

# Open Data kann der Forschung schaden

## Open data can hurt science

Stefan Eisebitt, Direktor | Director,  
Max-Born-Institut für Nichtlineare Optik und Kurzzeitspektroskopie

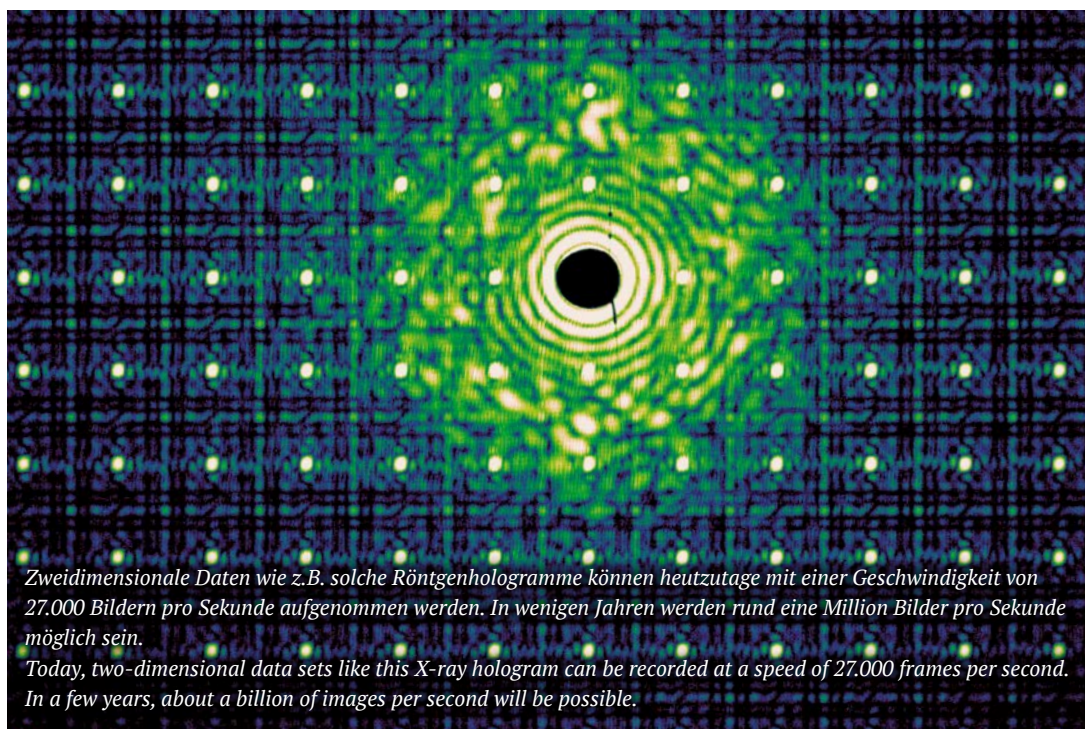
6

Die Bereitstellung von mit öffentlichen Mitteln generierten Daten – Open Data – ist in vielen Bereichen gesellschaftlich wichtig und eröffnet neue Geschäftsmodelle für Firmen und Dienstleistungen für Bürgerinnen und Bürger; man denke an Straßenkarten, Fahrpläne, Daten zur Luftqualität etc. Nun findet das Konzept Open Data im Kontext von „Open Science“ Eingang in immer mehr Absichtserklärungen auch in der Wissenschaftspolitik. Ist es gut, alle Daten öffentlich finanzierter Forschung allen Interessierten zugänglich zu machen?

Die Antwort darauf ist meines Erachtens so divers wie die Wissenschaft selbst - und lautet für manche Bereiche: Nein! Wissenschaft lebt von der Kreativität und dem Geschick

Making data generated with public money available to everyone—open data—is beneficial for many areas in society, creating new services for citizens and business models for companies. Think of street maps, itineraries or air quality data. Now, the concept of open data is being incorporated into science policies. Is it good to make data from publicly funded research available to all interested parties?

In my opinion, the answer to this question is as diverse as science itself—and for some scientific areas the answer is: No! Science requires the creativity and skill of the individual researcher. If we want to remain a leading knowledge society, it must be possible for young researchers to build their professional careers on successful research.



