

Yandex

Crowdsourcing Practice for Efficient Data Labeling: Aggregation, Incremental Relabeling, and Pricing

Alexey Drutsa, Valentina Fedorova, Dmitry Ustalov, Olga Megorskaya, Evfrosiniya Zerminova, Daria Baidakova

Part II

Introduction to Yandex.Toloka for requesters

Daria Baidakova,
Project Manager

Yandex.Toloka is a service of Swiss company Yandex Services AG

Tutorial schedule

Introduction: 15 min

Part I: 20 min
Main Components

Part II: 10 min
Introduction to
Crowd Platform

Part III: 15 min
Brainstorming
pipeline

Part IV: 60 min
Set & Run Projects

Part V: 25 min
Theory on
Aggregation

**Break:
30 min**

Part VI: 20 min
Set & Run Projects
cont.

Part VII: 10 min
Results &
Conclusions

Requester's dashboard

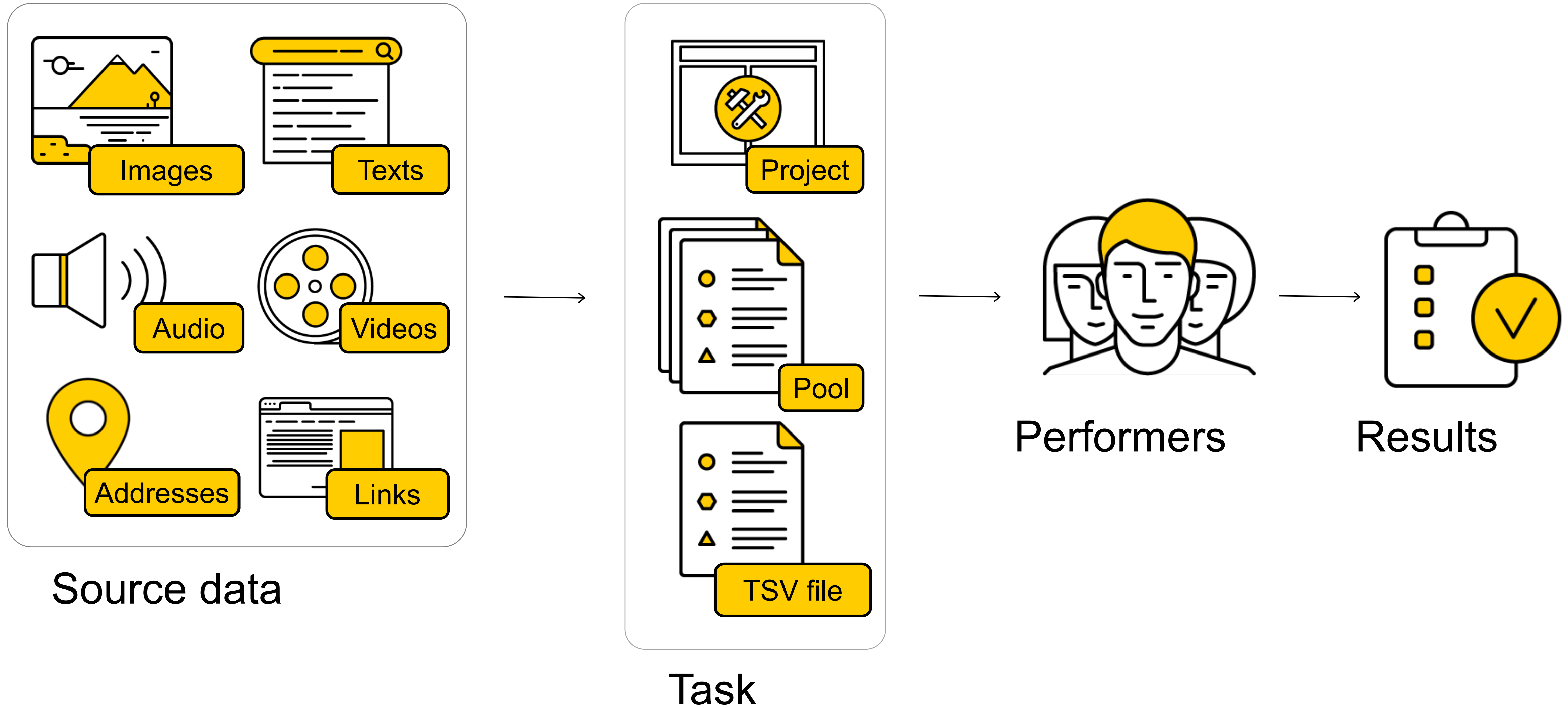
The image shows a screenshot of the Yandex Toloka Requester's dashboard with several yellow callout boxes and arrows pointing to specific UI elements:

