A Europe-wide Interoperable Virtual Research Environment
to Empower Multidisciplinary Research Communities and Accelerate Innovation and Collaboration

Security, Privacy and Trust Strategies for a Virtual Research Environment

Laura Hollink, Jacco van Ossenbruggen, Jan Wielemaker
Centrum Wiskunde & Informatica

CWI
Centrum Wiskunde & Informatica

VRE4EIC Project

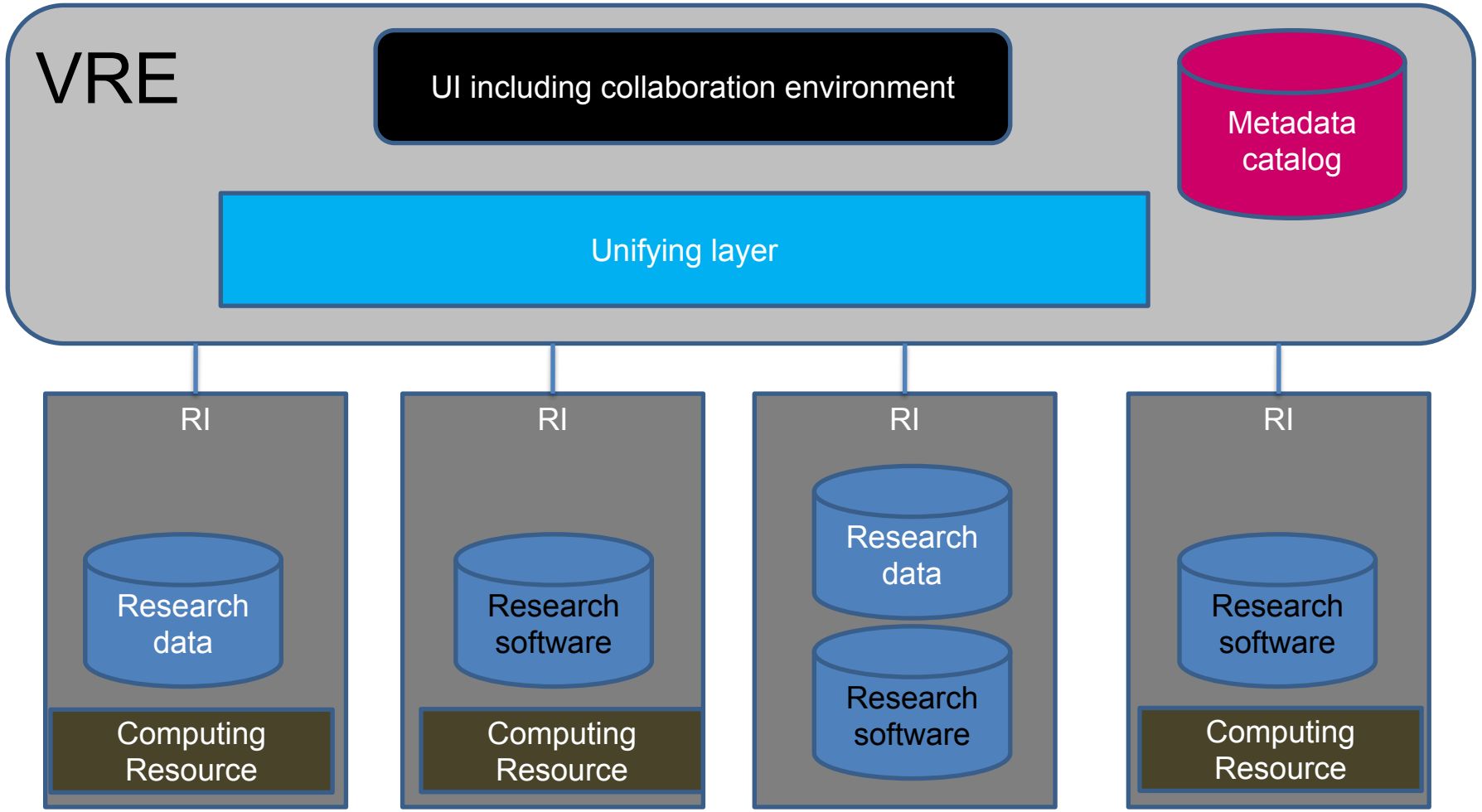
- A European-wide interoperatble Virtual Research Environment
- that bridges across silo RI's
- by providing a Reference architecture, Software components, Standardisation and Training.

