Outline for Data Management Plan for NSF Proposals

1. List the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project (narrative description):
   - Describe the materials to be produced during the project period
   - What will be shared? At what stage (raw, processed, analysed, etc.)?
   - Why are the materials of interest?

2. Data and metadata format and content standards (where existing standards are inapplicable/inadequate):
   - Document remedies to existing standards
   - Identify the formats of data files created
   - Select file formats for sharing that maximize potential for longevity; describe plans for conversion to these formats (if applicable)
   - Identify metadata standards, who will create metadata and at what stage; If metadata is proprietary, disclosure is not necessary, but justification is.

3. Describe policies for access and sharing data, incl. provisions for appropriate protection of privacy, confidentiality, security, intellectual property, etc.:
   - Trade secrets/commercial info/confidential data collected from subjects/etc. are not subject to data sharing requirement.
   - Describe practices that will be used to maximize security and protect privacy, if applicable
   - If data is shareable, explain why it may be of value

4. Disclosure of licensing on data, or restrictions the PI (or some other entity) is placing on use of the data.

5. Plans for archiving data, samples and other research products, and for preservation and access to them
   - 3-5 years advised for data retention
   - E.g. details of domain repository/external data storage and management resources, etc.
   - Details of how data can be made accessible, if necessary, during this time frame.
Data Management Plan

The data associated with this project will be systematically managed in the following manner.

Expected Data to be Managed: Data, in this case, includes software from algorithms, theoretical and experimental results in the form of publications, theses and reports including all figures and text, and experimental data. Excluded are trade secrets, commercial information, materials necessary to be held confidential until they are published, or any information protected under law.

Specifics of the NRI Proposal: We plan to open-source our human learning/fusion code as well as the experimental code uses in all human-robot experiments. In particular, the human node code will be developed and open-sourced for common devices such as phones, hand-holds and tablets that do not natively support .NET (e.g. Android, iPhone). Components will be modular and include information fusion and communication; a sketching library for tablet PCs; speech components that use both built-in MS speech for basic applications and SPHINX for more sophisticated ones; gesture software that exploit common sensors such as Kinect SDK. For the robots, we plan to release core learning/fusion algorithms that are platform independent; C, C# and MATLAB code will be leveraged for these tasks.

Data Formats, Storage and Preservation: Print data will be available from publishers or electronically in pdf format, subject to subscription or printing charges. Software and experimental data will be available in well known formats such as text, code, MATLAB, or spreadsheets. All data will also be stored electronically on servers within the Autonomous Systems Lab. The servers have a RAID style backup, along with an off-site backup for security. All software and papers are maintained with an SVN repository for explicit change control. All personnel on the project (PIs, CoPIs, post-docs, graduate students, undergraduate students) will have full access to all data behind Cornell Kerberos log-ins.

Period of Retention: Data will be retained for a minimum of three years after conclusion of the award or three years after public release (publication), whichever is later. Data related to a student’s research work will be retained for at least three years after the degree is awarded. Data that support patents will be retained for the entire term of the patent. Longer periods will apply when questions arise from inquiries or investigations with respect to research.

Sharing and Access of Primary Data: All analyzed primary data under this award will be published promptly. Published data such as articles, dissertations and book chapters are accessible from publishers or upon request to the senior and student researchers involved, subject to privacy, confidentiality, security, and intellectual property right policy of individual publishers. Subscription or printing charges may apply. Exceptions are data involving proprietary information or patent application. Public access to these data is restricted.