EGI Service Design Workshop

TUWIEN – OSSDIP
Open Source Secure Data Infrastructure and Processes

Andreas Rauber, Martin Weise
http://www.ifs.tuwien.ac.at/~andi/secure_data_infrastructure.html
Introduce the service: OSSDIP

- **Open Source Secure Data Infrastructure and Processes**
  - Secured data visiting services, data remains under owner’s control
  - Allows per-user access to data excerpts for data analytics / processing
  - Existing Platform: DEXHELPP: TRL 8-9 (operational for 10 years)
    Re-implementation: OSSDIP: TRL 4-5 (undergoing tests before roll-out)

- **Based entirely on Open Source technology, flexible components**
- **Operated as hosted service, with a focus of individual hosting**
- **Currently undergoing tests to replace another platform**
- **Pilot use cases with selected service providers in Dec / Jan**
- **Full roll-out in Feb of core modules, additional components (e.g. fingerprinting) to follow**
OSSDIP - Requirements

- Sensitive Data (privacy issues, commercial interests, …)
- Provide access for analysis, but ensure data is not leaked / misused
- **Data Visiting** instead of Data Sharing
- **Data owner maintains full control over data and use:**
  - **Who** to allow access,
  - over **which period of time**,
  - for **which subset of data**,
  - to answer **which research question / analysis goals**,
  - while monitoring what they are doing
- Full monitoring up to provable deletion
Secure Data Infrastructure @TUWIEN

- Set-up to support research in the wake of COVID19
- Primarily for commercially sensitive data
- Set-up and operational within 2 weeks
- To be enhanced…
- Based on experience from operating the DEXHELPP infrastructure for medical data (http://www.dexhelpp.at) at TU Wien over the past decade
- Same principles, focus on automation
- Desire by data owners to host themselves
- Ease set-up of data visiting services (modular, Open Source components)
- Reference implementation + hosted service
- Data stored on inaccessible data store
- Data provisioning on a per-user / per study basis
- Required data is extracted, optionally aggregated, fingerprinted and provided on shielded virtual machine (Compute VM) per use case
- Access by single researcher via Remote-Desktop VM only (media break, no tunnel) via VPN, from which an SSH connection is established to the Compute VM
- Extensive logging / monitoring
- Open-source: self-hosting or service provisioning
- Contractual agreements
- Defined Processes
OSSDIP Data Provisioning Process

- Data Owner signs data processing agreement
- Data Owner provides metadata (FAIRness)
- Data Owner receives account data
- Data Provider creates secured VM for data transfer (nullled storage, gateway access, shielded, …)
- Data Owner transfers data and notifies Data Provider
- Data Provider transfers data to secured storage and destroys VM
- Data Provider “publishes” metadata as agreed to allow researchers to find and request access
OSSDIP Data Access Process

- Analyst requests access to data, specifying research goals
- Data Owner verifies/approves request
- Contract on data access between Data Owner and Analyst
- Data Owner authorizes provisioning of specific subset of data (+ aggregation levels, + fingerprints, …) for specific time
- Analyst agrees to terms & conditions + monitoring regulations
- Data Processor creates shielded compute VM equipped with according extract of data + tools requested
- Data Processor provides access credentials to analyst
- Analyst accesses Remote Desktop VM via VPN (double encryption), from there SSH to Compute VM
- Results export only via Data Owner VM
Planned integration with EGI services
- AAI: trusted identity, 2-factor authentication
- Compute resources
- Data mgmt. system for Metadata publishing (FAIRNess)

Hosted service at TU Wien + Network of EGI partners that offer trusted infrastructure

Additional services for secure multiparty computation, federated machine learning, etc. to avoid data integration within EGI network
EGI Service Design Workshop

TUWIEN – OSSDIP

Open Source Secure Data Infrastructure and Processes

Andreas Rauber, Martin Weise

http://www.ifs.tuwien.ac.at/~andi/secure_data_infrastructure.html