's warranty of

23 habitability. We focus on co-compliance and

24 violations, not whether or not a building is

25 considered luxury, and we don't accept reviews from

2   tenants like a user generated review site. The

3   greatest use of our platform is that is distills

4   building histories into an easy-to-use public

5   interface that's accessible to anyone. It helps keep

6   renters from moving into bad buildings while also

7   giving recognition to good landlords, and it helps

8   make it easier for cities to understand their housing

9   stock and code enforcement needs. Today,

10   approximately 1 in 8 New Yorkers uses Rentlogic when

11   they're searching for a new apartment. This is

12   changing the market because renters are beginning to

13   seek out A rated buildings over ones with histories

14   of bad management, and it's worth noting that our

15   data shows that most of the landlord are actually

16   pretty good, but a few bad actors ruin it for the

17   community, the city and frankly for the reputation of

18   the industry as a whole. As previously mentioned,

19   the data or algorithm uses becomes primary—primarily

20   from HPD, DOB and ECB, and our work wouldn't be

21   possible with that open data initiatives because of

22   how the city tracks these violations like—like mold,

23   bed bugs and heat and hot water problems. These same

24   data sets are used by tens of thousands of companies

25   for—for completely different reasons, and while the

2 data is good, there is a lot of room for improvement.

3 You've probably heard before that it's not coded or

4 machine readable, but it should be almost like

5 numerically coded. So, instead of saying bed bugs,

6 it should say 0103 bed bugs. That's a lot easier for

7 a—for a piece of software to pick up because there's

8 10 different ways of spelling it. One problem that

9 you see with the open data as well, and as that's

10 just in housing is how it's being collected in the

11 first place, and with garbage in, you're going to end

12 up getting garbage out. So, in our platform if a

13 property owner has a low rating due to having many

14 violations, we actually send a third party state

15 licensed building inspector to conduct an inspection,

16 and—and verify that the problems have been fixed and

17 that the violations have been cured. We built an app

18 that-that can go on inspector's phones to insist—to

19 assist with inspecting the building and I don't

20 understand why the city doesn't do something like

21 this. It will make sure that the information that's

22 collected is more than just data. There can now be

23 media attached to it, and it will also force it to be

24 standardized, which will fix all the misspellings

25 and—and entry problems. While something like that

2   may potentially work for today in moving forward,

3   there is still all the data that's been collected in

4   the past. So, every startup at least that we've

5   spoke with that wants to work with city data—city

6   data has to put in a ton of tedious and expensive

7   work into just cleaning it up and to—to get it to a

8   place where you can use it. So, for us it took us a

9   year just to get the data into a format where we

10   could actually analyze it. So, I think the city

11   should—should figure out a way to clean up the

12   existing data, and there might be an easy way to do

13   it to work with the big apps or a Beta NYC or a

14   similar type of organization, and have like a

15   competition. It doesn't have to be for a lot of

16   money. People would actually jump at something to—

17   at—at the chance to work on a problem like this, and

18   in addition to that, thousands of apps that are out

19   there would more or less be improved over, and I, and

20   not to mention New York City would be recognized

21   worldwide and by the open data community, which could

22   possibly start to inspire other cities to start to do

23   similar things. So, in conclusion, we need the right

24   data to make the right decisions, and the hard part

25   most of it is actually already done, but if you were

2 to clean up the source you would see an immediate

3 positive over—positive effect almost overnight. Thank

4 you for the opportunity to testify.

5   CHAIRPERSON KOO: Thank you. Thank you

6 for all your testimony, and support.

7   RACHEL BYRD: Is it okay?

8   CHAIRPERSON KOO: Yeah one question.

9   RACHEL BYRD: It may not. Okay.

10   CHAIRPERSON KOO: Okay, if have

11 something, please.

12   RACHEL BYRD: Yes, I was able to share.

13   CHAIRPERSON KOO: Please share and finish

14 it, yes.

15   RACHEL BYRD: I am not the—I'm going to

16 discuss [coughs] the—in addition the confidentiality

17 that is—is touched on in the Speaker's Bill. The

18 access to agency data for purpose of better oversight

19 while emphasize—emphasizing the importance of

20 protecting confidentiality [pause] helps-helps—help

21 determine—I'm sorry—or personally identifiable

22 information by creating a new method and safeguards

23 for access to such records. PULP supports the

24 principle of broadening access to agency records

25 while protecting the confidentiality of those New

2 Yorkers to which such records refer. In the public

3 utility field protection of the personally

4 identifiable information or PII, of consumers is

5 strongly required by both state and federal law. It

6 is consistent with such safeguards. Therefore, for

7 the Speaker's bill to similarly take steps to protect

8 PII while allowing the underlying data to be

9 scrutinized as part of legislative oversight. For

10 example, anybody who receives one of eight public

11 assistance programs registered with HRA, is eligible

12 for Utility Consumer Discount or low-income discount,

13 but that information has to be carefully protected.