- Pending amount for tasks waiting for acceptance**: Points to the balance information in the top right header.
- Your balance**: Points to the balance information in the top right header.
- Requester name to show performers**: Points to the user name 'Alice' in the top right header.
- Your projects will be listed here**: Points to the 'No data' message in the projects table.
- Help and useful links**: Points to the help icon (question mark) in the top right header.

The dashboard interface includes the following elements:

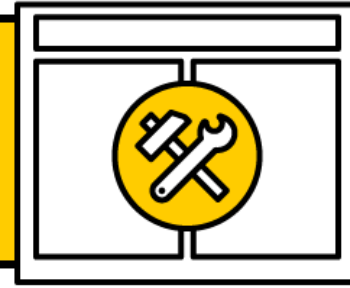
- Header:** Yandex Toloka logo, navigation links (Projects, Users, Skills, Profile, Messages), a help icon (?), balance (\$0.00 / \$16.95), a flag icon (UK), the user name 'Alice', and a profile picture.
- Projects Section:** A filter dropdown set to 'Active', a search input field containing '--', a 'Find project' button, and a '+ Create project' button.
- Table:** A table with columns: Title, Private comment, Created, and Open pools. The table content is empty, displaying 'No data'.
- Footer:** A 'Help and useful links' button and a pagination dropdown set to '100'.

How Toloka works



Tasks in Toloka

Project



Instructions

- examples
- acceptance criteria
- steps to follow

Specification

- input
- output

Task interface

- HTML
- JavaScript
- CSS

Pool



Parameters

- price
- overlap
- acceptance type
- time to complete

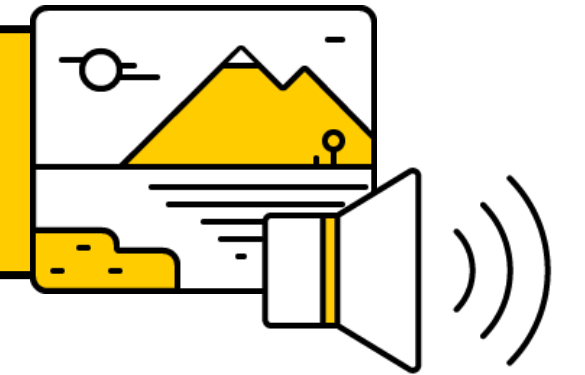
User filters

- language and region
- gender and age
- device and OS
- skill

Quality control

- fast responses
- captcha
- majority vote
- control tasks

TSV file



Data to use in the task

- texts
- links to audio, video or image files
- correct responses
- hints for training

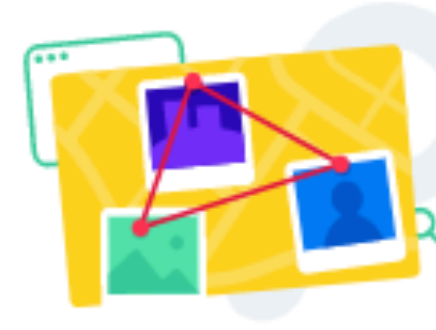
Task templates

Choose one or create your own



Offline data collection

Ask performers to complete tasks in the mobile app: collect data in the field, take photos, or record audio



