

Mechanical Turk (MTurk)

Purpose

The information in this guidance is for investigators conducting human subjects research using Amazon’s Mechanical Turk (MTurk). The aim is to provide suggestions for addressing privacy and confidentiality concerns, along with other ethical considerations when conducting human subjects research using the MTurk platform.

This guidance is based on information available in May 2020 - investigators are advised to carefully review information posted through Amazon mturk.com, as their policies/terms and conditions may change.

In this document

[What is MTurk?](#)

[Definitions](#)

[Is MTurk Anonymous?](#)

[Internal HITs](#)

[External HITs](#)

[Recommendations](#)

[Other Considerations](#)

[Payment/Purchasing](#)

[GDPR](#)

[Additional Information/References](#)

What is MTurk?

Amazon’s MTurk is a crowdsourced virtual labor market in which individual MTurk Workers complete online Human Intelligence Tasks (HITs) for pay. MTurk serves to match individual Workers with small “microtasks” and to provide a mechanism for compensating Workers. The platform appeals to human subjects researchers, as it facilitates access to a large population of motivated participants.

MTurk is a general-purpose and largely unregulated, crowdsourcing platform that is not expressly designed for the conduct of human subjects research. While MTurk’s security, privacy, and confidentiality measures may be sufficient for many business or personal tasks, ethical research is held to a higher standard. In addition to data security considerations, investigators are reminded that collection of data through this platform is subject to Amazon Services Conditions of Use agreement.

MTurk’s Acceptable Use Policy, Privacy Notice, Participation Agreement, and Amazon Services Conditions of Use Agreement should be reviewed in their entirety before initiating human subjects data collection plans using MTurk.

Definitions

Privacy and Confidentiality Definitions

Anonymous – Anonymous means that there is no possible way for *anyone* (not even the researcher) to identify or trace a subjects' identity from the data at any time.

Confidentiality – Confidentiality refers to the treatment of information that has been disclosed to a researcher in trust and with an expectation that it will not be shared in ways inconsistent with the original understanding or agreement.

Data Anonymization – Anonymization is the irreversible removal of any link(s) between an individual and his/her data to the degree that it would be virtually impossible to reestablish a link or identify an individual.

De-identification – Data management processes to separate, modify, and/or remove any personal identifying information from the data such that it would be difficult to re-establish a link

Identifiers / Personally Identifiable Information (PII)^[1] – PII includes any information about an individual that can be used to distinguish or trace an individual's identity, either alone (i.e. directly) or when combined with other personal or identifying information that is linked or linkable to a specific individual (i.e. indirectly).

Privacy- An individual's ability to control the extent, timing, and circumstances of sharing oneself with others. This includes access to their personal information, as well as access to their body (i.e. collection of biological specimens or being photographed).

MTurk Definitions

Crowdsourcing – A means of obtaining information/input on a task/project by enlisting a large number of people over the internet.

Internet Protocol (IP) address - A numeric address assigned to every device that connects to the internet or a network. As related to human subjects research, IP addresses are generally considered as personally identifiable information (PII).

Human Intelligence Tasks (HITs) - Jobs posted in Amazon's MTurk labor market web site

Internal HITs – Tasks that are completed using MTurk's internal *Survey Template* and run directly through Amazon.

External HITs- Workers are recruited and paid through MTurk, but through the use of a *Survey Link* data is collected through an external survey site (e.g., Qualtrics).

Requester – An individual who is registered on MTurk to post Human Intelligence Tasks (HITs) (e.g., an investigator, PI, researcher)

Worker – An individual, registered with MTurk to complete Human Intelligence Tasks (HITs) for pay (e.g., a participant or subject).

Worker ID – A unique code assigned by Amazon to identify MTurk Workers. This code is linked to an individual's information on the Amazon platform (i.e. profile, purchasing history, reviews, views, lists, Prime account information, etc.).

Is MTurk Anonymous?

There are two primary methods for structuring data collection using MTurk:

Internal HITs or External HITs.

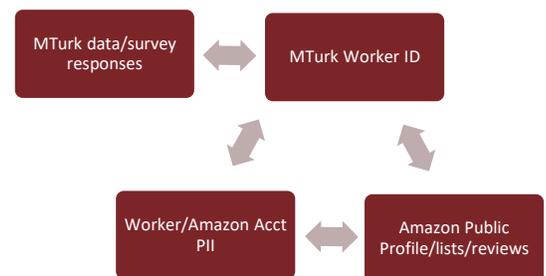
Internal HITs

Internal Human Intelligence Tasks (HITs) are those structured using MTurk's provided *Survey Template*, with the task/data collection occurring entirely within Amazon's platform. Data collected through an **Internal HITs** are **NOT anonymous** for a number of reasons.

