Panacea

People-centric cybersecurity in healthcare

Project Title	Protection and privAcy of hospital and health iNfrastructures with smArt Cyber sEcurity and cyber threat toolkit for dAta and people
Project Acronym	PANACEA
Project Number	826293
Type of instrument	Research and Innovation Action
Topic	SU-TDS-02-2018
Starting date of Project	01/01/2019
Duration of the project	36
Website	www.panacearesearch.eu

D2.4 “Secure Information Sharing, interconnectivity, cloud, authentication and interoperability aspects”

Work Package	WP2 – Research on advanced thread modelling, human factors, resilient response and secure interconnectivity
Lead author	Stelios Sfakianakis (FORTH)
Contributors	Silvia Bonomi (UROME), Claude Bauzou (IDEMIA), Aristodemos Pneymatikakis (iSPRINT), Maria Seretaki (iSPRINT), Claudio Ciccotelli (UROME), E. Spanakis (FORTH), V. Sakkalis (FORTH)
Peer reviewers	Matteo Merialdo (RHEA), Aristodemos Pneymatikakis (iSPRINT)
Version	V1.0
Due Date	31/03/2020
Submission Date	31/03/2020

Dissemination Level:

X	PU: Public
	CO: Confidential, only for members of the consortium (including the Commission)
	EU-RES. Classified Information: RESTREINT UE (Commission Decision 2005/444/EC)
	EU-CON. Classified Information: CONFIDENTIEL UE (Commission Decision 2005/444/EC)
	EU-SEC. Classified Information: SECRET UE (Commission Decision 2005/444/EC)

The work described in this document has been conducted within the PANACEA project. This project has received funding by the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 826293.

Version History

Revision	Date	Editor	Comments
0.0.1	14/1/2020	E. G. Spanakis (FORTH)	Owner/Creator
0.0.3	19/2/2020	A. Pneymatikakis, M. Seretaki (iSPRINT)	
0.0.4	22/2/2020	S. Sfakianakis (FORTH)	
0.1.0	4/3/2020	S. Sfakianakis (FORTH)	
0.2.0	9/3/2020	S. Bonomi, C. Ciccotelli (UROME)	
0.5.0	16/3/2020	S. Sfakianakis (FORTH)	
0.6.0	17/3/2020	S. Sfakianakis (FORTH)	
0.6.1	23/3/2020	A. Pneymatikakis (iSPRINT)	Revised section 5.2
0.7.0	24/3/2020	S. Bonomi (UROME)	Revised section 8 and 9
0.9.0	26/3/2020	S. Sfakianakis (FORTH)	
1.0.0	30/3/2020	S. Sfakianakis (FORTH), E. Spanakis (FORTH)	Final comments integration

List of Contributors

The list of contributors to this deliverable are presented in the following table:

Section	Author(s)
Section 8 & 9	UROME
Section 5.2	iSPRINT
Section 6	IDEMIA
Sections 4, 5.1, & 7	FORTH

Keywords

Information Sharing, Interoperability, Authentication, Biometrics, IoT, Standards, Blockchain

Disclaimer

Executive Summary

The aim of this deliverable is to provide information, guidelines, established practices and standards, and an extensive evaluation on new and promising technologies for the implementation of a Secure Information Sharing platform for health-related data. Information sharing is one of the core scenarios that the PANACEA toolkits target, and therefore it’s important to gather any relevant information that will facilitate the design, implementation, and deployment of the relevant platform. In this deliverable we focus strictly on the technical aspects and specifically on the sharing of health information. The reported information is the outcome of the following tasks:

- Task 2.5 “Innovative techniques for Secure Information Sharing and Cloud solutions” which aims at investigating alternative storage architecture supporting the Secure Information Sharing, possibly combining together different technologies to support Confidentiality, Integrity and Availability (CIA) properties for stored data. Especially, Blockchain technologies have gathered a lot of interest the last couple of years and therefore are one of the main areas that this task focuses on.
- Task 2.6 “Research on secure interconnectivity, authentication, and interoperability aspects” which focuses on the implementation of secure and interoperable mechanisms and relevant services to support the sharing of clinical information in a generic setting.

This document presents also additional research, evidence and argumentation regarding the following:

- With respect to clinical information sharing we argue that it is important to consider aspects such as the patients’ consent, the compliance with health IT standards for improved interoperability, and focus on the data protection and privacy as imposed by essential European regulation such as GDPR.
- For authentication, the use of biometrics, and more specifically the face characteristics, is important but especially when combined with a hardware token for added security and user friendliness.
- Cloud computing is now used everywhere and provides an important set of features, such as adaptive scalability, performance, and benefits from business perspective. Especially for the sharing of clinical information, cloud can be very advantageous especially in cases where central repositories or central coordination are needed. But organizations should also be wary about the data protection, privacy, and access control mechanisms that should be in place, either offered by the cloud provider or built in house, in order to properly handle sensitive data and comply with regulations such as GDPR.
- Blockchain is a highly interesting technology that could be put in good use in information sharing, more specifically to support data integrity verification and data auditability. Nevertheless, there are some major issues to be resolved, such as the compliance with GDPR’s “right to be forgotten” requirement which, unless the blockchain implementation is adapted, requires the deletion of data from the blockchain when the data owner asks to do it and this is not feasible, by design, in the “traditional” Blockchain implementations. Furthermore, there is also the additional complexity following by the huge heterogeneity of medical data (i.e., text, images, etc.) that do not fit exactly to the original design of Blockchain.

Table of Contents

List of figures

List of tables

1. Introduction

This deliverable reports on the outcome of Tasks 2.5 “Innovative techniques for Secure Information Sharing and Cloud solutions” and 2.6 “Research on secure interconnectivity, authentication, and interoperability aspects”. In these tasks, the focus is on the sharing of clinical information across systems, organizations, or even countries in order to support better healthcare and health delivery processes. Task 2.6 performed an analysis of existing solutions to support interconnectivity and sharing of data. Existing Information Technology (IT) solutions as long as specific standards and broadly adopted specifications in the clinical domain were considered and documented in this deliverable. On the other hand, Task 2.5 studied a range of deployment options and novel technologies like Blockchain for supporting the storage and management of the shared information.

1.1 Purpose

The purpose of this document is to provide detailed information regarding the following:

- Advanced and novel implementation strategies for a state-of-the-art information sharing environment
- The high-level requirements for the transfer of data between different health care organizations or cross-border.
- The available options and existing technologies to support the secure interconnectivity and trust between IT systems participating in a sharing-data “virtual community” (Health Information Organization).
- The different stakeholders and use cases in information sharing scenarios
- The standards, guidelines, and interoperability specifications for implementing a common understanding and integration in the sharing of clinical information
- The use of cloud computing and prospectively more advanced technologies such as Blockchain.
- User authentication and identification using biometrics and multiple factors

1.2 Quality assurance

1.2.1 Quality criteria

The QA in the PANACEA project relies on the assessment of a work product (i.e. deliverable) according to lists of QA checks (QA checklists – [QAPeer]) established with the QAM, validated at a project management level and centralized in the [PMP].

For the purpose of the QA of this deliverable, it has been assessed according the following checklists:

- PEER REVIEW (PR) QA CHECKLIST [QAPeer]: this deliverable is a report, it then requires a proper peer review according to the checks defined in this checklist;

1.2.2 Validation process

For the final validation of work products (i.e. deliverables) within the PANACEA project, a final QA review process MUST be used before the issuing of a final version. This QA validation process follows the Quality Review Procedure established with the QAM and validated at project management level in order to guarantee the high-quality level of work products and to validate its adequacy according to the defined quality criteria chosen and defined for each deliverable. The Quality Review Procedure itself and the selection of the QA Review Committee are described in the [PMP]. The QA validation process is scheduled in the QA Schedule [QASchedule] managed by the QAM.

1.3 Structure of the document

This document is structured as follows:

- SECTION 1. Introduction
- SECTION 2. Applicable and Reference Documents
- SECTION 3. Glossary of Acronyms
- SECTION 4. Information sharing in Healthcare
- SECTION 5. Advanced information sharing scenarios
- SECTION 6. User authentication protocols and identification
- SECTION 8. Secure Information Sharing: Deployment options
- SECTION 9. Evaluation of emerging technical solutions
- SECTION 10. Conclusions

2. Applicable and Reference Documents

2.1 Applicable Documents (ADs)

The following documents contain requirements applicable to the generation of this document:

Reference	Document Title	Document Reference	Version	Date
[PMP]	PANACEA Project Manager Plan		0.5	01/01/2019
[QAPeer]	PANACEA Peer Review QA Checklist		0.5	01/01/2019
[QAReqs]	PANACEA Requirements Review QA Checklist		0.5	01/01/2019
[QASchedule]	PANACEA QA Schedule		0.5	01/01/2019

Table 1: Applicable Documents

2.2 Reference Documents (RDs)

The following documents have been consulted for the generation of this document:

Reference	Document Title	Document Reference	Version	Date
[Aguilera11]	Marcos Kawazoe Aguilera, Idit Keidar, Dahlia Malkhi, Alexander Shraer. Dynamic atomic storage without consensus. J. ACM 58(2): 7:1-7:32 (2011)	https://doi.org/10.1145/1944345.1944348		2011
[Avizienis04]	Algirdas Avizienis, Jean-Claude Laprie, Brian Randell, Carl E. Landwehr. Basic Concepts and Taxonomy of Dependable and Secure Computing. IEEE Trans. Dependable Sec. Comput. 1(1): 11-33 (2004)	https://doi.org/10.1109/TDSC.2004.2		2004
[**Bigtable]**	F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A. Wallach, M. Burrows, T. Chandra, A. Fikes, and R. E. Gruber. Bigtable: A Distributed Storage System for Structured Data. ACM Trans. Comput. Syst. 26,	https://doi.org/10.1145/1365815.1365816		2008
[Bourquard17]	Bourquard, Karima; Orlova, Anna; Parisot, Charles. "Understanding	http://library.ahima.org/doc?oid=302159		2017

Reference	Document Title	Document Reference	Version	Date
	User Needs for Interoperability: Defining Use Cases in eHealth" Journal of AHIMA 88, no.6 (June 2017): 42-45.			
[Chockler01]	Gregory V. Chockler, Idit Keidar, Roman Vitenberg. Group communication specifications: a comprehensive study. ACM Comput. Surv. 33(4): 427-469 (2001)	https://doi.org/10.1145/503112.503113		2001
[Cock10]	Cock, P. J., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic acids research, 38(6), 1767-1771.	https://doi.org/10.1093/nar/gkp1137		2010
[Dolin01]	Dolin, R et al. The HL7 clinical document architecture. Journal of the American Medical Informatics Association, 8(6), 552-569.	https://doi.org/10.1136/jamia.2001.0080552		2001
[Dudley10]	Dudley, et al. "Translational bioinformatics in the cloud: an affordable alternative." *Genome medicine* 2.8 (2010): 51.			2010
[DynamoDB]	Amazon DynamoDB	https://aws.amazon.com/dynamodb		
[eHN16]	eHealth Network Guideline on the electronic exchange of health data under Cross-Border Directive 2011/24/EU Release 2 Patient Summary for unscheduled care	https://ec.europa.eu/health/sites/health/files/ehealth/docs/ev_20161121_co10_en.pdf		2016
[ENISA09]	ENISA, Cloud Computing Risk Assessment	https://www.enisa.europa.eu/publications/cloud-computing-risk-assessment		2009
[Estelrich14]	Estelrich, A., Solbrig, H., Cangioli, G., Melgara, M., & Chronaki, C. European Patient Summary Guideline and Continuity of Care Document: A	https://ieeexplore.ieee.org/abstract/document/7043084		2014

Reference	Document Title	Document Reference	Version	Date
	Comparison. InComputing in Cardiology Conference (CinC), 2014 (pp. 481-484). IEEE			
[Ethereum2014]	Wood, Gavin. Ethereum Project Yellow Paper, Ethereum: A secure decentralised generalised transaction ledger,	http://paper.gavwood.com/		2014
[EU11]	Directive 2011/24/EU of the European Parliament and of the Council of 9 March 2011 on the application of patients' rights in cross-border healthcare	http://data.europa.eu/eli/dir/2011/24/oj		2011
[EU12]	Regulation (EU) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation	http://data.europa.eu/eli/reg/2012/1025/2015-10-07		2012
[EU15]	Commission Decision (EU) 2015/1302 of 28 July 2015 on the identification of 'Integrating the Healthcare Enterprise' profiles for referencing in public procurement	http://data.europa.eu/eli/dec/2015/1302/oj		2015
[EU16]	Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)	http://data.europa.eu/eli/reg/2016/679/oj		2016
[EU19]	COMMISSION RECOMMENDATION of 6.2.2019 on a European Electronic Health Record exchange format	http://data.europa.eu/eli/reco/2019/243/oj		2019
[ElEmam13]	El Emam, K.. Guide to the de-identification of personal health information	https://doi.org/10.1201/b14764		2013
[Eckman07]	Barbara A. Eckman, Craig A. Bennett, James H. Kaufman, and Jeffrey W. Tenner. Varieties of interoperability in the	https://doi.org/10.1147/sj.461.0019		2007

Reference	Document Title	Document Reference	Version	Date
	transformation of the health-care information infrastructure			
[Fichman11]	Fichman, R. G., Kohli, R., & Krishnan, R. (Eds.). Editorial overview—the role of information systems in healthcare: current research and future trends	https://doi.org/10.1287/isre.1110.0382		2011
[FIDO]	FIDO Specifications	https://fidoalliance.org/specifications/		2020
[Fitbit20]	Fitbit	https://www.fitbit.com/eu/home		2020
[Galbally17]	Galbally, J., Coisel, I., & Sanchez, I. A new multimodal approach for password strength estimation—Part I: Theory and algorithms. IEEE Transactions on Information Forensics and Security, 12(12), 2829-2844.	https://doi.org/10.1109/TIFS.2016.2636092		2017
[Gandomi15]	Gandomi, A, and Murtaza H. "Beyond the hype: Big data concepts, methods, and analytics." *International journal of information management* 35.2 (2015): 137-144.	https://doi.org/10.1016/j.ijinfomgt.2014.10.007		2015
[Gartner]	Gartner Glossary: Big data	https://www.gartner.com/en/information-technology/glossary/big-data		2020
[Healthentia2020]	Driving Real World Outcomes in clinical research and patient care	https://healthentia.com/		2020
[HEART]	Health Relationship Trust	https://openid.net/wg/heart/		2019
[HGP19]	Reaching the Healthcare Mainstream: Wearables and Remote Patient Monitoring Market Map	https://hitconsultant.net/2019/06/21/wearables-and-remote-patient-monitoring-market-map/		2019
[HyperledgerFabricProj]	Linux Foundation. Hyperledger Fabric Project	https://www.hyperledger.org/projects/fabric		2020
[HyperledgerFabricArch]	Linux Foundation. Hyperledger Fabric Architecture	http://hyperledger-fabric.readthedocs.io/en/release-1.1/architecture.html		2017
[IHE-IT]	IHE IT Infrastructure Technical Framework	https://www.ihe.net/resources/technical_frameworks/#IT	Revision 16	July 2019
[IHE-MHDS]	IHE IT Infrastructure Technical Framework Supplement, Mobile Health Document Sharing (MHDS)	https://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_Suppl_MHDS_Rev2-0_PC_2020-03-05.pdf	Revision 2	March 2020
[IHE-Share]	IHE IT Infrastructure White Paper, Health	https://www.ihe.net/Technical_Framework/upload/IHE_I		2012

Reference	Document Title	Document Reference	Version	Date
	Information Exchange: Enabling Document Sharing Using IHE Profiles	TI_White-Paper_Enabling-doc-sharing-through-IHE-Profiles_Rev1-0_2012-01-24.pdf		
[ISA19]	Office of the National Coordinator for Health IT. 2019 Interoperability Standards Advisory	https://www.healthit.gov/isa/sites/isa/files/inline-files/2019ISAReferenceEdition.pdf		2019
[Kolodner08]	Robert M. Kolodner, Simon P. Cohn, and Charles P. Friedman. Health Information Technology: Strategic Initiatives, Real Progress	https://doi.org/10.1377/hlthaff.27.5.w391		2008
[Kuo11]	Kuo MH, Opportunities and Challenges of Cloud Computing to Improve Health Care Services, J Med Internet Res 2011;13(3):e67	https://doi.org/10.2196/jmir.1867		2011
[Larobina14]	Larobina, M., Murino, L. Medical Image File Formats. *J Digit Imaging* **27,** 200–206 (2014).	https://doi.org/10.1007/s10278-013-9657-9		2014
[Leinonen10]	Leinonen, R., Sugawara, H., Shumway, M., & International Nucleotide Sequence Database Collaboration. (2010). The sequence read archive. *Nucleic acids research*, *39*(suppl_1), D19-D21.	https://doi.org/10.1093/nar/gkq1019		2011
[LondonCN18]	Implementation of FreeStyle Libre® prescribing guidance across the NHS in London,	http://www.londonscn.nhs.uk/wp-content/uploads/2018/05/dia-London-FreeStyle-Libre-Implementation-Guidance-for-London052018.pdf		
[Michalas14]	Michalas et al. Security aspects of e-Health systems migration to the cloud. 2014 IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom)	https://doi.org/10.1109/healthcom.2014.7001843		2014
[Miller19]	Jen A. Miller, Healthtechmagazine. How Ochsner Health Uses the Apple Watch to Keep Patients Healthy,	https://healthtechmagazine.net/article/2019/12/how-ochsner-health-uses-apple-watch-keep-patients-healthy		2019
[Mobihealthnews 19]	Study: Proteus' digital pill system accurate, improved adherence	https://www.mobihealthnews.com/news/north-america/study-proteus-digital-pill-system-accurate-		2019

Reference	Document Title	Document Reference	Version	Date
	among California TB patients	improved-adherence-among-california-tb		
[Nakamoto2009]	S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2009			2009
[NIST11]	NIST, Guidelines on Security and Privacy in Public Cloud Computing	https://www.nist.gov/publications/guidelines-security-and-privacy-public-cloud-computing		2011
[NIST-auth]	NIST Special Publication 800-63B Digital Identity Guidelines	https://pages.nist.gov/800-63-3/sp800-63b.html		June 2017
[NIST-cloud]	The NIST Definition of Cloud Computing	https://csrc.nist.gov/publications/detail/sp/800-145/final		2011
[NIST-SOFA]	The Strength of Function for Authenticators - Biometrics (SOFA-B) Discussion Draft	https://pages.nist.gov/SOFA/SOFA.html	Draft	2020
[OAUTH]	J. Richer, A. Sanso. OAuth 2.0 in Action, Manning	ISBN: 9781617293276		2017
[ONC19]	A Shared Nationwide Interoperability Roadmap version 1.0	https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf		2019
[OTA95]	Office of Technology Assessment (OTA). Bringing Health Care Online: The Role of Information Technologies			1995
[Pereira19]	Pereira, C., Mesquita, J., Guimarães, D., Santos, F., Almeida, L., & Aguiar, A. Open IoT architecture for continuous patient monitoring in emergency wards. Electronics, 8(10), 1074.	ttps://doi.org/10.3390/electronics8101074		2019
[Perlman16]	Perlman, R., Kaufman, C., & Speciner, M. Network security: private communication in a public world	ISBN-13: 978-0130460196		2016
[Pharmavoice19]	Wearables In Clinical Trials	https://www.pharmavoice.com/article/2019-03-wearables/		2019
[Proteus20]	Proteus Discover	https://www.proteus.com/discover/		
[ReEIF15]	eHealth Network: Refined eHealth European Interoperability Framework	https://ec.europa.eu/health/sites/health/files/ehealth/docs/ev_20151123_co03_en.pdf		2015
[RFC6749]	D. Hardt. The OAuth 2.0 Authorization Framework	https://tools.ietf.org/html/rfc6749		2012
[RFC7525]	Sheffer, Yaron, Ralph Holz, and Peter Saint-Andre.	https://tools.ietf.org/html/rfc7525		May 2015