Data search

Collect and annotate any type of information found online



Side-by-side comparison

Users vote for the best option and explain why



Image or voice recognition

Get human feedback when automatic recognition fails



Data classification

Categorize large amounts of data



Content moderation

Check if content meets your guidelines



Content generation

Users submit a photo, audio recording, or original text



Relevance assessment

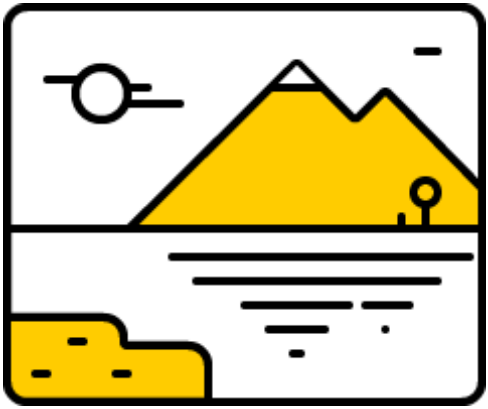
Evaluate results for search queries



Survey participation

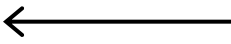
Get responses from a large audience

Creating a project



SPECIFICATIONS

Input data	Output data
<input type="text" value="image (URL)"/>	<input type="text" value="result (string)"/>
<input type="text" value=""/>	<input type="checkbox" value="like (boolean)"/>
<input type="button" value="Add field"/>	<input type="button" value="Add field"/>



Input data is what we show the performer in the task: links to images, text, video.

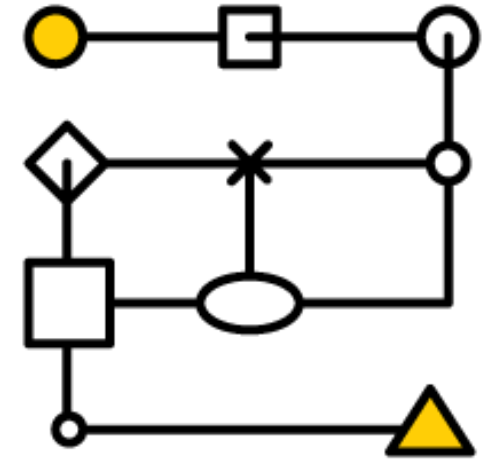
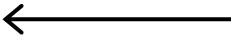
Output data is what we get back from the performer as a result of task completion: text, url, photo.



TASK INTERFACE

```
Preview html js css  
1 {{img src=image width="100%" height="400px"}}  
2 }}  
3 {{field type="radio" name="result" value="OK" label="Good" hotkey="1"}}  
4 {{field type="radio" name="result" value="BAD" label="Bad" hotkey="2"}}  
5 {{field type="radio" name="result" value="404" label="Failed to load" hotkey="3"}}  
6 <br>  
7 {{field type="checkbox" name="like" label="Do you like the photo?" hotkey="q"}}  
8 }}  
  
1 exports.Task = extend(TolokaHandlebarsTask, function (options) {  
2   TolokaHandlebarsTask.call(this, options);  
3 });
```

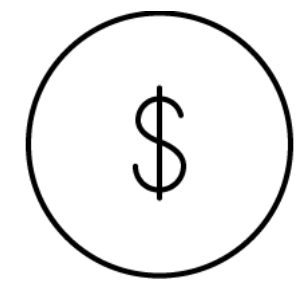
You can use **JavaScript** to configure validation of responses or show follow-up questions.



Task interface defines what the task looks like for the performer.

