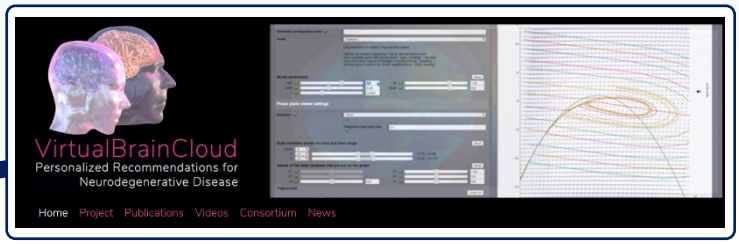
VirtualBrainCloud

Personalized Recommendations for Neurodegenerative Disease



www.VirtualBrainCloud-2020.eu

Public deliverable report

Deliverable 7.1

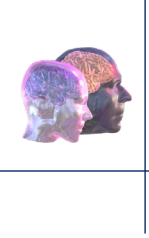
Report on Cloud Implementation Strategy applying „Trusted Cloud" Principles

Date	February 2020
Authors	Horst Schwichtenberg (SCAI), MalinRoth (SCAI), André Gemünd (SCAI), Sreeram Sadasivam (SCAI), Resham Kaur (FZJ), Michael Tarnawa (FZJ), Mariana Rissetto (UNIVIE), Catherine Altobelli (UNIVIE), Emily Johnson (UNIVIE),Michael Cepic (UNIVIE), Petra Ritter (CHARITE)

Dissemination level	**public**
Website	www.VirtualBrainCloud-2020.eu

This project has received funding from the **European Union´s Horizon 2020** research and innovation programme under **grant agreement No 826421**

Table of content



1. Introduction

A key concern of The Virtual Brain Cloud (TVB-Cloud) project is to develop a cloud-based architecture, termed the Cloud, for personalized prevention and treatment of neurodegenerative diseases. The Cloud will make use of existing local as well as pan-European services from the European Open Science Cloud (EOSC)[1]. The TVB-Cloud environment will leverage the potential of big data and high-performance computing (HPC). It will combine existing technologies and services to a cloud solution, particularly tailored to the requirements and demands of the healthcare sector.

Making use of external cloud computing services and infrastructures always requires trust in the involved cloud providers. Especially for clinics/medical institutes this represents a serious risk, since personal data has particular need for protection and is subject to a strict legal and regulatory framework. Therefore, in the next chapter, we focus on the term “Trusted Cloud” and specify what it implies in the scope of the TVB-Cloud project. After a short overview of existing definitions, we describe our core principles as well as derived tools and measures. Compliance to these principles and measures builds the basis for trust and consequently for the successful completion of the project.

In Chapter 3, the implementation strategy for the trusted cloud is presented. It is based on two different cloud solutions: a so-called “Public Community Cloud” and a “Secure Health Cloud”. After explanations of these solutions, the following section outlines the general approach to implement the Cloud. In section 3.3, we briefly draft the overall architecture, primarily highlighting relevant cloud services. Finally, the last section presents a short insight into related European infrastructure initiatives especially with respect to services planned to be implemented in TVB-Cloud. The conclusion will primarily expose open questions. Subsequently, the "Glossary" provides short definitions of several specialized terms used within the deliverable.

[1] https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

2. Trusted Cloud Principles

2.1. Definitions

There are many definitions for the term Trusted Cloud, perceived in various contexts. One of the major cloud providers - Microsoft - describes their infrastructure as Trusted Cloud and defines it on the basis of their four principles: security, privacy, compliance and transparency[2]. IBM similarly defines compliance, privacy and data security as the reasons why IBM Cloud would be a "trusted and proven enterprise cloud"[3]. IBM and Google for example even offer cloud solutions explicitly designed for healthcare[4][5][6].

These cloud providers provide organizational and standardized certifications following official audits conducted by third parties to verify compliance with their set of controls. These certificates serve to build trust. Google for example, lists the HITRUST Common Security Framework certificate amongst others, signalling compliance with ISO/IEC-27000 series and HIPAA standards.[7]

The German Federal Ministry for Economic Affairs and Energy (BMWi) used the above term in the scope of the technology program - Trusted Cloud[8]. This program was initiated to support the building of trust in cloud services, particularly in regard to medium-sized enterprises. Also in this context, certificates and standards were used to give users of cloud services an initial guidance and facilitate expansion into this technology[9] [10].

