Using Data Expedition as a Formative Assessment Tool in Data Science Education: Reasoning, Justification, and Evaluation

Olga Maksimenkova (✉), Alexey Neznanov
National Research University Higher School of Economics, Moscow, Russia
`omaksimenkova@hse.ru`

Irina Radchenko
ITMO University, St. Petersburg, Russia

Abstract—The paper addresses the questions of data science education of current importance. It aims to introduce and justify the framework that allows flexibly evaluate the processes of a data expedition and a digital media created during it. For these purposes, the authors explore features of digital media artefacts which are specific to data expeditions and are essential to accurate evaluation. The rubrics as a power but hardly formalizable evaluation method in application to digital media artefacts are also discussed. Moreover, the paper documents the experience of rubrics creation according to the suggested framework. The rubrics were successfully adopted to two data-driven journalism courses. The authors also formulate recommendations on data expedition evaluation which should take into consideration structural features of a data expedition, distinctive features of digital media, etc.

Keywords—Data expedition, data science, collaborative technologies, formative assessment, higher education

1   Introduction

It is well known that *Data Science* (DS) is relatively young and rapidly growing area. The discussions about DS and its understanding as a field are still heat topic of current interest [1]. To be clear in this paper we will follow the definition of DS from IBM Analytics [2]:

"*Data Science is an interdisciplinary field that combines machine learning, statistics, advanced analysis, and programming. It is a new form of art that draws out hidden insights and puts data to work in the cognitive era*".

Naturally, the youthfulness of DS is one the main reasons why *data science education* (DSE) is shaping today. However, several teaching and learning techniques and methods have been already introduced to the courses in this area [3, 4]. It is interesting that the actuality of DSE increases not only to data scientists but to the specialists of the other fields as well (e.g., education, medicine, journalism). This explains by

that millions of different information systems produce petabytes of data every hour all over the world and specialists and experts need skills in DS. Thus, many communities postulate that DS is a new foundation of Data Literacy (good example is [5]). The context of this paper lays in the field of DSE and it focuses on a data expedition DE as a collaborative formative assessment technique as it is follows from the definition by School of Data [6]:

"*Data expedition (DE) is a learning by doing computer-supported collaborative learning technique, which is applied to DSE*".

The purpose of any DE is to create a data-driven digital media artefact in 2-3 weeks by collaborating in small (2-3 persons) groups. A typical DE consists of several stages: data gathering/data collecting, data processing/data analyzing, data visualization and making a product based on data [7]. Naturally, this structure makes a DE an attractive and appropriate tool for complex skills in DS shaping. Moreover, the authors of this paper have great confident that a DE may successfully acts as a powerful collaborative assessment technique.

Surely, the idea of active learning and formative assessment implementation in DS courses is not new. Thus, these last few years practitioners of DSE began to document and reflex their experience in blended learning [8], flipped classroom [9], and project-based learning assessment [10] in implementation in DS courses. Moreover, several valuable examples of education DEs were introduced [7, 11]. However, the question of how to evaluate a digital media artefact obtained during a DE is still acute.

The paper explores a DE as a collaborative assessment technique in DS, focuses on its evaluation and introduces a framework for rubrics design.

2 Background

2.1 Data science education

As the authors share the position that an assessment should be agreed with Merrill's principles of instruction [12] and be designed according to learning objectives (outcomes) [13]. They also strongly believe that the clear understanding of data scientists' competencies is mandatory to design evaluation tool suitable to DSE. This section aims to distinguish crucial features of assessment in DSE. By reviewing the literature on DS curriculum design, teaching and learning approaches we collect the learning outcomes to be evaluated and generalize current teaching practices of DS-oriented programs.

Convenient to our goals data scientist's job descriptions are given in [8, 14, 15]. We will use a short-list of requirements to an ideal data scientist [8]. He/she must:

- Be of an analytical and exploratory mindset,
- Have a good understanding of how to do data-based (quantitative) research,
- Possess statistical skills and be comfortable handling diverse data,
- Be clear, effective communicators with the ability to interact across multiple business levels,

- Have a thorough understanding of their business context.

Anderson and colleagues in 2014 [16] claimed that difficulties with the curriculum design relate to multidisciplinary learning objectives of DS programs/courses. The valuable result of this work is the DS curriculum topic list. In 2016 Mills and colleagues [17] reported that the most of information systems programs begin offering DS-related courses. So, it is not surprising that in 2017 De Veaux and colleagues [18] specified the list by Anderson and colleagues, and highlighted that DS is not the direct sum of skills in statistics, computer science and mathematics, and introduced key competencies for bachelor in DS: computational and statistical thinking, mathematical foundations, model building and software foundation, data curation, and knowledge transference – communication and responsibility. These outcomes give us several clear directions in evaluation tool construction, but all these initiatives did not suggest any recipes of assessments.

