Statement to the IUPAP Council from the Working Group on communication in physics.

The working group were asked to consider the benefits and challenges to making research data open for wider reuse. The group recommends that to facilitate the discussions there should be a preferred definition to define data. We propose the following definitions:

Level 0 data – raw data, unprocessed

Level 1 – convert data to standard units; some initial calibrations

Level 2 – some data analysis, such as fit to curves, calibrations etc. Generally the data that will be supporting any figures in published articles and reports

Using this definition the group recommends that Level 2 data could be a good candidate for making openly available. Level 1 and Level 0 data require supporting information and formatting to be of most use and to facilitate accessibility.

There are many good examples of research communities sharing data well and integrating it into publication practices. Data supplementing articles is being published across disciplines, and in the life sciences mandatory publication of data for reproducibility already underpins several disciplines/journals. Research communities such as Astronomy and High Energy Physics have established formatting, linking and archiving protocols for data. However this is not the case across all areas of physics. The working group recognises that there are a number of initiatives and new publications emerging that help to bridge the gaps between the raw data classed as level 0 and the fully processed data at level 2 and that these should be monitored; new services emerging also provide suitable options for authors to index and store their data but the current landscape is still very fragmented.

In conclusion we recommend that IUPAP invite the physics community to provide, whenever and however possible, these data whilst recognising that this will be more complex in some areas than others, with additional supplementary information such as software, for example, required in some cases.

The publishing and library communities can play an instrumental role in this process in designing submission processes and guidelines together with linking mechanisms that can lead to more robust management, discoverability and archiving of the data. The benefits of this would contribute significantly to reducing duplication of effort at a later stage in the future.

We also recognise that by making data available researchers need some assurance that ethical practices will be adopted by others when making use of their data, abiding by any embargo periods or restrictions that may be imposed due to the nature of the data, and suitably acknowledging the original authors.