- MTurk Worker IDs are automatically embedded with survey responses in Internal HITs. All payment transactions are done using an assigned a 14-character alphanumeric code (i.e., MTurk Worker ID) that is unique to each Worker. This code is linked to the survey data to allow Requesters to review/approve work and to issue payment.

HIT ID	Worker ID	Output Data	Q5 Favorite	Comment	Q2 age	Q1 Gender	Q4 income	Q3 education	Select All
950ZYWXSМКCZP6ZH0T	A26A3WWGG1I286			... older shows... ...ally - cleaner and ...re...	40	Female	\$37,500 - \$49,999	Some college, no degree	
950ZYWXSМКCZP6ZH0T	A3KYHM91HB6XV			News	29	Female	\$25,000 - \$37,499	Some college, no degree	
950ZYWXSМКCZP6ZH0T	A10UFYPF709FLW			News I like this Hit...	35	Male	\$25,000 - \$37,499	Graduate degree (Masters, Doctorate, etc.)	

- Amazon has access to Worker's personally identifiable information (PII) (e.g., name, email address, and physical address) through a Worker's Amazon account and profile. Amazon Worker accounts are linked to other Amazon accounts and services (e.g., Prime, Shopping, Music, lists, reviews) via their Worker ID.
- While Requesters do not have direct access to Workers' PII, researchers have demonstrated (Lease et al., 2013) that re-identification of Workers can occur indirectly through triangulation. As noted above, Amazon's MTurk Worker ID is a shared identifier across Amazon services. This ID number is embedded in the URL of users' Amazon Profile pages. Therefore, for MTurk respondents who also maintain a publicly-available Amazon Profile, it may be possible to identify individuals based on the amount and type of information shared on his/her/their public profile.



External HITs

External Human Intelligence Tasks (HITs) involve recruitment through the MTurk platform, but link participants to a separate survey posted on an external site (i.e. Qualtrics). Upon completion of the Qualtrics survey participants are given a code, which is then entered into the MTurk HIT to cue payment. When structured in this manner, survey data from External HITs is not available to Amazon MTurk.

In general, data collected through External HITs is considered **confidential**. After data collection is complete and compensation is awarded, researchers should de-identify the data.

Data collected through external hits *may* be considered **anonymous** (i.e. no identifiers linked to data either directly or indirectly via a coding system), **if configured such that:**

- Survey questions do not ask for PII (including email or Worker IDs)¹
 - Note: this includes informed consent “*type name to agree*” if embedded in survey
- It is not possible to identify individuals based on their responses, or a combination of responses.
 - Note: specific demographic information such as exact dates of birth, zip codes, job titles, can all be PII depending on context and other data/information available.
- IP addresses are not collected by the external site (i.e. Qualtrics)
- Compensation codes in MTurk are configured such that they do not link individual MTurk Worker IDs to survey responses (i.e. each participant receives the same completion code rather than a unique code that can be linked to responses)

Note – The ability to verify/review and approve/reject individual Worker responses prior to awarding compensation is important to many investigators. For this reason, many elect to employ use of unique completion codes which allow a link between the survey data and the MTurk ID. In many instances having a temporary link is between the Worker ID and survey responses is acceptable. However, investigators are reminded that **whenever data are linked to identifiers (even temporarily) the data are NOT anonymous**. Investigators must accurately reflect these plans in their IRB application for approval and in consent information shared with participants.

Recommendations:

- **The use of External HITs**

Tasks should be configured as External HITs using **Qualtrics**. Qualtrics is the survey software approved and licensed for use at Iowa State University.

 - If a verification or review/approve process for Workers is needed, use unique completion codes to link survey data to Worker ID. The codes linking surveys and IDs should be deleted as soon as possible. While this method is NOT anonymous, for many studies this method provides sufficient confidentiality protection. Investigators are reminded to correctly report the collection of identifiers within their application for IRB approval.

¹ Collection of Worker PII by Requestors is a violation of MTurk’s Participation Agreement.