Reference	Document Title	Document Reference	Version	Date
	Recommendations for secure use of transport layer security (TLS) and datagram transport layer security (DTLS)			
[Rosenthal10]	Rosenthal, et al. "Cloud computing: a new business paradigm for biomedical information sharing." Journal of biomedical informatics 43.2 (2010): 342-353.	https://doi.org/10.1016/j.jbi.2009.08.014		2010
[SAML]	Security Assertion Markup Language (SAML) V2.0 Technical Overview	http://docs.oasis-open.org/security/saml/Post2.0/sstc-saml-tech-overview-2.0.html		2008
[Samsung20]	Samsung Collaboration with Kaiser Permanente Delivers Improvements in Home-Based Cardiac Rehabilitation	https://news.samsung.com/us/samsung-us-integrated-health-system-collaboration-home-based-cardiac-rehabilitation-results/		2020
[Staderini18]	Mirko Staderini, Enrico Schiavone, Andrea Bondavalli. A Requirements-Driven Methodology for the Proper Selection and Configuration of Blockchains. SRDS 2018: 201-206	https://doi.org/10.1109/SRDS.2018.00031		2018
[Technavio18]	Global Smart Healthcare Market 2018-2022	https://www.technavio.com/report/global-smart-healthcare-market-analysis-share-2018/		2018
[Zheng14]	Xiaochen Zheng, Joaquín Ordieres-Meré Development of a human movement monitoring system based on wearable devices,	https://www.researchgate.net/publication/276312289_Development_of_a_human_movement_monitoring_system_based_on_wearable_devices		
[Zhou10]	Zhou et al. "Services in the cloud computing era: a survey," in 4th International Universal Communication Symposium (IUCS), IEEE, Shanghai	https://doi.org/10.1109/IUCS.2010.5666772		2010
[UMA]	Maciej Machula, Justin Richer. User-Managed Access (UMA) 2.0 Grant for OAuth 2.0 Authorization	https://docs.kantarainitiative.org/uma/wg/rec-oauth-uma-grant-2.0.html	2.0	2018
[UMLS]	National Library of Medicine (US) UMLS reference manual. Bethesda: National Library of Medicine; 2009.	https://www.ncbi.nlm.nih.gov/books/NBK9676/		2009

Reference	Document Title	Document Reference	Version	Date
[Velsen16]	Lex van Velsen, Hermie Hermens, Wendy Oude-Nijeweme d’Hollosy. A Maturity Model for Interoperability in eHealth	https://doi.org/10.1109/HealthCom.2016.7749533		2016
[WS-Fed]	Web Services Federation Language (WS-Federation)	http://docs.oasis-open.org/wsfed/federation/v1.2/cd/ws-federation-1.2-spec-cd-01.pdf	1.2	2008

Table 2: Reference Documents

3. Glossary of Acronyms

Acronym	Description
2FA	Two Factor Authentication
API	Application Programming Interface
ATNA	Audit Trail & Node Identification
BPPC	Basic Patient Privacy Consents
CA	Certificate Authority
CDA	Clinical Document Architecture
CDISC	Clinical Data Interchange Standards Consortium
CDSS	Clinical Decision Support System
DICOM	Digital Imaging and Communications in Medicine
DPA	Data Processing Activity
eHDSI	eHealth Digital Service Infrastructure
EHR	Electronic Health Records
EIF	European Interoperability Framework
EMR	Electronic Medical Record
FAR	False Acceptance Rate
FHIR	Fast Healthcare Interoperability Resources
FIDO	Fast Identity Online
FRR	False Rejection Rate
GDPR	General Data Protection Regulation
HCO	Health Care Organization
HIO	Health Information Organization
HIT	Healthcare Information Technology
HITSP	Healthcare Information Technology Standards Panel
HL7	Health Leven Seven
IHE	Integrating the Healthcare Enterprise
IdP	Identity Provider
IoT	Internet of Things
ISO	International Standards Organization
IT	Information Technology
ITI	IHE IT Infrastructure
LOINC	Logical Observation Identifiers Names and Codes
MS	Member State
NCPeH	National Contact Point for eHealth
OASIS	Organization for the Advancement of Structured Information Standards
ONC	Office of the National Coordinator

OSI	Open Systems Interconnection
PACS	Picture Archiving and Communication System
PDQ	Patient Demographics Query
PIX	Patient Identifier Cross Referencing
RFD	Retrieve Form for Data Capture
RIM	Reference Information Model
RIS	Radiology Information System
SAML	Security Assertion Markup Language
SDO	Standards Developing Organization
SISP	Secure Information Sharing Platform
SNOMED	Systematized Nomenclature of Medicine
SNOMED-CT	Systematized Nomenclature of Medicine – Clinical Terms
SSO	Single Sign-On
TLS	Transport Layer Security
TTP	Trusted Third Party
U2F	Universal 2nd Factor
UMA	User-Managed Access
UMLS	Unified Medical Language System
XCA	Cross-Community Access
XDR	Cross-Enterprise Document Reliable Interchange
XDS	Cross Enterprise Document Sharing
WP	Work Package
WS	Web Service

Table 3. Table of acronyms

4. Information Sharing in Healthcare

Information Technology (IT) has long been identified as a cornerstone for the efficient, costless, timely, and reliable health care delivery [OTA95, Kolodner08]. The availability of health care information and patient records in digital form facilitates the persistence and posterity of valuable information and greatly support the decision-making process and even the extraction of new knowledge both at the individual and at the population levels.

Paraphrasing the famous words by John Donne, "*no IT system is an island, entire of itself*". In a highly connected world where geographic boundaries have largely eliminated and people can freely move between cities, states, countries, or continents, the requirement for two different information systems to exchange a person's clinical data or medical history becomes vital and persistent. Sharing health information (or Health Information Exchange – HIE) through electronic means greatly improves the cost, quality, and patient experience of the healthcare delivery. In simple terms[1]:

> *Health Information Exchange allows health care professionals and patients to appropriately access and securely share a patient's medical information electronically.*

To better secure the IT systems potential for interconnectivity and cooperation with other systems, the use of interoperable technologies and standards is needed. According to [ONC19]:

> *The vision is a learning health system where individuals are at the center of their care; where providers have a seamless ability to securely access and use health information from different sources; where an individual's health information is not limited to what is stored in electronic health records (EHRs), but includes information from many different sources (including technologies that individuals use) and portrays a longitudinal picture of their health, not just episodes of care; where diagnostic tests are only repeated when necessary, because the information is readily available; and where public health agencies and researchers can rapidly learn, develop, and deliver cutting edge treatments*

Depending on the extent and scope of the envisaged shared information space, there may be different levels of interoperability. Figure 4-1 shows a proposed "maturity" model for interoperability in eHealth [Velsen16]. The model consists of five levels that, subsequently, describe a more mature version of an interoperable infrastructure.

- At Level 1, an eHealth application consists of a single technology and is not connected to any other application. The single application does not change the nature of the task that it is supposed to serve, nor does it necessitate the redesign of medical procedures and/or protocols. Also, since a single eHealth application does not need to take into account means to communicate with other technologies, no standardization is required.
- At Level 2, a single eHealth application is directly linked to another application for simple data exchange. The transfer of data happens in a context where the involved parties have made simple agreements about a working procedure. Developers of both systems should make agreements on how to transmit data and its format. How data will be prepared before transmission, or processed afterwards, is up to the developers of each eHealth application.
- At Level 3, things become more interesting since we have distributed systems, either inside a single organization or even across organizations. These systems need to communicate either directly or through a central "mediator" but in any case, crossing the boundary between organization-bound and inter-organizational interoperability marks an important change in the need for standardization: agreement on protocols used, data formats, message exchange patterns, etc. The use of well-

[1] "What is HIE?" https://www.healthit.gov/topic/health-it-and-health-information-exchange-basics/what-hie

established and common standards facilitates the interoperability of the systems and allows their integration with future or third-party systems.

- Integration and interoperability are improved in Level 4, where eHealth applications from different suppliers that serve a common goal are linked, but the applications do not need to have common objectives. Often, these infrastructures make use of a Service Oriented Architecture (SOA) whereby specific services and their corresponding functionalities are used only when necessary. Data storage and processing will mostly be done at the level of the individual services. At this interoperability level, two sub-levels can be distinguished: a national level and an international level. The need for common standards is the same as in Level 3, but the security requirements can be more stringent and the medical processes can be quite different, especially in the case where citizens data from different countries are shared.
- Finally, at the “Universal” Level 5, open, interoperable infrastructure to which all eHealth applications are free to connect and disconnect. When connected, they are able to exchange data with all other applications that make use of the infrastructure. The eHealth applications do not have to serve a common goal and can span multiple countries.

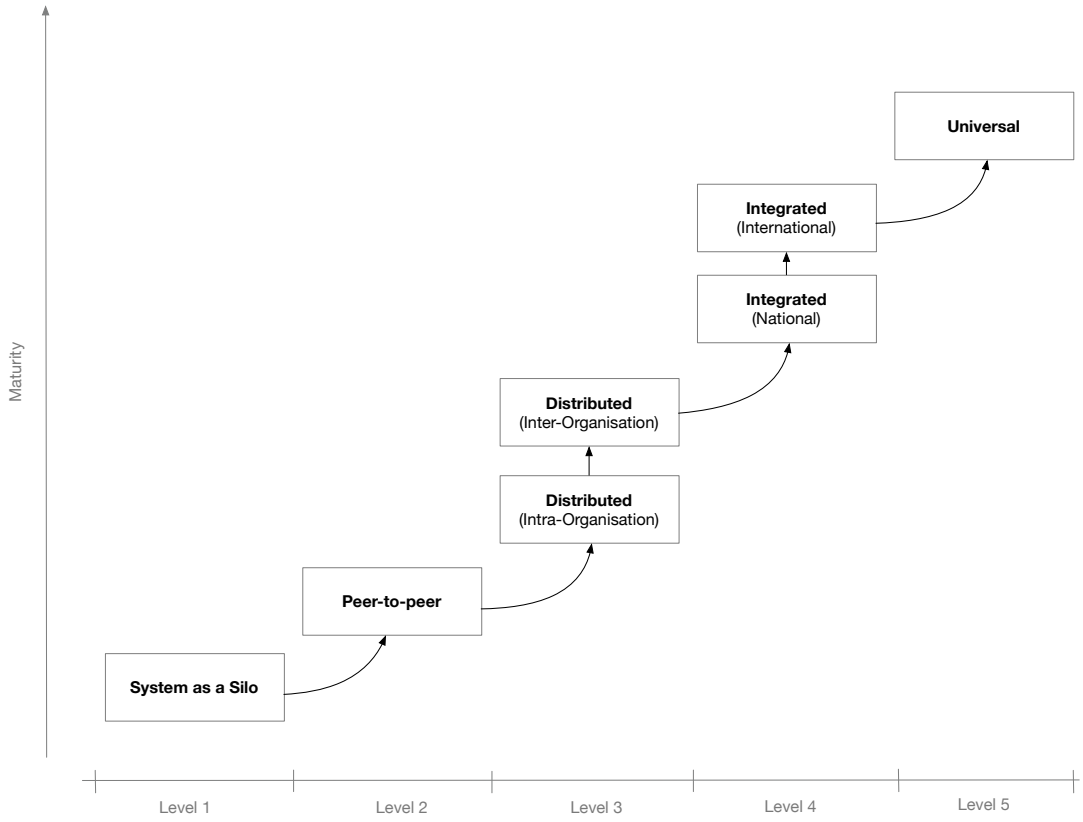

Figure 4-1 A maturity model for interoperability in eHealth (adapted from [Velsen16])

Nevertheless, increasing the level of interoperability and supporting information sharing in the health industry in an intrinsically challenging task [Fichman11]. Some reasons for this are the diversity of the disciplines involved (e.g. public health, epidemiology), the actors (e.g. doctors, nurses, insurers), the data formats, the processes and workflows, the multiplicity of stakeholders (e.g. patients, medical professionals, health organizations), etc. The fast pace of innovation in the basic sciences and the new modalities and treatment options (e.g. genomics and sequencing technologies) introduced in unprecedented rate further toughen the information sharing. Moreover, healthcare organizations are notorious for their reluctance to change existing processes, systems, etc. For example, even today a number of healthcare IT companies are using MUMPS (“Massachusetts General Hospital Utility Multi-Programming System”)[2], a general-purpose computer

[2] http://mumps.org/

programming language and database system originally designed in 1966 for the healthcare industry, or derivatives thereof.

To address these challenges, as described in the GA, the PANACEA Secure Information Sharing Platform (SISP) will provide an environment to coordinate and share information in near-real-time across the boundaries of an organization in a secure manner. The envisioned information sharing platform aims to support the exchange of data between organizations. In this section we provide an overview of the healthcare environment, together with existing standards, proposed solutions, and prior work.

4.1 The eHealth European Interoperability Framework

Information sharing involves more than one parties (healthcare providers, organizations etc.) that need to cooperate and agree on the way the exchange of information happens, and what are the rules and policies that govern it. Interoperability involves many different aspects, such as legislation and guidelines, contracts and agreements between exchanging parties, governance and maintenance, shareable workflows, standardized data elements, semantic and syntactic choices, applications, technical infrastructure, and safety and privacy issues. The Refined eHealth European Interoperability Framework (EIF) is a set of recommendations that specify the standards, protocols, procedures, and policies that when deployed can improve the interoperability of eHealth applications within the EU and across its Member States (MS) by providing specific recommendations for all these aspects [ReEIF15]. The following figure depicts how these aspects can be represented in interoperability “levels” that permit two different organizations to communicate:

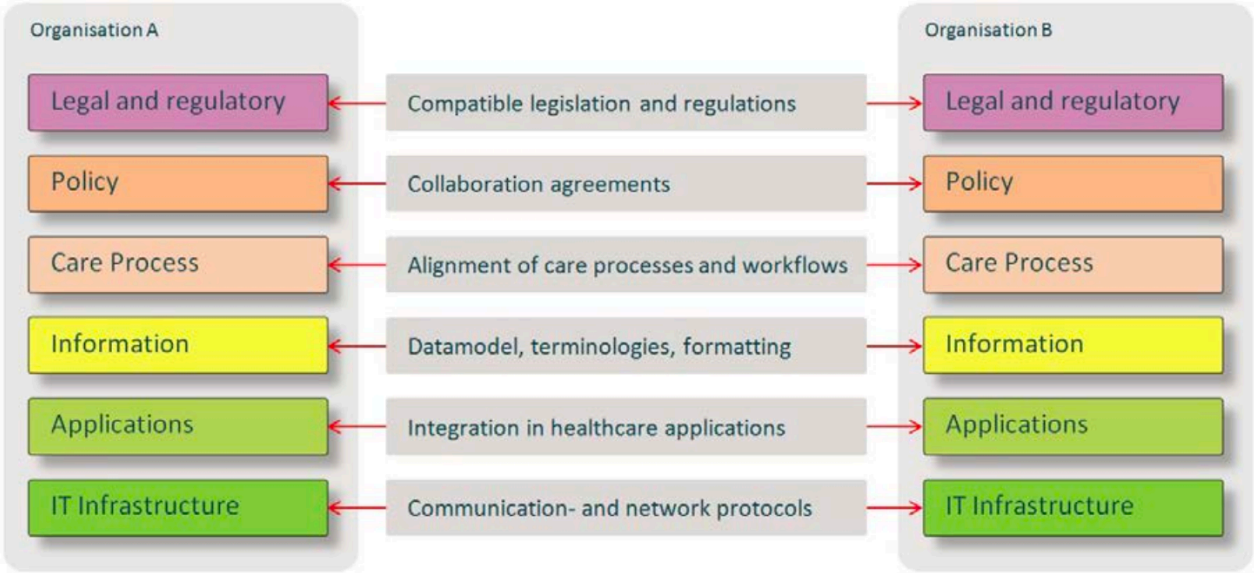

Figure 4-2: Alignment of interoperability levels between two communicating organizations [ReEIF15]

This framework provides a great overview of the needed “glue” so that two or more healthcare environments can collaborate and serves as a common, multi-level and multi-perspective model on the interoperability requirements. The six different levels of the (refined) European Interoperability Framework are the following:

- Legal and regulatory: Legislation and regulatory guidelines that define the boundaries for interoperability across borders, but also within a country or region.
- Policy, which represents the contracts and agreements between the sharing organizations so that trust is established and responsibilities are assigned
- Care process: shared workflows that define how the integrated care is delivered, and how these workflows are managed

- Information, which defines the data models, the concepts and their values, the terminologies and controlled vocabularies that cater for the common understanding of the exchanged electronic messages
- Applications, which define how the data are extracted from and imported to the healthcare information systems and how the transport of the data takes place using health-specific technologies and standards.
- IT infrastructure is at the lowest level and corresponds to general purpose communication and network protocols.

This generic eHealth Interoperability Framework is highly relevant for the use cases of the health information sharing since sharing is greatly facilitated between interoperable systems/organizations. In the following we provide an overview of the healthcare ecosystem (section 4.2) and the use cases for information sharing (section 4.3) and then we briefly cover the legal and regulatory level (section 4.4), the information level and the technical levels (applications and IT infrastructure, in sections 4.5 and 4.6) since for these there are well-established solutions relevant to the information sharing platform.

4.2 The Healthcare environment: Stakeholders, Systems, and Requirements

In order to have an overview of the scope of the healthcare domain where any data sharing solution may be introduced, it is important to identify the major stakeholders, that is the entities that operate (or are involved in any way) in this domain and which will be affected by any “disruption” or reform of the system. Some of the most important stakeholders or actors are therefore the following:

- The *patients* who are actually treated, or, in general, are the recipients of the health services.
- The *medical professionals* (physicians, medical personnel) to provide the medical care
- The *healthcare organizations* (Healthcare Providers) as represented by their director boards who actually administer the health delivery from a business perspective
- The medical device manufacturers that provide medical equipment to hospitals, organizations, and the patients.
- The insurance companies that provide health coverage plans
- The pharmaceutical companies that produce and market medications to be prescribed by physicians for the treatment of patients and also conduct clinical trials for the development of new drugs
- The *governments* and other *regulatory* parties who control, coordinate, and set the rules, rights, and obligations of any involved party

All these actors could have an influence in the design of a data sharing system and can also set important, and conflicting in some cases, requirements. For example, patients would like to have their medical record shared but only after their approval and only with specific authorized personnel in specific circumstances; a healthcare organization can be extremely cautious about sharing the data of their patients with another organization because they are concerned by the security and availability of their systems; governments of EU member states can impose strict laws about the transfer of their citizens in cross-border healthcare treatment scenarios; and medical professionals require fast and effortless access to a patient’s medical history in emergency situations, which cannot be the case if time consuming authorization processes are the norm. It is imperative, therefore, even though the objective is to design a *technical* solution for the sharing of clinical data, all these constraints and requirements are considered and addressed in a satisfying manner.

From a strictly technical point of view, the sharing platform may need to interoperate with a large number and diverse set of IT systems, each with their own protocols, data formats, etc. Some of the most important systems that manage patient related data and could be used as data sources for Information Sharing are:

- *Electronic Health Records* (EHR): These are patient-centered systems that store and manage clinical information, such as a patient’s medical history, diagnoses, medications, immunization dates, allergies, radiology images, and lab and test results. They are managed by authorized personnel, usually in the context of a single Health Care Organizations (HCO) although they can span more.
- *Personal Health Records* (PHR): These are electronic applications that are used by people managing their own health information in a private and confidential environment. They are simpler systems than

EHRs and in some cases they can be connected (temporarily or otherwise) to more enterprise level HCO systems (e.g. EHRs, or other hospital information systems)

- *Laboratory Information Systems* (LIS): used inside Hospitals and Clinics to record, manage, and store data for clinical laboratories in a patient-centric way (sending laboratory test orders to lab instruments, tracking those orders, and then recording the results in a searchable database)
- *Picture Archiving and Communication System* (PACS): These are systems used in a clinical setting for the storage and convenient access to medical images from multiple modalities (source machine types). DICOM (Digital Imaging and Communications in Medicine) is the standard format and suite of protocols for the storage and transfer of images from PACS services.