Funded by German Federal Ministry of Economics and Technology, one of the TrustedCloud projects was Cloud4Health[11]. In a consortium of five different partners, which included Fraunhofer SCAI, a cloud-based solution for secondary use of clinical data was developed. In short, a generic architecture was designed to extract structures and unstructured data from electric health recorded systems.[12]

2.2. Our Definition

In case of the TVB-Cloud project, we define the term "Trusted Cloud" on the basis of two core principles and related measures and tools. The two core principles "security" and "privacy" grow out of several goals envisaged to ensure information security and data protection.

---

[2] https://www.microsoft.com/en-ww/trust-center?market=af
[3] https://www.ibm.com/cloud/security
[4] https://www.ibm.com/cloud/healthcare
[5] A. Iyengar, A. Kundu, U. Sharma and P. Zhang, "A Trusted Healthcare Data Analytics Cloud Platform," *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, Vienna, 2018, pp. 1238-1249. doi: 10.1109/ICDCS.2018.00123
[6] https://cloud.google.com/solutions/healthcare/#securing-patient-data
[7] https://cloud.google.com/security/compliance/hitrust/
[8] https://www.bmwi.de/Redaktion/DE/Artikel/Digitale-Welt/trusted-cloud.html
[9] https://www.trusted-cloud.de/de/node/1083
[10] https://www.trusted-cloud.de/en/about-trusted-cloud
[11] http://cloud4health.de/projekt
[12] https://www.springer.com/de/book/9783319624570

One important thing to consider while characterizing the "Trusted Cloud" for the TVB-Cloud project is that the TVB-Cloud utilizes different kinds of data such as clinical, personal, pseudonymized, anonymized, research and open data. Open data in this context means data which anyone can access, use and share[13].

Due to the diversity in the mode of handling various data, it is very difficult to define a generic architecture. The TVB-Cloud project consists of diverse data providers, which means that, having an architecture congruent to the one used in Cloud4health is not feasible. Furthermore, the Cloud Infrastructure will not be restricted to one strictly secure environment. To begin with, work package (WP) 7 provides an open community cloud suitable for research purposes with open data (cf. chapter 3.1).

A challenge that arises in the context of specifying the TVB-Trusted-Cloud Infrastructure is the requirements mandated by the applicable European and national legal framework for the processing personal data, e.g. additional technical and organizational measures required for special categories of data namely data concerning health. During the course of the project, these regulations, depending upon the location of the partner sites, will be analysed in WP 2 and partly infrastructure solutions adhering these regulations are prepared.

Commercial cloud providers generate trust by their customers with auditable certifications to be compliant with the respective regulations (see for example AWS ISO Certifications and services to be compliant with ISO/IEC 270001:2013, 27017:2015,27018:2014, 9001:2015[14]). Research infrastructure providers in universities or research institutes follow technical, organisational and personnel aspects of information security (e.g. IT-Grundschutz [15]) compatible to ISO/IEC 27001 (e.g. Fraunhofer "Safety manual from- ISO27001), but they are not audited.

Most European research infrastructure providers are research institutions (e.g. Compute Provider Fraunhofer SCAI and Supercomputing Center Jülich) without such certification. For them it is difficult to provide verification. At first glance, this might appear to be a disadvantage compared to certified commercial cloud providers. In fact, this only leads to the need of building trust in other ways. On the one hand, the missing certification is compensated by the fact that public organisations and research institutes, compared to commercial cloud providers, are not interested in commercial use of user data.

On the other hand, explicit agreements (e.g. bilateral contracts) offer the opportunity to build the necessary trust between the various parties. Therefore, it is highly relevant for the TVB-Cloud to define and implement principles and security mechanisms that are agreed by the participating partners on a case-by-case basis. Furthermore, this approach makes it possible that not only the legal and regulatory framework are considered but also that for example individual provisions of the hospital data protection officers can be realized.