In 2016 Hicks and Irizarry [19] formulated several principles of teaching DS:

- organize the course around a set of diverse case studies
- integrate computing into every aspect of the course
- teach abstraction, but minimize reliance on mathematic notation
- structure course activities to realistically mimic a data scientist's experience
- demonstrate the importance of critical thinking/skepticism through examples

This requirements, frameworks and principles shaped the basis of practical DSE. Thus, Brunner and Kim in [20] described the design of an introductory DS course. Because the course was delivered through MOOC platform, the assessments' types were partly predefined by the platform: quizzes, programming assignment and peer assessment. Asamoah and colleagues [21] also reported on introductory DS course which is based on interdisciplinary approach. Both papers are quite circumstanced, and their results may be directly reused by the educators, but they mostly focused at courses structure and their agreement with curriculum and recommendations that were mentioned above [16, 18]. So, there were no special approaches to DSE or learning outcomes evaluation introduced in these papers.

These last few years several publications that let us feel enthusiastic about the future of active learning and formative assessment techniques in DSE have been appeared. This year Ryan in [8] called blended model of education a powerful approach to DSE. The model engages as industry as academia to educational processes. Ryan also reviled the impetuous grow of number of DS programs in different universities. Moreover, innovative educational technologies spreading among DS educators all other the world. Thus, Eybers and Hattingh in [9] reported on flipped class room approach implementation to post-graduate DS students. Despite the paper contains the impressive results it does not mention any assessment type which was utilized during the instructions. More relevant to our topic research by Giabbanelli and Mago [10] explored the role of teaching computational modeling during DSE and suggested a relevant course. Unlike Brunner and Kim [20] Giabbanelli and Mago used a project-based learning assessment to evaluate complex DS skills. The results of each projects had to be a research proposal and using real data was obligatory. Unfortunately, the

authors do not in detail developed the topic about features of projects, grading rules and students' reaction.

We can resume that a full-fledged assessment in DS courses should reflect complex nature of DS. So, taking into consideration growing interest to active learning techniques from DS educational society formative assessment seems to be suitable to DSE. We may also note that the complexity of assessment procedures in DSE grows from the interdisciplinary of DS and consequently from composite structure of DS courses and their learning outcomes.

2.2 What a data expedition is?

A data expedition is one of the most popular collaborative crowdsourcing techniques for making a digital media based on data. As usual, a data expedition begins from the clear purpose but without specific understanding of final digital media content and bounds [22]. Generally, this digital media contains such objects as structured links on data sets, text, scripts, data visualization. A data expedition may be successfully implemented as for educational as for research needs [23]. Moreover, in the second case, a data expedition serves as the powerful and the fast approach to clean, pre-process, briefly analyze data to support some preliminary hypothesis or to formulate them [24]. Thus, data expeditions are quite popular among data-activists all over the World. They ran on variety of purposes by different international organizations [22, 7] universities [8], and they are still sought-after by the data activists in Open Data Day [24]. The description of some data expeditions in Russian universities is given by Radchenko and Sakoyan in [7]. The first one data expedition in Russia was held in 2013. Since that time data expeditions headed by Irina Radchenko have been launching in various formats.

As far as this paper concerns on educational aspects, it is necessary to notice that data expeditions are quite popular as project-based educational events and mainly are held in two formats: distance (online) and blended (both online and offline). Online format is open for participation by everyone and it is provided via Internet. Blended format is a format for students of one educational institution (university or any other). This type of data expeditions was crystallized due to pedagogical reasons, after the authors had to break the ice between participants of the data expedition and to reduce their fears and anxiety.

An online data expedition has several remarkable advantages. First, each interested individual can participate in it. Second, a contribution from all participants is recorded in summary report. Third, the best projects are published in a collective blog. It allows to create a participant's portfolio. Moreover, the data expedition materials are publicly accessible, and they can be reused by all people without restrictions.

Along with advantages, we may affirm there is some disadvantages. Firstly, the absence of face-to-face lessons is the cause of psychological discomfort of participants. They are afraid of asking questions through messengers, email, and the other means of communication, not everyone has strong motivation to complete all tasks, etc. Secondly, an online data expedition has lower performance and poorer outcomes in

comparison to a blended data expedition. Lastly, as far as there no blended learning classes there no materials created from them.

