**Title**: Experts Study on Grand Challenges for Research Data and Information Management

**Type of project**: consensus study

**Draft task statement**:

A grand challenge is a fundamental unresolved research problem with broad applications, large scale, deep impact, and long-term benefits. Solving grand challenges necessitates creating specific technological or scientific innovations in order to remove critical barriers to progress.

The proposed study would identify grand challenges that are timely, and achievable in the medium-term, over the next five-to-ten years, with concerted, coordinated efforts, and appropriate research support. Specifically, this study involves four tasks:

1. Identify the critical scientific and technological barriers to progress in research data & information management.

2. Examine the underpinnings of these barriers, the research areas that are most relevant to these barriers, and the research initiatives that are yielding innovations with potential application to these barriers.

3. Assess the barriers and analyze their scope, impact, and timeliness -- and identify areas in which interventions to coordinate and catalyze research efforts are likely to yield substantial progress.

4. Produce a consensus report with findings and recommendations, taking into consideration a range of scientific and technological perspectives, that address items 1-3 above and articulate a set of grand challenge problems, their boundaries and impact, and preliminary approaches to them.

**Significance, and role of BRDI**:

Identifying and articulating grand challenges will help in stimulating high-impact research, coordinating disjoint research activities, and guiding allocation of resources.

Below are listed some significant barriers to optimal research data and information management that may reflect grand challenges:

- *Semantic representation and validation of research information*: Format migration remains a central technical strategy for long term information access, but is subject to information loss. What are approaches and methods for defining the critical semantic information contained in research data and verifying that such information is faithfully represented in future encodings?

- *Assessing the cost reliability of data management processes, technologies, and institutions:* What are valid and reliable methods to quantify and predict the real-world reliability and costs of long-term digital storage in the presence of correlated technical, economic and institutional threats?

- *Developing low cost storage which provides very-high long-term reliability:* Data storage represents a substantial cost for research data management. Furthermore, the absence of practical low-cost, low-maintenance storage that provides very high long-term reliability is a substantial threat to long-term access to scientific evidence.

- *Estimating the risks to information, costs of management, and future value of the information.* However, the future value of research data is notoriously difficult to quantify, and markets for research data and information are generally thin. Furthermore, because data is a quasi-public (“club”) good and will thus be under-provisioned both by pure market approaches, and by voluntary contribution, achieving optimal management of data requires active policy intervention. The development of economic models, methods, and empirical analysis that would lead to more rigorous, reliable, valid, and systemic evaluation of the value of research information potentially constitutes a grand challenge.

Identifying the right problems -- those which are broad, deep, and yet *tractable* -- is difficult, and requires the efforts of world class experts with broad perspectives and understandings of the field. Correctly framing these problems, delineating the boundaries of what is known, and providing insights into new sources of innovation is even more difficult. Since data management implicates so many different areas of research and practice, this study will require drawing upon and integrating expertise from multiple distinct fields.

The Board on Research Data and Information is uniquely positioned to conducting this study. The interdisciplinary expertise reflected by the board, and its ability to draw on experts in both research and practice in multiple disciplines are critical to the success of this consensus study. The reputation of NRC enables it to engage with the leaders in diverse research fields. And, its dissemination channels are well-suited to communicating the studies findings and recommendations to the target audience.

**Potential sponsors & specific audience**:

Grand challenges for data management are likely to implicate research in computer science, library and information science, information economics, and a variety of other fields -- making it of interest to researchers and funders in those fields.