	Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects	**Existing COUHES #**	10070039 44
		Date	**11/4/2013**

APPLICATION FOR CHANGES TO AN APPROVED PROTOCOL

Any change to a protocol that impacts human subjects must be approved by COUHES. If the change is minor and involves no more than minimal risks to subjects, an expedited review may be performed. All other changes are subject to a full Committee review.

1. Title of Study	
Scratch CDI: Preparing the Next Generation of Computational Thinkers	
2. Principal Investigator	
Name: Mitchel Resnick	Building and Room #: E14-445A
Title: Professor of Learning Research	Email: mres@media.mit.edu
Department: Media Arts and Sciences	Phone: 3-9873
3. Proposed Change. *Please provide a detailed description of the proposed changes, reason for this change, and indicate how these changes will affect the potential risks and benefits to the subjects. Define all abbreviations and use simple words. Unless justification is provided, this part of the application must not exceed 2 pages.* ***NOTE: when adding study personnel include MIT or outside affiliation, description of role in study and email address.***	

We propose to publish the data of the Scratch website, making it available as an academic dataset to share with other researchers for scholarly research. The academic dataset will include publicly visible data from the Scratch site published by Scratch users between 2007 and 2012.

RATIONALE: There are two main reasons to make this dataset more accessible: (1) to fulfill the intellectual merit goals of our NSF Cyber-enabled Discovery and Innovation grant, and (2) to lessen the need for other researchers to attempt to scrape the Scratch website for data. To expand on these two reasons: (1) The Scratch site was developed with NSF funding. In particular, the NSF CDI collaborative research grant was funded to advance understanding of online collaboration and learning on Scratch. During the final year of the grant (2014), we are organizing a conference for other researchers to learn more about the Scratch data. Providing a well-documented dataset will enable researchers from a wide range of disciplines to contribute to understanding of online collaboration and learning on Scratch. This contribution can also strengthen the case for future grants to study the change in participation and learning in Scratch 2.0. (2) Given that the Scratch site has been public since 2007, it would be trivial for any researcher to create automated mechanisms to harvest data from the website by "screen scraping" the site. While the data is available to scrape, existing methods of extracting Scratch data from the public site consume server resources and can negatively impact site performance.

DATA TO BE PUBLISHED: We propose to include the following in the dataset:

A) Public data visible on the website: This dataset includes projects, metadata, comments, forum posts, and interactions shared publicly. These data have four main elements:

1. Users. These are account usernames, which are public, and users are told not to use real names. To minimize risk, the usernames will be obfuscated in the dataset.
2. Projects. The artifacts publicly shared by the users.
3. Relationships and interactions. These are the public relationships between users and projects (i.e., project ownership), users and users (i.e, friendships, commenting), and projects with projects (e.g., groupings in galleries).
4. Site metadata. This includes data from the public source code repository and software and website documentation, as well as the public forums and wiki.

B) Interaction timing data. Some data in the Scratch database reflects the state of Scratch projects at particular points in time (e.g., the time a project received a "love-it"). Because these data have been made public implicitly and are not sensitive data, we will also data including this timing for all publicly visible events.

WHAT WILL NOT BE PUBLISHED: The dataset will NOT include data collected to administer the site that was never made public. This includes: (a) Users self-reported age and gender; (b) Email addresses of users; (c) IP addresses; (d) Data on which users viewed particular projects; (e) Information on which users participated in public interaction where the subject of the interaction was not made public (e.g., which user viewed a project or gave a "love-it"); (f) "Flags" or other messages sent privately between administrators and users. The dataset also will not include data removed by users. Because user-removed data may affect validity of results, we will publish summary statistics on the amount and type of data removed.

RELEASE AGREEMENT: A website will ask researchers to fill out a form giving us their name, contact information, and affiliations, and to read and agree to a Scratch Research Data Sharing Agreement (see accompanying file for this agreement).

Please attach copies of all revised material (consent form, study protocol (*PI & Dept. Head signature required),* recruitment, etc). Also, include a copy of all revised material with changes HIGHLIGHTED. Any incomplete application will be rejected and returned for completion.

Signature of Investigator _____ **Date** _____

Signature of Dept. Head _____ **Date** _____

Please submit a signed hard copy of this application including all revised material to the COUHES office at E25-143b.