



Blackbox

Data Collection Project

1. [Overview](#)
2. [Access to Data](#)
3. [Development Team](#)
4. [Frequently Asked Questions](#)
5. [List of Publications](#)

Overview

The BlueJ Blackbox data collection project is an initiative by the developers of BlueJ to collect data on how BlueJ is used, in order to increase understanding of how students learn to program. The data collected is for the purposes of academic research, and will only be used by computing education researchers.

Access to Data

For access to the data, contact blackbox-admin@bluej.org

Development Team

Michael Kölling
Ian Utting
Davin McCall
Neil Brown
Amjad Altadmri
Hamza Hamza

Frequently Asked Questions:

- [What data will be sent?](#)
- [Who will the data be sent to?](#)
- [How will the data be anonymised?](#)
- [How much traffic will this generate?](#)
- [How can I opt in/out?](#)
- [Why do I repeatedly get asked if I want to opt in?](#)
- [I'm a network administrator; how I disable participation for my users?](#)
- [My question isn't answered here](#)

What data will be sent?

The main data that will be sent is the ([anonymised](#)) source code from your projects. We also record the use of the BlueJ interface: for example, which methods are invoked, use of the codepad, use of the debugger, and how you use other features of BlueJ. No identifying information (e.g. username) will be sent with the data.

Who will the data be sent to?

The data will be sent to a server hosted at the University of Kent in the UK. Researchers at the University of Kent will have access to the data in order to analyse it, and access will also be provided to other recognised computing education researchers for the purpose of analysis.

How will the data be anonymised?

No identifying information (e.g. username, machine name) is sent to the server. The name of the project is sent, but the full path (which probably contains your username) is not. Source code is sent, but all comments before the class begins are blanked out -- that is, the top comment before your class will be blanked, as that typically contains your name. All other code (and comments) are sent to the server.

How much traffic will this generate?

The exact rate at which data will be sent is dependent on the actions you are performing and the size of your source code base. As an estimate, we believe for a handful of small classes (e.g. the projects accompanying the textbook) that the upload will be around 3-4 Megabytes each hour of continued use, and the download will be around 1 Megabyte each hour. As a quick point of comparison, loading the BBC news front page once involves downloading around 0.5 Megabytes of data and uploading 0.06 Megabytes.

How can I opt in/out?

To change your participation in this research, in BlueJ 3.1.0 and later, go to the Preferences window, and under the Miscellaneous tab there is an option to change your current participation.

Why do I repeatedly get asked if I want to opt in?

Your participation status is stored in BlueJ's properties file. This is stored in your user profile directory on your machine. For a home machine, or a school network which supports persistent profiles, you should be asked once, and this decision stored thereafter.

However if your network does not keep your profile, you will be asked every time you load BlueJ, because BlueJ cannot tell that you have been asked before. In this case, you will need to contact your network administrator and tell them to either let profiles persist (the ideal solution), or otherwise to alter the bluej.defs file supplied with BlueJ to include the line:

```
blackbox.uuid=optout
```

I'm a network administrator; how I do disable participation for my users?

If you want to opt-out your users by default, you can alter the bluej.defs file that is installed with BlueJ to include the line:

```
blackbox.uuid=optout
```

My question isn't answered here

If you are having a technical problem with BlueJ, even if it is related to the Blackbox project, please contact us via [our standard support form](#). If you have a question about the research side of the Blackbox project, you can contact us at blackbox-admin@bluej.org.

List of Publications

- Neil C. C. Brown, Amjad Altadmri, Sue Sentance, and Michael Kölling. 2018. *Blackbox, Five Years On: An Evaluation of a Large-scale Programming Data Collection Project*. In Proceedings of the 2018 ACM Conference on International Computing Education Research (ICER '18). ACM, New York, NY, USA, 196-204. DOI: <https://doi.org/10.1145/3230977.3230991>
- Becker, B. A., Murray, C., Tao, T., Song, C., McCartney, R., and Sanders, K., *Fix the First, Ignore the Rest: Dealing with Multiple Compiler Error Messages*, Proceedings of the 49th ACM Technical Symposium on Computer Science Education (SIGCSE '18). ACM, New York, NY, USA, 634-639, 2018. DOI: <https://doi.org/10.1145/3159450.3159453>
- Mirza, O. M., Joy, M., and Cosma, G., *Suitability of BlackBox dataset for style analysis in detection of source code plagiarism*, Seventh International Conference on Innovative Computing Technology (INTECH), Luton, pp. 90-94, 2017. DOI: <https://10.1109/INTECH.2017.8102424>
- Brown, N. C. C. and Altadmri, A., *Novice Java Programming Mistakes: Large-Scale Data vs. Educator Beliefs*, Trans. Comput. Educ. Volume 17, issue 2, Article 7 , 21 pages, 2017. DOI: <https://doi.org/10.1145/2994154>
- Keuning, H., Heeren, B., and Jeurig, J., *Code quality issues in student programs*, Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education, ser. ITiCSE '17. New York, NY, USA: ACM, pp. 110–115, 2017. DOI: <https://doi.org/10.1145/3059009.3059061>
- McCall, D., Kölling, M., *Meaningful categorisation of novice programmer errors*. In Frontiers In Education Conference , pages 2589-2596, 2014. DOI: <https://10.1109/FIE.2014.7044420>
- Murray, C., *A Comparative Study of Java Compiler Error Profiles Using the Blackbox Dataset*, Master's thesis, University College Dublin, 2016.
- Altadmri, A., and Brown, N.C.C., *Researching Programming Education with Blackbox (Abstract Only)* , In Proceedings of the 47th ACM Technical Symposium on Computing Science Education (SIGCSE '16), ACM, New York, NY, USA, 702-702. 2016. DOI: <https://doi.org/10.1145/2839509.2850479>
- Altadmri, A., and Brown, N., C.C., *37 Million Compilations: Investigating Novice Programming Mistakes in Large-Scale Student Data*, In Proceedings of the 46th ACM Technical Symposium on Computer Science Education (SIGCSE '15), ACM, New York, NY, USA, 522-527, 2015. DOI: <https://dx.doi.org/10.1145/2676723.2677258>
- Brown, N., C.C., Kölling, M., McCall, D., and Utting, I., *Blackbox: a large scale repository of novice programmers' activity*, In Proceedings of the 45th ACM technical symposium on Computer science education (SIGCSE '14), ACM, New York, NY, USA, 223-228, 2014. DOI: <http://dx.doi.org/10.1145/2538862.2538924>
- Brown, N., C.C. and Altadmri, A., *Investigating novice programming mistakes: educator beliefs vs. student data*. In Proceedings of the tenth annual conference on International computing education research (ICER '14), ACM, New York, NY, USA, 43-50, 2014. DOI: <https://dx.doi.org/10.1145/2632320.2632343>
- Brown, N., C.C., *Introduction to analysing the BlueJ blackbox data (abstract only)*, In Proceedings of the 45th ACM technical symposium on Computer science education(SIGCSE '14), ACM, New York, NY, USA, 748-748, 2014. DOI: <http://dx.doi.org/10.1145/2538862.2539012>