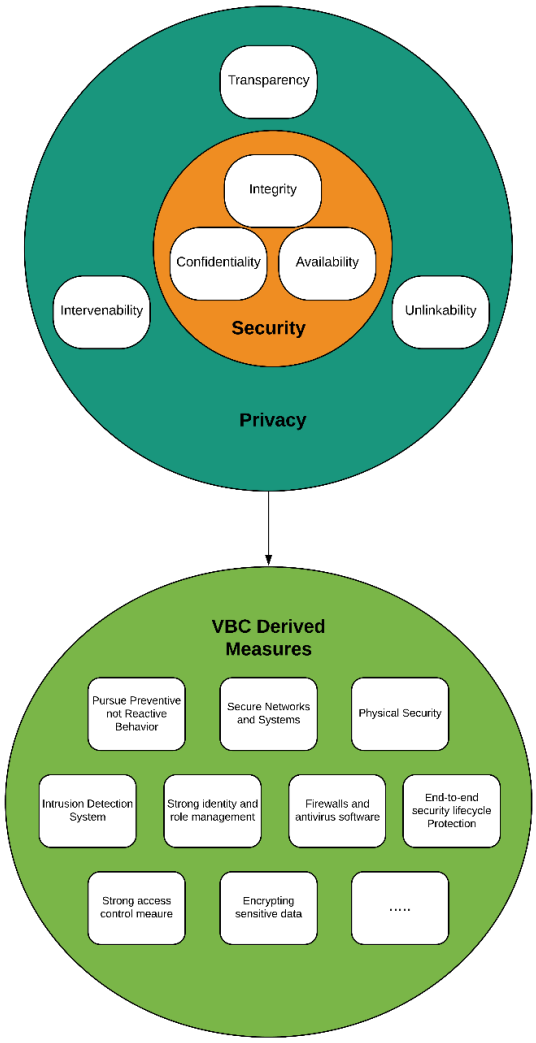
[13] https://www.europeandataportal.eu/elearning/en/module1/#/id/co-01
[14] https://aws.amazon.com/compliance/iso-certified/
[15] https://www.bsi.bund.de/EN/Topics/ITGrundschutz/itgrundschutz_node.html

2.3. Principles

When using outsourced cloud computing services, the transfer of responsibility and, control over information and system components lies with the cloud service provider. For the user this leads to a number of risks. Strong bilateral agreements are placed between individual partners of the project (e.g. a service provider like a High performance computing centre and a clinic) detailing the provisions relating to security and privacy being the key elements for a cooperation based on trust. In sum, “privacy” and “security” define the core principles for TVB-Trusted Cloud.

Confidentiality, integrity and availability, cumulated in the known CIA model[16], build the basis for the security principle. Confidentiality includes the implementation of measures that are designed to prevent sensitive information from reaching the wrong people, while making sure that the right people have access to it through a secure access management system. Integrity includes the integrity of systems and involves preserving the consistency, accuracy and trustworthiness of data throughout the entire lifecycle. It includes that data is securely retained and securely and timely destroyed at the end of a process. It also means that measures must be taken to ensure that data cannot be altered by unauthorized people. Availability means the provision of functioning systems and services as well as working communication channels.

[16] https://www.ibm.com/blogs/cloud-computing/2018/01/16/drive-compliance-cloud/



A data protection principle is transparency. Transparency implies that a user is always able to identify where his privacy-relevant data is stored, how it is processed, who is processing the data and who has access to it. Additionally, intervention by the user to the largest extent should be possible when it comes to data processing, which represents the aspect of Intervenability. Unlinkability is another aspiration. It means the inability of creating a relation between items that are constituted by a common purpose and context.

The essential difference between "security" and "privacy" can be summarized as follows: while security is of a rather technical nature, privacy is of a legal/regulatory and ethical nature. Consequently, it follows that privacy can only be ensured if security is guaranteed. While safeguarding security falls within the remit of WP7, fulfilling privacy is also dependent on results from WP2.

The following list presents a rough selection of possible measures and tools that will be supplemented on the basis of the requirements of the involved partners to create a basis of mutual trust.

Figure 1: Protection goals, trusted cloud principles and TVB-Cloud derived measures

1. Pursue preventive not reactive behavior
2. Provide intrusion detection systems
3. Use firewalls, antivirus software
4. Strong identity & role management
5. Strong access control measure
6. Encrypting personal sensitive data
7. Realize end-to end security-lifecycle protection of data
8. Create physical security
9. Ensure secure networks and systems
10. Enforce and be compliant to Information security policy and privacy policy
11. Data management: policies and controls.
12. Data quality assurance process

Figure 1 (modified from reference[17]) offers an overview of the protection goals, the Trusted Cloud principles and derived measures. It should be stressed that the protection goals and the two core principles apply for both, the provision of the open TVB-Cloud community solution as well as a secure cloud solution for the handling of sensitive data. Both solutions will be explained in more detail in chapter 3.1.

