Ownership OF DATA, INFORMATION AND KNOWLEDGE IN RESEARCH ORGANISATIONS

The Department of Science and Technology is responsible for our policy on access to data, and for developing the policy framework on intellectual property from public and funded research, which is very much the domain in which we operate. Data, information, knowledge, and their relation to intellectual property, access, equity of use, etc. are important to an organisation. The very nature of data and their acquisition have changed dramatically and continue to change in a very dynamic environment. One of the main reasons we support the Ministerial Declaration on the Organisation of Economic Cooperation and Development is to maximise the use of data in the public interest.

Data and data functioning have changed fundamentally in scientific research, causing open access to become important, since there is the possibility of the digital divide becoming a biological divide, a bio-informatics divide or a health informatics divide. These new divides are opening up because data are privatised, and it is becoming a potential problem for developing countries. In South Africa, where researchers see data as something they own, we need to create a culture where the data generated are for public use. There might be custodians of data for a period, who increasingly transfer the long-term management and curatorship of these data to other people, professionals in information and knowledge management.

At the **data** level, we are more and more convinced that if we use public assets to generate the data, there should be a reasonable set of ways in which that data become accessible to the research community. Obviously one must maintain control over ethical and confidentiality standards.

At the **information** level, we see a strong correlation between information and copyright. We support the global copyright conventions and believe that copyright is important. Institutions automatically think they control the copyright of the employees because they have to contract to control this copyright, which is often not understood. There is much confusion over the proper management of copyright and intellectual inputs at the time of authorship.

At the **knowledge** level, compared to the data level, we support increased protection as opposed to openness. This is because we live in a knowledge-based economy, where you have to privatise knowledge to make money out of it.

Business methods are not patented in South Africa, as is the case in the USA. This conservative approach extends to patent legislation, while at the same time we should use it to the greatest possible extent to protect the outcomes of public research. A good example is SAAVI, the South African Aids Vaccine Initiative, where we have actively patented some of the research findings not to make money, but to use these as a tradable asset to work with other people. If you do not patent your findings, people simply use your knowledge while you are denied the opportunity to use it as a lever to develop important things such as AIDS vaccines.

Patents, royalties, and licenses are indicators of intellectual capital that is held by a nation. South Africa is a very important patenting nation at a level of two patents per million people per year in the US patent office. Your entry ticket to the global knowledge economy is your ability to secure your knowledge by patenting; thus if you do not patent you do not play in the knowledge economy. If we do not learn how to play in that area we will simply see our published information become the public knowledge that other people will patent. We should be patenting our knowledge, and putting it into the knowledge economy, yet, somehow many of our academics feel it is morally wrong to patent knowledge. Is it morally right to give away your 'diamonds' so that other people can make money out of them? If you have been paid public money to generate knowledge, you have to put it to the best use in the public interest, which includes turning it into an asset through patenting.

Many academics are surprised when their knowledge appears to be useful to other people. If academic science is generated indicating the industrial potential that leads to global warming, how can you not be engaged in solving the global warming problem? The work of the independent investigator working for years on some pithy problem is over, but it has not necessarily penetrated our consciousness. The relevance question has never been more relevant.

Large-scale problems, such as HIV/AIDS, global climate change, ocean dynamics, massive extinction of plants and many more are happening. Invasion biology is changing in the world and many problems, which arise, seem to be accessible to scientists. A major drive is the change of computing power in digital storage and resources. The ability to store data has changed dramatically and has definitely improved. However, the quality of the data may not have improved as much.

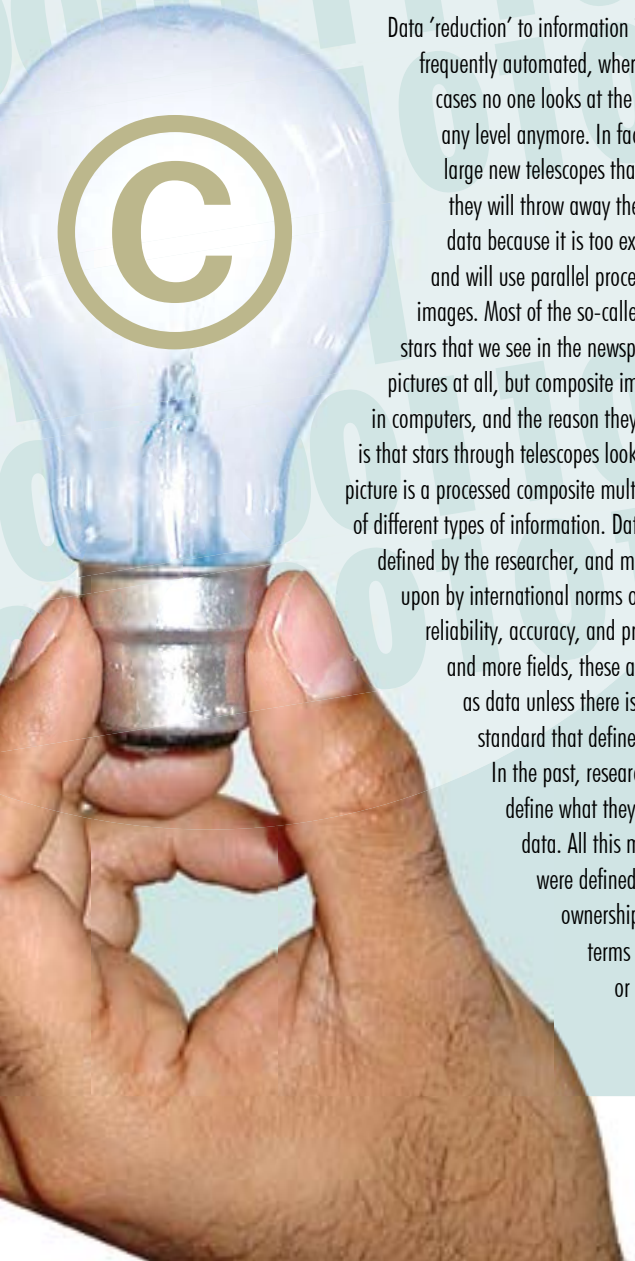
The original ownership of intellectual output is clearly with the individual when you do not work for an institution and you generate intellectual outputs. You can copyright it or, if you make an invention, you can patent it. Where people join institutions like public research organisations, an employment contract modifies a person's original ownership rights.

