CPI TALKS…

…with Paul Gilbert & Maurits Dolmans

In this month's edition of CPI Talks we have the pleasure of speaking with Paul Gilbert and Maurits Dolmans, of the London office of Cleary Gottlieb, Steen and Hamilton LLP. Maurits recently wrote articles entitled "Should We Disrupt Antitrust Law?" and "Pandora's Box of Online Ills." Paul has given presentations on "Competition Law and Big Data." Both have extensive experience in the IT sector. They have advised various clients in this area, but speak here on their own behalf.

Thank you, Paul and Maurits, for sharing your time for this interview with CPI.

1. **Recent months and years have seen intense discussion on "big data" and its role in competition in the digital economy. Much of this discussion relates to the question of whether, in the language of the competition rules, "big data" are a mere asset, a "barrier to entry," an "essential facility," or all of the above. Is the classic terminology misleading when applied to this new phenomenon? Rather, does this debate merely underline the need for a careful case-by-case approach to competition issues involving complex technological questions?**

We don't think the terms need changing. They are part of a framework that is used to differentiate those situations where competition intervention is needed from those where it is not. Intervening in the wrong cases will only undermine incentives to invest in collecting and analyzing data.

In most cases, data are an input for a product or service. It is not the service itself. So, is it right to ask whether data are an essential facility? Put another way, do competitors need access to that particular set of data for competition to exist in a downstream market? If not, intervention is unwarranted, and could do more harm than good. There may, in theory, be cases where data are an essential facility, because the data are indispensable and the source is no longer available. But they will be the exception rather than the rule.

Holding data is not like owning a harbor or controlling the electricity network. In those cases, anyone trying to operate a shipping service or supply electricity may have to have access to the infrastructure. Data are not "owned" or controlled in that way. I could record the color preference of every car buyer if I thought the information was going to be useful, and I may be only person to "own" that data. But that doesn't make that database an essential facility, and it doesn't stop anyone else doing exactly the same thing. Many types of data are "non-rivalrous goods," in the sense that they can be duplicated and recreated without limit.

It is also tempting to treat all types of data as the same, when they are not. Personal data are different from observed data, for instance. Observed data often has a short lifespan. For my car-color database to remain useful, I would probably have to repeat the exercise regularly. Otherwise the data will become stale and worthless. And I may not collect it at all if I had to give it away to all my competitors.

Personal data are different. If a consumer uploads their photographs and contacts to a social media platform, they could become locked in. If many users are locked in, that may create a barrier to entry for new social media platforms – in the sense that gathering an equivalent set of data is going to be more difficult or more expensive for the new entrant than it was for the first mover. In such a situation, ensuring that consumers can transfer their information freely and easily between different platforms can be pro-competitive. This, in turn, may require standard setting to ensure that data can be easily ported to new service providers.

So, the terminology is right. It just needs to be applied carefully, and seen in context.

2. Information scientists famously refer to the so-called "DIKW pyramid." In short, mere raw "data" provide "information" that experts can use to produce "knowledge" and (hopefully) "wisdom" that can yield valuable results. This leads some commentators to suggest that, at least in certain circumstances, data should not be considered to be a true "barrier to entry": the real barrier would be the technology and expertise needed to extract "wisdom." To what extent should competition practitioners, enforcers, and courts rely on expertise from the sciences to assess these and other questions in individual cases?

We agree. In this information pyramid, data are merely raw measurements, whereas information is an understanding of the relations between the data, knowledge is an understanding of the patterns that emerge, and intelligence or wisdom is an understanding of the principles that allow judgment, better decision making, and prediction.

Data are useless without the ability to analyze them intelligently and creatively. Take away the data from any online service provider (but leave the engineers) and they will quickly be back in business. Take away the engineers (but leave the data) and they will soon be in trouble.

There are lots of examples. Start-up companies such as WhatsApp, Hailo, BlaBlaCar, Snapchat, Instagram and Pinterest have all been able to build successful and innovative products with access to little or no data at the start. If data are a barrier, it can certainly be overcome.

Studies have shown that data quickly loses its incremental value after a certain point. According to the law of diminishing returns, once a critical volume is reached, the marginal value of additional data for statistical analysis is limited. There is no value in simply collecting more and more data.

Competition authorities and courts should take this into account. They need to ask what data is really needed, how much is needed, and how easily it can be collected. Data is everywhere, it is cheap to collect, can be bought and sold, and is non-rivalrous – it is not used up like oil.

So, it is right to think of intelligence and creativity as the real drivers of competition. As to whether they could become barriers to entry, it seems unlikely. There seems to be no shortage of either.

3. Recent reports (including the UK Furman Report, the EU Commission's Crémer Report, and the ACCC's Digital Platforms Report) discuss the treatment of data under the competition rules in some detail. Specifically, the Crémer Report underlines the need to adopt a careful approach to the "essential facilities" concept if applied to data. In your view, are there grounds for a modified approach to the "indispensability" and "new product" criteria under the classic essential facilities rules should they be applied to data?

Many competition authorities and regulators are grappling with the question whether we should be quicker to intervene in digital, data-driven markets. Should companies that collect data be required to give access to that data as a way of helping their competitors, even if they have not done anything that would otherwise be considered anticompetitive?

This is market engineering rather than antitrust, and is a risky path to follow. It means forcing Firm A to support Firm B even if Firm B does not need that support to compete. Forcing data-sharing in this way would pervert the normal competitive incentives of both. Firm A will be less inclined to invest in collecting data and designing innovative ways of doing so if it has to share the results with its competitors. Firm B will not have incentives to invest either, because it can free-ride on the efforts of Firm A.