- H2020 RIA over 3 years
- 8 partners from 4 countries:

TU Delft, CWI, CNR, FORTH, INGV, UvA,

EuroCRIS, ERCIM

VRE4EIC: VRE as bridge across RI's



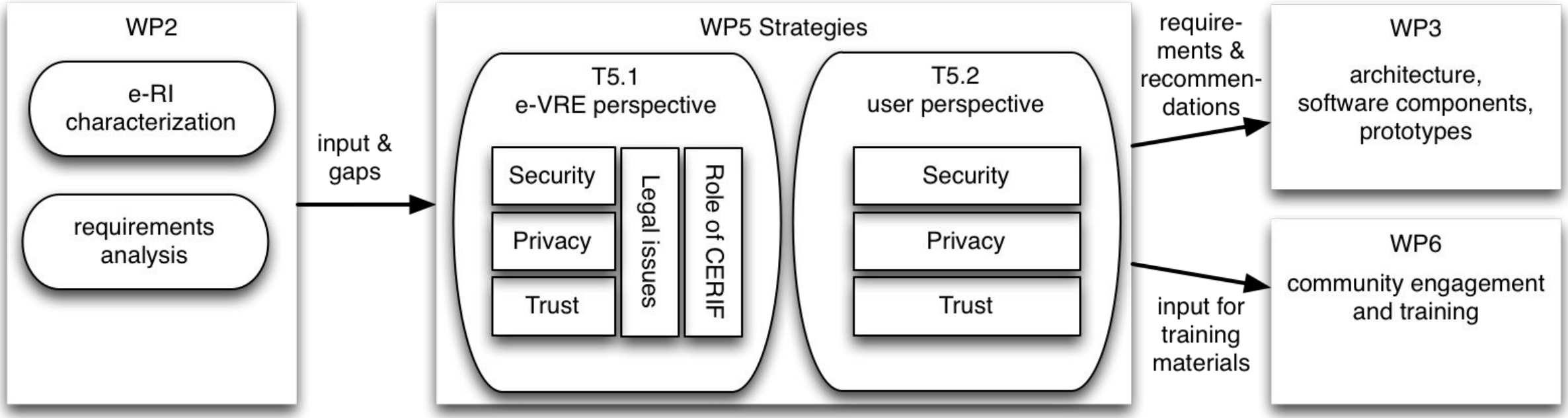
How to handle security, privacy and trust issues

The bridge function of e-VRE brings additional challenges

For example:

- laws and regulations vary
- Privacy risk of combining datasets
- trustworthiness of data, algorithms, software, people is harder to assess
- We need (single) sign-on with a wide variety of identity providers.

Main flow of information



Requirements, existing solutions and gaps from WP2 (highlights)

Trust:

- need to identify (and cite) datasets, including versions
- need for permanence of datasets
- need for provenance, logging, accounting (may conflict with privacy needs)

Privacy:

- needs to be guaranteed for both users and research data
- privacy levels e-RI not always strict enough for e-VRE (combi data)
- required functionality: delete data, the right to be forgotten

Security:

- need for secure access through various identity providers
- secure storage, backup and transmission of data (at e-RI and e-VRE level)
- need for metadata about level of security.

Trust strategy

- Users need **information** to be able to:
  - Assess the quality of the resources

- Role of the e-VRE:
  - provide access the trust-information of the RIs: provenance, permanence, citable IDs, permalinks to (guaranteed unchanged) versions of data, etc.
  - work on **interoperability** between the metadata formats in use at the underlying RIs
  - **Incentivise RIs** to implement advanced trust functionalities and to publish interoperable metadata

Trust – implications for users

- Trust in data:
  - Metadata standards to record provenance, versions, etc. may require training.
- Trust in people:
  - **reputation based trust** can come from user profiles listing resources (published papers, datasets, software, etc., including citations)
    - **"authority based trust"**: where available, VRE will use information from external organizations who provide quality information, e.g. Inspire website.
  - **Credential based trust** relies on a trusted identity provider.
- Trust in the VRE:
  - Relies on a secure system based on federated identities, with login at your own institute.