We have to understand that the view of the South African Government is that institutions that simply take these rights of ownership and do not grant benefit sharing to their employees are being irresponsible. There have to be benefit-sharing arrangements; someone is giving up a right in the employment contract. Until fairly recently the ownership of data was essentially relevant, with the main issues being traceability for the protection of rights. If you made an invention, you would have to have a signed notebook with the accurate data, particularly in the medical field. These original data should have been in a written form and you would have been able to follow your rights that way. When issues of integrity are at stake and you are perhaps accused of committing fraud, you would have to be able to provide proof of your original data.

Dr A. Paterson
Group Executive: Science & technology competitiveness: National Dept of Science & Technology, Pretoria

The synthesis of the data was what was respected, while the data were not treated with the same high degree of regard. Today data are much more important in research than used to be the case. The cost of capture, storage, and conservation has dropped, assuming you have a satisfactory infrastructure. One of the functions of the IKMD is to provide that infrastructure that allows for the more effective capture, storage, and conservation of data.

Regulatory regimes have insisted on primary data being available, which is very true in the medical domain. More and more primary data have to be available so that should something go wrong, then people can trace whatever issues are at hand. Another important point is the data generation 'moment' that has increasingly moved from the eye of the observer to automated instruments. Automation has huge implications on the amount of data, the quality of data, and the process ability of data. The data recognition event is now also automated.



Data 'reduction' to information is also being frequently automated, whereby in some cases no one looks at the primary data at any level anymore. In fact, in some of the large new telescopes that are being built they will throw away the observational data because it is too expensive to store, and will use parallel processing to generate images. Most of the so-called pictures of stars that we see in the newspaper are not pictures at all, but composite images made up in computers, and the reason they look attractive is that stars through telescopes look boring. The picture is a processed composite multi-layered image of different types of information. Data are less often defined by the researcher, and more often agreed upon by international norms or standards, reliability, accuracy, and precision. In more and more fields, these are not regarded as data unless there is an international standard that defines these data. In the past, researchers used to define what they thought were data. All this means public data were defined less in terms of ownership and more in terms of access or cost.

Is massive co-ordination as envisaged by the so-called enterprise possible with HIV/AIDS research? Will researchers sacrifice their individual data models to some generalised data models so that we can share data more efficiently? Unfortunately, the evidence is not good now. Is the publication the most reliable reward or promotion indicator in the face of global catastrophes and challenges? Imagine somebody keeping the tsunamis secret because they want to write a paper about it. Yet HIV/AIDS is a tsunami - it is just happening more slowly but dramatically in everybody's life. We say of course you cannot keep it secret, but there are HIV/AIDS researchers in this country and across the world keeping their data secret until they publish these. Have we properly defined the rewards in the knowledge economy for all knowledge workers? Careers are dependent on publications, which mean long periods of data latency and secrecy. There are very few norms and standards (apart from validation), which mean that after the work is done there is very low possibility of re-use.

Where do ownership obligations lie in the new architecture? Copyright is the protection of original work, which requires an interpreter. We live at the boundary of imagination and information and the potential for knowledge. It remains value-laden. Most things that are written down are not necessarily true - they are only true if they resonate with some truth that you already know.

Intellectual property of South African inventors is held by multinational companies and other offshore entities (39%), while Government, such as the CSIR, World Research Commission, and Medical Research Councils, holds 19%. Universities are brilliant producers of knowledge but they do not secure patents - only 6% of the total patents. One can start to analyse how good a country is by mapping patents and showing where they are happening. All this means is we not only have to change the regulation of how we treat data and knowledge, but we have to change the culture of how we treat data and knowledge. South Africa does not respect its inventors.

We often talk about data, information, and knowledge as a hierarchy. However, there is less evidence that these are defendable as a hierarchy, and we should start thinking of these as architecture. What we are talking about is the fullest possible access to primary data, underpinning published research, which seems inevitable (except where ethical and professional norms require otherwise).

Finally, knowledge is changing. In the world of information, we have to think about what is written down in the text, we have to think about what is written down in people's heads, and we have to think about the cost - yet the bottom line is, we have to think about the people.