06.01.2015, for immediate release

Implementation of Open Data and Young Scientists

- The YAE welcomes the move towards greater openness in science.
- We believe that research agencies should provide, finance and maintain infrastructure for safe, citable and transparent storage of (often sensitive) research data.
- We believe that suitable funding should be made available for costly data storage and curation, without jeopardizing existing research funding.
- We favor the Green or Diamond publication models, as they still provide open access to information without excessive financial burden on authors. Where possible, journals should be strongly pushed by funding agencies to reduce embargo periods.
- We are concerned about inclusion of Open Data in promotion and funding decisions. It should rather be encouraged as best practice as defined by the given research field.

As an Academy, the YAE commends the move towards greater openness and transparency. The increasing criticism of journal impact factors and the current peer review system, and the increasing number of retractions of scientific publications from high impact journals, with surrounding questions of reproducibility, clearly highlight the importance of working with new models of scientific research. These concerns were also reflected in DORA, the San Francisco Declaration on Research Assessment. Many stakeholders are currently responding to these concerns by experimenting with alternate publishing and review practices with varying degrees of success, and we face an exciting time of change in models of authorship and dissemination, when standards for the coming decades are likely to be set. These standards need to be defined for both Open Data, i.e. the sharing of relevant scientific data to the (scientific) public, and Open Access, where scientific papers are no longer behind a journal-subscription paywall. It should also be defined very clearly what the precise goals and intended outputs of the transition to Open Data should be, as the communal needs will vary widely both across and within disciplines.

The first of these relates to the legal and practical aspects of data sharing, including economical concerns in terms of burden to researchers from the cost of meaningful curation of data. We believe that if each scientist has to provide their own data management plan, with no legal training and background, this would be a huge duplication of effort, with young scien-

tists in particular running a high risk of unintentionally running foul of the law. It would put another burden on the young scientists, who are already in the busiest time of their life with starting to organize teaching, grant writing for survival, research and family. Rather, we would strongly favor individual disciplines and fields to define and shape best practice policies for their areas, based on interaction with researchers and knowledge of the field, which are then followed by practitioners in that area. Here, it is important that these guidelines are formulated taking into account also trans-disciplinary aspects so that data can be readable by scientists working in other disciplines, as we believe that the rules for one discipline do not clearly transfer to another. Clearly, an important step is for training relevant legal and practical aspects such as data curation to become part of the training of scientists as young as PhD level or even earlier, to ease the transition towards "Open Science". Under no circumstances should curation costs come at the expense of ever-decreasing research budgets from national and European sources.

Tying in with this, while Open Data is overall commendable, we believe it is important to have data protection policies in place for sensitive data such as medical data that could violate patient privacy, or dual use data that may have originally had innocent intentions but that can be exploited for military or terrorist applications, or for commercial applications for which the author(s) is/are given no credit. Training in these issues needs to be provided from an early stage, and as Open Science takes shape, guidelines need to be established in parallel on how to define and to protect sensitive data. Additionally, young scientists often compete very hard to attain decreasingly available permanent positions. While data should clearly be shared by the time of publication, the individual scientists should decide the exact point of release during the publication process so that data sharing requirements can be calibrated to the needs of a specific field in terms of a balance between rapid sharing of data and maintaining privacy pre-publication.

Secondly, there is clearly a major cost involved with storing and curation of data, as well as a question of quality and long-term sustainability of data. There have been discussions about research budgets allowing for data management work packages, but not all funding agencies yet allow this. Ideally, data should be held in safe, citable repositories, and we believe that nationally or European-wide infrastructure in line with other major investments such as the Partnership for Advanced Computing in Europe (PRACE) is critical for facilitating this. A big step forward for Open Data would be the creation of central data repositories by funding agencies like ERC and national research councils, which would provide a great reduction in

costs because of the economy of scale and would avoid duplication of efforts among multiple repositories. These repositories shall be complementary to those already present in some research fields (CERN, climate science), but will provide a trusted resource for all other fields.

Moreover, it would allow for a centralized effort by the research agencies to facilitate, maintain, optimize and finance data curation, and leave the researchers to be maximally creative. At the same time, it is not necessary that all data be released (which would lead to large volume data dumps), but only the most relevant to analysis and reproducibility. The consideration of what is relevant and the optimal timing of publication differs for the different research fields, and therefore there should not be one standard model for all fields. Quality control of the data will become an important issue which should be facilitated through searchability, listability, and tagging the data. The use of Open Data will be enhanced if young scientists have a safe, citable way to store the data, and will likely increase as the positive aspects are witnessed first hand. It is also of utmost importance that this is done in a framework of proper attribution to create a safe climate for data sharing. We believe this will be one of the strongest motivators among young scientists to move towards Open Science, without damaging research careers. Additionally, we do not believe that the generation of large volumes of data should be used as an assessment point in tenure and promotion decisions, but rather, such data should, if used in assessment at all, only be evaluated based on its contribution to the research community as measured by utilization by other scientists. Therefore, EU and member states should provide resources for data storage, account for integrity, transparency, citability and sustainability of the data.

The move towards Open Access publications is commendable given the current situation where scientists deliver the scientific output and publishers make money from their papers. In part this is working well for the scientists because the publishers are advertising and promoting the research and applications, but only to those institutions that are paying the subscription fees for their journals. The drive for changing from a reader-pays model to an author-pays model is therefore laudable, but brings forward its own problems. This is especially true for some fields, like the humanities, and young researchers in general, who do not have the economic means to pay the publication costs, either with Open Access journals or through the Gold route at traditional publishers. Young scientists generally welcome support in managing financial relationships with publishers in order to responsibly use taxpayer money, in the interests of the Citizens. Support for this should be provided by funding agencies in grant contracts, in cases where open access is mandated, and publishers should be completely transparent about their costing models and any additional charges they may intend to impose for

data curation. Additionally, the only acceptable model for Open Access would be the Green route where the paper is published in the traditional way, and after an embargo period can be made available Open Access in a non-edited version in a repository. Several publishers have already accepted this preferred route and have reduced the embargo period significantly. This allows both rapid dissemination and protects research and university budgets.

The drive for Open Access has also led to a new view on the peer-review system, which remains effective in its traditional closed form, but in some research fields is now performed openly as well. The latter route has the potential for quality control, guarantees that ideas are not stolen during the review process and rewards reviewers in a more visible way. This in contrast to the closed route where the hard work of reviewers is being done for free, and invisibly. Initiatives such as the scheme by the American Chemical Society (ACS) to reward authors with "Open Access Credits" when they publish in an ACS journal, and which gives them the opportunity to make future papers available in Open Access, would be better aimed at rewarding reviewers for their work. This approach is indeed pursued by other journals such as PeerJ, who provide an entirely free Open Access publication coupon for any reviewer who submits their review on time.

Finally, it is important to remember the long-term scope of science. We live in a digital instant gratification society, where we want answers at once. Many scientific findings show their merit over decades and in humanities even centuries. We must not forget that the road to scientific discoveries is long and arduous, and it is easy to get lost in a rush to share large volumes of data without proper analysis and context.

Prepared by the Young Academy of Europe, "Open Science" Task Force

Monica Brinzei

Michel Crucifix

Lynn Kamerlin (Chair)

Marcel Swart

Yoav Tsori

on behalf of the Young Academy.