Moreover, in a health ecosystem there may be additional systems, for example for the management of insurance claims, clinical decision support systems (CDSS), etc.

4.3 Use cases

In the European context, the initial set of health information domains for health information exchange in Europe (focusing on cross-border scenarios, but nevertheless indicative) includes the following [EU19]:

- Patient Summaries, i.e. short summaries of patients’ medical history
- ePrescriptions/eDispensations of medicinal products
- Laboratory reports
- Medical images and reports
- Hospital discharge reports

These health information domains have been prioritized in alignment with the eHealth Network established priorities, on the basis of current work under the eHealth Digital Services Infrastructure and clinical relevance for cross-border healthcare. This list of course is limited but we can expect that will be expanded in the future to cover for example telemonitoring, homecare and medical devices at home, clinical research, etc.

Two EU projects that had pivotal role in the definition of interoperability use cases are the Antilope project[3] and the eStandards project[4] that have developed an initial set of interoperability use cases that can be used as the basis for deployment at the European, national, or even regional level, shown in the following table. Wherever applicable and useful, several variants of these use cases are given, to support the different deployment scales. More details about each use case and the complete realization scenarios, based on available profiles and standards, can be found in the use case repository at https://usecase-repository.ihe-europe.net/ .

#	Medical domain	Description	Scale
1	Medication	e-Prescription and e-Dispensing	1a) Cross-border 1b) National/Regional 1c) Intra-organizational 1d) Homecare
2	Radiology	Request and results sharing workflow for radiology	2a) National/Regional 2b) Intra-organizational

[3] https://www.antilope-project.eu/
[4] https://www.estandards-project.eu/

3	Laboratory	Request and results sharing workflow for laboratory	3a) National/Regional 3b) Intra-organizational
4	Patient Summary	Patient Summary sharing	4a) Cross-border/International 4b) National/regional 4c) Homecare
5	Referral- and Discharge reporting	Cross-enterprise Referral and Discharge Reporting	National /Regional 5a) Referral of patient from primary to secondary care 5b) Discharge report from secondary care
6	Participatory healthcare	Involvement by chronic patients in electronic documentation of healthcare information	Homecare
7	Telemonitoring	Remote monitoring and care of people at home or on the move using sensor devices	Homecare
8	Multidisciplinary consultation	Medical Board Review	National/Regional
9	Public Health	Immunization	National/regional Intra Organizational
10	Antenatal care	Antenatal care	National/Regional

Table 4 – Use cases defined by Antilope and eStandards EU projects

More generally, and taking into consideration the efforts of the Healthcare Information Technology Standards Panel (HITSP) in the United States, the "user needs" for clinical information sharing can be categorized in many different application domains, but the concepts and terminology differ among the various initiatives. Concepts like breakthrough areas, business cases, use cases, realization scenarios, technical use cases, and storyboards represent different levels of granularity in modeling the relationship between user problem and correspondent technical solution. The following table provides some examples of these terms and a mapping between different projects in the US and EU [Bourquard17]:

US HITSP[5]	EU Antilope Project	EU eStandards Project
Breakthrough (Business Cases)	**Business Cases**	**Use Cases/Realization Scenarios**
EHR laboratory result reporting	Laboratory	Request and results sharing laboratory workflow

[5] http://www.hitsp.org

US HITSP[5]	EU Antilope Project	EU eStandards Project
Biosurveillance	Referral and discharge reporting	Referral from primary to secondary care
Emergency response	Patient summary	Exchange of patient summaries across border Exchange of patient summaries across Atlantic
Consultation and transfer of care	Multi-disciplinary consultations	Healthcare provider directory
Medical home	Participatory healthcare (chronic diseases)	Workflow care plan management
Remote monitoring	Telemonitoring	Mobile services to empower patients with heart failure
Quality	Radiology	Request and results sharing workflow for radiology
Medication management	Medication	ePrescribing and eDispensing on national/regional scale
Maternal and child health	Neonatal care management	Neonatal care plan management at the local or regional scale
Immunization	Immunization	Immunization information sharing at the local, regional, or national levels
Consumer empowerment		
Patient–provider secure messaging		
Public health case reporting		
Newborn screening		
Clinical research		

Table 5: Examples of Business Cases, Use Cases, and Realization Scenarios by Project

Taking into consideration all the above it is important that the design of an information sharing platform is defined around a specific set of use cases in healthcare. Different use cases may correspond to different workflows and processes, which are represented by the Care Process level in the eHealth European Interoperability Framework, and it is extremely difficult to build a genetic, interoperable, complete, and flexible solution for all intents and purposes.

4.4 Governance in Information Sharing

Clinical information sharing typically involves the participation of two or more parties that agree on a governance structure and comply with the policies enforced. Information governance in health care is defined as the structures, policies, and relevant procedures initiated and adhered to by health care providers and medical insurance companies to collect, organize, utilize and secure data. Usually technical solutions are agnostic to the governance and its policies and are equally applicable to the case where the participants are just a clinic and a physician's office and in cases where multiple hospitals are sharing medical records. In any case, some legal agreement should be in place to control and provide the rules of information sharing.

There are situations though that the governance structure imposes the need for specific technological solutions to be in place and has an influence on the way the interactions and the transfer of data take place. A characteristic example can be the cross-border or cross-continent information sharing, where multiple constraints must be satisfied, such as the unique identification of the persons and roles involved, the compliance with diverse national laws and processes, etc. In Europe, Directive 2011/24/EU covers a large range of issues for the cross-border delivery of healthcare [EU11]. This directive establishes that the patients'

fundamental rights on data protection should be safeguarded and “*the right for individuals to have access to their personal data concerning their health, for example the data in their medical records containing such information as diagnosis, examination results, assessments by treating physicians and any treatment or interventions provided*”. Another contribution from this Directive is related with the recognition of prescriptions issued in another Member State (Article 14). The Directive provides special rules on how to identify the medicine prescribed (whose designation may vary amongst Member States) and how to identify the prescriber (Article 11). In addition, it also regulates matters as reimbursement, international cooperation between healthcare entities and cross border enforcement of patients’ rights (namely Articles 3/d, 7/7 and 14.

4.4.1 Patient Consent

As described in the eHealth Task Force Report as the 1st lever of change (“My data, my decisions”), “individuals are the owners and controllers of their own health data, with the right to make decisions over access to the data and to be informed about how it will be used.” [EC12] This is further reinforced by the Data Protection Directive (Directive 95/46/EC) and GDPR [EU16] where it is explicitly stated that data should not be disclosed without the patient’s consent and that the patient has always the right to request *erasure* of personal data. GDPR defines “personal data” as “any information relating to an identified or identifiable person” and therefore health data are also subject to the same restrictions and rights.

In order to get a person’s consent, it is important to provide her/him with all the necessary information “in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language” (Article 7 of GDPR[EU16]). Additionally, the patient should be informed about the purpose of the data gathered (diagnosis, treatment) and their processing, the retention period, and the third party (e.g. another practitioner) that their data will be shared with. All these considered, the patient’s *informed consent* will then need be recorded in a “clear unambiguous and affirmative action” such as by written statement, or electronic means. This means that all consent should be “opt-in”, i.e. a positive action or indication: silent, “opt-out” consent is forbidden by GDPR. Nevertheless, sensitive data processing can still be lawful, even without consent, in some particular contexts, as for instance ‘preventive medicine, medical diagnosis, the provision of care or treatment or the management of health-care services’ or for reasons relating to public health (Article 8/3 of Data Protection Directive and Article 9/2 of GDPR).

Acquiring the patient’s consent in a clear and unambiguous way is linked to the requirement for the data controller to provide evidence that the patient has indeed consented to processing of his or her personal data. Furthermore, the patients have the right to withdraw their consent at any time, and it be as easy to withdraw as to give consent. After the withdrawal of the consent, the data should not be further processed for the purposes of the consent initially provided, and therefore should be deleted (unless in the context of another legal ground).

4.5 HIE Architectural Considerations

Two are the main architectural approaches for the implementation of an information sharing platform: Centralized and Federated [Eckman07]. In the centralized approach (Figure 4-3) a central data warehouse and accompanied services act as middlemen for the exchange of information and a single source of patient data that are shared among the participating organizations.

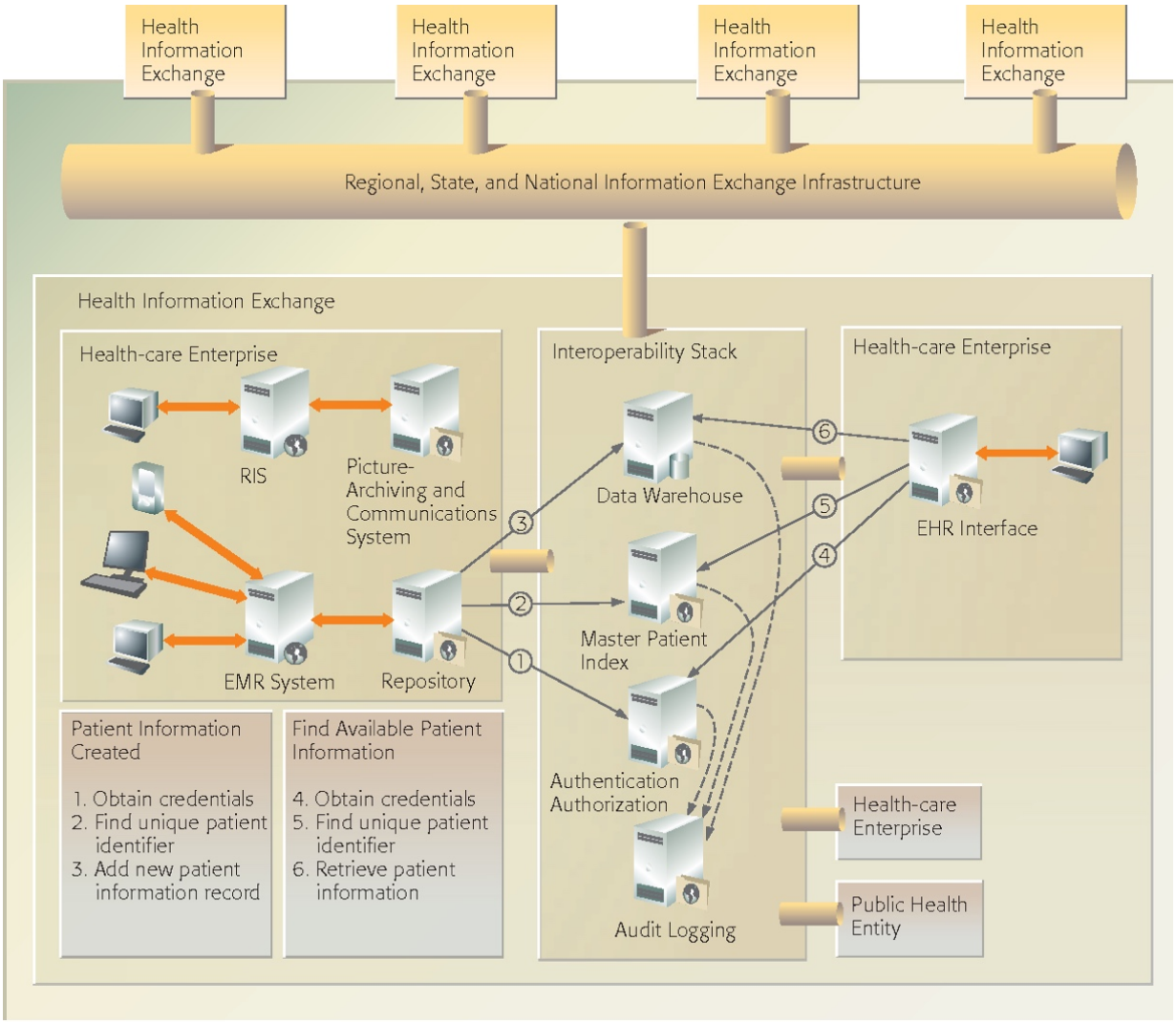

Figure 4-3 Central warehouse architecture for HIE [Eckman07]

In the federated architecture (Figure 4-4) a central infrastructure is also in place but, in this case, it merely acts as a facilitator for locating the data sources (in Figure 4-4 the Record Locator Service takes the place of the Data Warehouse shown in Figure 4-3). An example of this case would be a common registry that stores only the *links* to the original patient records, medical images, etc. while the linked data are not transferred outside their primary premises unless explicitly requested by any interested client system.

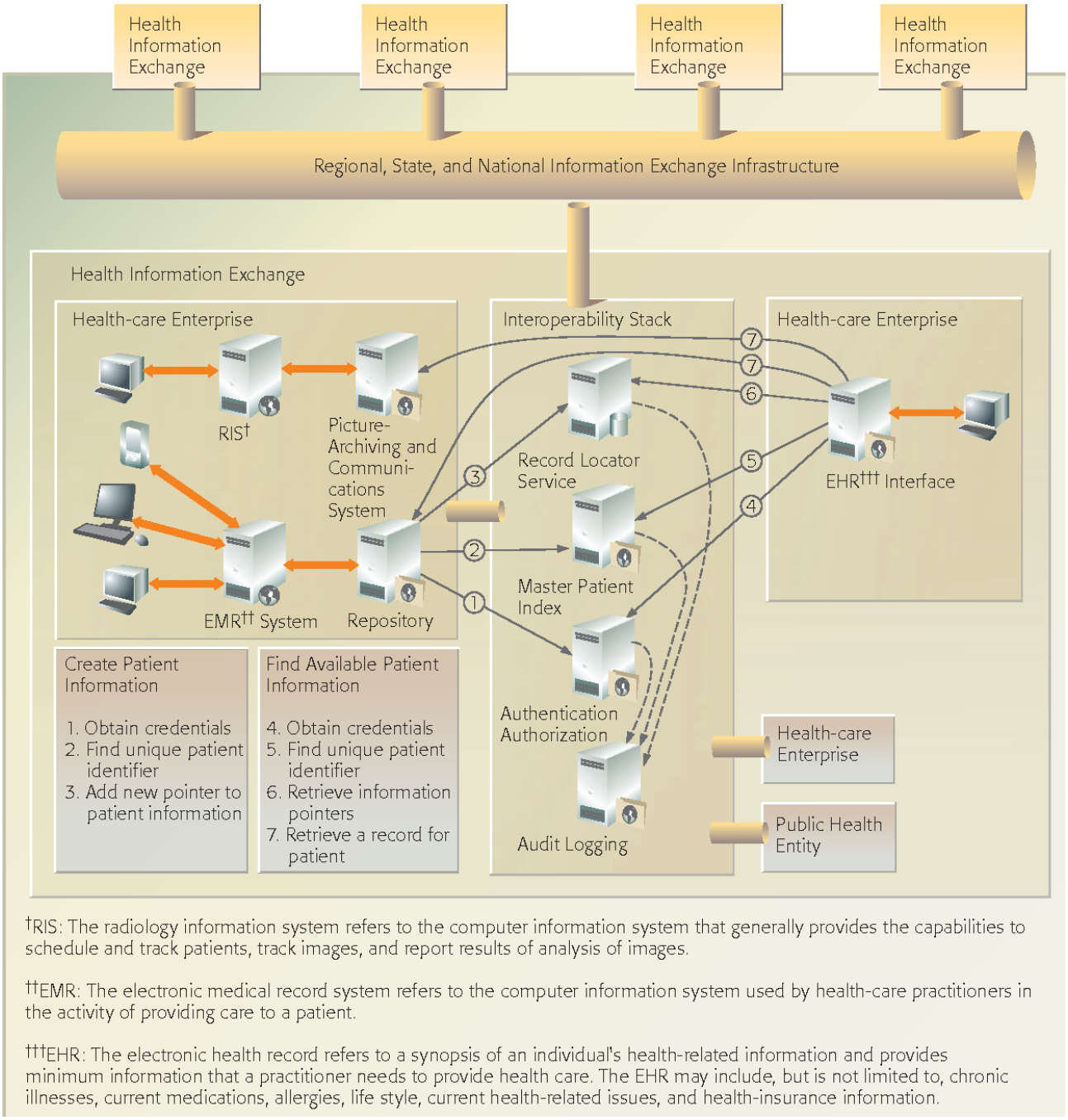

Figure 4-4 A federated architecture for HIE [Eckman07]

In addition to these opposite approaches for designing a distributed information sharing platform one can introduce various hybrid options, such as using messaging with “publish-subscribe” communication that can be introduced to complement either the centralized or the federated architectures.

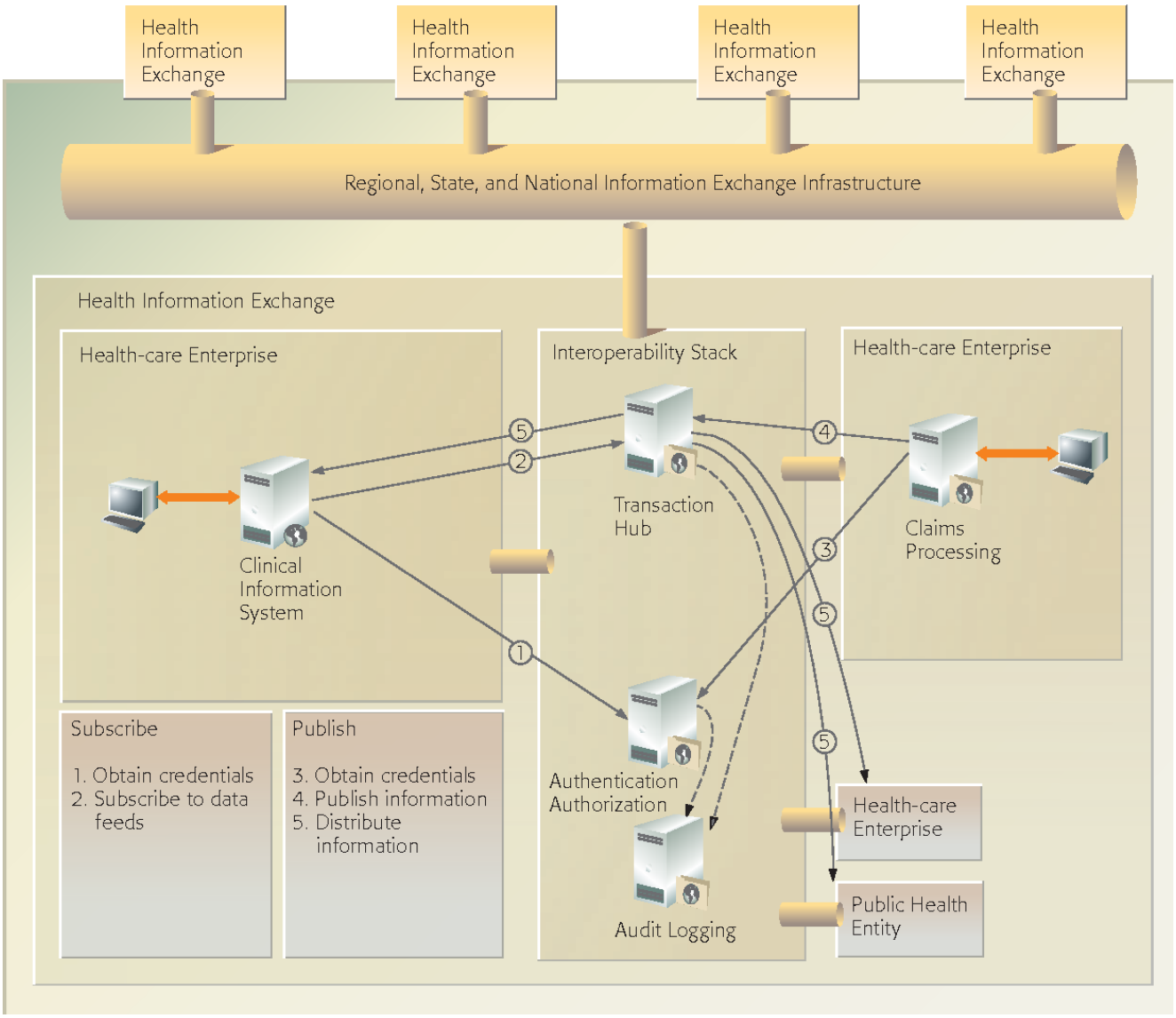

Figure 4-5 Message-based (publish-subscribe) HIE Architecture [Eckman07]

There are advantages and disadvantages in all of the abovementioned options. For example, in the federated approach there are more strong concerns about the privacy, security, and the availability of the data shared and their original sources. The operation of a mission-critical Radiology Information System (RIS) in a hospital can be severely affected if multiple peers request DICOM images from its Picture Archiving and Communication Server (PACS), and this poses an additional burden and cost for the acquisition and management of adequate infrastructure in the source organization. On the other hand, a centralized strategy allows for easy access to the whole information shared but also leads to a concentration of the costs for maintaining the infrastructure needed and can be problematic at the operation level (a “single point of failure”). Furthermore, there are more costs on integrating the different data sets under a common “schema”, resolving conflicts, or even supporting the timely update of the persisted information when a source system acquires new or modified data. More detailed comparison about the alternative deployment options, also from a GDPR and cybersecurity perspective, is provided in Chapter 8.