Privacy strategy

- e-VRE users need their usage data and access credentials protected
- people whose personal data appears in datasets need to be protected
- data owners need information about the consequences of sharing data with the e-VRE community (risks of combining datasets, e-VRE privacy policy, AAAI options)

- Strategy of the VRE:
  1. Conform to the EU data protection rules when collecting and storing usage data:

Privacy strategy

- e-VRE users need their usage data and access credentials protected
- pec
- dat
  the
  opt
- Stra
  1.

EU General Data Protection Regulation Key Points:

A. Easy access: data subjects are guaranteed to have free and easy access to their personal data and get understandable information about how their data is being processed.
B. Consent: data subjects will be asked for their consent explicitly.
C. The right to be forgotten: data subjects have the right to request erasure of personal data.
D. Data portability: data subjects have the right to transfer their personal data between service providers.
E. Breach: in case of a data breach, organisations are required to notify both individuals and the relevant data protection authority.
F. Responsibility and accountability: data protection must be designed into the business processes for products and services, and privacy settings are set at a high level by default.

http://ec.europa.eu/justice/data- protection/document/factsheets_2016/factsheet_dp_reform_citizens_rights_2016_en.pdf

Privacy strategy

- e-VRE users need their usage data and access credentials protected
- people whose personal data appears in research data need to be protected
- data owners need information about the consequences of sharing data with the e-VRE community (risks of combining datasets, e-VRE privacy policy, AAAI options)

- Strategy of the e-VRE:
  1. Conform to the EU data protection rules when collecting and storing usage data
  2. Provide data owners with information to set the appropriate access levels (again**: metadata interoperability** is important)
  3. Privacy relies on security issue: storage, backup and transmission of data

Privacy strategy

- e-VRE users need their usage data and access credentials protected
- people whose personal data appears in
- data owners need information about the
  the e-VRE community (risks of combinin
  options)

This means an increased burden on users:
- to understand what personal data is collected
- to understand different privacy concerns of the underlying RI's

This can be addressed by documentation and training

- Strategy of the e-VRE:
  1. Conform to the EU data protection
     usage data
  2. Provide data owners with information to set the appropriate access levels (again**: metadata interoperability** is important)
  3. Privacy relies on security issue: storage, backup and transmission of data

Security strategy (1/2)

- VRE cannot provide security where an RI fails to do so (e.g. regarding secure storage)

- International AAAI solutions are still unfinished (attribute management, organizational aspects of federated login) so we can:
  - Track developments of AAAI projects, such as AARC2.
  - Implement a local solution in such a way that RIs can easily switch to larger-scale solutions in the future

Security strategy (2/2)

- Multi-factor authentication is recommended
  - Which one depends on:

security (i.e. encryption), privacy (i.e. storage of additional user information) and ease-of-use (i.e. the need for additional devices).

- AAAI metadata use for ‘Role-based access control’ in the VRE should be trustworthy.
  - Who guarantees this? Personnel department? Project manager? IT?
  - It is recommended that the responsibility lies with the same department that is responsible for the organisation’s own AAAI records.

Security – implications for users

- Users need to understand the complex e-VRE environment
  - Access settings, privacy policy, licenses, etc.
  - People may need training.
- Users need an account with an identity provider.
- Users need to trust the AAAI component of the e-VRE, and the identity provider.

Strategy with respect to legal issues (IPR, Licensing, Accounting)

- Users need **information** on:
  - licenses, terms and conditions, and intellectual property rights of resources.
    - e.g. for attribution, or to determine how they can (re-)use the resource.

- Strategy of the e-VRE:
  - As for trust: Provide **functionality** to access legal info, work on **interoperability** of metadata on legal info, **incentivise** e-RIs to provide interoperable metadata.
  - Implement **accounting services** to keep track of all user actions.
  - Be **conservative** when implementing access mechanisms to ensure that all terms and conditions of the e-RIs are met, that neither the e-VRE or the user become liable.
  - Agreements will have be made per RI.