14 So when someone is found eligible when they apply, it

15 is—that information without the background

16 information, the details is conveyed to the utility

17 who then puts the consumer into a low-income discount

18 program. So, but there had to be extensive

19 negotiations to protect their privacy in that

20 process. The Speaker's Bill is also critically

21 important in the sustained Council to fully explicate

22 agencies' budgets when making their decisions during

23 budget consideration. Certainly in 2018, Council

24 Members repeatedly requested details of programs

25 asking agencies to be more transparent in their

2 submitted budget—budgets. For example, when

3 something called a one-shot was referred to during

4 one hearing, it was discussed solely in terms of

5 rental assistance. While critically important to New

6 Yorkers, one-shots also refer to a state required

7 utility assistance program under PSL 131S that

8 provides an essential safety net for low-income New

9 Yorkers. New York City is approx.—is appropriately

10 responsive to consumers in need of these funds often

11 the difference between warmth and light and cold and

12 dark, and hopefully are more rigorous and aggressive

13 pursuit of utility practices could actually save the

14 city money because they would not need to draw on the

15 131S funding. How much, however, is not evident

16 given the absence of such detail from the budget. We

17 have tremendous resource in DORIS and I'll leave that

18 into the written part of the statement. I'll—I'll

19 pass on the rest of it. I appreciate your time and

20 patience with my technology problems.

21   CHAIRPERSON KOO: Thank you. So, Council

22 Member Kallos, you have a short question, right?

23   COUNCIL MEMBER KALLOS: I just wanted to

24 just thank Yale Fox over at Rentlogic. I have gone

25 through that same dataset. I would love to share my

2 pivot table of every single different way we write

3 violations with you. Ultimately, I'm happy to

4 collaborate with everyone on the table around

5 another—we shouldn't have to do a bill on this, but

6 it would be really amazing if we had information

7 architects in the city of New York who could do

8 something, which is like the first thing I learned

9 which is normalizing data so that the data that goes

10 in is normalized. We've been working with Beta NYC

11 and Noel and trying to get them to normalize the data

12 going into the city records so that the data could be

13 useful. I—I just want to ask to anyone on the panel,

14 but in particular Yale and Noel but like if you—if

15 you were a city employee, and they let you in that

16 white room with all the city's data, how would that

17 be different that just being able to request an

18 individual piece of data that you know already

19 exists?

20   YALE FOX: I have to be wearing a

21 different hat, Council Member. Like if—if—I don't

22 think the general—necessarily the general public

23 would be allowed into the type of that the Speaker--

24   COUNCIL MEMBER KALLOS: [interposing] I'm

25 asking if you were a city employee what could you do

2 with being in that clean room versus just knowing

3 that maybe there's this piece of information you

4 could request.

5   YALE FOX: I would be—so as a cyclist, I

6 would be looking at culminating kind of a then

7 diagram of 3 to 1 block bike lane complaints versus

8 actually looking at the moving violations and the

9 different types of summonses that would have been

10 written in regards to bike—bike lane blockage or

11 illegal parking. Currently, we're—we are not allowed

12 to look at NYPD moving violations and so as we've

13 seen through the Vision Zero, people who tend to

14 block lanes are atrocious drivers, and I know that

15 Council Member Lander has introduced some legislation

16 to look at bad drivers, and this is a way that we can

17 start calling out behavior change in regards to

18 drivers or also corporations like deliver groups

19 that—delivery companies that are routinely using bike

20 lane as staging areas for your amazon packages. So,

21 that would be like one—one—one then diagram. The

22 other one would be looking at deteriorating housing

23 conditions. It's to be able to pull together various

24 DOB databases. We'll-we have some write-up of it on—

25 on our website, but we're essentially looking at how

2   311 quality of life service complaints like Yale is

3   integrating, potentially leveraging DOB permits to

4   see exactly how many permits have been applied to a

5   building that has been under tenant protection, since

6   the start to identify what are the most vulnerable

7   buildings that currently are receiving some type of

8   construction and yet have had a high quality of life

9   service requests meaning that there's an immediate

10   threat to the tenants in those buildings around the

11   loss of affordable housing.

12   RACHEL BYRD: We have a—a then diagram.

13   The medical examiner does not have any requirement to

14   report suspicious deaths to the State Public Service

15   Commission, and we recently had a tragedy this winter

16   where a gentleman died from hypothermia. He had his

17   heat cut off, his gas from non-payment, and the

18   Public Service Commission keeps or the state keeps

19   detailed records of deaths due to—or mishaps or

20   disconnections, turn-offs, shut-off notices, and the

21   Medical Examiner has records of suspicious deaths or

22   questionable deaths, but there is not connection made

23   between the two unless it's highly publicized where

24   this particular incident was. So, we would like to

25   see some connection between those—between those data.

2   CHAIRPERSON KOO: Alright, thank you all.

3 [off mic] And we are also joined—[on mic] we are also

4 joined by Council Member Ulrich. Thank you. Yeah.

5 So, thank you all for coming to participate in this

6 public hearing. This meeting is adjourned. [gavel]

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

C E R T I F I C A T E

World Wide Dictation certifies that the foregoing transcript is a true and accurate record of the proceedings. We further certify that there is no relation to any of the parties to this action by blood or marriage, and that there is interest in the outcome of this matter.

Date ____ October 2, 2018 _____