The Crémer Report recognizes precisely this concern. It recognizes the risks that a lower threshold for intervention would have on commercial incentives. It says that "*a thorough analysis will be required*" to determine whether access to data is "*truly indispensable*" in order to compete in neighboring markets. It also recommends that competition law principles should be central to any sector-specific regulation that is designed to mandate data access.

We agree. It is easy to lose sight of the benefits that come from companies investing in collecting and analyzing data in ever more inventive ways. Undermining these incentives could be hugely damaging. As we said before, with respect to personal data, the better option is to give individuals the right to port their data between different services providers, so as to encourage the latter to keep improving their offering to persuade the users to stay with them.

4. One of the most controversial forms of "big data" relates to personal information used to provide services such as online social media, search, and advertising. Individuals' rights concerning such data are regulated separately under rules such as the EU GDPR and the California Consumer Privacy Act. Certain competition remedies (notably mandated access) would raise obvious potential conflicts with these specific regulations. How should this circle be squared? Is this an argument for non-intervention by competition enforcers in cases of such potential conflict, or at least careful remedy design (e.g. anonymization of user data if disclosure is mandated)? Is enough being done by data protection regulators and competition enforcers to coordinate in this regard?

It would be perverse for competition law remedies to override data protection. Central to GDPR and other data protection rules is the principle that individuals should be able to control what happens to their personal data – information about them.

I may be happy to share my personal information, health records and banks details with one company that I trust, but not with others. That is undermined if the company is then forced to share my details with someone else that I don't trust, or have perhaps never heard of. When a consumer shares their personal data, they have to be confident that it will be protected and won't be shared without their consent.

A good example of how the system can work is in Open Banking. If I want to use a new service – perhaps a FinTech start-up company – that company will be able to access my bank account details securely, but only if I give my consent.

Many of the largest technology companies (including Apple, Facebook, Google, Microsoft and Twitter) are currently working on a similar initiative, called the Data Transfer Project. The purpose is to allow consumers to transfer their information freely between providers using open source coding. It will mean that users do not have to download their data from one service before uploading to another. This is procompetitive. It helps users to switch even more quickly between providers and try out new services, as well as backing up their data on multiple services. And the user always remains in control. This approach to "portable data" creates new opportunities for intermediary services or portals that help users transfer data, or manage multiple platform settings.

This type of initiative facilitates competition and new entry, while still respecting consumers' rights and data protection rules.

5. A common complaint levied against companies possessing large datasets is that their existing "scale" renders it impossible for new entrants to compete. Are there alternative means that regulators, policymakers, and industry could use to make large-scale data available to new entrants (e.g. databases of photo, video, or anonymized personal data that could be used to train machine learning algorithms)? Would such an approach be feasible as a remedy in a hypothetical competition case? Or is such an approach best left to bespoke industry or government-mandated initiatives?

It is wrong to assume that scale is necessarily critical or a barrier to entry. This idea ignores the law of diminishing returns. Experience in many areas where data are used as an input, like machine translation, image processing, speech recognition, or teaching artificial intelligence systems, the first million data are much more important than the second and so forth. Huge amounts of open source datasets are increasingly available for AI training (from sources as various as kaggle.com, FiveThirtyEight, Google Public Data Explorer, or from EU Open Data Portal, Data.gov. WHO Open Data repository, and World Bank Open Data, etc.). Open source software and cloud computing makes it easier for start-ups than ever before to run complex computations on these open data.

We have already mentioned several examples of companies that have been able to grow quickly without having access to large datasets when they started. In each of these cases, "scale" was not a barrier. If anything, they show that innovation is critical and that data are easy to gather.

The European Commission has also considered this question many times in merger cases. Whenever it has looked at markets in detail, it has found that rivals are able collect the data they need. In *Facebook/WhatsApp*, for example, it found that "*there will continue to be a large amount of Internet user data that are valuable for advertising purposes.*"

In *Microsoft/Skype*, it found that barriers to entry were low, pointing to the "*immediate success*" of new entrants Viber, Fring, and Tango. Viber was downloaded more than a million times within three days of its launch, 10 million times within two months, and 15 million times within six months.

There may be cases where access to an existing database is critical for competition to exist. It will depend on the type of data and whether the database is replicable. In the UK, for example, the National Health Service provides access to huge amount of anonymized healthcare data for analysis. That data wouldn't be available otherwise and cannot be replicated. But this is an exception. Online user data are very unlikely to fall into this category.

In the EU, we have a tendency to blame U.S.-based online firms (ignoring China for the time being) for the lack of EU online powerhouses, and for the difficulties of digital disruption. They are conveniently big targets. But a focus on data sharing is not the solution. The societal problems resulting from digital disruption should be solved by targeted and proportional regulation, rather than forcing firms to share data across the board. If we want to grow EU-based innovative powerhouses, the answer is a policy of creating an environment where innovation can flourish: innovation hubs combining academic centers of excellence and existing technology businesses with carefully targeted Government-sponsored open projects, which together can fertilize start-up businesses by providing talent and risk capital, a culture of entrepreneurship, critical thinking, and creativity, nurtured in an environment of fair and balanced tax laws, IPRs strong enough to encourage innovation but not so strict as to block new entry, and competition rules that preserve opportunities for new entrants without punishing them once they become successful. There is plenty of big data; what is scarcer is technical talent and business creativity, and wise regulators.

![CPI Competition Policy International logo]

CPI Subscriptions

CPI reaches more than 35,000 readers in over 150 countries every day. Our online library houses over 23,000 papers, articles and interviews.

Visit competitionpolicyinternational.com today to see our available plans and join CPI's global community of antitrust experts.