Round table *"Planning of future activities on "Data in complex systems" to be performed within the GIACS coordination action."* Palermo, Wednesday April 9th 2008.

The round table considered again the talks presented at the Conference and the points raised during the first round table held on Tuesday April 8th. The discussion therefore evolved along the following lines (synthetically reported).

Shavitt: Perhaps one should limit the focus of the present discussion. For example just considering valuable tools for time series analysis can be useful in complex systems research.

Monti: I am still convinced that the problems related to the access, storage, data mining and analysis of biological data and social data are intrinsically different.

Lillo: There are three main topics to be discussed. They are (i) tools, (ii) technological aspects and (iii) legal aspects. Tools are needed for a useful and successful analysis of complex systems data. They need to be shared, benchmarked and validated by different groups. Technological aspects such as data storage solutions and data standards are crucial for an efficient access to data and for an easy exchange among different research groups. Legal aspects are key issues with respect to all the procedures and phases of the research.

Garas: It should be attempted to provide a sort of guideline for researchers, which will be exposed to the need of access to proprietary data.

Liljeros: Indeed it should be disseminated and shared a code of conduct about the use of data sensitive to privacy problems and/or having a proprietary nature.

Monti: Indeed a convergence of genomic, personal and social data can be foreseen for the near future. Moreover the integration of genomic data with biomedical data might soon rise quite strong privacy issues.

Eubank: The validity of the use of artificial data sets as surrogated data depends on the research community that will use those data and on the characteristics that they might have. Concerning real data, in many countries government institutions might be reluctant to share data because they can show limits of the governmental regulations.

Mueller: Manipulated data and anonymized data always do some distortion in research analyses.

Shavitt: The most efficient approach will be within the same research community. It is difficult to consider aspects, which are useful for all research areas.

Monti: One should start from a concrete situation. Meetings like this one can help to understand how problems emerge in different research areas and how concrete approaches should be attempted within each specific discipline. It would be interesting to list a set of questions that might be encountered in different disciplines that require to find specific solutions for each different discipline.

Lancet: Let us use the contacts through GIACS we are developing thanks to this conference to put some attention on problems related with data that might be of

interest for the different research communities. For example, as a first goal we could attempt to contribute to the definition of aspects to be included in a standard contract a research group would be willing to sign with a company, which is proprietary of data. It is something challenging but probably it might be accomplished in successive small steps.

Eubank: Why this problem should be limited to the complex systems community? It seems to me that it is a quite general issue that might actually be encountered by any research group.

Shavitt: The judging of multidisciplinary research should be done within the core discipline. For example there is a lot of research about Internet outside the networking community, which is irrelevant.

Eubank: If one enters a new field and accesses data of the field, it is instrumental for the success of the research to do joint efforts with researchers already in the field.

In conclusion the discussion raised the difficulties related with an approach aiming to tackle the problems at a multidisciplinary or transdisciplinary level. It was therefore suggested to maintain moments of common discussion about these topics, list the problems, which are considered relevant by the entire community and plan future activities of more specialized research communities. Organizing in the near future two short working sessions of thematic small groups will do this attempt. One working session will be about (i) large databases in biomedical research and the other will be about (ii) large databases in social and economic systems.