Pool settings

Parameters



Price



Time to complete



Overlap

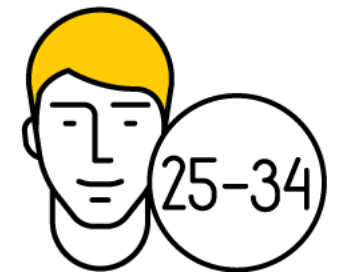


Acceptance type

User filters



Language and region



Age



Gender



Device and OS

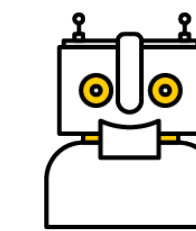


Skill

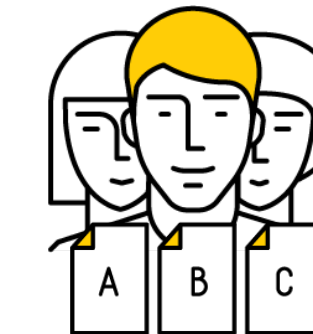
Quality control



Fast responses



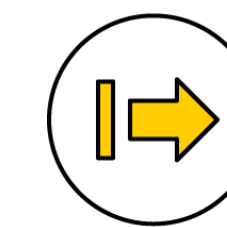
Captcha



Majority vote



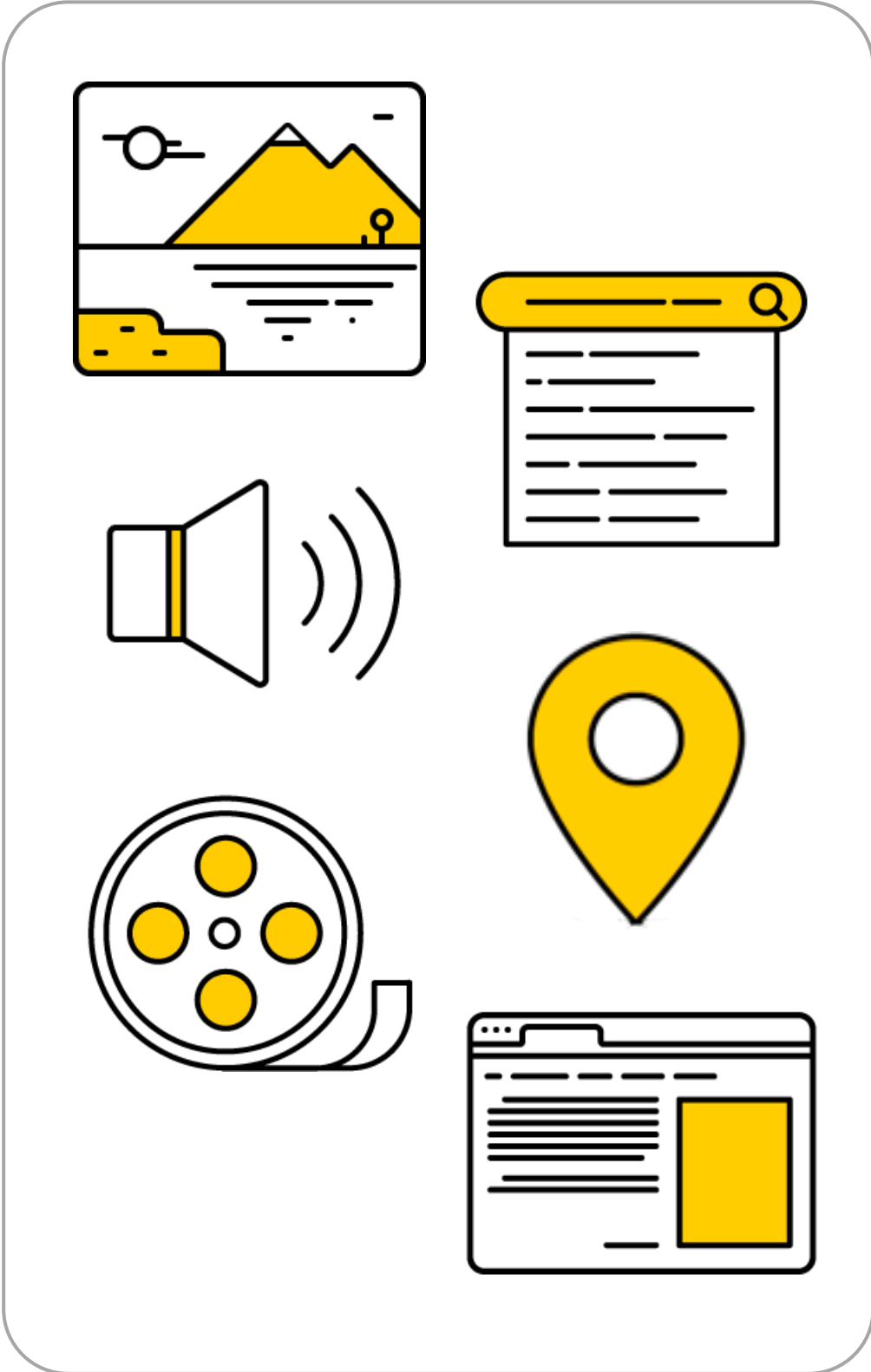
Control tasks



Skipped assignments

and more...

TSV file with tasks



Create a TSV file with links to images, audio, video, texts, or coordinates

Upload the TSV file to the pool

Start the pool to assign tasks to performers

Task types

Main



INPUT:image

Tasks that you don't have answers for, and you want to get responses from performers.

Control (golden)



INPUT:image

GOLDEN:result

Tasks that you know the correct answers to. Used for checking performers.