The advantages of a blended data expedition may be listed as follows: high efficiency, lots of questions from participants, and consequently quality feedback from experts and the other participants, additional materials which are created during the data expedition as responses on a request from participants, better motivation and involvement in the data expedition, more interesting topics for data-driven projects.

The blended format itself brings several disadvantages. First is the limited number of participants. Second one is a lack of scalability i.e. we do not have a unified frame so far. At last a blended data expedition is always bonded to certain time and physical location.
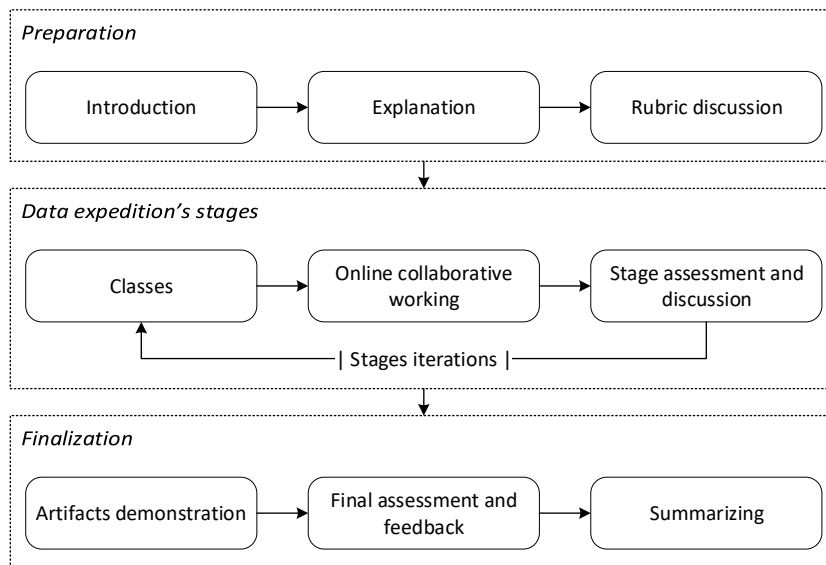

Fig. 1. Stages of the common framework for a data expedition

All the stages have different durations. Preparation is a very important and the longest stage. It includes a large amount of work on organization of the data expedition workplace, the collection of scaffolding materials, the creation of an outline, a plan for the classes, and assessment tools. Next stage is an introduction for students. It should meet students with the aim of a data expedition, its tools, methods and outline. Explanation is the stage which helps to adopt students with data expedition tools. This is an important stage, because many of students do not have any skills in working with supportive software used in a data expedition. Missing this stage leads to further discontentment and incomprehension among students. Next three stages are the stages, that can be repeated several times. They may contain face-to-face classes and online-working. This is the main part of a data expedition and it lays in the context of blended learning concept. Assessment allows to keep a feedback from students and to correct the direction of the data expedition. Final assessment shows a picture of the

data expedition in general. It allows to see outcomes from all projects of participants created during the data expedition. It also gives a feedback to students that motivates them to continue working with open data. Summarizing stage is a stage for collecting all materials from the data expedition. This stage includes publishing the best students work in collective blog as well.

The methodological recommendations for data expeditions was described in [1] context of blended learning concept with an emphasis on open educational resources, open data and open educational practices. We transposed these principles to education in university and applied these methodological recommendations to formal education. It shows more efficient results as an implementation in educational process at university. In this case data expedition contained mandatory assessment which could be used as a credit for a course in general.

3 Educational Data Expedition from the Evaluation Point of View

3.1 Applying the styles to an existing paper

A data-driven digital media artefact acts as a main objective result of a DE. From the significant properties of DEs we can deduce requirements on digital media artefacts. But for educational DE we should consider additional aspects. The short-list of what should be obligatory taken into consideration during evaluation is follows.

- The processes of digital media artefact creation. They should be fixed in DE diary by participants of a collaboration. Automatic software logs gathering is not enough!
- The quality of data citation. All the data should be cited according to data citation principles [11].
- Correctness and reproducibility. We need the measures that is taken to ensure correctness and reproducibility of data transformations – it is very hard task.
- The complexity, interactivity, and design of final artefacts. We give an advantage to good storytelling [25] with interactive data dashboards [26] and infographics.

3.2 Data expedition evaluation criteria

As it is follows from [6] the evaluation of data expedition reflects interdisciplinary nature of DS and consequently is rather complicated. Wherefore, as far as educational data expedition is a prolonged in time procedure we formulated several criteria which responds as to process as to result. Because each criterion may be expressed on various levels, we developed rubrics. Most of the criteria have five levels with comprehensive description. The list of criteria is follows.

