INFORMATION SHEET FOR PARTICIPANTS

REC Reference Number: LRS-16/17-4786

YOU WILL BE GIVEN A COPY OF THIS INFORMATION SHEET

Title of study  Public Provenance Services

Invitation Paragraph

The provenance services are public services provided freely to the community to transform, store, share, analyse and process provenance. Services are available from https://openprovenance.org. Usage of the services, and therefore participation in the study, is totally voluntary.

What is the purpose of the study?

As part of the efforts to encourage adoption of provenance technology, we provide the provenance services freely to the community. In order to gain an understanding of provenance in the wild and to improve the services, we collect statistics, referred to as Collected Data, about the provenance data submitted to our services by the users.

The services, including their design, algorithms, implementation, usability and performance, are the subject of provenance-oriented software engineering and data management research. Collected Data is used by researchers to improve the design and performance of the services, and advance research on provenance analytics techniques, integrity and trust of provenance.

Specifically, the purpose of Collected Data is:

- to improve the design and performance of the services,
- to advance research on provenance analytics techniques, integrity and trust of provenance,
- to publish about the services and technology underpinning them.

Why have I been invited to take part?

We do not actively identify participants for this research. We publicise the availability of the provenance services on our web pages/blog posts or in our papers.

Do I have to take part?

Anyone can use the provenance services and, by so doing, becomes a participant in our research.

What will happen to me if I take part?

There are two modes of interactions with the provenance services.
Registered Mode

You will need to register for a user account in order to upload your provenance documents for storage and some analysis. For this, you need to provide your email address (so that you can use it to reset your password in case it is lost or forgotten). Once you have an account, you can start to store your provenance documents on the provenance services, referred to as Stored Data. By default, uploaded provenance documents are private; you can decide to make them public. Please be aware that all documents contributed to the repository that are marked “public” will be publicly available to anyone who accesses the repository. You will, however, be able to later change or specify access control for each document and remove those that you no longer wish to share or store at any time.

Unregistered Mode

Alternatively, you may decide to process provenance data, without having to persist it for the long term. In that case, you do not need to register for an account but can just upload your provenance data for processing. In such circumstances, the provenance data will not persist on our services.

Data Collection

The services collect and store the following data, referred to as Collected Data:

- Web access log.
- Metrics of provenance data posted to the services (e.g. size, statistics and topology measures) inspired by our Provenance Analytics Paper [http://dx.doi.org/10.1007/s10618-017-0549-3].
- Validation reports that describe places where provenance is invalid, incomplete or incorrect.
- Digest of provenance data posted to the services (e.g. signature of provenance and timestamps).
- Anonymous summaries of provenance data, inspired by our GAM'15 paper [http://dx.doi.org/10.4204/EPTCS.181.9].

In addition, for registered mode, we also require your email address (to order password resets) besides your Stored Data.

We note that some provenance data that will be cached temporarily for the duration of a user session to improve the responsiveness and usability of the services and further data is collected indirectly by third-party “web analytics” services such as Google Analytics, New Relic, Sentry, and User Voice.

Incentives (where relevant)

No incentives are provided beyond the intrinsic functionality of the services.

What are the possible benefits and risks of taking part?

You can use the provenance services

- to process provenance, and integrate it in your own applications,
- to create provenance,
- to share provenance,
• to store provenance,
• to sign provenance and check its integrity,
• to visualise provenance,
• to edit, design, check and validate provenance, and
• to analyse provenance data.

Risks:

Data protection

In order to protect your data, all communications with the services are automatically encrypted using the industry standard Transport Layer Security, which is supported by all modern web browsers.

Stored Data: As with any public data store, please make sure you check your data for any sensitive information before making it public. In any case, you always retain control of the data. If you later decide the data is no longer suitable to be kept in the data store, you can delete them from our repository yourself after logging on to the service. Your data will be stored on a server that is protected behind a firewall and can only be accessed by authorised staff from inside our network. Your email address and password are never stored in plain text. As a result, we cannot reveal your password in case it is lost or forgotten. You will need to use your email address to request a password reset.

Service availability and backup arrangements

The services databases will be regularly backed up. However, please note that the backup strategy is best effort. No guarantee is provided about the reliability of backups, the ability to restore data following a crash, and the ability to restore data promptly if at all possible.

We reserve the right to remove data deemed not appropriate for the service. Note that all stored provenance, in particular stored as public, is attributable.

From time to time, services and algorithms will be updated, and we offer no guarantee that data will be preserved through such upgrades.

The service is available for an undefined period of time, subject to the ability of the research team to resource its maintenance.

Qualitative Properties

There is no guarantee provided in terms of asserting qualitative properties of provenance such as its validity, its authenticity, its quality or its trustworthiness. As algorithms are researched and deployed, we offer no guarantee of consistency of such properties over time.

Will my taking part be kept confidential?

All the data you upload to the server will be private by default unless you choose to make them public. Your email address and password will always be kept private.
How is the project being funded?

The project is self-funded.

What will happen to the results of the study?

Results will be published in scientific journals and conferences. Anonymised Collected Data underpinning the results will also be published to promote scientific reproducibility.

What if I changed my mind?

If you change your mind, you can log in to your account to remove any document you have uploaded. Explicit deletion of Stored Data results in such data being permanently erased.

If you wish to completely delete your user account from the repository, please contact us using the details below. User account deletion results in all Stored Data by this user to be permanently erased.

Note that explicitly Stored Data will also eventually disappear from backups as they are rotated on a regular basis.

However, Collected Data persist in an anonymized format and, hence, we provide no mechanism by which Collected Data can be erased.

Who should I contact for further information?

If you have any questions or require more information about this study, please contact me using the following contact details:

Professor Luc Moreau  
Department of Informatics  
King’s College London  
The Strand

What if I have further questions, or if something goes wrong?

If this study has harmed you in any way or if you wish to make a complaint about the conduct of the study you can contact King’s College London using the details below for further advice and information:

The Chair, BDM Research Ethics Panel, rec@kcl.ac.uk

Thank you for reading this information sheet and for considering taking part in this research.