4.6 Existing HealthIT solutions and standards

The healthcare industry has developed a large number of standards, data formats, terminologies, etc. in order to support the design and building of interoperable IT systems. In this section we provide an overview of the most important solutions that are relevant for the sharing of medical data

4.6.1 Standards Developing Organizations

Perhaps the most well-known and most important standards are the ones introduced by HL7[6] and SNOMED CT[7], which can be used as a foundation for the development of data exchange standards among eHealth systems [Benson12]. Health Level Seven (HL7) International is an organization that provides a comprehensive framework and related standards for the exchange, integration, sharing, and retrieval of electronic health information. Where HL7 is a standard for exchanging information among different health systems, SNOMED CT is used to encode this information. SNOMED CT is an international, comprehensive, multilingual clinical healthcare terminology, owned by the SNOMED International, that is also mapped to other international standards, such as ICD10[8] and LOINC[9]. With SNOMED CT it is possible to also express the meaning of the information, thereby enabling semantic interoperability. Because of this, SNOMED also supports the operation of clinical decision supports systems that can use data from different sources that use SNOMED CT as the interoperability standard for encoding health information.

Two more standards organizations are IHE[10], which is highly important and we describe it in more detail in Section 4.7.4, and the Clinical Data Interchange Standards Consortium (CDISC). CDISC[11] produced the Operational Data Model to “facilitate the regulatory-compliant acquisition, archive and interchange of metadata and data for clinical research studies”. ODM is XML-based format that provides a number of constructs for modeling electronic Case Report Forms (CRFs) and can also be used in sending forms data from a clinical trial system to an electronic health record (EHR) system.

Finally, in the area of medical devices, the Continua Health Alliance[12] is a non-profit, open industry coalition of healthcare and technology companies working to establish a system of interoperable personal health solutions. Continua has set out to develop an ecosystem of connected technologies, devices and services that will enable the more efficient exchange of fitness, health and well-ness information. The foundation of this ecosystem is a set of interoperability guidelines that specify how systems and devices made by different companies can work together. The first set of Continua standards includes specifications for using existing standards such as Bluetooth, USB, medical devices (IEEE 1173) and HL7 to enable people to use home-based devices to monitor their weight, blood pressure, glucose and blood oxygen levels and share this with their healthcare professionals.

4.6.2 Data formats, syntax, and semantics

The Office of the National Coordinator (ONC) in the United States provided the comparative report “Interoperability Standards Advisory” on the best available standards to meet various health information interoperability needs [ISA19]. The study compared standards with regard to the following criteria: the maturity of the standard, its adoption level in USA, its cost and availability. The report shows that there is a significant standardization effort for the formal representation of health information. However, there is still a lack of standards for representing knowledge in certain medical domains (gaps) while there are multiple alternative

[6] https://www.hl7.org/
[7] https://www.snomed.org/
[8] https://www.who.int/classifications/icd/en/
[9] https://loinc.org/
[10] https://www.ihe.net/
[11] https://www.cdisc.org/
[12] http://www.continuaalliance.org/

standards for others (overlaps). An efficient linkage to external health data sources should include policies to resolve issues related to gaps and overlaps in the data representation standards utilized by the sources.

The exchanged information can be in multiple of data formats based on the type of data, device category, etc. Some of the most common formats based on the health applications using them are the following:

- For medical imaging the use of DICOM is almost universal. According to the DICOM standard, a DICOM file format (the format is recognized by the “.dcm” extension) consists of (i) a header that includes the patient metadata (e.g., patient identification number, age, gender) and the specifications of the image acquisition protocol (e.g., the scanner parameters) and (ii) the pixel data, where a pixel value is represented as an 8-bit, 16-bit, or 32-bit integer value through a linear transformation with a certain slope and intercept [Larobina14]. The DICOM standard supports image compression that is useful in the case where a series of DICOM images is produced by the scanner.
- In the area of DNA Sequencing and other -omics data formats include FASTQ file format [Cock14] that is used to store sequence information and the standard flowgram format (SFF) that is used to encode sequence reads [Leinonen10].
- The majority of the EHR systems adopt the Health Level 7 (HL7) standard clinical document architecture (CDA) as the interoperable data format [Dolin01]. CDA is based on a referenced information model (RIM) that serves as a semantic model that consists of a set of structural components (e.g., classes with data types) and semantic relations (e.g., a healthcare provider “belongs to” an organization where the healthcare provider is an actor who is represented by a class and has a name, a surname, etc.) that are used to represent clinical notes in the form of an extensible markup language document [Dolin01]. The CDA is part of the HL7 version 3 family and uses the RIM as an object-oriented model to define the actors (e.g., clinicians, healthcare providers), the clinical document standard terms, and the targets (i.e., patients). CDA can also include multimedia content, such as medical images. All this information is organized into a human readable and standardized format that enables the interlinking of EHRs among different clinical centers.

The use of controlled vocabularies allows for the unambiguous representation of important value sets, such as the diagnosis, the medicines, etc. The following is a list of such terminologies:

- The ICD (International Classification of Diseases) provides a common language for reporting and monitoring diseases, used throughout the world to compare and share data in a consistent standard way between hospitals, regions and countries and over periods of time[13]. It is used to classify diseases and other problems for payment, management and research, as recorded on many types of health records including medical records and death certificates. ICD-11 is the latest version of it, whereas ICD-10 (released in 1993) remains widely used.
- SNOMED CT, already mentioned above, is the most comprehensive multilingual clinical healthcare terminology available. It is used in electronic health record systems to facilitate clinical documentation and reporting and to retrieve and analyze clinical data. SNOMED CT is both a coding scheme, identifying concepts and terms, and a multi-dimensional classification, enabling concepts to be related to each other, grouped and analyzed according to different criteria. SNOMED CT is superior to ICD-10 for clinical representations due to its controlled focus on clinical concepts and multi-axial structure. While ICD-10 is designed as a hierarchical statistical classification system, SNOMED CT is represented by multiple levels of granularity. Some of these levels include: Body Structure, Clinical Finding, Event, Procedure, Substance, etc.
- LOINC (logical observation identifiers names and codes) provides a set of universal identifiers for medical laboratory observations. LOINC provides codes for the observation names (e.g. eye color), not the observation finding (e.g. blue eyes). LOINC therefore provides codes for questions and where needed, other vocabularies, such as SNOMED-CT, provide codes for the answers. LOINC is a community-built universal code system that facilitates the exchange, pooling, and processing of laboratory and other clinical observations. It is a controlled terminology that contains unique identifiers and “fully specified” names built using a formal structure, which distinguishes among tests and observations that are clinically different.

[13] WHO. ICD-10. International statistical classification of diseases and related health problems. Geneva: WHO; 1992

- The Unified Medical Language System (UMLS) as an important terminology resource, intended for use mainly by developers of health information systems. UMLS is a large multipurpose multilingual vocabulary database that contains over five million terms and one million concepts covering information about health and biomedical concepts, their names and relationships between them. The UMLS "Metathesaurus" is built from over 100 different source vocabularies and seeks to reflect and preserve the meanings concept names and relationships from these sources. For example, if two source vocabularies use the same name for different concepts or define the same concept in different ways, the Metathesaurus represents both sets of meanings and relationships, and indicates which meaning is used in which source vocabulary. It is therefore a valuable resource for the translation between the different source vocabularies.

4.6.3 Applications level interfaces

HL7’s name comes from “Level Seven”, which, according to the Open System Interconnection (OSI) model that standardizes communication functionality in IT, corresponds to *application layer*. This layer specifies the shared communication protocols used by interconnected systems in order to exchange information. From its establishment in the late ‘80s, HL7 was therefore focused on exchanging information within hospitals. The focus remains almost the same today, but HL7 has progressed from different paradigms over the years, in order to describe the structure, semantics, and management of the exchanged information.

There are two main versions of the standards HL7 produces: version 2 and version 3, with different underlying technical foundations. Version 2 (HL7v2) is the most widely used healthcare interoperability standard and it is modelled around *messages*. HL7v2 messages are plain text messages composed of “segments” in a specific order and each segment contains *fields* on a specific manner, which have values of different data types. Data types are the building blocks of the fields and may be simple, with a single value, or complex, with multiple components. These components themselves have data types, which can be simple or complex, leading to sub-components. All this hierarchical structure is built around a set of delimiters, which are characters like ‘^’ and ‘&’ that separate segment, fields, components, and sub-components. Each message also carries a *trigger* event that indicates what caused the message to be generated, e.g. the admission or discharge of a patient. Even if this seems complex at first, in fact everything in HL7v2 follows the same rules and therefore it is easy for one to familiarize with. Having said that, there are still rules and classification of the different message types in a large set of categories, such as patient administration, financial management, order entry, etc. that leads to a huge standard containing all possible interactions between health information systems, e.g. patient administration systems, EHRs, laboratory information systems, etc.

The development of HL7 version 3 (HL7v3) on the other hand started around 1995 in order to overcome some lack of consistency between the HL7v2 implementations and somehow constrain v2’s flexibility and lots of optionality. HL7v3 strives to be definitive, comprehensive, and testable, using an object-oriented development methodology based on a Reference Information Model (RIM). The RIM is an essential part of the HL7 Version 3 development methodology and provides an explicit representation of the semantic and lexical connections that exist between the information carried in the fields of HL7 messages. RIM used as an abstract format permits the generation of the “wire format” of HL7v3 which is actually XML messages that can be validated by schemas. The use of RIM allows for domain specific models and the incremental refinement of those models into design models that are specific to the problem area. This, in turn, allows for a more constrained interoperability among health information systems.

4.6.4 Integrating the Health Enterprise (IHE)

The Integrating the Health Enterprise (IHE) initiative has developed a set of standards, called integration profiles, which are detailed specifications for communication among systems to address key clinical use cases, all based on established standards. According to IHE, a profile is a guideline for implementation of a specific process, by providing precise definitions of how standards can be implemented to meet specific clinical needs. IHE Profiles organize and leverage the integration capabilities that can be achieved by coordinated implementation of communication standards, such as DICOM, HL7, W3C and security standards. IHE Profiles address critical interoperability issues related to information access for care providers and patients, clinical workflow, security, administration and information infrastructure. Each profile defines the *actors*, *transactions*

and information content required to address clinical use cases by referencing appropriate standards. IHE defines the following core concepts in order to specify the addressed clinical scenarios in a consistent manner:

- An *Actor* is an essential component of an IHE Integration Profile that is an abstraction of the endpoint responsible for the initiation or response to a Transaction. Systems implement one or more Actors (Grouped) as declared in the systems Integration Statement. 1) A functional component of a communicating healthcare IT system and device. 2) Actors are information systems or components of information systems that produce, manage, or act on information associated with operational activities in the enterprise.
- An IHE *Integration Profile* specifies a coordinated set of interactions exchanged between the functional components of communicating healthcare IT systems and devices. These functional components are called IHE actors. An IHE Integration Profile specifies their interactions in terms of a set of coordinated, standards-based transactions. An IHE Integration Profile is a reusable specification that defines the Interoperability solution to a healthcare workflow that requires two or more systems to work together.
- A *Transaction* is a pre-defined interaction between Actors. IHE Transaction defines the network semantics, trigger events, and expected actions 1) Transactions are interactions between actors that communicate the required information through standards-based messages. 2) Transactions are interactions between actors that transfer the required information through standards-based messages.
- A *Use Case* is the defined healthcare workflow that outlines the interoperability problem that is the focus of an Integration Profile. A textual and graphical depiction of the actors and operations that addresses information exchange in the context of a set of specific tasks performed by different systems or devices.

Some of the IHE integration profiles that might be interesting in the context of interfacing health information systems and sharing of clinical information are:

- Audit Trail & Node Identification (ATNA): Basic security through (a) functional access controls by limiting network access between nodes and limiting access to each node to authorized users (b) defined security audit logging and (c) secure network communications, using underlying standards such as Transport Layer Security (TLS) with stronger ciphersuites.
- Consistent Time (CT): Enables system clocks and time stamps of computers in a network to be synchronized (median error less than 1 second).
- Cross Enterprise Document Sharing (XDS): Share and discover electronic health record documents between healthcare enterprises, physician offices, and clinics, acute care in-patient facilities and personal health records.
- Patient Demographics Query (PDQ): Enables applications to query by patient demographics (e.g. name) for patient identity from a central patient information server.
- Patient Identifier Cross Referencing (PIX): Allows applications to query for patient identity cross-references between hospitals, sites, health information exchange networks, etc.
- Patient Demographics Query HL7 v3 (PDQv3): Extends the Patient Demographics Query profile leveraging HL7 version 3.
- Patient Identifier Cross Referencing: Extends the Patient Identifier Cross-Reference profile leveraging HL7 version 3.
- Basic Patient Privacy Consents (BPPC): provides mechanisms to record patient privacy consent and to enforce it.
- Retrieve Form for Data Capture (RFD): Enables EHR applications to directly request forms from clinical trial sponsors and public health reporting.
- Clinical Research Document (CRD): Clinical Research Document describes the content pertinent to the clinical research use case required within the Retrieve Form for Data-Capture (RFD) pre-population parameter.
- Cross-Community Access (XCA): Allows to query and retrieve patient electronic health records held by other communities.
- Cross-Enterprise Document Reliable Interchange (XDR): Exchanges health documents between health enterprises using a web-service based point-to-point push network communication.
- Clinical Research Process Content: This profile is based on Retrieve Process for Execution profile, which provides general framework for messaging interaction. It further details the content which is specific to research domain.

In the following paragraphs we describe in more details some of the above IHE profiles that are relevant for the cross-enterprise or cross-border sharing of health information.

Cross Enterprise Document Sharing (XDS)

Cross-Enterprise Document Sharing enables a number of healthcare delivery organizations belonging to an XDS Affinity Domain to cooperate in the care of a patient by sharing clinical records in the form of documents as they proceed with their patients’ care delivery activities (Figure 4-6). According to IHE, an XDS Affinity Domain is “*a group of healthcare enterprises that have agreed to work together using a common set of policies and share a common infrastructure*” (e.g. a community of care). Federated document repositories and a document registry create a longitudinal record of information about a patient within a given XDS Affinity Domain. The documents themselves are provided by “Document Sources” actors, which can be any healthcare provider. XDS as well as the rest of IHE integration profiles is document format neutral, any type of clinical information is supported. This profile is based on ebXML Registry standards[14]. Document consumers (i.e. other medical systems, or healthcare organizations) can query the document registry; however, the query is defined in the registry. The implementation of queries is up to the document registry, no specific querying mechanisms are defined by IHE, and consumer has to transmit only the query id and query parameters to invoke it. This makes querying on registry flexible and easily customizable, as well as protecting the registry from malicious queries.

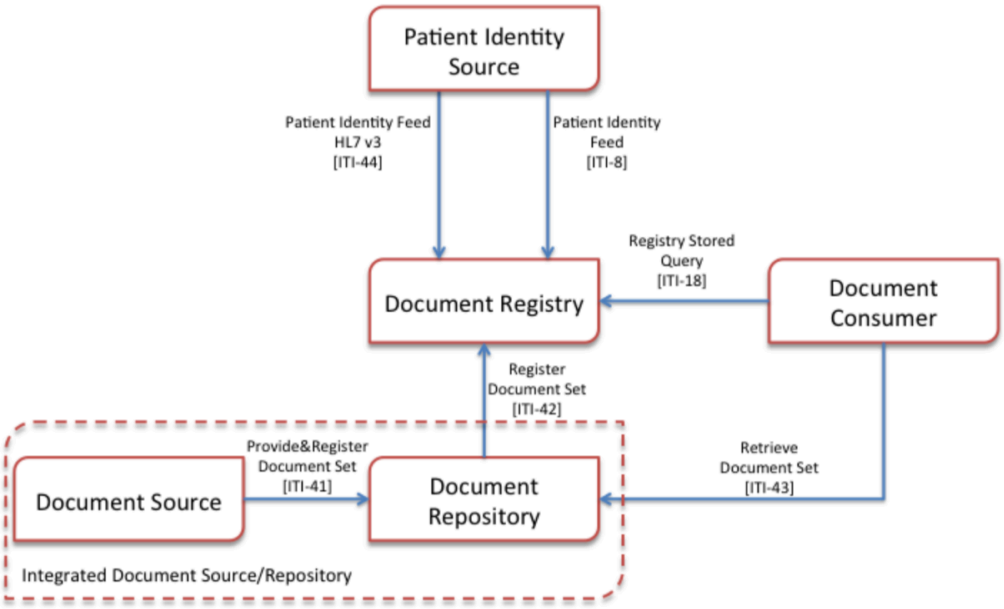

Figure 4-6 Cross Enterprise Sharing Actors and Transactions [IHE-IT]

This profile heavily relies on ATNA and CT profiles, as well as on existing IT infrastructure within healthcare enterprise (document repository, document registry). Therefore, the implementation effort is high.It can also be combined with the BPPC profile in order to capture a patient's acknowledgment and/or signature of one or more of the policies agreed in the context of an XDS affinity domain. The consent can then be captured using an CDA document with optionally a scanned copy or optionally a digitally signature that are shared in the affinity domain. A Document Consumer is then obliged to enforce patients’ consents (or dissents) when a clinical document is acquired.

Especially for imaging, XDS-I (“XDS for imaging”) is an extension of XDS supporting the exchange of medical images by incorporating specific characteristics of the DICOM standard. With XDS-I, the DICOM images are not saved in the Repository, but they remain in the source PACS system. In the XDS Repository, only metadata

[14] http://www.ebxml.org/specs/

are stored that allow indexing and efficient retrieval from the PACS server. The DICOM images therefore are physically in the PACS, but can be accessed from the XDS infrastructure.

Cross-Community Access (XCA)

The Cross-Community Access profile supports the means to query and retrieve patient relevant medical data held by other communities (Figure 4-7). A community is defined as a coupling of facilities/enterprises that have agreed to work together using a common set of policies for the purpose of sharing clinical information via an established mechanism. Facilities/enterprises may host any type of healthcare application such as EHR, PHR, etc. A community is identifiable by a globally unique id called the *homeCommunityId*. Membership of a facility/enterprise in one community does not preclude it from being a member in another community. Such communities may be XDS Affinity Domains, which define document sharing using the XDS profile or any other communities, no matter what their internal sharing structure. The communication and information exchange take place between the Gateways and not directly between the organizations of the two (or more) communities.

