Working Group for Data Security and Trust (WGDST)

Version	Date	Edited/Reviewed by	Comment
0.1	14.12.2015	Stefan Pröll	Initial document
0.2	18.01.2016	Stefan Pröll	Use Cases
0.3	16.02.2016	Stefan Pröll	Collaboration
0.4	23.05.2016	Alessandra Scicchitano	Use Case
0.5	28.02.2017	Becca Wilson	Redraft for more concise focus.
0.6	10.03.2017	Becca Wilson	Merge feedback from founding WG members and co-chairs

Problem Statement and Scope

The exchange of information is based on a very fundamental concept: mutual trust. Trust is difficult to establish and easy to lose. In order to be able to trust each other, research facilities, companies and institutions need to agree on and ensure minimum data protection standards. This requires a verifiable layer of trustworthiness, which goes beyond paying lip service. Only if the partners can agree on data authenticity and integrity standards, only if they can show that the protection of sensitive data meets the policy needs and if access can be granted and revoked based on valid criteria, trustworthy data exchange is possible.

So far the topics of data security and trust are dealt with in isolation in the existing RDA working groups, with some essential aspects potentially not covered. What we want is a common arena for understanding and harmonizing various notions of data security and trust across research domains, which should facilitate common agreements on standards, best practices and policies. This, in turn, should allow access to, and/or exchange, of sensitive data without disclosure to unauthorised parties and according to clearly defined and verifiable protocols.

It is clear that the RDA Working Group for Data Security and Trust (WGDST) cannot address all relevant questions from the very start. Our aim is to focus on a few practical topics which immediately allow to improve security and identify opportunities for the controlled sharing of sensitive research data. These topics initially include:

- Operational (technical) policies on data access and data release for research data that is deemed sensitive because of privacy or commercial considerations, or other concerns
- Authentication and authorisation protocols for data access
- Protocols for data integrity and authenticity
- Secure and privacy aware data processing
- Data sharing (processing data offsite) and code exchange (processing onsite)

Deliverables and Work Plan

Within the 18 month time frame, the WGDST will produce the following items:

D1 - (Virtual) Kick-Off Meeting

In the first meeting we will discuss and agree on a timeline for this WG and set monthly web-conferences.

Output: The minutes and the agreed timeline will be published on the WG wiki.

D2 - Investigating the Existing Standards, Recommendations Best Practices, Processes and Solutions

We will identify the different aspects of research data security and trust and create an overview of the area. We will investigate the existing standards, recommendations and best practices to distill the common questions and issues that are related to the protection and verification of data - mapping them to the different aspects of research data security previously identified. We will identify exemplar projects that demonstrate how research data can be exchanged in a secure way, how trust can be established and maintained and what pitfalls need to be avoided.

This overview does not only help us to limit the scope of the working group, but also is an important output providing insights to the security requirements of the research community.

Output: Make available a summary of D2 findings on the best practices (including examples), standards, recommendations, process and solutions currently available.

D2 - Scenarios, Use Cases and Potential Adopters

We will collect use cases and scenarios that can demonstrate the applicability of any guidelines we develop. These use cases will additionally provide insight into current data security and trust practices as well as highlight any potential challenges.

The following set of initial use cases, potential adopters and solutions were presented during the BoF session at P6.

- Access and Use of Confidential Microdata in Social and Economic Sciences: Mike Priddy (DANS, Netherlands)
- Big Facilities for Small Science: Vasily Bunakov (STFC, UK)
- Data Exchange in Biomarker Research: Peter Kieseberg (SBA Research, Germany)
- DataSHIELD: Prof Paul Burton, Dr Becca Wilson (University of Bristol, UK)
- DEXHELPP: Rudolf Mayer, Stefan Pröll (SBA Research, Germany)

During P7 we added a new initial use case:

- E-infrastructure and Distributed Data: Allesandra Scicchitano (GEANT, Netherlands)

Output: Maintain a list of use cases and potential adopters that will grow through the engagement activities of the working group.

D4 - Gap Analysis

Based on the findings in D2 and D3, we will create a gap analysis of the security landscape in relation to research data protection. We aim to identify areas where there is a lack of best practices, protocols and standards. These results may serve as a basis for identifying research questions for further projects and allow to raise the awareness for potential data leaks and security breaches.

Output: Make available a summary identifying areas lacking best practice, protocols and standards with respect to data security and trust.

D5 - Guidelines

Combining the exploratory work from D2-D4 we will produce guidelines - harmonised across domains - to facilitate the sharing of research data in a secure, usable, trustworthy and systematic way. This set of guidelines will enable data creators to look up common data sharing scenarios and identify the best practice processes, solutions and examples accordingly. These guidelines will reassure data creators in their data sharing practices and provide support for those not yet familiar with data security practices.

Output: The WG guidelines for data security and trust in data sharing will be made available for potential adopters.

D6 - Final Meeting

We will present the guidelines produced by the working group for each of the treated topics, which facilitate data creators to employ appropriate processes for exchanging data in a secure way with respect to their own requirements. The meeting will serve as a mechanism to identify potential adopters of the guidelines for future evaluation.

The meeting will also discuss any further WG/IG for future RDA activity related to data security/trust including e.g.

- the long term view of encryption standards
- security audit of data repositories
- collaboration in research environments that involve sensitive data
- the definition, the application and the verification of machine-executable policies on data access and data release

Engagement with Existing RDA Work and Groups

Several RDA activities deal with security in a broader context, which highlights the importance of the topic for the research data community. We are currently aware of the following list of related groups under the umbrella of the RDA:

- RDA/CODATA Legal Interoperability IG
- RDA/NISO Privacy Implications of Research Data Sets IG
- Ethics and Social Aspects of Data IG
- BioSharing Registry: connecting data policies, standards & databases in life sciences WG
- Health Data Interest Group

We are actively asking for collaboration and looking for synergies to exploit and aim to avoid duplication. Our goal is to link existing work within the groups and collaborate and exchange with other active RDA initiatives in the area of data security and trust. We will also seek to engage further examples of best practice, use case scenarios and potential adopters of our outputs.

Value Proposition

We expect that various stakeholders will benefit from the outputs of this proposed working group, examples include:

Data users: the outputs will work towards making research data available for (re-)use, rather than being siloed.
Data creators, owners and holders: can adopt the WG guidelines to facilitate data security and trust when data sharing and also examine exemplar projects that demonstrate best practice.
Research and Research-related communities: can make use of a knowledge base of current processes, protocols, best practice, solutions and challenges surrounding data security and trust in research data sharing.

Chairs and Founding Members

Name	Institution
Stefan Pröll	SBA Research, Austria
Rudolf Mayer (co-chair)	SBA Research, Austria
Peter Kieseberg	SBA Research, Austria
Andreas Rauber	TU Wien, Austria
Vasily Bunakov	STFC, UK
Paul Burton	Newcastle University, UK
Mike Priddy	DANS, Netherlands
Becca Wilson (co-chair)	University of Bristol / Newcastle University, UK
Alessandra Scicchitano (co-chair)	GEANT, Netherlands
Mary O'Brien Uhlmansiek (co-chair)	Washington University in St Louis, USA

Conclusions

The aim of this WG is to provide usable data security guidelines for data creators, allowing them to exchange data in a secure way. The guidelines provided by this WG are based upon examples of best practice and explained using concrete examples. Besides delivering the necessary security know-how, the group also raises the awareness for security measurements in the domain of research data.