3. Implementation Strategy

3.1. TVB-Cloud Solutions

The TVB-Cloud Infrastructure will be built on a cloud architecture with services in a private cloud, external services hosted at different locations and managed by external entities, and front-end applications running on local clients. To structure the overall TVB-Cloud-Infrastructure, we introduce and define the terms “Public Community Cloud” and” Secure Health Cloud”.

From day one, WP7 starts providing services available in an open cloud infrastructure, which offers all kind of users the possibility to do research with non-sensitive data. For such a cloud solution, we use the term “Public Community Cloud”.

Public Community Cloud

Without going into too much detail the cloud-based environment will use EOSC resources and provide access to HPC services to support the overall community. One of the first steps is to port and host available applications of “The Virtual Brain”[18] to operate first services.

The management of federated identity and authorization will be based on existing EOSC services. Repositories delivering information about available data (meta- and real data) as well as providing information about access regulations, roles etc., will be setup in close cooperation with the other work-packages. Subsequently, we will provide services supporting data transfer and services enabling access to available compute and storage cloud resources. Additional platform services may be deployed or made available via this infrastructure.

Due to global and local regulations in case of data concerning health, its processing needs an implementation of a variety of technical security measures agreed on by the participating organisations. An infrastructure providing such security mechanism will be called “Secure Health Cloud”.

Secure Health Cloud

Parallel to the public Community Cloud, we build a prototypical environment including all relevant measures to fulfil security and privacy conditions for health data, accepted by the security and data officers of the individual participating clinics.

[17] HANSEN, Marit; JENSEN, Meiko; ROST, Martin. Protection goals for privacy engineering. In: *2015 IEEE Security and Privacy Workshops*. IEEE, 2015. S. 159-166.
[18] https://thevirtualbrain.org/tvb/zwei

An aspect to consider in the "Secure Health Cloud" are the applicable legal and regulatory frameworks in regard to the physical transfer of personal data, and in particular, data concerning health, e.g. health-records. Moreover, data subjects and healthcare entities must stay informed of where and how electronic personal data is moved or stored and physical controls may be necessary (refer GDPR). In this case, the secure cloud environment can be established locally at the clinics possessing own storage and HPC resources or having access to individual partner sites (on trust basis).

Security measures for this infrastructure strongly depend on results of WP 2 and regulations provided at partner sites regarding the handling of health data. The technical realization will be accompanied by individual risk assessments and service level agreements (SLAs).

The following presents three different scenarios to do computing with distributed sensitive data:

1. *Transfer data to single secure or trusted system for processing:* In this first scenario, anonymized or de-identified data is transferred to a HPC/Supercomputer site. There it is stored for pipeline processing. This typical scenario will be supported from the beginning to offer a quick solution for using supercomputer resources at Forschungszentrum Jülich. We will begin to develop a technical implementation with a pilot user.

2. *Compute where the data is and combine the aggregate of non-sensitive results for transfer:* In contrast to the first scenario, this "federated processing" will help us when transfer of data is not allowed by an institution in any way. This assumes that local compute resources are available and that the institution trusts a user and allows access to their local resources. The partner AMU and UNIG are potential candidates for such a solution. All relevant tools for this scenario are provided by the Cloud services (e.g. mapping of user credentials, secure transfer). An alternative here would be a mechanism providing a TVB-Cloud user the possibility to send analysis code to the sealed site.

3. Processing on demand by streaming data: In this scenario, data is streamed on-demand to a single secure system for processing purposes and the external resources only keep the data during processing. Data is never stored and all conceivable marks are removed. That means data sanitizing of all data related components, including VMs. Infrastructure supporting such a scenario has been developed by the Cloud4health project. The overall process can be monitored and controlled by the user from a clinic. The resource provider is a trusted compute site and the infrastructure concept needs to be signed by the user sites, as well as the compute provider. Even when the cloud provider site has no cloud and security certifications, it has to demonstrate that it is able to fulfil the agreed criteria. It is planned to build up such a dedicated infrastructure.