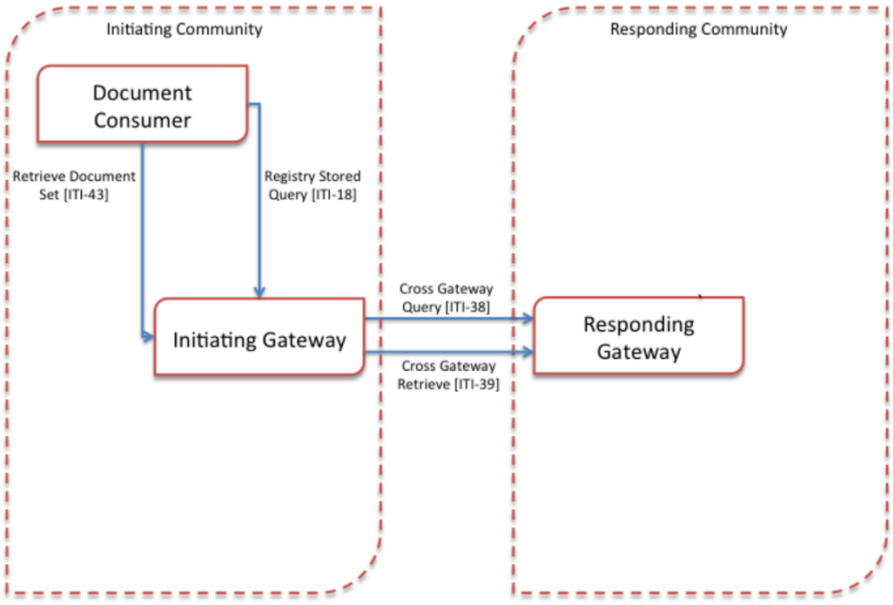

Figure 4-7 Cross Community Sharing Actors and Transactions [IHE-IT]

Cross-Enterprise Document Reliable Interchange (XDR)

Cross-Enterprise Document Reliable Interchange provides document interchange using a reliable messaging system. This permits direct document interchange between EHRs, PHRs, and other healthcare IT systems in the absence of a document-sharing infrastructure such as XDS Registry and Repositories. XDR requires less infrastructure comparing to XDS, however combination with PDQ/PIX for querying and cross-referencing patient identifiers is required to make sharing solution more usable.

Figure 4-8 XDR Actors and Transactions [IHE-IT]

Retrieve Form for Data Capture

The Retrieve Form for Data Capture Profile (RFD) provides a method for gathering data within a user’s current application to meet the requirements of an external system. RFD supports the retrieval of forms from a form source, display and completion of a form, and return of instance data from the display application to the source application.

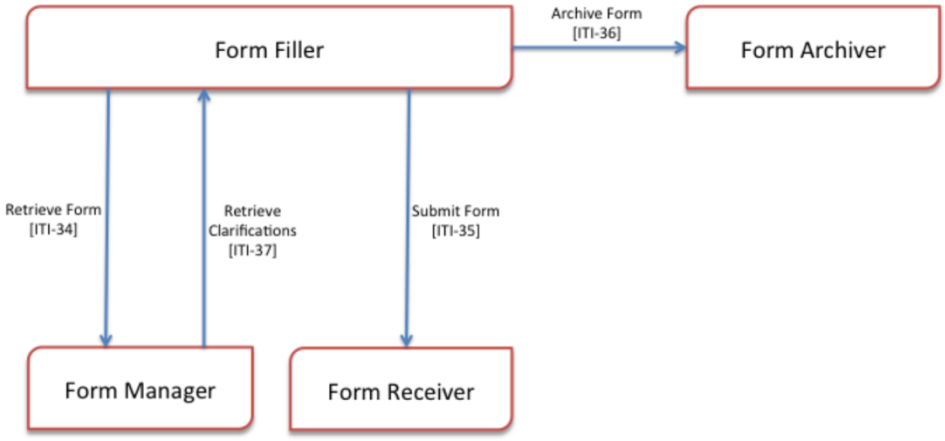

Figure 4-9 RFD Actors and Transactions [IHE-IT]

This profile describes manual filling out of the form, while allowing some pre-population of data from the local EHR system. The form itself is XHTML based, which supports submit transaction. Optionally Form Manager can support XForms[15]. Upon user’s request from form filler application (local EHR for example) the form is retrieved from the external form manager and displayed within form filler application. Some of the data might be prepopulated, format of prepopulated Data is defined by domain-specific IHE content profiles, is application specific and out of the scope of RFD integration profile. Semantics of the form itself is out of the scope of the profile as well. As of now IHE specifies following domains:

- Cardiology
- Dental
- Endoscopy
- Eye Care
- Laboratory
- Anatomic Pathology
- Patient Care Coordination
- Pharmacy
- Quality, Research and Public Health
- Radiation Oncology
- Radiology

This profile covers one patient at a time only (as the rest of the IHE profiles actually), so to get data for multiple patients this profile has to be combined with patient demographics query (PDQ) and patient cross-referencing (PIX) profiles.

Mobile Health Document Sharing (MHDS)

Mobile Health Document Sharing is a new IHE profile that describes how to build a Document Sharing Exchange using IHE profiled FHIR standard, rather than the legacy IHE profiles that is dominated by XDS and HL7 v2. This is therefore a comprehensive profile combining several IHE profiles that support patient

[15] XForms 1.1, W3C Working Draft. http://www.w3.org/TR/2004/WD-xforms11-20041115/

identification, health document location and retrieval, provider directories, etc. in order to provide a standards-based interoperable approach to health information sharing built on HL7 FHIR.

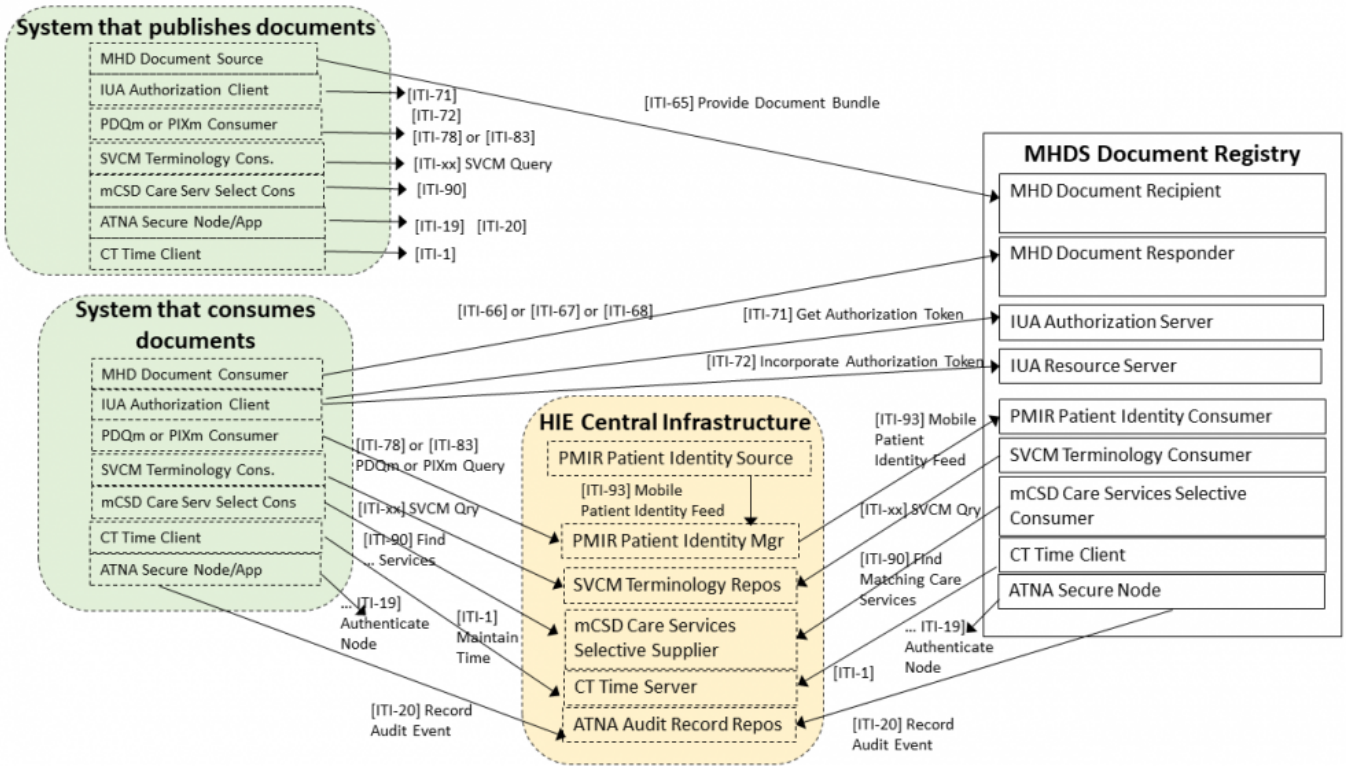

Figure 4-10 Actors and transactions in MHDS [IHE-MHDS]

At the time of writing, this profile is in the “public comment” phase of development before proceeding for trial implementation.

Information sharing with IHE profiles

The wide range of IHE profiles covering a large spectrum of interactions among healthcare systems and other actors allows for the implementation of an information sharing platform in a flexible way. The strength of the Document Sharing profiles such as XDS is that they enable effective sharing of data among multiple, disparate systems in a way that minimizes the burden that data sharing imposes on those systems. These profiles may be categorized according to three different data sharing models [IHE-Share]:

- Direct Push – supports point-to-point push of documents where content is sent directly to the intended recipient found through manual means or infrastructure enabled directory
- Centralized Discovery and Retrieve (XDS Affinity Domain) – a community of sharing partners agrees to use a common infrastructure to enable health document sharing. A document source will publish the existence of documents to a location (registry) that is accessible to other systems. Then, document consumers can discover document locations that have been previously published and pull a copy of the document.
- Federated Discovery and Retrieve – content is pulled directly from the content holder who is found through manual means or a directory

These models for document sharing are in essence similar to the architectural options we have list in section 4.5 and can support different use cases. The Direct Push model can be relatively simple but it cannot satisfy all use cases because it relies on the source of documents to know where those documents will be needed. The centralized discovery model allows for better scalability but in a single organizational context (“XDS affinity

domain”), while the federated discovery and retrieve model is able to sustain more distribution, local autonomy, and better security when different communities/affinity domains are linked together.

The implementation of these models using the IHE profiles is as follows:

- Direct Push – Cross-Enterprise Document Reliable Interchange (XDR) and Cross-Enterprise Document Media Interchange (XDM)
- Centralized Discovery and Retrieve (XDS Affinity Domain) – Cross-Enterprise Document Sharing (XDS)
- Federated Discovery and Retrieve – Cross-Community Access (XCA)

Figure 4-11 illustrates these models and how the different IHE profiles are used in each of them.

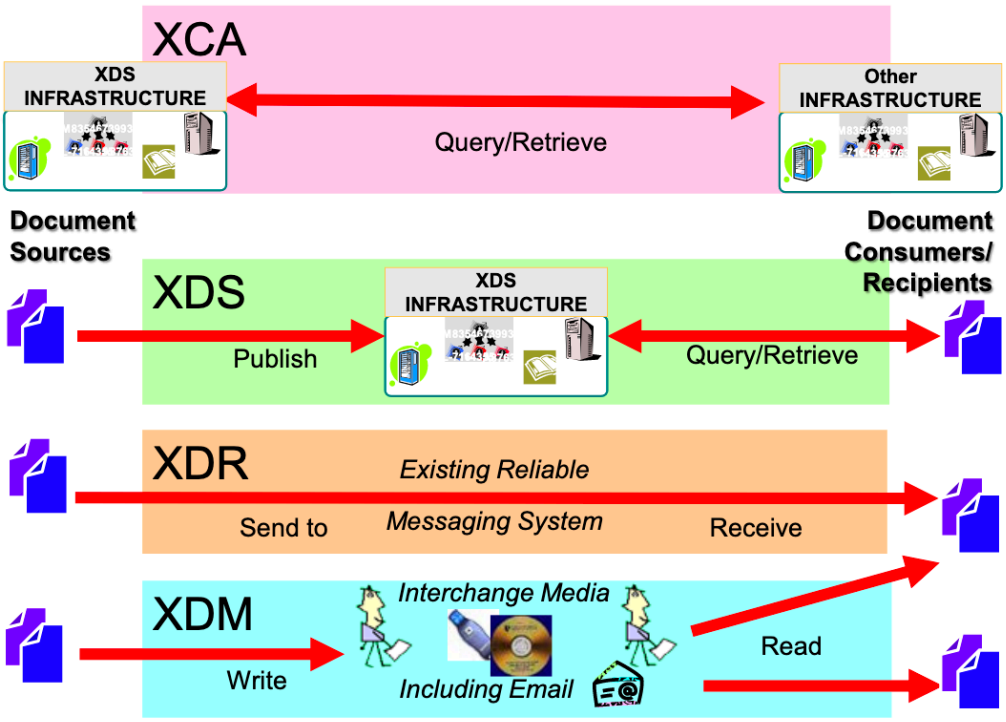

Figure 4-11 Implementation of different document sharing architectures using IHE profiles [IHE-Share]

4.6.5 Secure communication over the Internet

Transport Layer Security are widely used to secure the exchange of data in the Internet [RFC7525]. Public Key Infrastructure (PKI) and the hierarchical level of trust based on X509 certificates signed by Certificate Authorities (CA) are also supplementary (and intertwined) mechanisms to establish authenticity and trust among communicating peers [Perlman2016].

In an Information Sharing Platform these technologies are requisite but can also be adapted and deployed using custom settings. For example, specific, more secure, encryption algorithms (“ciphers”) may be enforced or a platform specific CA can be put in place and certificates signed by this authority should be trusted by anyone and any system participating in the share community.

5. Advanced information sharing scenarios

In Deliverable 1.4 “Relevant user scenarios, use cases and KPIs for Panacea Toolkit validation” we have identified a number of concrete user scenarios covering the whole feature set of the PANACEA toolkits, from technical aspects to users’ behavior and governance. In this section, we focus on information sharing and we document existing approaches, initiatives, and products for two specific use cases: the cross-border information sharing in the European context, and the Remote Health Monitoring with IoT[16].

5.1 Cross-border information sharing

Cross-border healthcare in Europe has been recognized as of 2011 with Directive 2011/24/EU which established patients’ rights to access safe and high-quality healthcare, including across national borders within the EU, and their right to be reimbursed for such healthcare [EU11]. From a technical point of view, the cross-border sharing of clinical information is complex scenario due to the fact that data need to be transferred between different countries and therefore requires overcoming barriers such as establishing a common trust framework, uniquely identifying citizens, translating between different schemas and terminologies, etc. The epSOS project (2008-2014), epSOS meaning “Smart Open Services for European Patients”, was an EU co-funded so-called Large-Scale Pilot aimed to contribute to the interoperability of eHealth services for an integrated healthcare system at the European level. The main objective of epSOS was to provide concrete cross border services like e-Prescriptions and Patient Summaries (PS) for the secure and efficient medical treatment of patients when travelling across the Europe.

Figure 5-1illustrates an exemplary use case from epSOS, sharing of patient summaries, with the following steps:

- Patient who is citizen of Country A (country of origin) visits a health care practitioner in Country B (country of treatment)
- Since the patient is a citizen of another country, his/her Patient Summary (PS) is requested at the national contact point (NCP) of country A (country of origin). The patient summaries retrieved through the NCP contain essential information needed for the continuity of care such as the most important clinical patient data (e.g. allergies, current medical problems, medical implants, or major surgical procedures), a list of current medication including all prescribed medication that the patient is currently taking, etc.
- PS translated by semantic services, so that it can be understood by the practitioner and the systems ate Country B
- PS is sent to NCP of country B, and then becomes accessible to practitioner

[16] D1.4 includes two use-cases dealing with cross-border information sharing, to validate the PANACEA Secure Information Sharing Platform [7C/FPG (Italy)-7HRC (Greece)_Cross-border information sharing (Dialysis) and 8C/FPG (Italy)-HSE (Ireland)_Cross-border information sharing (Emergency)] and one use-case dealing with remote monitoring, to validate the PANACEA Machine-to-Machine authentication solution [9B/Healthentia-QTrobot relationship],

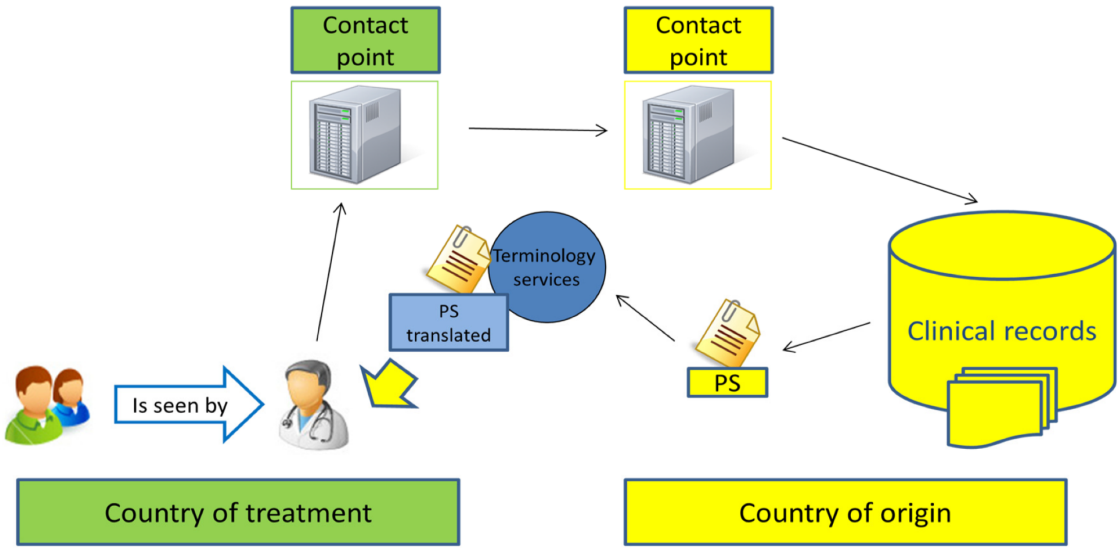

Figure 5-1 The primary use case for epSOS based cross-border treatment of a patient

As it is evident from the description above the role of national contact points (NCPs) are critical NCPs are organizations at the national level (e.g. in Germany) and act as a bidirectional technical, organisational and legal interface between the existing different national health providers and infrastructures in order to fulfil the epSOS use cases. At the same time, epSOS does not specify any infrastructure or impose any constraints either in the country of origin or in the country of treatment.

In terms of the actual content of clinical information that is exchanged in the epSOS domain, the Patient Summary (PS) consists of all the needed information in the context of a resident of one Member State (country A) visiting another MS (country B) and seeking for Health Care. The PS made available to the Health Care professional of country B should contain updated and reliable information. The primary application of electronic Patient Summary is therefore to provide the Health Care professional with a dataset of essential and understandable health information at the point of care to deliver safe patient care during unscheduled care and planned care with its maximal impact in the unscheduled care. In practice, the patient summaries can be expanded to include rich clinical data that are optional epSOS summaries. In fact, the Health Care Encounter Report (HCER), which was added later to the palette of epSOS document types, is closer to the cross-border discharge summary concept and can be used to provide more accurate description of the content of a patient’s clinical records. It is important also to note that epSOS PatientSummary is compatible with the HL7 Clinical Document Architecture (CDA) and in fact is defined as a CDA document template. The epSOS architecture is generic enough to accommodate the sharing of different document types, such as the Consolidated CDA and the Continuity of Care Documents [Estelrich14].