- Data expedition dairy (experiment log) for research data expeditions. The criterion reflects the reproducibility of the results of completed data expedition.

- Using scripts written in one of the programming languages. The criterion allows to evaluate the laboriousness of data preparation/visualization/analysis processes.
- Using special software for data preprocessing or analysis. The criterion stands for ability of implementation data analysis by means of professional tools.
- Programming code. The criterion allows to validate the independency of scripts' development.
- Data tiding. The criterion allows to estimate the accuracy of quantitative statistical results.
- Data analysis. The criterion reflects the level of data analytics skills of a team such as settling a hypothesis, testing hypothesis, etc.
- Data visualization. The criterion allows to estimate the accuracy of quantitative statistical results and skill of a team in preparing interpretable results of data analysis.

Given criteria are quite general and give us a frame to evaluate educational DE. Surely, every DE requires either extra criteria or detailing of specified above.

3.3   Supportive software

Software using during each DE play a significant role in evaluation. At first, it supports collaboration and helps to identify and measure an author's contribution to processes and results. At second, special tools allow teams to collect, version, and share artifacts of their work as inside as outside of a team. So, a DE uses a lot of software for collaborative work, data sources access, data processing, and publishing data-driven products. Supportive software may be grouped as followed (with some examples in brackets).

1. Integration, communication, and management tools
   (a) Rich collaborative environments (Microsoft Office 365, Google G Suite)
   (b) Project management tools (Trello, Microsoft Teams, MeisterTask)
   (c) Online communication tools (Microsoft Skype, Telegram)
2. Search and external discussions tools
   (a) Universal search engines (Google Search, Microsoft Bing)
   (b) Specialized search engines (Wolfram Alpha)
   (c) Social Media Services (Facebook, Twitter, Instagram)
   (d) QA services (Data Science Stack Exchange, Public Lab, Quora)
3. Data processing tools
   (a) Universal cloud data storages (DropBox, Microsoft OneDrive, Google Drive, Yandex Disk)
   (b) Code repositories (GitHub, BitBucket, GitLab)
   (c) Data analysis and visualization tools (R Studio, Jupyter Notebook, Orange, Microsoft Excel, Microsoft Machine Learning Studio, Tableau, OpenRefine, Infogram, Plotly)
   (d) Open data platforms (CKAN, Zenodo)

(e) Specialized data extraction tools (Import.io, Kimono, Scrapy, ABBYY FineReader)

(f) Data validation tools (CSVLint, Data Package Validator, Data Hub LOD Validator)

4. Media publishing platforms (Medium, Microsoft Sway, Tilda, GitBook, Wordpress, Blogger)

We should note that collaborative technologies utilization is one of the main trends in software development. It is amazingly corresponded with the ability to evaluate a collaborative work of a team

4 Data Expedition Implementation in Universities

The DE was adopted to two educational activities in parallel. Both were implemented to data-driven journalism master-students of different universities. So, the digital media artefact was a data-driven article. The type of article was defined within a DE and the rubrics were expanded by extra requirements.

Note that the prerequisites for these groups, the rules of their shaping, and instructional design were different. So, the paper does not provide any comparison between these groups and their results.

4.1 Instructional design

Data expeditions were implemented twice in the spring term of 2016/2017 academic year. Two groups of students from different universities took part in educational data expeditions. First group was first-year students of "Data Journalism" master program at National Research University Higher School of Economics, Moscow (HSE). Second group consisted mostly of students of European University, Saint-Petersburg (EU) and selected researchers who could join the event. After we will address these groups as HSE-students and EU-participants.

Both data expeditions were given as a summative assessment at the end of educational modules. HSE-students were engaged into data expedition after a part of Scientific Research Seminar which followed several blocks of trainings in DS. So, the students were familiar with the basics of Python-programming, mathematical statistics, data analysis and open data retrieving. EU-participants did not work with open data previously, and they had basic skills in statistics and data processing only.

4.2 The data expedition in HSE: Features and schedule

The data expedition in HSE was conducting at the same time as in EU but was slightly different. At first, only HSE-students of the first-year "Data Journalism" master program was involved into this activity. At second, the participation was obligatory because the DE acted as an assessment after several modules of the Scientific Research Seminar, "Open data" and "Introductory programming" courses. This explains

by that the DE was targeted to evaluate complex skills in data-driven journalism which should have been shaped by the end of the first year of education. At third, DE was preceded by short in-class session where students divided into working teams and took part in discussion on evaluation rubrics. At fourth, collaboration between participants and administrators was supported by Microsoft Office 365 Education (using OneNote Class Notebook [27]).