Zweidimensionale Daten wie z.B. solche Röntgenhologramme können heutzutage mit einer Geschwindigkeit von 27.000 Bildern pro Sekunde aufgenommen werden. In wenigen Jahren werden rund eine Million Bilder pro Sekunde möglich sein.  
Today, two-dimensional data sets like this X-ray hologram can be recorded at a speed of 27.000 frames per second. In a few years, about a billion of images per second will be possible.

der Forschenden. Wenn wir auch langfristig eine erfolgreiche Wissensgesellschaft bleiben wollen, müssen sich Karrieren auf erfolgreicher Forschung aufbauen lassen, denn nur so werden wir talentierten Nachwuchs für die Forschung gewinnen können. Daraus ergeben sich mehrere Konsequenzen für Open Data-Regelungen in der Wissenschaft: (i) Das Erstverwertungsrecht der Forschenden muss unbedingt sichergestellt werden, damit sich erfolgreiche Forschung in wissenschaftliche Reputation ummünzen lässt. Die erforderlichen Embargozeiten hängen von der Fachdisziplin ab. Die durchschnittliche Dauer einer Promotion ist die absolute Mindestzeit für eine Embargofrist. (ii) Daten sind nutzlos ohne die Metadaten, die ihre Entstehung beschreiben und sie nachvollziehbar machen. Es gibt Fachgebiete, in denen die Entwicklung eines sehr speziellen Experimentes Kern der Arbeit ist. So dauert es in vielen Bereichen der Experimentalphysik Monate und Jahre, bis ausgehend von einer Idee ein Messkonzept entwickelt und einzigartige Hard- und Software erstellt wird, um dann ein Experiment durchführen zu können. Da es sich nicht um standardisierte Messungen handelt, ist hier der Aufwand der Erzeugung von Metadaten gigantisch und würde u.U. ebenso viel Zeit verschlingen wie das Experiment selbst. Würde dies erzwungen, wären die Forschenden umgehend international nicht mehr konkurrenzfähig. (iii) In den Naturwissenschaften sind die anfallenden Datenmengen teilweise enorm: beispielsweise können an einem einzigen Instrument des Europäischen Röntgenlasers XFEL pro Minute zwei Terrabyte an Daten entstehen! Die Kosten für die IT-Infrastruktur, um Open Data in Deutschland auch für solche Experimente umzusetzen, wären immens. Ginge dies zu Lasten der der Forschung selbst zur Verfügung stehenden Mittel, hätte man ihr damit einen Bärendienst erwiesen.

In vielen Bereichen kann die Wissenschaftsgemeinde von Open Data sehr profitieren, für manch andere wären Standardregelungen desaströs. Es ist wichtig, dass eine differenzierte Betrachtung in mögliche Regelungen Eingang findet.



Several implications for open data regulations in science ensue: (i) the researcher must be able to exploit his or her data first. The required embargo periods depend on the research field. The average duration of research towards a PhD thesis is an absolute minimum time. (ii) Data are useless without their metadata, making transparent how the data was obtained and can be interpreted. In some scientific disciplines, the development of a particular experiment is the core activity. In many areas of experimental physics, for example, it takes months and years to progress from an idea to a concept for an experiment and to develop the required hard- and software to actually carry out that experiment. As these are not standardized measurements, the effort to generate metadata can be tremendous and possibly as time-consuming as developing and carrying out the experiment in the first place. If the generation of useful metadata was enforced, the scientist would immediately become less competitive in comparison with his or her international peers. (iii) In some areas of the natural sciences, large amounts of data are generated at ever-increasing rates, because this improves the experiments, and it is becoming possible due to technological advances. For example, a single experiment at the European X-Ray Free Electron Laser XFEL can generate two terabyte of data per minute! The cost for the IT infrastructure required in Germany to make all data from such experiments open and publicly accessible would be tremendous. If, in turn, these expenses would reduce the funding available for the research itself, we would be doing science a great disservice.

In many areas, the scientific community can clearly profit from an open data approach. In others, having to follow standard rules not adapted to the peculiarities of the field could be disastrous. It is pivotal that future policies are flexible enough to deal with these differences.