3.2. General Implementation Strategy

Since there are various parties involved in the provision of the Cloud solution (e.g. clinics, universities, HPC resource providers, EOSC e-infrastructure providers and other private cloud providers), there is no common technical solution available right at the beginning, which would be comparable to a "in-house" platform solution. Technical solutions and SLAs (based on "trusted principles") between partner sites have to go hand in hand. Here we are dependent on other work packages.

For this reason, WP7 pursues the flexible strategy of initially deploying an open so-called community cloud, which will be extended with security mechanisms to improve the security level and result in the final, trusted Cloud solution. The idea behind the community cloud is to create a cloud that is linked to the HPC resources provided by FZJ, the local cloud resources provided by SCAI and the European Open Science Cloud EOSC. This community cloud will provide all relevant and required services including all services of the TVB platform (c.f. D6.1.). Precisely, this means that in this scenario a user who is working with non-critical open data (e.g. medical data of mice) could use the public community cloud for his research. Virtual machines are provided for the tools developed by the other work packages of the project. Separate standalone tools like JupyterHubs for example equipped with the TVB library are provided for direct usage.

In parallel, all three scenarios (described in chapter 3.1) will be implemented in close cooperation with individual pilot users. The technical implementation of the first scenario will primarily be supported by FZJ. The solutions that were presented in the second and third scenario will be realized by SCAI in close cooperation with FZJ and the individual involved parties. The central issue in these scenarios is the development of the local gateways and staging areas in the private protected zones at the sites to connect internal with external cloud services.

Altogether, we will follow a flexible implementation strategy to deploy infrastructures offering for example a secured end-to-end infrastructure to process medical data that is at least anonymized and de-identified, and a community cloud for other data. The overall Cloud environment will be successively extended with infrastructural components and provide the opportunity (e.g. realized by providing VMs) for individual work packages to implement and configure TVB-Cloud components like e.g. a XNAT platform. For both solutions, the TVB-Cloud users and partners need to trust that all infrastructure providers (also of EOSC) are compliant to the predefined principles.