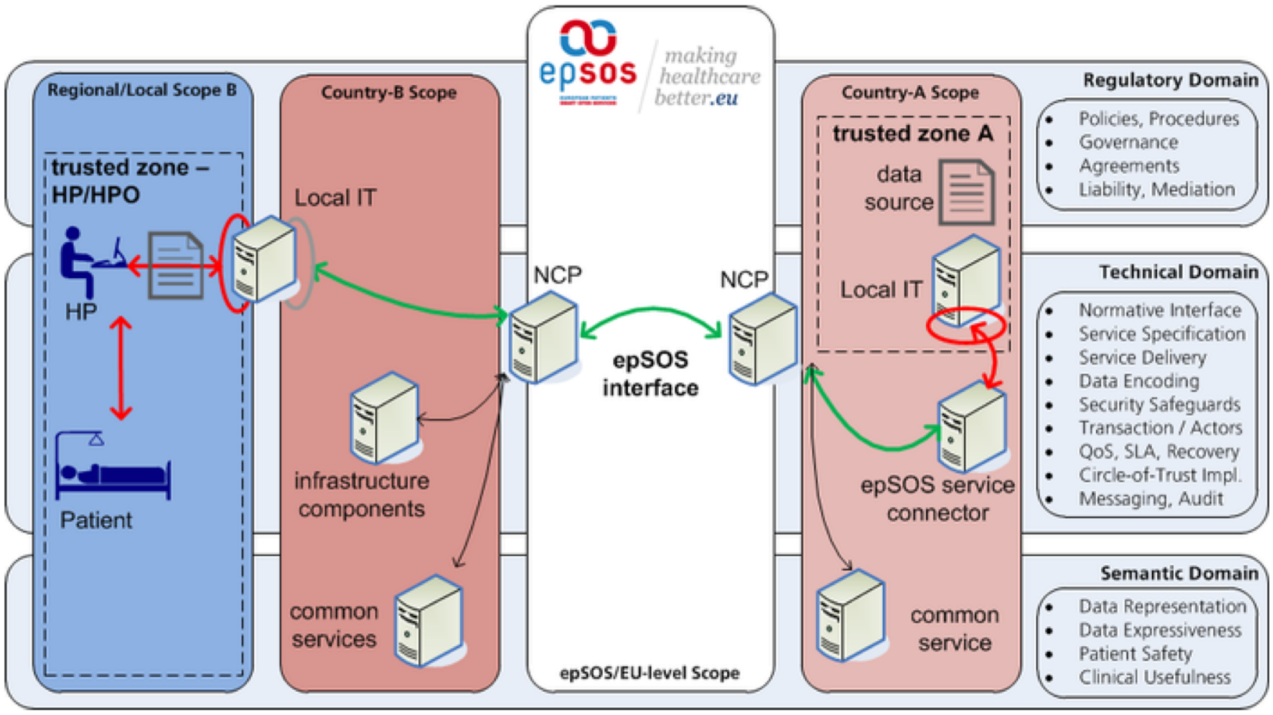

Figure 5-2 The epSOS architecture using the National Contact Points (NCP) as gateways

Figure 5-2 depicts, from a technical point of view, the flow of the clinical records patient data through the epSOS NCP gateways. As an outcome from the epSOS project, OpenNCP is the reference implementation and open source set of software components that can be used to build a complete NCP node. The provided IT infrastructure and functionality is adequate to support both internal (in the same country) or cross-border (across countries) communication for the exchange of patient data.

Since epSOS, the Commission, together with Member States, is building an EU-wide eHealth Digital Service Infrastructure or eHDSI, aimed at allowing the exchange of limited health data – specifically ePrescriptions and patient summaries – across national borders. Member States can connect their health systems to the eHDSI through a dedicated national contact point for eHealth (NCPeH) which are actually specialized NCP for health data exchange. Especially, for patient summaries there are extensions such as the “patient summary for unscheduled care” use case that addresses cases where the patient is a frequent visitor in the country of treatment and has already some information from previous encounters, from eHealth Network [eHN16].

5.2 Remote Health Monitoring

When designing the PANACEA cybersecurity toolkit, it is important to study cases where IoT systems (wearable devices and companion mobile apps) have been used in remote health monitoring. The outcome is an understanding of the mechanisms behind capturing of Real-World Data (measurements) by IoT systems and the transmission of those measurements to a healthcare data management platform.

The global smart healthcare market is expected to reach $169.30 billion by 2020 with a prominent role for remote monitoring [Technavio18] (Figure 5-3).

Figure 5-3 Smart healthcare market forecast and the role of remote health monitoring by technavio [Technavio18].

Currently, about 10% to 15% of trials are incorporating wearable devices, primarily to collect data as exploratory endpoints [PharmaVoice19], while Kaiser Associates, based on research conducted by Intel, estimates that by 2025, 70% of clinical trials will incorporate sensors. There is an increased utilization of wearables by consumer for monitoring their health, with devices from about 500 companies providing wearables and patient monitoring solutions (Figure 5-4) [HGP19].

Figure 5-4 Wearables and remote patient monitoring market map provided by Healthcare Growth Partners [HGP19].

In the following paragraphs, use cases from some of these devices, representative of key remote sensing categories, are presented.

5.2.1 Smart pills: Proteus Discover

Proteus Discover [Proteus20] is an IoT system comprising of a pill hosting ingestible sensors, a small wearable sensor patch, an application on a mobile device and a web portal (Figure 5-5). Once activated, Proteus Discover provides insight into patient health patterns and medication treatment effectiveness, leading to more informed healthcare decisions for everyone involved

Figure 5-5 Proteus discover system for remote monitoring of medication adherence [Proteus20].

These sensor-enabled pills are combined with the regular prescribed medication. These pills, once in the patient’s stomach, monitor activity patterns of the patient. The wearable patch receives all the recorded information, providing a clear image about the treatment’s process. The patient can monitor the process through the mobile app, while the doctor utilizes the web portal. Apart from receiving the medical signals, the patch also measures steps, rest levels, heart rate and blood pressure.

Proteus Discover has been a joint research activity of University of California - San Diego, Stanford University, HHS and Orange County Health Care Agency. An independent study suggested that wirelessly observed therapy using Proteus' system could be considered equivalent to in-person medication adherence programs, but is much better received by patients taking oral medications [Mobihealthnews19].

5.2.2 Smart pill bottles: AdhereTech

Adhere Smart pill bottles [AdhereTech19] (Figure 5-6) carry the necessary sensors to detect when a patient removes a pill or some volume of liquid of the medication inside them. This ability helps investigators and doctors to secure the treatment process of their patients. Notifications are sent to patients if the smart bottle detects any medication non-adherence.

Figure 5-6 Adhere smart bottles [AdhereTech19].

The bottles are also collecting and transmitting real world data, to be analyzed and used in populating the AdhereTech’s secure dashboard.

AdhereTech’s programs result in patients staying on the medication longer, with increased adherence while they are on therapy. This drives additional fills of medication per patient per year (PPPY), as compared to the baseline metrics. The system is being successfully used [AdhereTech19], amongst others, by:

- Avella Specialty Pharmacy (UnitedHealth Group), achieving 1-2 additional fills of medication per patient per year, time on therapy increase by 26%, fill rates increase by 9% and dose-level adherence increase by 15%.
- Diplomat Pharmacy (largest independent pharmacy in US), achieving 1+ additional fills of medication per patient per year, patient retention increases by 12%, significant increase in adherence and persistence and gap days’ reduction by 19 days per patient per year.
- US Bioservices (AmerisourceBergen), achieving 1+ additional fills of medication per patient per year, time on therapy extension by 1.5 months, patients new to the medication having 1.3 additional fills and those already on the medication having 1.2 additional fills
- Clinical Use-Case: Avella Specialty Pharmacy (United Health Group), achieving median adherence rate of 100% for AdhereTech patients, AdhereTech patients taking nearly 4x as many doses in on-time window vs. control group, very positive feedback from patients and reduction of the need for costly interventions by 90%

5.2.3 Continuous Glucose Monitors: Freestyle Libre

One in eleven adults, aged between 20-79 years, have diabetes, and one in two adults with diabetes are undiagnosed [DiabetesAtlas20]. On account of that, investigators are researching, developing and utilising Continuous Glucose Monitors so as to observe glucose levels of patients avoiding taking blood sample.

The FreeStyle Libre system measures glucose levels through a small sensor (the size of two stacked quarters) applied to the back of the upper arm. It provides real-time glucose readings for up to 10 days, both day and night. The sensor can also read glucose levels through clothes, making testing discreet and convenient. It is accompanied by a touch-screen reader holding up to 90 days of data, which allows people to track their glucose levels over time (Figure 5-7).

Figure 5-7 FreeStyle Libre Continuous Glucose Monitor [DiabetesAtlas20].

Studies show that FreeStyle Libre users who scan more frequently spend less time in hypoglycemia and experience improved average glucose levels. According to a study published in The Lancet, people using the FreeStyle Libre system spent 38% less time within hypoglycemia as compared with those who managed their glucose with traditional self-monitoring glucose system [Abbott20].

Freestyle Libre was is with prescribing guidance across the NHS in London. NHS London Procurement Partnership (LPP) were asked to facilitate the production of a pan-London clinical consensus for the use of FreeStyle Libre in the NHS [LondonCN18].

5.2.4 eClinical Systems: Healthentia & Fitbit

Healthentia is an eClinical platform [Healthentia20] responding to two of the biggest challenges in Clinical Research: patient retention and data integrity by improving patient experience and communication between stakeholders in the eClinical market. It captures patient and clinical outcomes from mobile, medical and IoT devices, using a patient-centric app and offers research services to Sponsors (Pharma), Investigators & Contract Research Organizations (CROs). Healthentia employs the Fitbit [Fitbit20] wearables to measure behavioral Real-World Data of the patients.

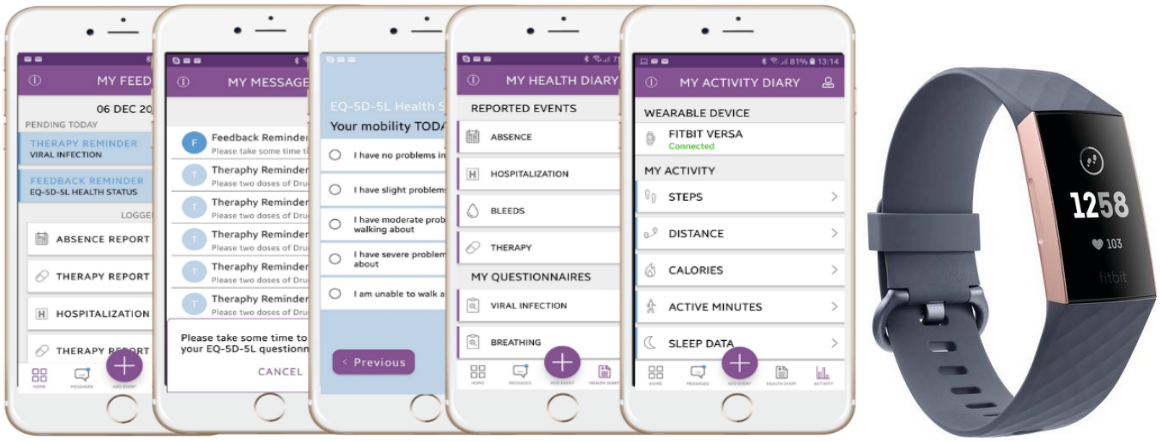

Figure 5-8 Healthentia used to monitor patients in clinical trials using a Fitbit tracker [Healthentia20].

Patients and Investigators which use the Healthentia solution, can monitor the clinical process through an eCOA and ePRO given environment. Real World Data are measured by the wearables and are reported online by the patients. The patients to monitor their physical condition and their trial process from the mobile app, while investigators monitor multiple patients remotely from the web portal. Healthentia is being used by CrosNT and Arithmos.

Fitbit is not the only activity tracker employed in patient monitoring; an increasing number of them and smartwatches are used beyond wellness devices, entering the realm of medical ones:

- The Samsung Gear S3 and Galaxy Watch have been used by Kaiser Permanente to monitor patients in home-based cardiac rehabilitation [Samsung20].
- The Apple watch has been used from Ochsner Health System in New Orleans to help monitor and manage chronic diseases [Miller19].
- The Pebble watch has been tested in four different experiments to demonstrate the usefulness of monitoring and collecting physical data [Zheng14].

IoT devices have penetrated patient monitoring, opening up remote and unobtrusive monitoring opportunities. PANACEA will be providing the means to secure the exchange of the data of the patients between such devices and the healthcare platforms by means of guidelines for Machine-to-Machine identification. These guidelines will be implemented on a use case employing Healthentia and an assistive device, QTrobot

6. User Authentication Protocols and Identification

In this chapter, we present user authentication strategies in a general way. First, we will discuss authentication methods and their suitability for use in healthcare. Then we present how user authentication can be embedded into the overall user right enforcement of the hospital, as we will demonstrate in PANACEA. We also present FIDO, which is very suitable in healthcare mostly at the patient level, when patient have to be authenticated to make sure they are associated to the suitable patient record, or in the context of telemedicine. This is beyond the scope of PANACEA in term of what will be demonstrated. Finally, the HEART protocol, which is currently being developed to take care of the authorization that patients may give to medical personnel to access their medical record, is presented.

6.1 User Authentication Methods

In this section, we discuss the different authentication methods, and, as much as possible, compare them. omparison is in fact clearly dependent of the usage considered, and our focus is for a usage in Health.

There are mostly 3 ways of checking people identity, using

- What they know (password)
- What they have (e.g. a telephone, a badge)
- What they are (biometrics)

Therefore, in this section we look at the authentication methods one by one (password, hardware, biometric), with a bias on usage by medical personnel, and will also briefly address multi factor authentication. This work is intended to provide support to Task 4.2, ‘Secure identification and authentication systems.

For usage in the medical domain, we believe that

- Ease of use is extremely important, especially nothing difficult to remember or to do
- Speed is key as medical personnel may work in emergency situation
- In case there are accuracy trade off like in biometrics, it is tolerable to accept some false positive (e.g. someone not allowed to use a medical device will get access to an account with this permission), rather than not giving access to someone who has a vital need for it

Previous considerations have been reinforced by the stakeholder analysis performed during Year 1 of the project and collected in D1.2 and further confirmed by PANACEA HCO end-users.

6.1.1 Password Authentication

Passwords are the easiest authentication means: they do not require specific hardware and their verification is extremely easy. However, passwords are suffering a number of vulnerabilities [Galbally17].

There are a number of “intrinsic” vulnerabilities that even extra safe behaviours of the user will never get rid of them, such as:

- As passwords are typed, they can be seen by someone behind the user – even if their exact characters are hidden (e.g. appearing as “stars”), just by looking at the keys pressed
- Passwords are easily forgotten, and unreliable ways of password recovery are proposed (e.g. questions such as “name of your first pet” or “first name of your grandmother” are not safe). Password

loss management has high cost, either in financial terms (e.g. around 70$[17]), or loss in working time (e.g. 1 hour of work per user and year[18])

There are also “content” vulnerabilities when the user selects a password that is not robust:

- Password is too common like *1122334455*
- Password is too short and therefore easier to find by brute force attack
- Password is not mixing up all available characters on the keyboard
- Password can be found in a dictionary, or as a concatenation of words of a dictionary, not modified, or modified only by very traditional “swap” methods (e.g. @ for a, € for e, 1 for i, $ for s, etc.)
- Password contains data that can be found by social engineering (e.g. the data of birth of your mother or the name of your pet)

Finally, when the way a user handles his or her password is not safe (“Management” vulnerabilities):

- Passwords are same for all the services used and sites visited (e.g. for my bank and for my favorite book store)
- Passwords are not changed or not changed often enough, or changed using a “sequential marker”
- Passwords are written in readable form on paper or in a computer file without any encryption
- Passwords are shared with colleagues or relatives “to make work easier”
- Passwords are managed through a social network, e.g. Facebook, and you are not sure what they do with it
- Passwords are stored in your browser(s) and can be found by an attacker

In PANACEA, partner IDEMIA has tentatively defined 3 levels of robustness for authentication relying on passwords:

- Lower level: Passwords content and management are against the rules that makes them more secure (see above)
- Average level: Password content and password management “rules” are followed, but relying on the user’s memory only (very unfriendly and high risk of loss)
- Fair level: Average level + passwords are managed automatically by a password management system that generates strong passwords, remind you to change them very often (or change them itself), and keep them safely encrypted (but you still need to remember and change at least two passwords: the password management system password and your computer login password: “unfriendly but manageable”)

As a conclusion, passwords are at a first glance simple and convenient, and do not generate a lot of cost, but a closer look shows that getting even an average level of security from passwords is complex and is not for free.

6.1.2 Hardware based authentication

Hardware is convenient because you carry in your pocket a number of hardware pieces that can be used as tokens: your telephone, its SIM card, your cars key, a badge to enter your office facility.

---

17 http://www.sparkhound.com/blog/does-one-password-reset-cost-your-company-7-or-70-every-time-the-password-is

18 https://specopssoft.com/blog/the-true-cost-of-password-resets/

However, hardware alone is not a proof of identity: it is easily stolen, easily lost, and also easily transferred to another person. Therefore, hardware can be used mostly in multi factors authentication scheme. Most of the time it is used is in “one-time passwords” solution, but this approach is not attractive for busy medical practitioners.

A 3-level graduation was proposed by NIST for digital identity “tokens” [NIST-auth] – however they are not just tokens, they include one or more authentication factors. This is very similar to what is required for securing exchanges with IoTs.

- Level 1: no security: the token has something like an identifier that is associated to a person and it can be checked this identifier is known and (but the identifier can be copied, or the token can be reproduced)
- Level 2: the exchange of the identifier (or any kind of secret) is done with state-of-the-art cryptographic techniques (using keys, could be symmetric or asymmetric cryptography).
- Level 3: the same as level 2, however the keys are in a “hardware wallet” (also known as a secure element) within the token, therefore they cannot be altered.

Obviously when hardware tokens are used for authentication, a token cannot be considered in isolation. The check involves

- The token reader
- The computer that identify and authenticate user
- The network(s) in between

Therefore, the token robustness is not really a token issue but an end to end issue. And key management on the computer side must also be secured for level 3 (using a software wallet or a hardware wallet).

6.1.3 Biometrics

Biometrics is an interesting authentication mean since a) cannot be forgotten, and b) cannot be lost.

However, it is not perfectly safe:

- No biometric algorithm is perfect….
    - There may be mistakes
        - Tell you “you are not Jane” – while you are actually Jane (False non match aka false reject)
        - Tell you “You are Jane” – while you are Barbara (False match also known as false accept)
    - They may not acquire the biometric characteristic properly (failure to enroll), e.g. do not find ridges in a finger image because the finger belongs to someone handling caustic material
- Biometrics may not be available e.g. fingerprints in the professional domain for people who have to wear gloves at all time, or finger injured hidden by a bandage
- Biometrics may be stolen (e.g. the skin of your finger “peeled” or your finger cut) but this is extremely exceptional
- Biometrics may be “counterfeited” (e.g. false finger made of silicon, fake mask, etc.)
- Biometric devices can be spoofed by stolen or counterfeited biometrics

The usual graphic to represent biometric devices accuracy trade-offs is presented below:

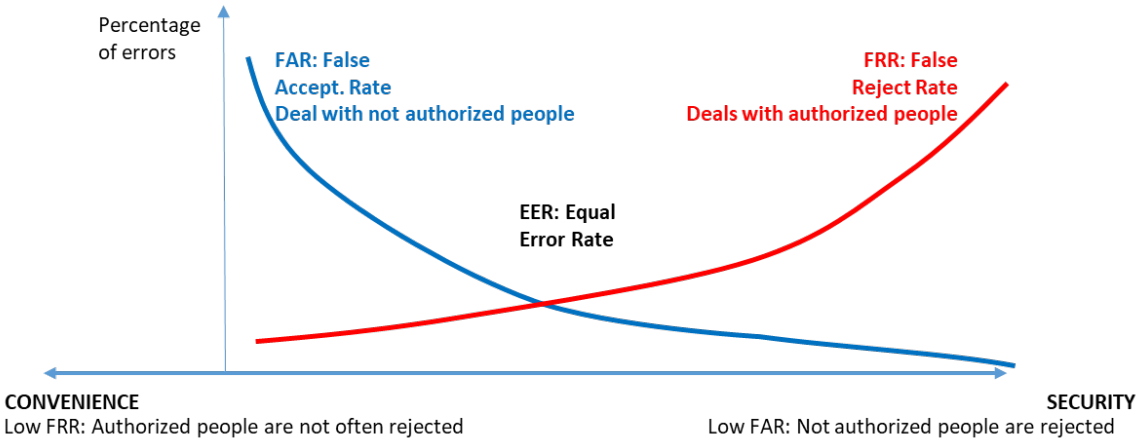

Figure 6-1 Biometric accuracy trade-offs

Overall ranking of biometric authentication robustness in simple categories, as for instance with passwords, is hazardous because the “weakness” criteria are mostly continuous.