7. Metadata strategy

- Interoperable metadata is key for trust, privacy and security.

- Strategy for the e-VRE:
  - CERIF serves as a hubb to which several local metadata formats can be linked.
  - CERIF metadata model provides Role-Based Access Control (RBAC)
    - One or many roles can be given to users (based on groups, organisations, other/personal characteristics)
    - Access permissions are given based on roles.

Architecture and software components / Recommendation	UI	AAAI components	Cryptographic interface	Interoperability Manager	Metadata Manager	Linked Data Manager	e-VRE Web Services
PR2: agree to privacy policy	x	x					
PR3: privacy of data + usage		x	x				
PR5: collecting usage data				x			
TR1: interoperable metadata					x		
TR3: metadata creation					x		
SR1: VRE more secure than RI		x		x			
SR3: various ID providers		x		x			
SR4: credentials VRE-RI		x		x			
SR5: RI usage restrictions		x		x			
SR6: include new RIs		x		x			
SR7: RBAC		x			x	x	x

Full list of recommendations in public deliverable (D5.1) on www.vre4eic.eu

- PR1: The e-VRE should have a privacy policy that conforms to the European Data Protection Directive.
- PR2: The e-VRE user should be aware of and agree to the privacy policy of the e-VRE.
- PR3: The e-VRE should guarantee the privacy of both users of the e-VRE (authentication and access logs) and of sensitive research data that is stored through the e-VRE.
- PR4: Privacy recommendations with respect to research data management (skipped)
- PR5: Privacy recommendations with respect to e-VRE usage data
- TR1: The task of the e-VRE is to provide (CERIF) metadata related to trust. At the e-VRE level, the main requirement is to correctly convey the information that is already present at the e- RI level (incl. data ownership, permanence, licensing and liability) of each dataset.
- TR2: An e-VRE must conform with the IPR policies of the e-RIs that it provides a service layer for.
- TR3: An e-VRE trust policy should take into account that there is a high cost associated with the creation and maintenance of extensive metadata and provenance information. The preferred e-VRE strategy is to collect this metadata automatically as much as possible while allowing users to manually add metadata if they estimate that this is cost-effective.
- SR1: The e-RIs form the baseline for security, privacy and trust for the research data they manage; the e-VRE must guarantee standards that are at least as strong as the e-RI.
- SR2: The e-VRE security policy should make explicit who is liable in case of different types of security breaches. In addition, a protocol is necessary regarding the actions to be taken in the event of a security breach.
- SR3: A successful e-VRE is compatible with a wide variety of identity providers in order to suit the needs of associated e-RIs.
- SR4: The e-VRE should be able to pass on security credentials from the e-VRE users to the e- RI.
- SR5: The e-VRE should ensure that its own operations do not violate usage restrictions of resources of the e-RIs.
- SR6: The e-VRE should be compatible with several external access mechanisms and be able to include new ones when new e-RIs connect to the e-VRE, and allow unrestricted access to open data.
- SR7: The e-VRE provides Role-Based Access Control (RBAC) to separate several layers: users and groups on one hand, and roles, permissions and resources (or types of resources) on the other hand. This needs to be enforced for all interfaces.

8. Mapping of recommendations
to architecture and software components

- UI, AAAI component and interfacing components:
  - Authentication interface; Accounting interface; Authorisation interface; Cryptographic interface

Summary and conclusions

- Interoperability of metadata from the various e-RIs
  - security credentials need be passed on between the e-VRE and the e-RI
  - CERIF as metadata model for information about users, groups, roles, permissions, resources and types of resources
- Documentation/Training is needed for uses to handle the increased complexity.
- E-VRE / e-RI relationship
  - strategies are based on incentives
  - need for explicit documentation
- AAAI is focal point of e-VRE design
  - support as many AAAI solutions as possible
  - need for information about the access rights of the user

Acknowledgement

The VRE4EIC project has received funding
from the European Union's Horizon 2020 research and innovation
programme under grant agreement No 676247