3.3. TVB-Cloud Services - resources available today

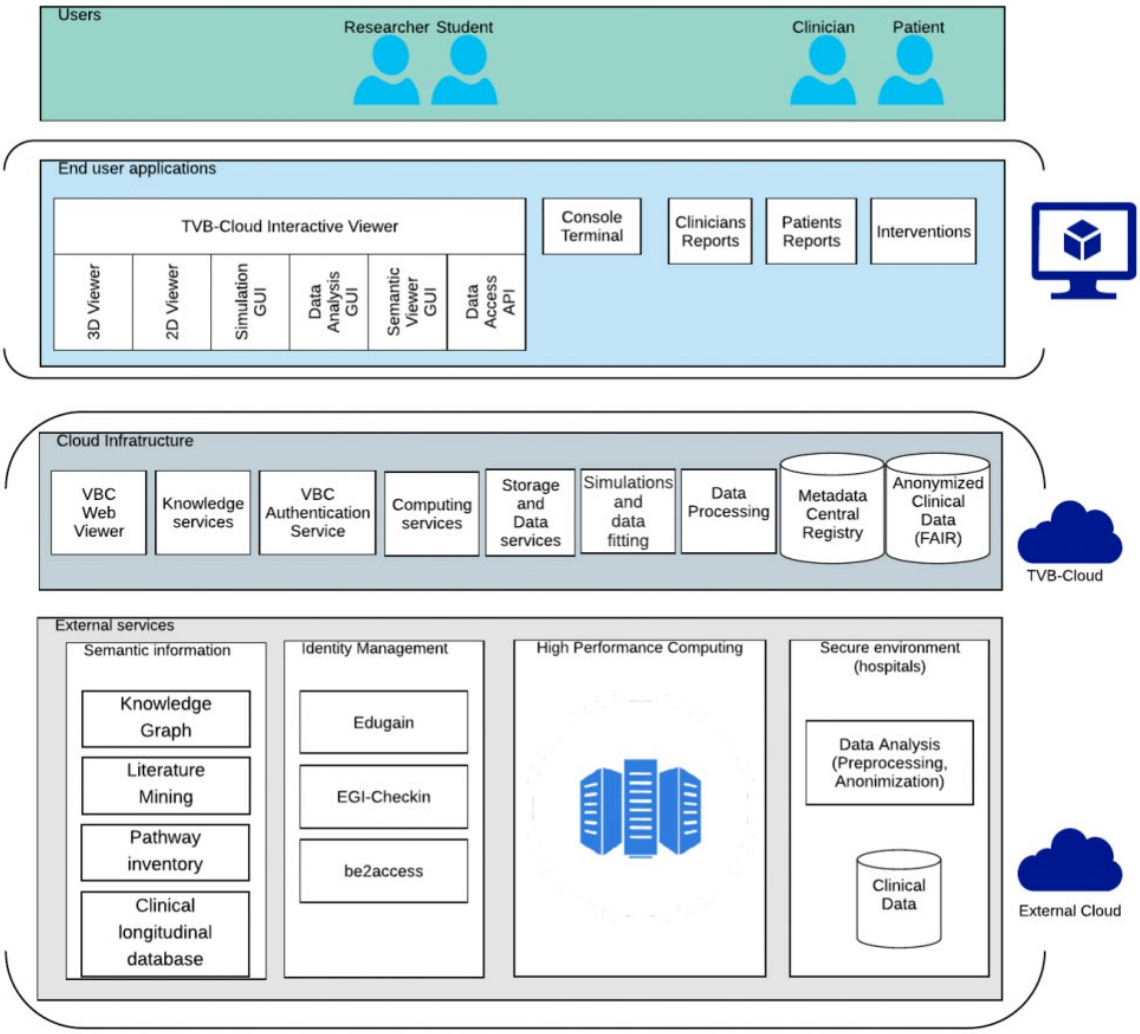


Figure 2: TVB-Cloud architecture

Figure 2, originating from deliverable D6.1, shows the currently envisioned services of the Cloud infrastructure. The cloud-based infrastructure and integration services will be implemented by Fraunhofer SCAI and Forschungszentrum Jülich/JSC (Jülich SupercomputerCenter). It will provide a secure and reliable platform for the TVB-Cloud services and the technical means to utilize supercomputing / HPC resources for compute-intensive workflows. As the realization of the platform is dependent on the results of the other work packages and the cloud-infrastructure architecture will be described in detail with the next deliverable, only existing basic functionalities that are definitely used are listed here.

Relevant services building the basis for TVB-Cloud are from the e-infrastructures for Collobarative Data Infrastructure EUDAT and the European Grid Infrastructure (EGI). Both are nowadays under the umbrella of the European Open Science Cloud via the EOSC-Hub project. JSC and SCAI are service providers and partners of EGI. Furthermore, JSC is strongly involved in EUDAT.

The HPC services for computation are locally provided by JSC and SCAI. JSC offers tier-0 systems included in PRACE (Partnership for Advanced Computing in Europe) and SCAI hosts smaller HPC clusters. Normally, the login nodes (the only externally accessible servers comprising the external connectivity for users) of the supercomputers at JSC are accessed via a SecureShell (SSH) connection. The HPC-specific solutions SLURM and PBS are used as the job scheduling and resource management utilities for accessing the compute nodes (backend servers running computations). The access is regulated by personal user accounts, which are created after a manual application at PRACE. The application will be peer reviewed and a quota of CPU hours is granted individually. Applications need to fulfil performance guarantees and the involved computational projects have to scale well over N nodes (parallelity). The jobs do not have internet access. In other words, the data needs to be staged (copied) in and out of the system before and after the computations.

The UNICORE[19] system allows seamless access to compute and data resources at the JSC site and will be extended with TVB-Cloud components when necessary. It also allows users to access supercomputing resources from participating sites. The users can create, submit and monitor supercomputer jobs using UNICORE as an interface. Secure data transfer from and to the cloud is a major priority in the TVB-Cloud project and the UNICORE gateway will be extended to fulfil appropriate security measures. The service is in operation and TVB-Cloud can rely on the following features:

- Data transfer & management: Stage data in/out of HPC environment, manage HPC storage
- Desktop Client and Portal (Web-)GUI
- RESTful API services
- Support for the PRACE Research Infrastructure
- HDFS Hadoop, Amazon S3 and CDMI Interfaces

The UNICORE Gateway supports the OAuth2/OpenID-Connect based federation of existing Identity Providers. In particular, the following Proxy services are supported and could be used for the integration in the Cloud (c.f. AARC Blueprint Architecture[20]):