The module of the Scientific Research Seminar was delivered by the authors and had took 8 academic hours. First 4 hours were spent on practice in cleaning data. The last 4 hours were reserved for rubrics introduction and discussion, group shaping and data-expedition introduction.

The purposes of the data expedition were

- To engage students in collaborative work with real open data
- To push students to generalize their knowledge and experience in programming, data analysis and data science tools
- To show students directions for future activities by giving formative feedback
- To give a feedback on learning outcomes achievements and troubles with the instructional design of the master program to academic and administrative supervisors

For these purposes and according to these purposes evaluation criteria from Section 3.1 were specified by adding:

- A journal of an experiment to evaluate processes of DE
- A structure and a specification for the content of a paper to evaluate the quality of result digital media
- The requirements to journalistic, popular-science or science text
- This year total number of students who was engaged to the data expedition was 28. They were divided into 9 groups without any randomization and with the only limitation on a group size. The group size is recommended [7] to be not more than 3 persons

4.3 The data expedition in EU: Features and schedule

Participants was invited to the DE through the announcement which was posted in Internet[1] and everyone who was interested in work with open data was asked to send his/her CV with an indication of his/her scientific interests and experience in scientific research. So, the participants had different levels of knowledge and skills.

The data expedition was running for two weeks. All the participants (14) were divided into small groups of two people. They were immediately provided with educational materials, posted on GitHub[2]. Participants were also invited to contribute their

---

[1] https://eu.spb.ru/forthcoming-events/17625-opencitydataworkshop
[2] https://github.com/iradche/Data-Expedition-in-EU

materials on GitHub. A digital media which was expected at the end of DE was a draft of a scientific paper based on data.

During the DE all the participants attended lessons which were targeted except others to fill the gaps in their knowledge and skills. Each lesson included the following sections:

- lesson plan
- Tasks
- Assessments
- Supportive information and additional material
- Instructions
- Useful links
- The list of tools and software
- Summary
- The assignment(s) for the next lesson

Assignments to work online were given after every class. A link to rubrics allowed to improve students' comprehension on assessment criteria. It also contains variants of assessment for each criterion3.

The criteria introduced in Section 3.1 were modified and we presented 8 criteria in total: the data expedition participation's diary, data processing software, data analysis tools, data cleaning, data analysis, data visualization, paper's structure and paper's content. Every criterion had 4 discrete levels: 0, 0.25, 0.75, and 1 (where 0 – nothing has been done, 1 – a criterion is completely fulfillment).

The progress of the participants' team work was collected in a Google Spreadsheet.

5 Results and Discussion

5.1 Data collections

As it is easy to see from Section 3.1, in our study we have two general populations because data expeditions were started independently in two higher schools and had some differences. The first general population consists of the first year "Data journalism" master students of NRU HSE, Moscow. All the participants were asked to fill in an electronic post-survey. The volume of a research sample (HSE-students) is 9 observations. The sample differs from the population, because each group delegates a member to provide feedback via the post-survey.

The second general population included EU-students and several researchers who joined the data expedition. In this case, post-survey was taken place too and the volume of the sample (EU-participants) is 14 observations.

---

[3] https://docs.google.com/spreadsheets/d/1fDgEbYeI1P87ob_MaS-oWcbro_BV0oj-Ofr9PeUXgn4/pubhtml

5.2 Tools and methods

EU-participants were asked to pass a pre-questionnaire that was presented as a Google Form with eight closed-ended questions about students experience in working with open data. After the DE students were asked to participate in a post-survey. It contained four closed-ended questions and three open-ended questions about their impressions, behavior and intentions on further work with open data.

Before final grades publication all the HSE-students were invited to take part in a post-survey. The post-survey was implemented as a Windows Form hosted at the Microsoft Office 365 Education platform. The questionnaire contained 3 multiple choice questions and 2 open-ended questions. The questions were targeted to get the information about learners' behavior during the data expedition and satisfaction of assessment.

As far as it was a pilot implementation and both research samples are rather small we use simple learning analytics for data analysis and interpretation.

5.3 The data expedition in HSE

The results of the post-survey demonstrate students that found their participation in data expedition very useful. From the first row of Table 1 it is easy to see that none of them reported on useless of data expedition. This also evidently proved by open answers, for example, see Responses #1 and #3 in Table 2. The data from the second row in Table 1 allow us to conclude that prolongated nature of a data expedition is a problem to master students. Probably, we may reduce this problem if students are prepared to this type of assessment during the instructions. The last row demonstrates that the third part of students feel themselves inconvenient working in group.