Depending on the usage security can be favored over user friendliness or vice versa [NIST-SOFA]

- Security: very low False Acceptance Rate (FAR, or False Match Rate), difficult to spoof
- User friendliness: low False Rejection Rate (FRR, or False Non Match Rate), low failure to enroll

Obviously in the hospital, as explained in the introduction of this section, user friendliness is very important because medical staff has first to save lives.

Tentatively we propose the following classification when user friendliness is a “must have” criteria:

- Lower level: acceptable user friendliness (e.g. less than 2% FRR and less than 2% failure to enroll in operational condition)
- Average level: Lower level + low False Accept Rate (less than 1% FAR)
- Good level: Average level + difficult spoofing attack (benefit of spoofing on the average not worth the effort needed to succeed)

Different modalities have different characteristics:

Modality Criteria	Iris	Face	Finger
Easy to acquire?	Not so easy	Very easy	Easy
FAR / FRR tradeoff?	Very good FAR/FRR trade-off can be reached	Fair FAR / FRR trade-offs can be achieved, still progressing fast	Good FRR/FAR trade-offs can be obtained, very good with multiple fingers
Easy to spoof?	Difficult	Feasible although liveness detection is progressing	Depends on sensor: sensors exist with which spoofing is extremely difficult

6.1.4 Multi factor authentication

Since the end of 2019, in EU it is now mandatory to use a two factors authentication approach for financial transactions (PSD2[19]). The reason is that it is proven that two factors authentication is definitely better than one factor only. And obviously two factors authentication is also preferable for health-related applications.

Implementing two factors authentication does not require that factors are independent in their management: mostly when hardware tokens are involved, the two factors are interdependent.

For the medical domain, it seems that passwords are quite inappropriate because you need to type them, something which takes time, and mostly because medical practitioners would need to remember potentially numerous passwords.

Multiple biometrics is a challenge in Europe. Due to GDPR, and following the rule of “proportionality”, requesting the use of two different biometrics would correspond to extremely high security level requirement, probably higher than what hospitals actually require.

Therefore, our proposal for authentication of medical staff is to use a hardware token and one biometric characteristic. Face seems the most appropriate one, because it is compatible with wearing gloves, is evolving to be robust to medical masks, and because it is far easier to acquire than iris.

6.2 Single Sign-On and Federated Identification

User authentication is challenging in a distributed environment such as the one assumed by the Information Sharing platform since the same user needs to contact (directly or indirectly) many different systems and organizations.

The following are the typical components (actors, roles) that we can identify during such interactions:

- An Identity Provider (IdP) is responsible for
  - Managing the identities of people allowed to access all or part of the services of the part of environment it controls, and
  - Authenticating users, when they claim an identity, i.e. checking that they are who they claim to be, and
  - Providing identity assertion to other services called service providers.
- A service provider is an accessible service within an architecture that serves protected resources.
- A policy decision point is an entity that takes authorization decisions.

Direct and independent authentication of users on each service (service provider) is not practical or advisable for the following reasons:

- Users using different kind of services from the same access point would need to re authenticate each time they access another service
- Services would implement a diversity of authentication methods (e.g. one service using fingerprint, another one using a certain password and a third one using another password), causing cost at the equipment and confusion among the users.

[19] https://ec.europa.eu/newsroom/fisma/item-detail.cfm?item_id=658958&utm_source=fisma_newsroom&utm_medium=Website&utm_campaign=fisma&utm_content=Payment%20services&lang=en

- If authentication becomes a nightmare, users would finally use workaround and make it inefficient.

In a “brokered” authentication scheme an authentication broker (such as an IdP) is responsible for authenticating the users and issuing identity tokens to all relevant services (Figure 6-2). The user can then use such identity tokens to access the third-party services. To avoid that the user needs to provide his credentials each time he accesses a different service, the authentication broker can keep the authenticated session open. This results in new identity tokens being issued automatically (Single Sign-on, SSO) for each service the user accesses as long as the authentication broker's authenticated session is still active.

However, when users are requesting access to different physical machines (e.g. medical devices), then they need to be authenticated again for each physical device they want to get access too: what matters here is that they use the same authentication means at all time.

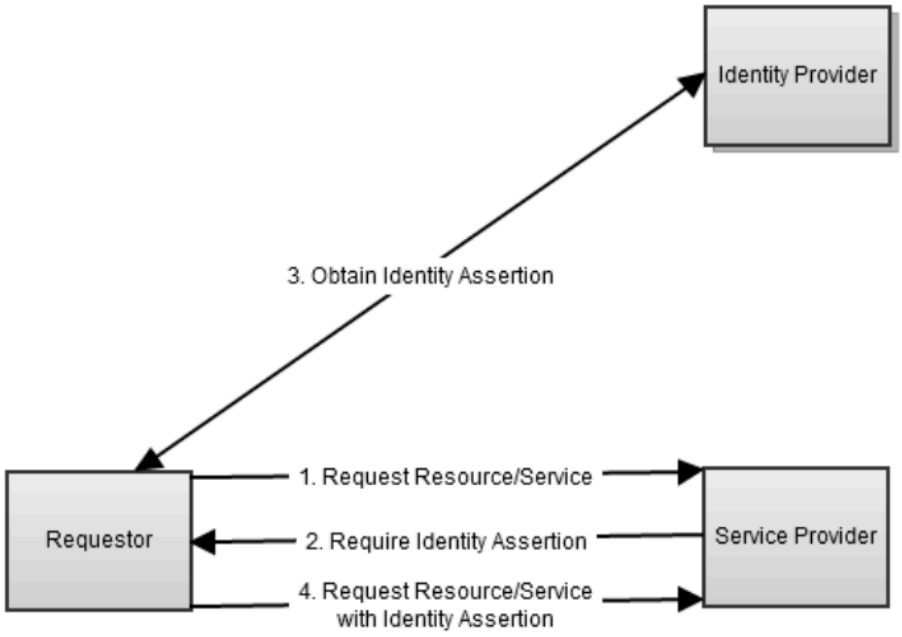

Figure 6-2 Brokered authentication [WS-Fed]

Brokered authentication (Figure 6-2) is initiated when a user (the requestor) requests a protected resource from a service provider (1). The Service provider replies that he needs an identity assertion (2) to be able to verify whether the user is allowed to access the requested resource. The requestor authenticates on a supported identity provider and requests an identity assertion (3). The requestor sends the assertion to the service provider as claim who he is and, if he gets access, receives the requested resource (4).

As a design pattern this authentication flow allows for user’s *federated identity* and Single Sign-On (SSO) since the users acquire authentication tickets (*tokens*) that are shared across multiple IT systems or even organizations. While SSO allows a single authentication credential to access different systems within a single organization, a federated identity management system provides single access to multiple systems across different enterprises. Examples of brokered authentication include OpenID and OAuth, as well as Shibboleth[20], which is based on OASIS SAML [SAML].

[20] https://www.shibboleth.net

6.3 FIDO

The Fast Identity Online (FIDO[21]) alliance is an industry specification group that was founded a few years ago and has now grown to more than 250-member companies. The goal of the alliance is to define an interoperable specification for mobile authentication and to overcome existing fragmentation and silos.

Technically, FIDO concentrates on authentication and explicitly excludes identity and ID federation. It can however be embedded into identity schemes and combined with ID federation, although not directly supported by the FIDO protocol. Steps like the initial identification and the identity or attribute attestation are not part of the FIDO protocol but can be performed on top of the FIDO authentication step (i.e. prior or afterwards)

FIDO comes with two flavours of the protocol, the universal 2nd factor (U2F) protocol for two-factor authentication and the UAF-protocol (Universal Authentication Framework) for password-less authentication (e.g. using mobile device biometrics) and transaction signing. Both protocols are summarized under the FIDO 1.x specifications and have been unified in the FIDO 2.0 specification. The approach of FIDO is to balance security on the one hand with usability on the other hand and to overcome existing technology silos by creating an open landscape for authenticators.

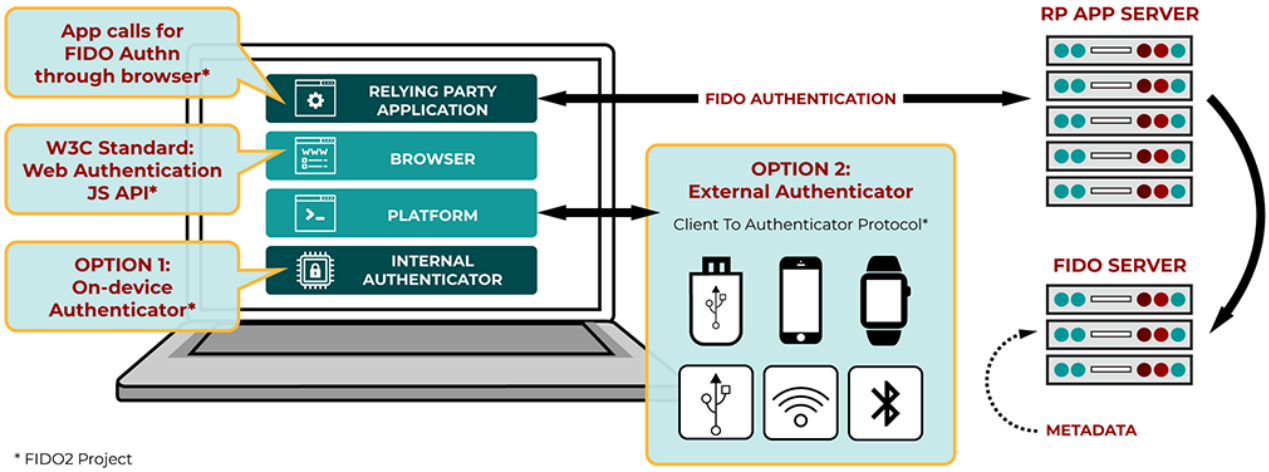

Figure 6-3 FIDO components and specifications [FIDO]

The principle of FIDO is based on simple challenge-response protocols using asymmetric keys. In contrast to previous PKI-based systems FIDO wants to explicitly reduce complexity by restricting PKI to the absolute minimum. As a consequence, the user-centric registration triggers the generation of the FIDO key pair and exports the public key to the service provider while the private key is kept on the user side. No further PKI is used in the authentication step. However, especially in the context of identity applications it would be possible to provide further PKI (e.g. a certificate over the public FIDO key) on top of the standard FIDO concept.

6.4 Health Relationship Trust (HEART)

Health Relationship Trust is a still-in-progress effort[22] to define the interoperable process for systems to exchange patient-authorized healthcare data enabling patients to control how, when, and with whom their clinical data is shared [HEART]. It is based on existing modern (state-of-the-art) security protocols and adds additional components to ensure that patient clinical data is securely exchanged. The basis is the following open standards:

[21] https://fidoalliance.org/
[22] https://openid.net/wg/heart/

- HL7 FHIR (Fast Healthcare Interoperability Resources) “RESTful” application programming interface (API)
- OAuth 2.0, which is an open standard supporting delegated authorization, i.e. allowing a “resource owner” (user, patient) authorize a client software to access protected data on their behalf, effectively removing the need to share their access credentials with the client software. [RFC6749]
- OpenID Connect which provides an open authentication protocol, more lightweight than SAML, that provides SSO, session management, and identity claims retrieval built on top of OAuth 2.0 [OAUTH]
- User Managed Access (UMA) is a protocol built on top of OAuth 2.0 that allows the resource owner rich control over access to their resources through the use of an authorization server of the resource owner’s choosing, either to software that they control or to software that another user may control [UMA]. The UMA protocol allows therefore user-to-user delegation, and handling of multiple authorization servers per resource server. [OAUTH]

In the specific use case of health information sharing HEART with its UMA infrastructure will allow the patient, being the “resource owner” of own medical history, to use his/her own authorization server and introduce it to the health information system (“resource server”) that manages his/her data. A health care professional, being a “requesting party” in UMA’s parlance, can discover the patient’s authorization server once s/he tries to access patient’s data using the web-based FHIR API. HEART therefore allow patients to set up permissions and authorizations for sharing their clinical data to ensure that their data is only shared with individuals, institutions, and apps that they choose and thusly enables patient-mediated interoperability for FHIR APIs.

7. Cloud Computing, Big Data, and IoT in Healthcare

Cloud computing and big data technologies have been pervasive in various industries such as marketing, finance, insurance, e-commerce, etc. They have been slowly adopted in the healthcare domain where there are more stringent requirements regarding trust and data privacy, availability, and legal and ethical regulations. In this chapter we explore both technologies together with the Internet of Things (IoT) focusing on their application in healthcare.

7.1 Cloud Computing

Information used for health care today largely comes from self-report surveys and frequent doctor hospital consultations and treatments. Personalized health systems and pervasive mobile monitoring technology expands the scope of potential cases making it possible to sense, mine and learn human behaviours and intentions in order to provide personalized feedback fine-tuned to individual users, corresponding timing and location. The term "Cloud Computing" implies the provisioning of computing resources, both hardware and software, in the form of a service [NIST-cloud]. These services are characterized by certain qualities, such as availability on-demand, self-service, broad-band access, high availability and rapid elasticity. The “cloud” infrastructure technology is the evolution of traditional technologies for sharing and managing large sets of IT resources, such as data networks, data centres and computation clusters, and is usually characterized and distinguished from such technologies by the dynamic allocation.

In principle, we can distinguish the following three major service models [Zhou10]:

- **Infrastructure as a Service** (IaaS): IaaS is the supply of hardware as a service, that is, servers, network technology, storage or computation, as well as basic characteristics such as Operating Systems and virtualization of hardware. Commercial IaaS platforms are offered by Amazon, Microsoft, Oracle, etc.
- **Platform as a Service** (PaaS): At the PaaS level, the provider supplies more than just infrastructure; i.e. an integrated set of software with all the stuff that a developer needs to build applications, both for the developing and for the execution stages. Examples of commercially available PaaS platforms include Google App Engine, Microsoft Azure Services, etc.
- **Software as a Service** (SaaS): In this service model the provider offers software as a service and this was one of the first implementations of Cloud services. Examples of SaaS platforms include the Google App, Microsoft Dynamics CRM online, and others.

Currently, there are also three deployment models for cloud computing. Based on the location and who manages it, a cloud can be defined as:

- **Private cloud**: It is deployed within an organization's infrastructure and the resources are dedicated to the organization itself. Management and resource allocation are also controlled in-house by the organization.
- **Public clouds**: are open to public, but resources and infrastructure are owned by the organization providing the cloud service. Although public cloud providers often ensure client’s data security and integrity, data control, especially for sensitive data, can always be an issue.
- **Hybrid clouds**: they combine both public and private clouds using data and application migration techniques.

The value brought by the cloud computing especially in the biomedical domain can be immense but there are certain trade-offs to be considered [Rosenthal10]:

- From the business perspective Cloud can reduce the operational costs of the management of large data sets, the system administration costs, maintenance of the infrastructure, keeping backups, etc. The flexible provision of computation and storage resources according to dynamic needs result in better utilization of the resources, which could not be the case when a healthcare organization manages own hardware and software on premises.

- Cloud provides a lot of flexibility in allocating resources according to demand, and a health care organization or research consortium can request more resources for better performance and scalability [Dudley10]. Nevertheless, this flexibility should always be quantified in terms of pricing, because there may be a certain point where having all the infrastructure in house is more beneficial than using a commercial cloud.
- High availability, resilience, fault tolerance, and low-level security are aspects taken care by the cloud provider, and health care organizations should have a service level agreement and relevant contracts in place to specify these requirements.
- Collaboration and sharing between healthcare organizations become easier and faster, in the case of course that the different parties agree on the use of a common cloud provider and the associated infrastructure. Unified access policies, authentication mechanisms, and other security related dimensions are more easily designed and implemented. Even for a single organization, applications available through the cloud have the benefit to be accessible from anywhere, subject only to the internal policies of the organization. On the other hand, this potentially opens the attack surface since now the data are transferred and stored elsewhere and for example physical boundaries such as the organization’s walls are not in place in more.
- Legal and legislation-compliance issues need to be addressed, especially since GDPR has been established throughout Europe. Again, with cloud’s easy access to the data it can be difficult to keep track who has access to the data and under which circumstances so healthcare organizations must have steadfast rules governing information sharing and access. The cloud can be supportive, especially in case of PaaS and SaaS, and at least all major cloud providers like Amazon have data centres in Europe so that data of the European citizens are not hosted overseas. But a healthcare organization adopting a cloud solution should be aware that by uploading their data to the cloud they are relinquishing direct control, and should be assured that data protection mechanisms are in place, along with transparent information about possible security incidents. Also cloud providers should be checked that important rights, such as the right to erasure (i.e. removing patients’ data on their request), are fully respected.

In PANACEA we foresee that new care models incorporating advanced information and communication technologies have the potential to provide service platforms able to improve health care, personalization, inclusion and empowerment of the individual. A major challenge thus is related to caring patients is the early detection of exacerbations of the disease providing personalized, accurate and fully automated emergency alerting systems that smoothly interact with the personal health professional, regardless of their physical location in order to ensure in time intervention in case of an emergency. Integrated scalable cloud systems have the power to enable advanced monitoring of people able to detect possible pathological conditions. The cloud therefore offers some important advantages but it also introduces challenges [Kuo11] [NIST11] so it’s essential that organizations should perform a proper cost benefit analysis and a thorough risk assessment before adopting it [ENISA09]. Apart from the control of the data and general security aspects [Michalas14], a public cloud also requires reliable network communication for the transfer of the data, some infrastructure to be in place in the case an organization wants to automatize the linking of internal systems with the cloud, and of course it requires the adoption of the “pay-as-you-go” (adaptive but may be difficult to predict) billing method. Some of these concerns can be alleviated by adopting a *private cloud* solution. This would mean that organizations do not cede control of their data but there should be an important upfront investment in acquiring hardware, recruiting specialized staff, etc. So, this solution most probably makes sense for specialized research consortia or organizations where the cost is distributed and the primary use case is the sharing of data.

7.2 Internet of Things

Crucial innovation is needed in the process of making and deploying large scale Information and Communication Technologies, empowering end-user services to be usable, trusted and. This will require multi-domain, multilevel, trans-disciplinary work, grounded in theory but driven by citizens’ and healthcare professionals’ needs, expectations and capabilities and matched by business ability to bring innovation to the market. Drawing from the literature, like the situation that is current in Europe with the pandemic of the coronavirus, modern societies must be prepared for critical infrastructure protection focusing on the need for resilience, examining strengths and weaknesses of traditional approaches risk assessment and preparation and on-time response.