- EGI Check-In: EGI check-in acts as a central hub server between Federated Identity Providers and EGI Service Providers. It allows users to select a preferred identity provider from multiple federated authentication sources.
- EOSC-Portal Check-in: EOSC portal check-in provides access to EOSC services through a central access portal. It allows various Federated Identity Providers to authenticate users in order to give them access to EOSC services.
- EUDAT B2ACCESS: It is an EUDAT provided authentication and authorization platform which can be integrated with any service. Users can login using different authentication methods like home organization Identity Provider, Google account and EUDAT ID. However, EUDAT ID is created by B2ACCESS on registration. So, in this case, B2ACCESS acts as an Identity Provider itself. It accepts Edugain, IGTF and social accounts (lower assurance level).

[19] https://www.unicore.eu/
[20] https://aarc-project.eu/architecture/

The use of ipython notebooks via JupyterHub is very common in research communities like the TVB-Cloud one. They will be made available for different contextual scenarios. An instance of JupytherHub for HPC services is available at JSC and will be made available for TVB-Cloud users. Other instances of JupyterHub, for example configured for training TVB tools are in preparation to be operational TVB-Cloud-services on EOSC/EGI as well as local cloud resources.

3.4. European infrastructure initiatives

Making it easier to access, share, and combine health and other "sensitive" datasets across different disciplines and locations is also a central concern of other initiatives under the umbrella of the EOSC, supporting FAIR data principles. As described in EU publication "Turning FAIR into reality"[21], "data should be Findable, Accessible, Interoperable and Reusable to the greatest extent possible. FAIR is a significant concept in its own right since it offers a set of principles to enhance the usefulness of data" and FAIR data is not necessarily open data. "The FAIR principles apply equally to data that remain restricted or internal to a given organization: data will be more usable and have greater value if they are FAIR". This means that also for the TVB-Cloud the FAIR data principles need to be considered to allow for the sharing of data. An example for an ESFRI[22] with a FAIR culture is the Life Science Community ELIXIR (the European research infrastructure for life science information). To provide the necessary technical framework that enables the implementation of FAIR principles in the scope of the TVB-Cloud, existing solutions will be evaluated. The technical implementations that will be realized in the Cloud will be based on SLA agreements arranged by WPs 1 and 2.

Often mentioned in the context of research with sensitive data are the Nordic countries Norway, Denmark, Finland and Sweden. In the years between 2012 and 2014 they have built national secure servers for sensitive data[23] [24] in research and performed projects (NeIC/Tryggve) to facilitate access to sensitive data services across borders[25]. It is important to mention here that these countries benefit from a long history of technical and scientific collaboration, similar legislations, as well as geographical and economical proximity. Many of the Nordic actors are also involved in EOSC related initiatives. For example, in the EOSC-hub project, that aims to build a single contact point in the scope of data-driven research and includes activities referring to the handling of sensitive data.

In the scope of the TVB-Cloud, we will monitor developments in this direction and assess whether possible solutions could also be reasonably implemented in our TVB-Cloud solution. EOSC-hub project e.g. aims to integrate ePouta (secure servers for sensitive data in research in Finland) and the EOSC service B2SHARE. More concretely they will work out a Secure B2SHARE environment at the supercomputing centre CSC in Finland, which should be able to store sensitive data & make non-sensitive metadata searchable while ensuring controlled access based on permissions decided by the data owner. The secure B2SHARE is an extension of the service provided by EUDAT and in operation at

[21] https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283
[22] https://www.esfri.eu
[23] https://research.csc.fi/epouta
[24] https://www.uio.no/english/services/it/research/sensitive-data/
[25] https://indico.neic.no/event/18/contributions/249/attachments/80/139/1545_NEIC2019-copenhavn-sensitive_data_Francesca_Iozzi.pdf

FZJ. It is a good practical example of an EOSC service that will be implemented in the TVB-Cloud infrastructure.

The developments and results from the EOSC-hub initiative working on data anonymization in TSD (secure servers for sensitive data in research in Norway) are a foundation for data protection measures and tools in the TVB-Cloud.