Naturally, that a pilot reviled several problems in integration of data expeditions to daily educational practice. For example, Response #4 shows that a student was expected new data and did not understand that he/she had to demonstrate some skills in working with data. Similar picture is observed in Responses #5 and #6. Respondents did not recognize their experience in programming, data analysis and mathematical statistics as the preliminary to the data expedition. This clearly shows the power of the data expedition as a formative assessment tool. It punishes students to clue their previous learning experience altogether.

Table 1. HSE post-survey. Multiple choice answers distribution

Question	33%	67%
Was your participation to the data expedition useful for you?	Yes	Probably Yes
How was your time allocated during the data expedition?	Little by little, but regularly	I worked not regularly. Sometimes for a long, sometimes there were no time at all
Give a characteristic of your interaction with the other participants?	I prefer working alone	I had enough interaction, I am satisfied

Table 2. HSE post-survey. Selected answers to open questions

Question	Answer
What were useful for you in data expedition?	**Response #1**: *"I have created universal code which I am going to use in the other projects"*
	Response #2: *"Data expedition is a completely new for me type of a project. So, to accomplish the task the time had to be distributed in a separate way. But, each project goes to "Experience" storage and this is good"*
	Response #3: *"Practical experience in writing data-driven article"*
What were not satisfactory for you in data expedition?	**Response #4**: *"On the one hand, we were free in the selection of topic but on the other the result would have been limited stricter. I mean clear requirements to our work. Moreover, it would be better to work with new not previously publicized data (not only links to open data portals which are we familiar with and have no interest). But the freedom in topic selection should be kept"*
	Response #5: *"I did not like that we immediately got down to independent work. At the lessons before the data expeditions we only clean data. It would be better to pass through all the stages with teachers' control"*
	Response #6: *"I wish more practice before a data expedition"*

5.4 The data expedition in EU

In this section, we present the main results of questionnaires and DEs for EU-participants. Table 3 shows the partial results of pre-questionnaire in EU.

Table 3. Results of pre-questionnaire in EU, percent

Question	Yes, %	No, %
Have you previously worked with open data?	28,6	71,7
Have you previously worked with online services for open data visualization?	7,1	92,9
Have you previously written scientific papers in your native language?	92,9	7,1
Have you previously written scientific papers in English?	35,7	64,3
Is it interesting for you to learn about Open Science and its implementation for research?	92,9	7,1

The results of this questionnaire allow to make a conclusion that participants had not experience with open data and open data visualization. Most of them (92,9 percent) already have written scientific papers in Russian, but only 64,3 percent of them have written scientific papers in English. The question about Open Science was contained in both questionnaires. Participants answered positively both times (92,9 percent in first questionnaire and 100 percent in second). All participants noticed that this data expedition was quite useful for them in second questionnaire. They noted a broad outlook and awareness on all the questions of data expedition organizer. Participants emphasized the importance of practice in following programs: *HackPad*, *OpenRefine*, *Tableau Public*, and *GitHub*. All participants wanted to take part in further data expeditions.

According to participants' opinion, there was not enough time for work in the data expedition for all of them. Simultaneously all participants pointed out the sufficiency of interaction with each other.

Figure 2 shows the results of questionnaire in European University, question "Have you previously used collaborative tools?" Some of the respondents answered only

"Yes" without details. In total, 85.7 percent of participants already had an experience with different collaborative tools. Table 4 shows the results of pre-questionnaire in European University.
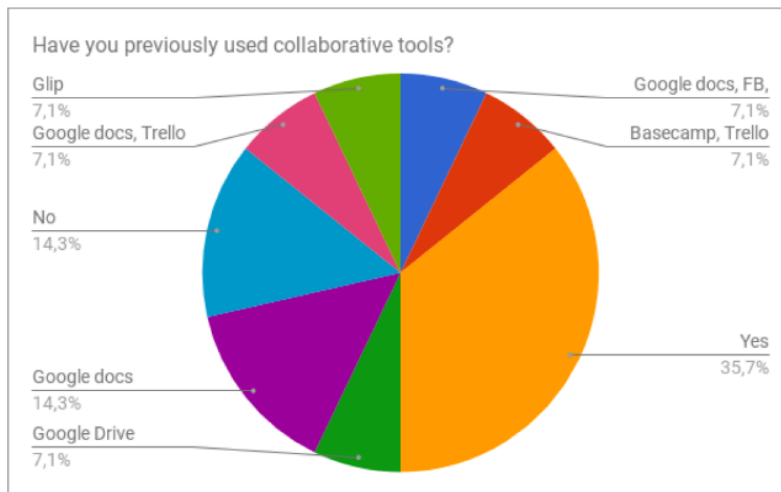

Fig. 2. The results of first questionnaire in European University,
question "Have you previously used collaborative tools?"

