WLCG Privacy

David Kelsey STFC-UK Research and Innovation
(Presented to WLCG GDB, CERN, 10 July 2019)

eosc-hub.eu

@EOSC_eu

Dissemination level: Public

EOSC-hub receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 777536.

Contents

Personal Data Protection and Privacy

- WLCG Privacy Notice
  - https://docs.google.com/document/d/1Fs_7OMq2Ck-viZnYw823dzuNfYj0cr7vd-42EMrZt_s/edit
- Processing of Personal Data Policy Framework
  - https://docs.google.com/document/d/1X6m8FhLHH3qe5plWOZQYoYJKt_z6oHJo6euDB1Y3e6s/edit

- Much discussion at June GDB and since then
  - Some suggestions were easy to fix/address
  - Other issues more contentious – let's discuss!
- General question: *Do we define our aspiration or do we describe reality? (Answer: We have to abide by GDPR)*

WLCG Privacy Notice

Issues

- Must we enforce the deletion or anonymization of accounting logs at 18 months?
- Why do we need to keep the user registration data for 36 months?
- Access only to people authorized by WLCG
  - What about world-readable dashboards showing names?
- User rights to access/erase/modify – contentious
- Minor – define/change “participant”?

Why? - Access logs - max 18 months

- *"Access logs and accounting records are kept for up to 18 months before being anonymised or deleted"*
- GDPR Article 5 (e)
  - … kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed;
- WLCG/EGI **POLICY ON THE HANDLING OF USER-LEVEL JOB ACCOUNTING**
  - https://edms.cern.ch/document/855382
  - Adopted in August 2009

Old Accounting Policy – retention period

- The Sites are responsible for deleting the local accounting records according to local personal data retention policy. This needs to be long enough to ensure that all records have been successfully transferred to the ADC database.
- The ADC is responsible for deleting the copies of the individual accounting records in the central database, or for removing or anonymising personal identifying information, e.g. the CommonName or e-mail components of subject DNs, from these records, at the latest 18 months after receipt of the data in the ADC. Personal identifying information, e.g. the CommonName component, contained in aggregated data must be treated in the same way.

Guidance on retention

- Recital 39 says:
  - the period for which the personal data is stored should be limited to a strict minimum and that time limits should be established by the data controller for deletion of the records (referred to as erasure in the GDPR) or for a periodic review.
- Guidance (from UK Data Protection Network)
  - if you need to retain some data for particularly lengthy periods of time, then consider anonymising the data first
  - make sure you have at least *some* concrete justifications for why you keep data for the periods you do, rather than a vague "because it might be useful some day" type argument.
  - Even with these justifications, your view of why it is "necessary" to keep data will likely not always align with what a regulator, data subject, or court considers "necessary"

Why concerns? What should we do?

- “accounting systems may keep records for much longer than 18 months. And this statement would require us to either throw accounting away after 18 months, or rewrite the records. I don't think either is desirable.”
- A proposal
  - We have to define a retention period (GDPR)
  - This is primarily for the central accounting database
  - In EGI we have been anonymising at 18 months for 10 years
    - This has not caused problems (AFAIK)
  - Services which cannot comply should have their own Privacy Notice explaining how long and why they need that time

User registration data – max 36 months

- WLCG Privacy Notice says:
  - *"WLCG will keep your user registration data for as long as you remain a registered member of the VO plus the maximum accounting record retention period. In order to enable WLCG to support the user employment life cycle and unless you explicitly request otherwise, WLCG may keep your registration data for up to 36 months after you leave."*
- It says "up to"
- Need to keep enough personal data to confirm it is the same person who returns

Other comments on Data Retention

- “We need to remember the history of the experiment. When a problem was found, when solved, by whom. When a given sample was produced with which configuration, and see 100 y later if he was a "trusted one". ”
- DaveK answered: “OK. Yes. But do we need to the identity of an individual who ran a job?”
- Response “well, the name of the individual ends up in the path of the storage for the output, so either we change FILENAMES after 18 months, or we delete the file, or the answer is yes”

Who has access?

- WLCG Privacy Notice says
  - *WLCG will make your personal data accessible only to those authorised by WLCG, and only for the purposes described above.*
- Comment: “I don't think this is a "true" statement as of today and would be misleading to users. It would actually be important to systematically review how in practice user data is kept as of today and published, and how "private" it actually is. For instance, there are dashboards where the user's full name is not scrambled and anyone can find out what a particular user is doing by looking at their grid activities.
- PROPOSAL
  - Dashboards should either STOP showing usernames or
  - If they need to do so, then describe the reason why in their own Privacy Notice

User access?