Other projects related to the topic of healthcare are for example the EOSC-Life project[26] and the Clinical Research Initiative for Global Health[27]. The EOSC-Life project aims to connect 13 biological and medical ESFRI research infrastructures to create an open collaborative space for all aspects of life science research and all life science domains. The collaboration aims to provide a database following the FAIR principles and reusable tools and workflows. The Clinical Research Initiative for Global Health was launched in late 2016, aiming to support and optimize international cooperation on clinical research. It expands upon the European Clinical Research Infrastructure Network (ECRIN) for establishing global standards in clinical research and encouraging global communication. While this project rather relates to other work packages of TVB-Cloud project, WP7 will evaluate the utilized tools and workflows, which may be used in the Cloud and thus need to be supported by the infrastructure.

Since EOSC is built up in the next years, WP7 needs to pursue upcoming technical solutions and newly available services on the way to a complete EOSC integrated cloud.

[26] https://www.ecrin.org/news/press-release-eosc-life-project-develops-open-collaborative-space-digital-biology-europe
[27] https://crigh.org/

4. Conclusions

The present deliverable presents a flexible strategy applied to implement the trusted Cloud. For this reason, initially the two core concepts “security” and “privacy” were determined and several measures presented, ensuring the adequate handling of sensitive data. These build the foundation for trust between the individual cloud providers and other partners. The terms “Public Community Cloud” and “Secure Health Cloud” were differentiated and finally set into relation to show the big picture of the trusted Cloud. Subsequently, the Cloud was characterized by presenting basic services and access to HPC. The last chapter shortly provided the status quo relating cloud services for healthcare in European Science Clouds.

The development and implementation of the trusted Cloud however depends on the individual requirements of all involved partners and the output of several other work packages. Currently, there are still some open questions. Hence, WP7 needs to be fully informed about detailed requirements and conditions for the handling for sensitive data by all involved partners, particularly clinics. This will e.g. enable the work package to assess on what level of encryption and security the data protection officers will accept the transfer and authorize the processing of sensitive data on external resources. For this reason, detailed and suitable SLAs and conventions need to be agreed on and be contractually bound by the TVB-Cloud project or bilaterally between partners. It will be necessary input for WP7 to develop an infrastructure solution.

Another question that remains open is a clearer demarcation of the TVB-platform, not including the different infrastructural components that could be locally installed in individual clinics of the Cloud environment. An important question in this context is where to position the metadata service providing information about individual data sets as well as access regulations, user roles, etc. One possibility could be to set up a central and public cloud service located at one of the cloud providers. Another option would be that each individual site operates their own metadata service. In this case, a timely and exact synchronization or federation would be necessary. Also, the integration into the TVB-platform could be possible. Furthermore, the handling of credentials needs to be considered. A proper mapping between e.g. credentials valid within a clinical environment and the credentials valid for the infrastructures provided by other involved players like EOSC, JSC, SCAI etc. must be implemented. The next step will be to discuss these questions with involved partners to find and implement suitable solutions.

Complementary to this document, D7.4.1 will focus on the TVB-Cloud architecture and EOSC services, (M12) while in D7.2 we will define more detailed technical security measures (M14). In month 36 of the project we will provide a new version of this deliverable highlighting the implementation strategy applying EOSC cloud principles.

5. Glossary

TVB-Cloud	The Virtual Brain Cloud Project
Cloud	The Virtual Brain Cloud
EOSC	European Open Science Cloud
XNAT	XNAT is an open source imaging informatics platform developed by the Neuroinformatics Research Group at Washington University.
VM	Virtual machine
JupyterHub	A JupyterHub serves Jupyter notebook (python) for multiple users
Anonymization	Irreversible removal of the link between the individual and his or her medical record data to the degree that it would be virtually impossible to reestablish the link.
Pseudonymization	Substitutes the identity of the data subject in such a way that additional information is required to re-identify the data subject.
De-identification	Removal or replacement of personal identifiers so that it would be difficult to reestablish a link between the individual and his or her data
Encryption	Process of transforming data using an algorithm to make it unreadable except to the intended recipient
FAIR	c.f. https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_0.pdf
Tier-0	Structured provision of European HPC facilities • Tier-0: European Centres • Tier-1: National Centres • Tier-2: Regional/University Centres
eduGAIN	eduGAIN is an international interfederation service interconnecting research and education identity federations.
IGTF	Interoperable Global Trust Federation supporting distributed IT infrastructures for research
SCAI	Fraunhofer institute for algorithms and scientific computing
FZJ	Forschungszentrum Jülich
JSC	Jülich Supercomputing Centre (part of FZJ)