



Digital Public Goods

SHANE GREENSTEIN

Kellogg School of Management

.....Precisely how does the online world provide public goods? That is the question for this column.

Public goods in the digital world contain some of the same features as those in the offline world. Yet, there are some key differences in the boundaries between public and private, and that shapes what arises and what does not.

That will need an explanation.

Definitions

Let's start with a breezy description of the economics behind a public good, and translate it to the digital world.

According to standard textbook definition, a public good possesses two traits—nonrivalry and nonexcludability. Nonrivalry stresses that one person's consumption does not preclude another's. For example, one person breathing fresh air does not prevent another person from doing the same. Nonexcludability focuses on the absence of gateways on consumption. For example, fresh air is nonexcludable, because it is impossible to stop several people in the same area from breathing the same fresh air.

Needless to say, all of this is easier to get right in theory than in practice. Nonrivalry justifies a single supplier, either a private organization or government entity. Lack of excludability provides the justification (or excuse) for government taxes, which (sometimes) provide a useful way to fund public goods that cannot be funded through a standard pricing

mechanism. No wonder arguments over such matters are never settled.

Now consider the application of these definitions in the online world. What parts of the digital world possess the features of a public good—namely, nonrivalry and nonexcludability?

Consider technical knowledge. It is a public good offline, and online, too. In both cases it results from effort and investment in research. As illustration, think of the protocols that make the Internet operate, as defined by the Internet Engineering Task Force (IETF). Obviously, one person's use of a newly invented protocol does not preclude another. Moreover, the group of inventors and its large community of users set it up so the IETF does not restrict access. No gateway governs use of the protocols.

Other common technologies, such as those emerging from the World Wide Web Consortium, fit a similar definition. So do the standards for 802.11, also known as Wi-Fi, as defined by the IEEE. The same goes for many other standards that few mainstream users have ever heard of, such as 3GP, DOC-SIS, and open source projects such as Apache.

Many of these public goods are interesting because they are not supported by government taxes today, though government funding invented them or got them started. The participating firms today pay their dues through paying

their personnel to support these efforts and improve the technology. Although nobody has ever figured it out precisely, it seems pretty obvious in retrospect that the gains to society from inventing these technologies far exceeded the costs to the government.

Perhaps more interesting in their origins, some online public goods—notably, the W3C, or a working group in the IEEE, or Apache—emerged from somebody's initiative. It took enormous effort to make those initiatives effective, and to create an effective technology. Yet, if a nonprofit organization offers the technologies, the government's involvement extends only to providing legal frameworks in which these organizations function.

Those examples also illustrate something about what is not a public good. For example, many firms operate private consortia to develop proprietary bilateral standards for their digital goods, especially software standards. Those consortia are not public goods when the technologies stay with the consortia participants, excluding others. Many firms also employ cross-licensing and other contracting arrangements that share technologies. Again, these are not public goods if others are excluded.

Several examples blur these boundaries. WebKit and Hadoop, for example, take inputs from private firms and open source, and possess nonrivalrous and nonexclusive aspects in their most common forms today.

Information services

Consider Wikipedia. It fits the definition of a public good. There is no exclusivity, and one person's use does not prevent another's use. For similar reasons, Brewster Kahle's Internet Archive also looks like a public good. The same goes for the Sloan Digital Sky Survey, which provides pictures of virtually every star and constellation (if you have never seen it, it is very cool, by the way).

Countless public goods of this type serve niche audiences. Every museum supports a webpage—the Smithsonian, the Met, the Charles Babbage Institute, and many more with educational missions. These sites do not want to be exclusive or rivalrous. Every nonprofit with an educational mission also provides a public good, educating others about the details behind many topics, such as climate change, preventing genocide, feeding starving children in the developing world, and a bazillion other subjects.

Archives are a variant, and as long as they are nonexclusive, they act like public goods. A well-known academic archive is the Social Science Research Network, or SSRN. At last count, it had half a million abstracts in its records. Another known archive, arXiv, is a little older and even larger, and it gets frequently used in the hard sciences, particularly physics. At last count, it exceeded half a million papers. Both archives receive thousands of new papers a week. These organizations also do not use taxes. They live off donations or foundation support or other forms of in-kind goodwill.

These examples also illustrate another boundary between the public and private. Many firms—journal publishers, for example—maintain archives and databases. Though these are nonrivalrous, many publishers maintain paywalls and charge for access, perhaps offering limited nonexclusion as a teaser to generate interest.

That gets us closer to a well-known situation—namely, what happens in news. Many major news organizations find themselves on the boundary between

using a paywall or not. *The New York Times*, *The Wall Street Journal*, and *Financial Times* attract readers with free access to single articles, but charge for accessing more from the archive, albeit to different degrees. What to call that—a public good at small scale only?

Some news organizations take another approach, eschewing the paywalls and instead making the content—and ads!—accessible to every reader. Advertising can be sold and targeted based on the demography of visitors or click-throughs, even if the content next to the advertising is nonexcludable to users. Hence, in the last decade there has been an explosion of news and media sites with missions that resemble public goods and, yet, they have revenue-generating activities.

It is not just news. Consider a ratings site. Yelp is a good example to illustrate a broader trend. Specifically, Yelp helps aggregate user ratings for new readers, and user behavior helps the site target its ads. Sites focused on user-generated content also work with this combination. Similar remarks could be made about IMDb, Reddit, and countless more sites using this well-trodden business model.

The model extends to video, too. Think about YouTube; it is nonexclusive and nonrivalrous, and it sells ads. In short, many sites with resemblance to public goods are all over the Internet.

Hybrids

As noted, a number of businesses blur the boundary between the public and private. For example, governments collect weather data from satellites and many private firms repackage it, and then sell ads next to the weather map. The ads pay for the reformatting and updating. Taxpayers paid for the satellites, and the government data effectively lowers the costs of running firms displaying the data.

The US patent database provides another example. That data is available online for anyone to use on a site sponsored by the United States Patent and Trademark Office. Google Patent reformatted it to make it easier to search

(in my humble opinion), which keeps users on their sites.

Here is one last hybrid that really blurs the boundary. Local governments collect data about traffic in (typically) limited efforts, using traffic sensors on highways, making the results available on rudimentary websites. Private firms also could collect such data (and do, often from aggregating cellphone signals). Most interesting, such data are repackaged by firms, such as Navteq, in more useful forms for display both online and reselling in packaged formats.

To be sure, some hybrids on these boundaries can flounder if they run afoul of legal limits. Google Books contained the promise to be a public archive, for example, but the initiative could not get past the copyright hurdles and, despite the potential, remains shuttered today.


Underprovision?

Circling back to the main question, what about online public goods seems at all new? It's the expansive hybrids and the ad-support sites that go into new territory. Sure, there has always been some sort of hybrid offline, too, but the range and extent looks new to me.

That still leaves some questions about whether all public goods are provided. Surely the answer is no. There are still plenty of public goods waiting for more effective means to provide them.

More to the point, there are still unresolved issues requiring new digital public goods. Here, as with other online activities, look for innovation in the nontechnical arrangement as well as the technical aspects. MICRO

Shane Greenstein is the Kellogg Chair in Information Technology at the Kellogg School of Management, Northwestern University. Contact him at greenstein@kellogg.northwestern.edu.

 Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.