We used Google Sheets with marks of students' progress as milestones assessment in data expeditions. It allowed to keep whole educational process in transparent view. Moreover, the students could see not only their own progress but all picture in general and the other students' outcomes and progress in details.

The project revealed that participants of data expedition have a low level of skills in operating with open data and statistics. But simultaneously a data expedition initiates huge interest in working with open data.

The pilot demonstrated that the developed framework for rubrics reduces laboriousness of grading rules design. Grading rules for the first DE were similar to the second one and contained all the criteria from Section 3.1. The main differences were in data analysis methods, application field knowledge for data analysis interpretation, experience in software tools.

6 Recommendations and Conclusion

The context of the paper lays in the field of data science education. The paper contributed to data science education, overviewed the implementation of educational DEs and focused on their assessment role. As a primary result the paper introduced the framework for rubrics design.

Based on results of the pilot the authors have formulated the recommendations to educational DEs evaluation:

1. Evaluation should take into consideration two aspects of a DE: orientation to the predefined goal and multi-stage nature of main DE processes.
   (a) Predefined goal allows one to check correspondence between task description and final digital media artefacts
   (b) Multi-stage nature of DE processes allows one to assess involvement and impact of participants based on DE diary and logs
2. The significant features of digital media artefact which is created during a DE should be reflected in grading rules
3. Proposed rubrics should be discussed with DE's participants for common understanding of goals and better involvement in the methodology of DE
4. Grading of data citation, reproducibility and provenance should follow available guidelines (see [28] as an example)

Some responses on the pilot's post-survey allow us to carry out the directions of the future work on educational DEs' methodology:

1. Assessment in form of DE should be implemented after a short supportive block which aims to refresh relevant knowledge and skills
2. Working groups should be completed according to rubrics. For example, if programming skills mentioned in grading rules, the groups should contain at least one programmer

Using DE is highly beneficial for interdisciplinary educational programs where data science meets domain knowledge. In this work the authors shared the experience of using DE in data journalism master-programs. Evaluation criteria for DEs were developed according to the professional requirements to data scientists' competencies which are documented in recommendations to DS curriculum design and jobs' descriptions. Our practice proved that a DE is a good example of an interactive collaborative learning technique and it effectively works as a formative assessment tool. The nearest future we plan to improve the assessment quality, to expand it into related domains, and to gather additional statistics and opinions of students with different background.

7 Acknowledgement

8 References

[1] Anderson, P., Bowring, J., McCauley, R., Pothering, G., & Starr, C. (2014). An Undergraduate Degree in Data Science: Curriculum and a Decade of Implementation

Experience. Proceedings of the 45th ACM technical symposium on Computer science education, (pp. 145-150). https://doi.org/10.1145/2538862.2538936

[2] Anderson, P., Nash, T., & McCauley, R. (2015). Faciliatating Programming Success in Data Science Courses through Gamified Scafollding and Learn2Mine. Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education, (pp. 99-104). https://doi.org/10.1145/2729094.2742597

[3] Asamoah, D., Doran, D., & Schiller, S. (2015). Teaching the Foundations of Data Science: An Interdisciplinary Approach. Pre-ICIS Business Analytics Congress Conference, (pp. 1-9).

[4] Bhargava, R., Deahl, E., Letouzé, E., Noonan, A., Sangokoya, D., & Shoup, N. (2015). Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data. Retrieved 06 26, 2017, from Data-Pop Alliance.

[5] Brunner, R., & Kim, E. (2016). Teaching Data Science. Procedia Computer Science , 80, 1947-1956.

[6] Data Citation Synthesis Group. (2014). Joint Declaration of Data Citation Principles. (M. Martone, Ed.) San Diego CA: FORCE11.

[7] Data Science. (2017). Retrieved July 14, 2017, from IBM Analitics: http://www.ibm.com/analytics/us/en/technology/data-science/

[8] De Veaux, R., Agarwal, M., Averett, M., Baumer, B., Bray, A., Bressoud, T., Sondjaja, M. (2017). Curriculum Guidlines for Undergraduate Programs in Data Science. Annu. Rev. Stat. Appl., 4, 2.1-2.16. doi:10.1146/annurev-statistics-060116-053930

[9] Eybers, S., & Hattingh, M. (2016). Teaching data science to post graduate students: a preliminary study using a "F-L-I-P" class room approach. International Conferences ITS, ICEduTech and STE 2016 , (pp. 189-196).

[10] Fink, L. (2003). Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses. Wiley.

