



MANAGED CROWD SOURCING

Leveraging a distributed labor network to decrease turnaround times and reduce costs for repetitive data tasks without sacrificing quality

Last updated: January 11, 2012

Information Evolution and Managed Crowd Sourcing

IEI assists information firms by handling data tasks via managed crowd sourcing. We use the Amazon Mechanical Turk workforce of 500K+ independent researchers (“turkers”) and employ the following processes to ensure accurate results. The advantages to the customer are that it drastically decreases data processing turnaround times and costs less than in-house or even traditional outsourced resources.

IEI consults with the customer on the tasks to be handled to ensure mutual agreement on tactics and goals. Then the initial testing work is undertaken, an instructional video is recorded, and the project is posted. Results go through a vetting process to identify quality labor resources and the project is submitted. Finally, the results are delivered to the customer.

Specifically How the Process Works

Ensure project task suitability for managed crowd sourcing

Project must be able to be broken down into simple, quick tasks that can be accomplished via Internet research and have only one correct answer.

Task cannot be accomplished via data harvesting, which is typically faster and cheaper.

Structure tasks appropriately

These projects must be broken down into individual tasks, such as verification or appends, and not run concurrently. (Certain individual tasks need to run in sequence, for example, first getting a missing piece of data and then submitting that found data to an online database.)

Structure tasks appropriately

Tasks are structured to ensure the retrieval of high-quality data

- Assign each task to three workers and only trust responses when at least two of the workers agree.
- Ask turkers to obtain an arbitrary piece of data on the Internet, even if it won't be used, in order to verify that workers personally complete the task and are not providing random responses. These responses are known as “Gold” responses and serve as a measurement of worker data quality against a solid, verifiable piece of data.

Structure tasks appropriately

Make instructions as descriptive and clear as possible

- Record brief, narrated, instructional videos of how tasks should be handled.
- Explain the ultimate goal of the project.
- Explain what to do in case of anomalies. For instance, if research yields no results, partial results, or suspicious results, then workers use an automated, pre-formatted default response value (such as “NIF” for “no information found”). Anomalies will occur so categorizing all types of errors is important and standardized results for “dry holes” allows easier back-end processing.
- Mention how the work will be reviewed (e.g., format, accuracy via comparison to other results).
- Note that work will be rejected if “bad” or completely irrelevant data is submitted.
- Note that turkers who submit high-quality results will be paid more for follow-on tasks than those who do not.

Structure tasks appropriately

Test tasks in-house before assignment

- A set of tasks are completed in-house based on the instructional video. The larger the data set, the more records are tested.
- Throughput times are noted for the task to judge compensation value.
- Types of anomalies are noted.
- Instructions are adjusted based on anomalies discovered and pricing is set based on the time taken to obtain accurate results.

Assign and Manage the Task

Task is posted with all required data

Note for the requesting organization:

- If you have a recognized brand name and can use it, that's great because you can attract better resources and build a reputation with turkers.
- However, you can protect your anonymity, if necessary, by running sensitive work under the Information Evolution name.

Assign and Manage the Task

- Task expiration date is not important since tasks are typically completed extremely quickly
- Time allotted per HIT (this must be consistent with price offered)
- Keywords: company name (if desired), “data,” “fast,” and “easy”
- Payment rates: 90% of tasks pay \$0.01-\$0.10

Assign and Manage the Task

- Allow only workers with a 95 percent accuracy rate or higher to view tasks.
- Once qualified turkers are identified and their email addresses are obtained then the rate paid to these qualified turkers for follow-on work is substantially increased (e.g., doubled) to increase throughput times.
- Geographical restrictions for turkers can be added if the task requires specific understanding of American culture or slang (e.g., transcriptions of U.S. audio files).

Assign and Manage the Task

- Descriptive text is added to requests: “If you would be interested in receiving email updates about forthcoming HITs from us, enter your email address here.”
- Project is posted at night or over a weekend for quicker response rates.
- A control set of data is posted first (the set for which results were obtained in-house).

Control set results are reviewed

- Results are reviewed immediately upon completion.
- Correct results are identified and these turkers are targeted to do the second set of records.
- If results identify additional issues then instructions are adjusted.
- Payments are made. Payments are not rejected unless results are clearly junk (not just inaccurate). This is to protect the poster's reputation among turkers.

Second task is submitted

- The second set of data is a larger set of data that has not been handled in-house.
- The task is submitted *only* to the vetted turkers.
- Each task is submitted four times.
- Results are reviewed again:
 - Only those results where 2 of 3 results match each other are accepted as accurate.
 - Any other potential anomalous issues are identified and instructions are adjusted again to clarify any issues that are discovered.
- Turkers are paid.

The bulk of the task is submitted

- The task is submitted *only* to vetted turkers.
- Only those results are accepted where 3 of 4 results match each other.

IEI follows up with the researchers professionally

Turkers are thanked and paid promptly.

Costs

Net costs to the customer for a typical project

- Initial Consultation: \$1,000
- Process Engineering, Instructions, Testing, Payments to Resources: \$2,000
- Research cost to complete each task: $\$0.05/\text{task}^* \times \text{four iterations} = \$0.20/\text{record}$
- Tasks per record: average of 2
- Research cost per record: approximately \$0.40/record

**IEI charges the customer \$0.02/task on top of the rate paid to the research resources to cover process management.*

Costs

Final cost-per-record for customer for typical projects:

- 10K records: \$7,000 (\$0.70/record)
- 30K records: \$15,000 (\$0.50/record)
- 100K records: \$43,000 (\$0.43/record)

INFORMATION EVOLUTION

Matt Manning, President
Information Evolution, Inc.
1601 East Fifth, #109
Austin, Texas 78702
Tel: 512.203.4431
Email: mmanning@informationevolution.com
Web: www.informationevolution.com



Drop us an email today to discuss your upcoming data acquisition, updating, and information management challenges.