**Title**: *Trustworthy Distributed Digital Curation*

**Type of project**: consensus study

**Draft task statement**:

Organizationally and technically distributed and diversified approaches to managing research data are beginning to emerge, but have not yet been deployed at large scale. This is in large part because of the difficulties of defining, establishing, and verifying trustworthiness of organizations, procedures and content. To accelerate progress in this area this study will:

1. Identify the various technical, legal, economic, and organizational models and practices related to distributed digital curation; and the major organizations engaged in distributed digital curation and its development and promotion.

2. Examine the explicit and implicit threat models, reliability assumptions, trust models, and trust criteria related to distributed digital curation organizations, practices, and assessment. And analyze the strategies and mechanisms that are engineered for, or contribute to, the establishment, maintenance. and validation of trustworthy organizations, curation processes, and digital content for distributed digital curation.

3. Assess the state of the art and practice focusing on gaps and opportunities, including: community practices and technologies that are potentially ripe for wider-scale adoption; significant pervasive vulnerabilities in current distributed digital curation approaches; discoveries from research fields such as psychology, organizational behavior, computer science, privacy, and information security that have the potential to be productively applied to trustworthy digital curation; and areas in which additional research is urgently needed.

4. Produce a consensus report with findings and recommendations, taking into consideration the various stakeholder groups in the digital curation community, that address items 1-3 above.

**Significance, and role of BRDI**:

Research data (and digital information generally) are subject to many types of threats. These include media failure, hardware failure, software failure, communication errors, network failure, media and hardware obsolescence, software obsolescence, operator error, natural disaster, external attack, internal attack, economic failure, and organizational failure. Many of these threats can be substantially mitigated only through technically, physically, and organizationally diversification and distributions.

Failures of individual replicas are inevitable, and regular verification of distributed content is thus essential to ensure that failures and corruption of individual replicas is detected and repaired before permanent loss can occur. Furthermore, strategies of diversification and distribution are dependent on explicit and implicit models of trust and threats, many of which poorly understood. Moreover, the increase in the volume of data being produced and managed implies that individual organizations are hard-pressed to maintain copies of all the data of interest – and increasingly rely, at least in part, on collaborating institutions to manage data of scientific interest.

 Frameworks for evaluating the trustworthiness of data repositories are emerging and becoming widely recognized. However these frameworks focus on evaluating and certifying single organizations and archival processes. As important, building the virtual organizations and organizational collaborations that are needed for a robust data management strategy requires fostering, maintaining, and validating trust – a level of trust engineering that goes well beyond the mitigation of specific threats, or the certification of single organizations.

 As awareness of data management become more widespread, the need for robust distributed and diversified approaches to curation have become urgent. The Board on Research Data and Information is well-positioned to identify gaps and opportunities in this area. 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 a consensus study. In addition, the strong connection of the board with stakeholders in the areas of research, curation, and funding will help to ensure that the gaps and opportunities identified are both meaningful and actionable. Finally, the dissemination channels available to NRC are widely read and respected by the target audience.

**Potential sponsors & specific audience**:

The consensus study will develop findings and make recommendations for research, practice, and policy. Because of the interdisciplinary nature of both data management and trust engineering, this will appeal to a diverse audience, including:

· Research funders that produce data; that fund data-producing projects; and/or that fund research in digital curation, information security, or organizational behavior. These funders have a strong interest in ensuring that the data management approaches taken by fundees, and that the organization that both funders and fundees rely upon for data management are trustworthy.

· Leadership in organizations that have significant data-management responsibilities. These leaders need to understand the major areas of risk to their digital content; the ways in which a distributed and diversified organizational approach can mitigate these risks; and appropriate methods to approach and evaluate organizational collaborations and virtual organizations.

· Stewards of digital content, who wish to identify and apply good practice.

· Researchers in the field of digital curation.