[11] Giabbanelli, P., & Mago, V. (2016). Teaching computational modeling in the data science era. Procedia Computer Science, 80, 1968-1977. https://doi.org/10.1016/j.procs.2016.05.517

[12] Guidance. Data scientist: role description. (2017). Retrieved July 14, 2017, from Gov.uk: http://www.gov.uk/government/publications/data-scientist-role-description/data-scientist-role-description

[13] Hicks, S., & Irizarry, R. (2016). A Guide to Teaching Data Science. Retrieved July 13, 2017, from Cornell University Library: http://arxiv.org/vc/arxiv/papers/1612/1612.07140v1.pdf

[14] Jormanainen, I., & Sutinen, E. (2013). An open approach for learning educational data mining. Koli Calling'13, (pp. 203-204). Koli. https://doi.org/10.1145/2526968.2526997

[15] Kosara, R., & Mackinlay, J. (2013). Storytelling: The Next Step for Visualization. IEEE Computer, 5, 44–50. https://doi.org/10.1109/mc.2013.36

[16] Logi Analytics. (2017). The Definitive Guide to Dashboard Design. Retrieved 07 18, 2017, from http://logianalytics.com/dashboarddesignguide/

[17] Merrill, M. (2002). First principles of instruction. Educational Technology, Research and Development, 50(3), 43-59.

[18] Microsoft. OneNote Class Notebook. (Microsoft) Retrieved 06 23, 2017, from http://www.onenote.com/classnotebook

[19] Mills, R., Chudoba, K., & Olsen, D. (2016). IS Programs Responding to Industry Demands for Data Scientists: A Comparison between 2011 - 2016. Journal of Information Systems Education, 27(2), 131-140.

[20] MIT. (2015). Data Scientist Position Description. Retrieved July 14, 2017, from Massachusetts Institute of Technology: http://ist.mit.edu/sites/default/files/about/org/roles/Data_Scientist_Position_Description_v4.pdf

[21] Open Data Day (Data expedition session) - focusing on Infrastructure. (2015). Retrieved July 16, 2017, from Outbox blog: http://www.outbox.co.ug/blog/open-data-day-data-expedition-session-barabara-team-focusing-infrastructure

[22] Open Data Day 2016: Data Expedition. (2016). Retrieved July 16, 2017, from Sinar Project. Empowering Malasian Citizens: http://sinarproject.org/en/updates/open-data-day-2016-data-expedition

[23] Plaue, C., & Cook, L. (2015). Data Journalism: Lessons Learned While Designing an Interdisciplinary Service Course. Proceedings of the 46th ACM Technical Symposium on Computer Science Education, (pp. 126-131). https://doi.org/10.1145/2676723.2677263

[24] Radchenko, I., & Sakoyan, A. (2016). On Some Russian Educational Projects in Open Data and Data Journalism. In D. Mouromtsev, & M. d'Aquin (Eds.), Open Data for Education. Linked, Shared, and Reusable Data for Teaching and Learning (pp. 153-165). Springer. https://doi.org/10.1007/978-3-319-30493-9_8

[25] Rauber, A., Asmi, A., van Uytvanck, D., & Pröll, S. (2016). Identification of Reproducible Subsets for Data Citation, Sharing and Re-Use. Retrieved 05 17, 2017, from http://www.rd-alliance.org/system/files/documents/TCDL-RDA-Guidelines_160411.pdf

[26] Ryan, L. (2017). The data science education and leadership landscape. Elsevier.

[27] School of Data. (2013). Data expeditions. Retrieved 05 29, 2017, from School of Data: http://schoolofdata.org/data-expeditions/

[28] What is Data Science? (2013). Retrieved July 14, 2017, from NYU Center for Data: http://cds.nyu.edu/research/

9 Authors

Olga V. Maksimenkova is a research fellow at International Laboratory for Intelligent Systems and Structural Analysis, Faculty of Computer Science, NRU HSE (Moscow, Russia). She is a member of the Association for Computing Machinery (ACM).

Alexey A. Neznanov is a senior research at International Laboratory for Intelligent Systems and Structural Analysis, Faculty of Computer Science, NRU HSE (Moscow, Russia). Alexey is also an associate professor at Department for Data Analysis and Artificial Intelligence, Faculty of Computer Science, NRU HSE. He is a member of the Institute of Electrical and Electronics Engineers (IEEE).

Irina A. Radchenko is an associate professor at ITMO University (St.Petersburg, Russia). Irina is also an associate professor at NRU HSE (Moscow). She is an ambassador of the FREYA project (http://www.project-freya.eu).