Microsoft

November 17, 2017

[Addressee]

Re: *Microsoft's Comments re Big data and Innovation: Implications for Competition Policy in Canada*

Dear:

Thank you for the opportunity to offer comments on the Competition Bureau's September 18, 2017 draft for consultation on "*Big data and Innovation: Implications for Competition Policy in Canada.*" We appreciate the Bureau's preparation of this draft discussion paper and are in broad agreement with its treatment of the challenges posed by the application of the Competition Act to transactions or conduct involving "big data". We also commend the Bureau for taking steps to more fully illuminate the Bureau's general principles concerning the enforcement of the Competition Act to the accumulation, acquisition and use of big data in a constructive way that will help to guide businesses to navigate intersection of data-driven technologies and competition law generally.

We appreciate that the Draft emphasizes that the purpose of competition law in Canada is not to regulate prices, profits, market shares, or for that matter, the amount of data that companies gather and use. We agree that competitive market forces are indeed the best means of ensuring an innovative, efficient, and prosperous economy. Reinforcing this principle provides important balance and guidance as competition in the big data era may lead firms with access to superior data or algorithms to grow at the expense of others, perhaps rapidly, which is not assumed to be "bad" for competition in general.

Modern computing power has indeed expanded the ability to collect, store, process and analyze data on a large scale, raising complex questions about the commercial nature of the accumulated big data and the implications for competition in numerous industries across the global economy. Because artificial intelligence (AI), machine learning (ML) and the Internet of things (IoT) promise to make big data analytics a central feature of virtually every area of commerce, we also appreciate that the Bureau understands that advances using big data are occurring rapidly and in unpredictable ways that could be slowed or even stopped without careful attention to the risks of "overenforcement" (*i.e*., acting where there may be no genuine harm to competition).

The challenge – which the Draft goes a long way toward addressing – is for the Bureau and the courts to separate cases requiring closer scrutiny from the bulk of cases where data ownership and usage is economically beneficial, drives innovation and is competitively benign. Ultimately, agencies should apply traditional tools and avoid acting on models of competition analysis that do not rely on hard evidence about the nature and use of the subject data. Thus, we commend the Bureau's general conclusion that current analytical frameworks can be applied usefully to cases involving big data, while inviting a discussion with stake holders in unique circumstances where it appears that new or more specialized theories might be required.

We have only a few comments and suggested revisions to the draft that the Bureau may wish to consider.

Section II.B definition of big data, and the "value of data". We agree that the value of big data may not result from its being sold at some price; instead, value can result from its use as an input into other products and into marketing campaigns. However, we respectfully submit that it also bears emphasizing that large sets of structured, unstructured, labeled or unlabeled data *may have no known value* in the development of ML or AI applications and may in fact not have any significant value. Data is thus best viewed as an "asset" that may or may not have value just like any other asset depending on its commercial utility.

The key question in each case is what model is being developed in each case and what data specifically needs to be deployed to predict? In each instance, many different types of data (including "proprietary" data or commercially available sources) are used as inputs for modelling, and in many cases the relative value of a given data set cannot be determined with any certainty until an algorithm has been successfully trained and then tested using separate portions of the data set.

Section III.B challenges with assessing market power of data in mergers. Microsoft agrees that there are many indicia of market power that could theoretically be addressed in each case. We respectfully submit, however, that it is possible and easier to assess relevant indicia of data market power within the structure of this simple four-step framework:

- ***Do the parties own or control the relevant data?*** Data that is only processed by a company, and not owned or controlled by it (within the meaning of data protection laws) is not likely relevant from a competition perspective. A data processor generally can access the data only to accomplish the agreed data processing on behalf of the customer (controller) and not for its own use. Moreover, simply providing tools that customers can use to store or analyze (process) their data does not convey any special "data asset" to the vendor of those tools. A data controller, on the other hand, can use its own data and, typically with the users' consent, volunteered or observed data for specified purposes that include AI for improving its own software and services.

- ***Is the relevant data a proprietary input?*** Data that is not sold as a product or not otherwise available to competitors is generally unlikely to represent a unique critical input in any competitive sense. In fact, if the data controller has downstream competitors (those that develop a product using the data), that strongly suggests that those competitors have access to comparable internal or external data sets. Using proprietary data to seek competitive advantage is economically efficient behavior that drives innovation.

- ***Do reasonably available substitutes for the relevant data exist or is it truly "unique"?*** In most cases, the data is useful but not essential to compete or there are reasonable substitutes such that the way in which the owner or controller may choose to leverage that data should not raise a significant competition issue. Many data already have or could have full or partial substitutes that can be combined or can be collected by starting new lines of business. Only in rare cases is a given data set necessary for competition in each market and without reasonable substitutes, *i.e.,* truly "unique." To the extent that a firm can acquire or combine data from other sources (or if customers in the relevant market "multihome") or recreate relevant data to train an algorithm by other means, even if with some effort, those data sources may be able to provide the substitute necessary for competition in that relevant market.

- ***Is the data a critical input for products of others?*** Simply having more data than anyone else does not protect a company from competition. Another critically important aspect of ML and AI is that, at their core, they rely on dynamic experimentation. Indeed, "big data" is increasingly being understood to mean the application of a *multiplicity of signals* rather than just sheer scale of data. Increasingly, it is the multidimensionality of the data inputs

that matters. Moreover, having many signals in the same set of data is not the goal; rather, the goal is to include a *greater number of diverse signals*, which provides more explanatory power because the results are a better fit for the intended use. This is not the same as having just more data – because the benefits of sheer scale can diminish rapidly – but better data suited to enhancing given products or services.

It is true that AI and ML systems often leverage multiple massive data sets for their creation, operation, and improvement. Even if a data set is a "must have" data set without reasonable substitutes, *it should not raise competition concerns unless the company controlling that data is dominant in the upstream and downstream market, and it realistically can withhold the data from use to foreclose competition in a downstream market, i.e.,* a market for a product that is built using that data set.

Section III.F.2 Quasi-structural and behavioural remedies. In this section, the Bureau notes that "beyond requiring a firm to cease a particular course of conduct, an appropriate remedy may, in exceptional cases, require that data be made available to competitors for use as an input, for example through the compulsory licensing of intellectual property, which is not new to antitrust. In such cases, data may be considered to be an "essential facility," without access to which a firm may not be able to compete effectively in a downstream market. Providing access to the data may be an appropriate quasi-structural remedy allowing potential competitors in the downstream market to overcome their main barrier to entry. However, considerations related to intellectual property rights and the structure of the relationship with data providers may pose challenges to the design and implementation of data remedies. The Bureau is mindful that mandating licensing of data can potentially chill incentives to innovate".

We would respectfully submit that more could be said about the "exceptional cases" where such quasi-structural remedies might be considered. Specifically, "a scenario may exist where a dominant firm (i) controls a data set, (ii) uses the data as an input for its dominant product, (iii) has the legal (data protection and privacy) right to transfer the data set to others but has not done so, (iv) that data input is necessary for competition by others in a downstream market where the firm is also dominant, and (iv) there are no reasonable substitutes. Such a scenario would raise a competition law concern. But all things considered, this scenario will be *very rare*."

Very truly yours,

Greg Sivinski
Assistant General Counsel
Microsoft Corporation