Privacy by design services and DataTags
Addressing personal data handling in academic research

Francisco Romero Pastrana and Esther Hoon
Data Federation Hub - Human Subject Research Group

Research data FAIR-ness and the GDPR
Currently it is unclear for researchers and research support what needs to be done to make research data FAIR and to comply with the upcoming GDPR (General Data Protection Regulation) when such data contains sensitive personal information. This lack of clarity and appropriate information hampers responsible handling and sharing of sensitive data. Fortunately, research occupies a privileged position within the GDPR. The GDPR aims to encourage research, as long as organizations make an objective risk assessment balancing the public interest in research and the rights of participant and implement the appropriate safeguards. It is the obligation of the data controller to properly implement safeguards during the whole research life cycle. Importantly, this responsibility continues when making the data available for re-use (FAIR).

Safeguarding through Privacy by Design (PbD) services
The GDPR explicitly mentions the methods of a ‘privacy impact assessment’ and ‘privacy by design’ as means to build in safeguards in an early stage. As such, organizations must incorporate privacy into the design, operation, and management of a given system or design specification. In order to translate the PbD philosophy and legal GDPR requirements into proper system requirements, ‘privacy design strategies’, and their associated ‘tactics’ and patterns, have recently been proposed. For research projects, privacy must then at least be incorporated in the design of the data management plan, the supporting research services, and the operational protocols.

Proposed Privacy by design services:
• Pseudonymization / TTP services
• Privacy Impact Assessment support
• Virtual research environments
• Customized data workflows: iRODS
• Sharing sensitive data: DataTags
• More to come...

Integrated Rule-Oriented Data System (iRODS)
iRODS is open source data management software. It virtualizes data storage resources, and supports microservices, storage systems, authentication, networking, databases, rule engines, and an extensible API. Importantly, it can implement powerful customized workflows where organizational data management policy can be captured in an automated, auditable fashion.

iRODS Capabilities

Data Protection
• Unified namespace
• Storage Tiering
• Access control
• Data Integrity

Subject’s Rights
• Data Discovery
• Automated Ingest
• Publishing

Accountability
• Provenance
• Auditing
• Compliance

DataTags: Interoperable rights expressions on re-use of sensitive data
DataTags are human-readable and machine-actionable labels that express conditions under which files can be stored, used, transmitted and re-used. A dataset repository that uses the DataTags approach, iRODS metadata and its rule engine, can automate sensitive dataset uptake, archiving and sharing. Importantly, it can request specific dataset handling compliance to dataset recipients, like proper recipient authentication and data retrieval, and legally-binding data user agreements (DUA) to enforce data recipients GDPR obligations.

The Data Federation Hub / Human Data
The University Medical Center Groningen and the University of Groningen joined efforts to set up an integrated research support platform: the Data Federation Hub/Human Data. The DFH/Human Data support the whole research data lifecycle while ensuring data security and efficiently protecting the privacy of participants.

www.rug.nl/dfh    dfh@rug.nl    e.hoorn@rug.nl    f.romero.pastrana@rug.nl