Training



INPUT:image

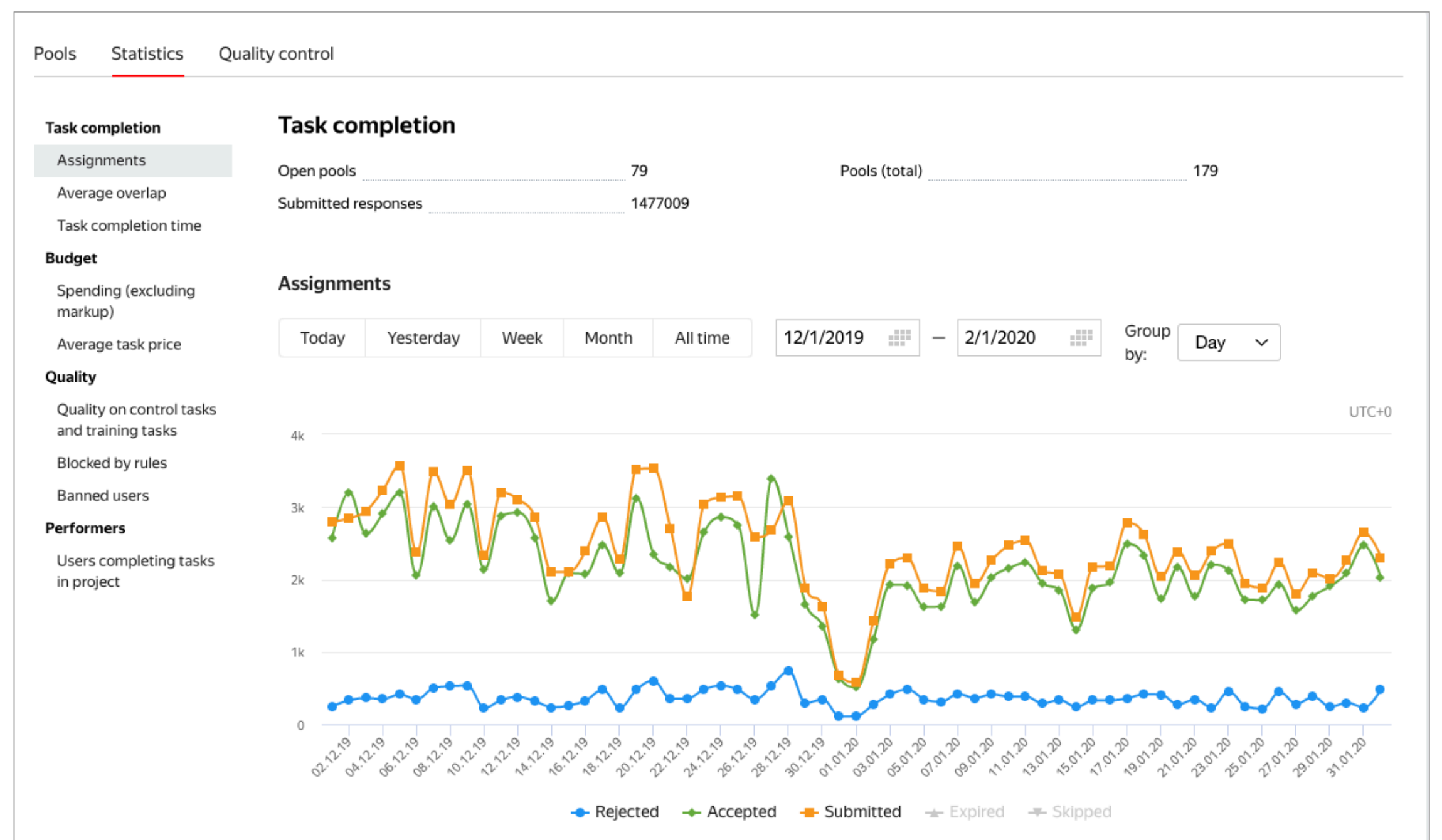