- User having access is a GDPR right
- Guidance from the UK ICO
  - You must provide data subjects with:
    - confirmation their data is being processed;
    - access to their personal data; and
    - other supplementary information.
  - You must comply with any subject access request within one month of receipt.
  - You cannot charge a fee unless the request is “manifestly unfounded or excessive”.

User access?

Guidance (continued)

- Where you process a large quantity of information you can ask the data subject to specify the information they want access to.
- You may refuse to comply with a subject access request where this is “manifestly unfounded or excessive”.
- PROPOSAL
  - Can we limit access to User Registration Databases and central Accounting Databases?
  - Log files stored at sites are too many – or ask user which site does she want the details?

User's rights to erasure

Guidance from UK ICO

- Data subjects have the right for their data to be erased where:
  - the personal data is no longer necessary in relation to the purpose for which it was collected/processed;
  - the data subject withdraws their consent or objects to the processing and there are no overriding legitimate interest to continue processing;
  - the personal data was unlawfully processed or has to be erased in order to comply with a legal obligation; or
  - the personal data is processed in relation to the offer of information society services to a child.

- This does not seem to apply to WLCG

User Objection/Rectification

Guidance

- Data subjects have the right to object to:
  - processing based on legitimate interests, the performance of a task in the public interest or the exercise of official authority (including profiling);
- Where a data subject otherwise objects to you processing their personal data then you must comply with this request unless you can demonstrate overriding compelling legitimate grounds to continue processing
- Data subjects can have their personal data rectified if it is inaccurate or incomplete.
- You must comply with any request to rectify within one month of receipt. This can be extended to 2 months where the request is complex.
- Issues: What is "inaccurate" (now or when recorded?). What is "incomplete"

Define “participant”

WLCG Privacy Notice - To whom do we transfer your data?

- *Your personal data may be transferred only to the following parties, and only as far as is necessary to provide the WLCG and VO services that you make use of:*
  - *authorised WLCG participants,*
  - *third parties whose data privacy and protection policies are equal to or more restrictive than the WLCG policy.*
- Who are the “participants”? Other sites/services
- Shall we just say “authorised WLCG services”?

WLCG Policy on the Processing of Personal Data

Issues

- Why do we say “fairly, open and transparent”?
- Retention length (18 months)?
- User rights of access
- Where do we report data breaches?
- Are audits required?
- Dual meaning of word “processing”

Fair, open and transparent

- Why do we say: *"The End User whose Personal Data is being Processed shall be treated fairly and in an open and transparent manner."?*

GDPR Guidance

- You must use personal data in a way that is fair. This means you must not process the data in a way that is unduly detrimental, unexpected or misleading to the individuals concerned.
- You must be clear, open and honest with people from the start about how you will use their personal data.

Reporting Data Breaches

- *Infrastructure participants shall*
  - *Respond to suspected breaches of this Policy promptly and effectively and take the appropriate action where a breach is found to have occurred;*
- What is appropriate action? Can we give examples?
- DaveK: There is a legal GDPR requirement for an entity suffering the data loss to report this to their local Data Protection authority without delay. (For those without a local authority) guess we should also require reporting to wlcg-privacy@cern.ch?

Are audits required?

- Policy says: *"Perform periodic audits of compliance to this Policy and make available the results of such audits to other Infrastructure Participants upon their request."*
- Who actually does periodic audits?
- Self assessments are an important part of Monitoring in a Code of Conduct and is best practice to show compliance
- Proposal: We could do this via an example assessment spreadsheet in the same way as we do for IGTF and WISE/SCI assessment

Minor problem - Dual meaning of "Processing"

- In the Policy Definitions:

Infrastructure

- *The bounded collection of universities, laboratories, institutions or similar entities, which adhere to a common set of policies [R 2] and together offer data processing and data storage services to End Users.*

Processing (Processed)

- *Any operation or set of operations, including collection and storage, which is performed upon Personal Data [R 1].*
- **PROPOSAL: Change Infrastructure to say "offer data analysis and data storage services"**

Thank you
for your attention!

Questions?

EOSC-hub

eosc-hub.eu @EOSC_eu