- **IP Addresses**
 - Configure program settings such that IP addresses are not obtained if the aim is to collect data anonymously.
 - If IP addresses are needed for location or duplication reasons investigators should remove these identifiers as promptly as possible. If IP address are obtained, the study cannot be represented to the IRB or participants as “anonymous.”
- **Informed Consent**
 - Do not have participants enter/type their names as a means of confirming agreement; rather use a “check box” or “by continuing you agree...”
 - Accurately inform participants about how Worker IDs will be used and kept confidential.
 - Do not represent the study as “anonymous” if it involves internal HITs or if any PII (including Worker ID or IP address) is connected to the data - even temporarily.
- **Minimize collection of PII** whenever possible. Consider how information collected might be combined to indirectly identify participants. Minimize collection of specific information if higher level/“zoomed out” information will answer the research question.
For example:
 - Avoid collecting exact dates of birth (DOB) if an age or year of birth is sufficient.
 - Do not collect address information (i.e. street address, zip codes); if geographic location is necessary, collect at a larger level (e.g., state or county).
- **Delete Worker IDs once compensation is awarded.**
- **Respecting Participants/MTurk Workers.** The majority of MTurk Worker complaints received by the IRB relate to rejections, often resulting from unfair or poorly constructed attention checks.
 - **Attention check** questions can be a helpful tool to aid in identification of low-quality data from participants speeding through studies or from bots. Proper attention check questions should be straightforward and not ambiguous. These questions should not be designed to “trick” participants, but rather to ensure participants are human (not bots) and are putting forth reasonable efforts.
 - Configure attention check questions such that if failed, data collection will stop/exit to avoid further wasting the participant’s time. It is best to include the checks early and at reasonable intervals such that the participant does not invest a significant amount of time only to be ejected at the end.
 - Inform participants if attention check questions will be used and consequences for failed attention checks.
 - **Implications of rejecting HITS.** Rejections remain with Workers forever, negatively impacting their ability to get future work on MTurk. Take care to avoid arbitrary rejections or rejecting work in error. If rejections are made in error, work with MTurk to promptly correct.

- **Clearly describe eligibility criteria for HITs.** Ask screening questions early to avoid participation by ineligible Workers.
 - Note: HIT descriptions/posting information and HIT instructions are considered recruitment material. For non-exempt studies, these materials require IRB review and approval. Please
 - submit this information with your IRB application.
- **Pay Workers promptly.**

Other MTurk Considerations

Payment by ISU

Prior to posting HITs, researchers should work with the appropriate departmental contacts for accounting and/or procurement to ensure that reimbursement plans are correctly configured and align with ISU procedures.

GDPR

MTurk Workers can be located anywhere on the globe with an internet connection. Unless a HIT explicitly restricts data collection to specific IP Address locations (e.g., only United States IP Addresses) it is assumed that data collection activities may be completed by participant Workers located in Europe for which additional privacy and confidentiality regulations may apply.

The General Data Protection Regulation (GDPR) is a European Union (EU) and European Economic Area (EEA) regulation on data protection and privacy that went into effect on May 25, 2018. GDPR applies to all individuals physically located in within the EU/EEA regardless of citizenship. Investigators collecting will collect or use Personal Data from research subjects located in the EU/EEA are required to comply with GDPR requirements.

GDPR uses the term “Personal Data” to refer to any information relating to an individual person who can be directly or indirectly identified. Examples of Personal Data include (but are not limited to):

- First and last name
- Email address
- ID number
- Location information such as home address or GPS coordinates
- Online identifiers such as **IP address**, browser or Internet cookies
- **Demographic, behavioral, or health-related information that could identify directly, or indirectly through a combination of data points person**

Whenever data will be collected from MTurk Workers located in the EU/EEA in a way that is not anonymous, GDPR requires that:

- (a) specific information be provided to subjects regarding collection and use of their Personal Data, and
- (b) for some types of information deemed especially sensitive, explicit consent from subjects be obtained.

Investigators are advised to include GDPR-mandated information in the informed consent form(s) to be presented to subjects.

Projects subject GDPR requirements are routed through Iowa State University Counsel and Information Technology Services to ensure that all GDPR mandates are met.

Additional Information:

Iowa State University Links:

[GDPR General Data Protection Regulation](#)
IT Security

[GDPR: Compliance with the European Union General Data Protection Regulation](#)
Iowa State University Policy Library

[Research Participant Payment Process](#)
Controllers Department

[Use of Amazon Mechanical Turk](#)
Office of the Vice President for Research

Amazon MTurk

[Amazon Privacy Notice](#)

[MTurk Acceptable Use Policy](#)

[MTurk Participation Agreement](#)

[MTurk Privacy Notice](#)

Citations:

Huichuan Xia, Yang Wang, Yun Huang, and Anuj Shah. "Our Privacy Needs to be Protected at All Costs: Crowd Workers' Privacy Experiences on Amazon Mechanical Turk." *Proc. ACM Human-Computer Interaction*. 1, 2, Article 113 (November 2017) <https://doi.org/10.1145/3134748>

Lease, Matthew and Hullman, Jessica and Bigam, Jeffrey and Bernstein, Michael and Kim, Juho and Lasecki, Walter and Bakhshi, Saeideh and Mitra, Tanushree and Miller, Robert. 2013. "Mechanical Turk is Not Anonymous" SSRN <http://dx.doi.org/10.2139/ssrn.2228728>

U.S. Office of Management and Budget. 2016. "Managing Information as a Strategic Resource." *Federal Register*. July 28. Accessed April 29, 2020. <https://www.federalregister.gov/d/2016-17872>.

Document History

Created/Approved: 6/2/2020