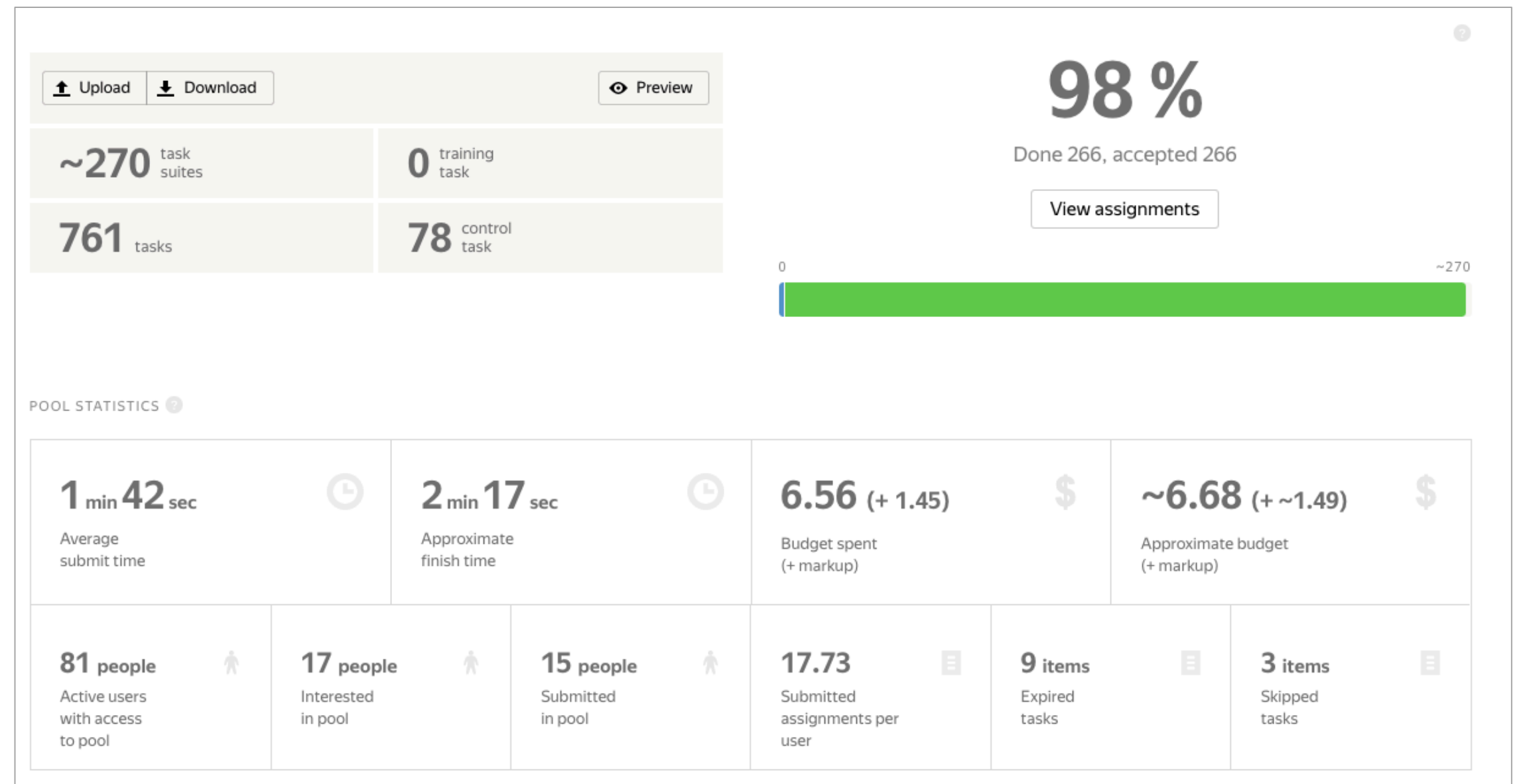
GOLDEN:result

HINT:text

Tasks with correct answers and hints. Used for training performers to complete the task.

Task completion

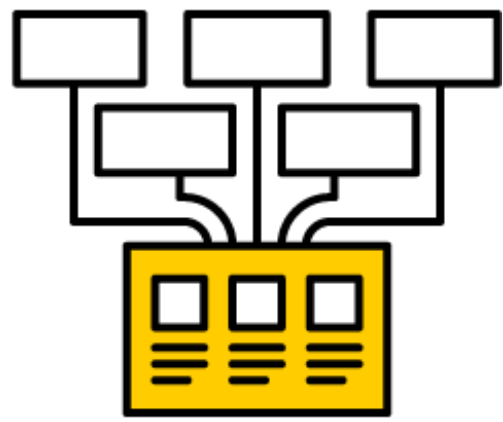
Track progress on your tasks on the project page, in the pool, and in **Statistics**



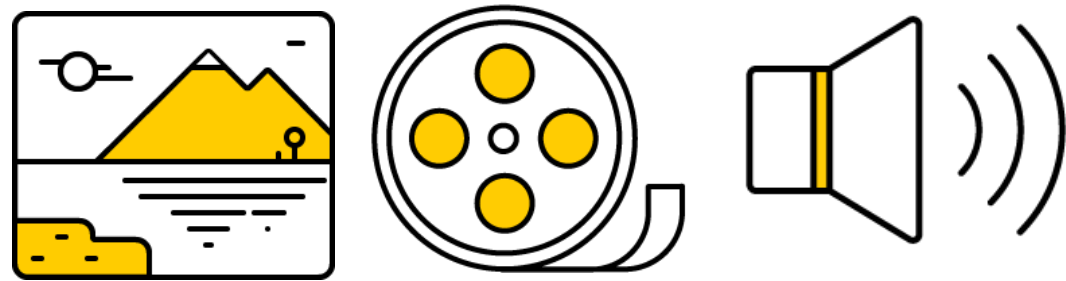
Results



TSV file with responses



Aggregated results



Photo, video, audio

The screenshot shows a web interface for results. At the top, there are buttons for 'Statistics', 'Download results' (with a download icon), and 'Edit' (with a dropdown arrow). A dropdown menu is open, listing 'View operations', 'Dawid-Skene aggregation model', 'Aggregation by skill', and 'Download attachments'. Below the menu, a large '100%' is displayed, followed by 'Done 18, accepted 18' and a 'View assignments' button. At the bottom, a green progress bar is shown, spanning from 0 to 18.

Yandex

**Thank you!
Questions?**

Daria Baidakova

Project Manager



dbaidakova@yandex-team.ru



<https://research.yandex.com/tutorials/crowd/sigmod-2020>