With the advent of ubiquitous computing and communication technology, people are able to use information access technologies at any time and from anywhere through smartphones or tablet computers. New community-based arrangements and novel technologies can empower individuals to be active participants in their health maintenance by enabling them to self-regulate and control their health and wellness and make better health and lifestyle related decisions using community-based resources and services. Mobile sensing technology and health systems, responsive to individual profiles, supported by intelligent networking infrastructures and combined with cloud/IoT computing can expand innovation for new types of interoperable services that are consumer oriented and community based. This could fuel a paradigm shift in the way health care can be, or should be, provided and received, while lessening the burden over exhausted health and social care systems. Advances have been identified in the fields of systems for personalized health monitoring, such as smartphone platforms and intelligent ubiquitous smart applications; activity profiling and lifestyle capturing, enabling the recognition of human activity and lifestyles and the deployment of platforms for health assistance and interaction support. Delivering innovative ubiquitous eHealth and mHealth services, including citizen-cantered wellness goes well beyond the development of technical solutions.

Today the Internet of Things is not only a technological revolution that influences our daily lives, but also a great opportunity for a large number of players in industrial domains. To do so one of the most promising ideas is to realize a vision of an open service framework for IoT. This will allow IoT to become not only a reality but an actual commodity by encouraging key industrial players/stakeholders to participate and establish an ecosystem from which they will benefit. It will also allow developers, service providers, software developers, and network operators to develop and expand in terms of business.

As an example, we reference the OpenIoT (http://www.openiot.eu/) project that was first released in 2012. OpenIoT aims to create an open source middleware for getting information from sensor clouds, without having to worry about what exact sensors are used [Pereira19]. It also aims to explore efficient ways to use and manage cloud environments for IoT “entities” and resources (such as sensors, actuators and smart devices) and offering utility-based (i.e. pay-as-you-go) IoT services. As an output OpenIoT provide instantiations of cloud-based and utility-based sensing services enabling the concept of “Sensing-as-a-Service”, via an adaptive middleware framework for deploying and providing services in cloud environments. The main outcomes are the open source middleware with the same name OpenIoT and the OpenIoT Virtual Development Kit (VDK) that is a ready-to-use version for academic and training purposes. OpenIoT is as reference semantic-based platform in the current IoT-EPI (European Platform Initiative).

OpenIoT proposes an open service framework for the IoT – where a common service layer is created facilitating a uniform – necessary secure - entrance into the IoT-related mass market, establishing a new ecosystem enabling: device developers to produce IoT devices and register for the public/developer community, software development of Apps or Web-service programs for mobile devices, tablet, or desktop computers, which can connect and control IoT devices through networks, and finally register them on an App store site like Apple, service providers purchase IoT devices and register them on the open service framework, where the large and fast growing number of IoT devices could be monitored and managed efficiently-similar to the way network operator manage, manipulate and control passive networking equipment, network operator could capitalize on their mobile and wireless communication technologies, consumers could easily find, connect, and control them using IoT device searching and browsing service.

Toward that PANACEA will emphasize on the design, implementation and usage of IoT in the medical field. Medical IoT, sometimes referenced as healthcare IoT, indicates a growing number of IoT uses in the medical industry. These generate a wide range of IoT applications and devices specifically designed for healthcare needs and settings, such as sensors and apps for remote healthcare monitoring, consultation, and delivery. Medical IoT supports life-changing improvements to traditional medical devices, such as the smart inhaler for people with asthma. Broader IoT use cases can also apply to healthcare, such as leveraging IoT connectivity to monitor critical medical devices and equipment, and to receive alerts when they require maintenance or replacement. Below we summarize a picture including a map of connected medical devices – what is common OpenIoT does not actually apply! The only common in – between these examples is the model of business, communication capabilities and certification processes (similar to any medical equipment).

Figure 7-1 Medical IoT – map

From a security point of view any solution implemented must simply guarantee protection of any connected device -used for any medical purpose- from cyber-attacks throughout the entire product lifecycle, based on security analysis automation and real-time protection technologies.

A well written study from ENISA[23] states clearly that at the end everything will become connected- IoT Ecosystem. It is thus important not only to understand threats, asset but also consider the context of use. In that respect there are specific recommendation that must be follows and specific “good practices” to follow.

[23] https://nis-summer-school.enisa.europa.eu/2018/cources/IOT/ENISA_NIS_Summer_School_IoT_Security_1_2018.pdf

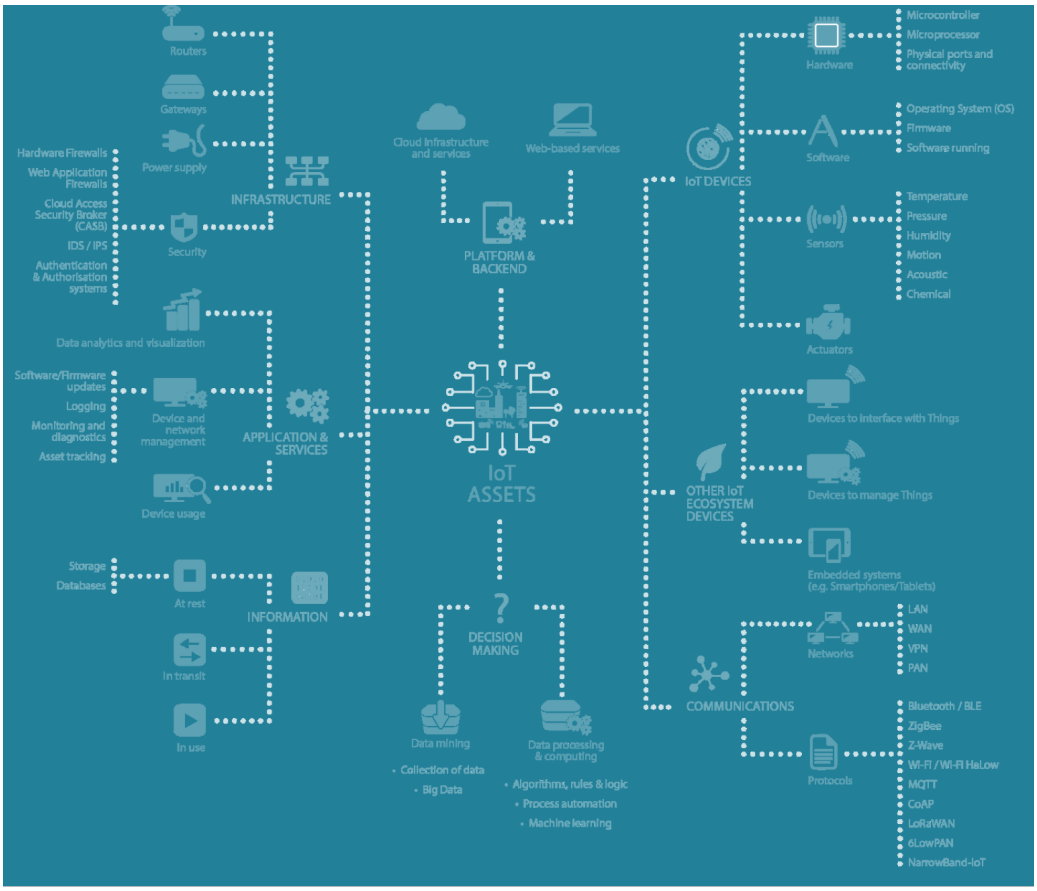

Figure 7-2 Medical IoT - context of use

From that analysis we see the following scenarios in respect of medical IoT and security threats (see figure below):

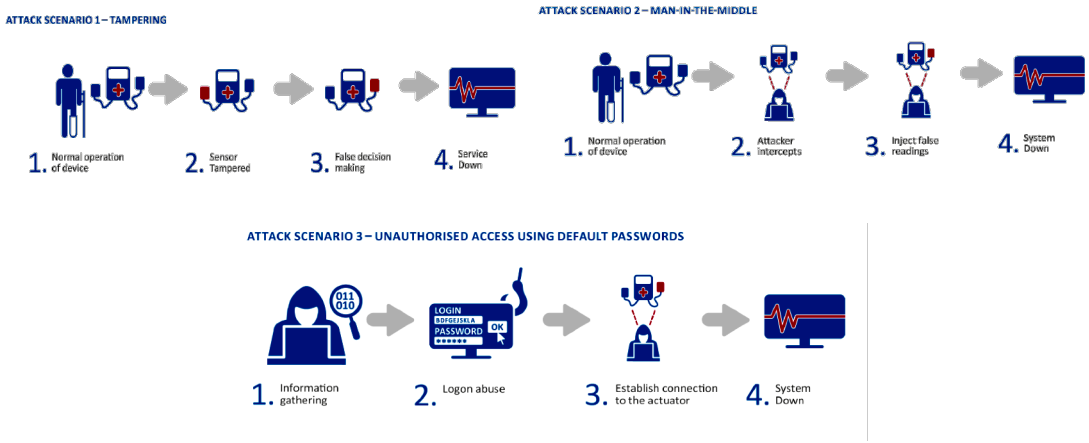

Figure 7-3 Medical IoT – attack scenarios

From all the above we realize that healthcare is a vast ecosystem, making applications for the Internet of Things in healthcare to be endless. It is a fact that smart devices and IoT have infiltrated into healthcare spaces. However, critical success factors of IoT adoption include not only the explosion of demand for services but also the response from vendors and provides. OpenIoT is a living result of building on top of many requirements an open architecture for IoT. On the other hand, medical IoT is another world where for sure we

want to apply similar principles of IoT to a healthcare model using examples as the Open IoT Platform. The ambition is to create an Internet of Medical Things-IoMT ecosystem able to empower patient/citizens in their daily care activities making them feel safer and be healthier, and improve how physicians deliver healthcare. To do so we know that in a vast expanding marker (see figure below) networked medical devices and applications in healthcare IT must change future strategies for healthcare organizations, and not only affect diagnostics, treatments and patient health management but also work around the following fact: more connected devices ➔ larger attack surface ➔ security is a significant challenge for healthcare organizations (where security is not suboptimal).

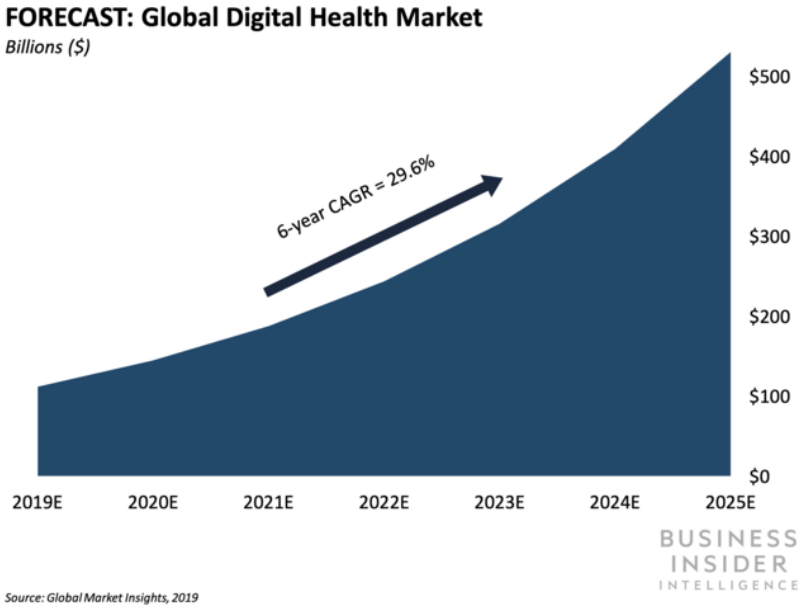

Figure 7-4 IoT Healthcare in 2020 market expectations (https://i.insider.com/5cffa94f6fc920124f62cc72)

One of the most important efforts in order to regulate that is the EU Cybersecurity Act[24] came into force on 27th June 2019. In a shift towards a role that adds more value to the European Union, ENISA, which will henceforth be known as the EU Agency for Cybersecurity and will receive a permanent mandate. In order to scale up the EU’s response to cyber-attacks, improve cyber resilience and increase trust in the Digital Single Market, the EU Cybersecurity Act: i) Strengthens ENISA, the European Union Agency for Cybersecurity to improve the coordination and cooperation in cybersecurity across EU Member States and EU institutions, agencies and bodies; and ii) Establishes an EU cybersecurity certification framework that will allow the emergence of tailored certification schemes for specific categories of ICT products, processes and services. Companies will be able to certify their products, processes and services only once and obtain certificates that are valid across the EU. This certification can play a critical role in increasing trust and security in products and services that are crucial for the Market. Without a common framework for EU-wide valid cybersecurity certificate schemes, there is an increasing risk of fragmentation and barriers in the single market. The EU Agency for Cybersecurity, ENISA, with the help of national experts will prepare the technical ground for the certification schemes that will then be adopted by the European Commission through implementing acts. The EU-wide certification framework creates a comprehensive set of rules, technical requirements, standards and procedures to agree each scheme. Each scheme will be based on agreement for the evaluation of the security properties of a specific ICT -based product or service e.g. smart cards. This certificate will attest that ICT products and services that have been certified in accordance with such a scheme comply with specified cybersecurity requirements. The resulting certificate will be recognized in all EU Member States, making it easier for businesses to trade across borders and for purchasers to understand the security features of the product or service.

[24] https://ec.europa.eu/digital-single-market/en/eu-cybersecurity-act

7.3 Big Data

Big data is usually defined as “high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation” [Gartner] but the definitions abound [Gandomi15]. In simple terms, big data have those characteristics that make their storage, management, and processing impossible for a single machine. From this simplifying definition it’s clear that the “bigness” of data is relative and moving as the developments in hardware and software proceed.

The healthcare is a typical domain that generates big volumes of data: the “traditional” electronic health records (EHR) that aim to store a patient’s whole history of examinations, operations, drugs, etc; the multi-gigabyte medical images (e.g. digital pathology); and the genomics and transcriptomics data sets that are now produced by “next generation sequencing” technologies are some examples. Sensor-enabled mobile phones have the potential to collect in situ continuous sensor data that can dramatically change the way health and wellness are assessed and monitored, as well as how self-management of health conditions is made, and care and treatment are delivered. New classes of applications are being explored both in academic- and industry- based research centres. Bringing together and correlating data among different and heterogeneous sensors would allow inference of new knowledge from these sources. There is a huge potential for time critical and computational heavy applications to routinely conduct large scale computations on distributed resources using edge serverless computing.

Figure 7-5 – Healthcare providers connected more closely with their patients[25]

In our effort to capture the existing vast amounts of big data and putting it to work for health care applications in order to support the digitalization of healthcare that is already underway, yet, because of the slower tempo of technological adoption by healthcare insiders, as compared to other industries, digitalization has not been so obvious. In that respect at PANACEA we emphasize in areas of research that we believe important to healthcare professionals and patients in years to come and we must ensure cybersecurity and acceptance.

[25] https://ai.myesr.org/healthcare/embracing-healthcare-4-0-digitalizing-healthcare-as-a-key-enabler-for-high-value-care/

- Internet of Medical Things (IoMT): Driving the next generation of connected healthcare through secure devices in terms of manufacturing (PANACEA: Security by design) and secure human interaction (PANACEA: user authentication and identification)
- Artificial Intelligence (AI) & Healthcare 4.0: The next digital revolution sweeping over the healthcare industry, this is all about capturing the vast amounts of data and putting it to work in applications, thus making healthcare management decisions better informed, while allowing for significant gains in efficiency and cost control. Also, AI can help reduce the amount of work, and keep the clinicians focused on the patients. (PANACEA: Risk assessment toolkit)

Most of that Cybersecurity for these kind of interventions is important and aims in managing the risks in this new digital age. With the monetization of patient records for nefarious reasons on the rise, cybersecurity has taken the spotlight in the information age. It will be up to suppliers and the hospitals to work together to deliver security to the patient's identity.

Health care organizations must practice good 'cyber hygiene' in today's digital world

In that respect the rapid growth present today has brought an increase in different types but also frequency of cyber-attacks. Many well-known cybersecurity solutions are in place to counteract these attacks. However, the generation of Big Data over computer networks is rapidly rendering these traditional solutions obsolete. Within PANACEA in order to cater for this, we focus on Security Analytics and Risk Assessment to assist in monitoring and assess of malicious and suspicious (outlying) patterns dangerous for healthcare organizations. Such a behavior is envisioned to encompass and enhance all traditional security techniques. In WP6 we will be able to demonstrate the integrated system including all different modules of PANACEA.

8. Secure Information Sharing: Deployment options

In this section, we describe and discuss some options to be considered when deciding how to deploy a system able to support Secure Information Sharing between Heath Care Organizations (HCOs).

Information sharing can be defined as the act of two (or more) parties to establish an agreement on passing information each other to increase the overall individual knowledge. As a consequence, it is fundamental that parties involved in the sharing explicitly manifest their interest in participating into the sharing process.

When Information Sharing has to be supported by a system, it is worth to consider the common *dependability* attributes. According to [Avizienis04], dependability is an integrating concept that encompasses the following attributes:

- *availability*: readiness for correct service,
- *reliability*: continuity of correct service,
- *safety*: absence of catastrophic consequences on the user(s) and the environment,
- *integrity*: absence of improper system alterations.
- *maintainability*: ability to undergo modifications and repairs.

A common technique to achieve service’s availability, reliability and integrity is to deploy it using replication techniques while concerning maintainability a typical approach is to use a modular design approach.

When adding considering also the security attribute we need to consider that it is a composite attribute made of the three CIA basic attributes i.e., Confidentiality, Integrity and Availability. The direct consequence is that to get security we need, in addition to reliability, safety and maintainability also the concurrent existence of the following conditions:

1) absence of unauthorized disclosure of information (i.e., confidentiality)
2) absence of unauthorized system alterations (i.e., integrity)
3) availability for authorized actions only.

Let us note that, when dealing with confidentiality of medical data (regarding EU patients), we need to consider also requirements imposed on the system by specific regulations and laws and in particular by GDPR.

Thus, before moving to the description of some possible deployment option, we will recap the key actors in the Information Sharing process with specific focus on medical data, we will describe the actual sharing pattern and then we will try to promote an innovative view.

8.1 Information Sharing in the Health Care: Key Ingredients

When considering Information sharing it is fundamental to understand:

- *who* the parties involved in the sharing are,
- *what* the parties want to share, and
- if the object of the sharing imposes particular constraints on the sharing process e.g., personal data require to be shared in compliance with GDPR.

Additional aspects may include:

- the *context* of the sharing, e.g. during a patient’s visit to a physician,
- the access needed, e.g. read-only or also modification, etc.

Concerning the parties, it is interesting to consider sharing between:

- Two (or more) parties within the same organizational entity (e.g., between parties inside the same organization but at different sites)
- Two (or more) parties belonging to different organizational entity (e.g., between parties in two different organizations but collaborating to reach a common goal).

Concerning the information to be shared and following GDPR, we can classify them mainly in two classes:

- *Personal data*: any data related to an identified or identifiable person, where “identifiable” means that the data allow to directly or indirectly identify the person they refer to (e.g., an e-mail address). Among the different categories of personal data, some are considered particularly sensitive and, thus, their processing is regulated differently from other personal data (cfr. art. 9, GDPR). Obviously, such categories of sensitive data include health-related and medical data.
- *Non-personal data*: any other data not encompassed by the previous definition.

Concerning non-personal data sharing, there exist solutions building up on distributed shared storages [Cassandra, DynamoDB, Bigtable] (also using blockchain technologies [Nakamoto2009, Ethereum2014, HyperledgerFabricProj]) possibly enriched with access control mechanism to enforce security requirements. When dealing with personal data, such solutions need to be adapted as there is an additional important aspect that must be taken in to account: the *data processing consent* provided by the data subject that regulates explicitly who is authorized to process data, the purposes for which the processing is allowed and the allowed duration of the processing.

Thus, in this chapter, we will focus our attention only on the sharing of personal data (including Sensitive Medical Information) requiring data processing consent. Let us note that, in some specific cases, medical information (e.g., data concerning health or a person's sex life or sexual orientation) do not require an explicit consent for their processing as they are specifically regulated by GDPR (cfr. art. 9, par. 2, GDPR). In order to simplify our presentation, we will assume, for those types of data, that a default consent exists so that we can consider them as a particular case in our analysis.

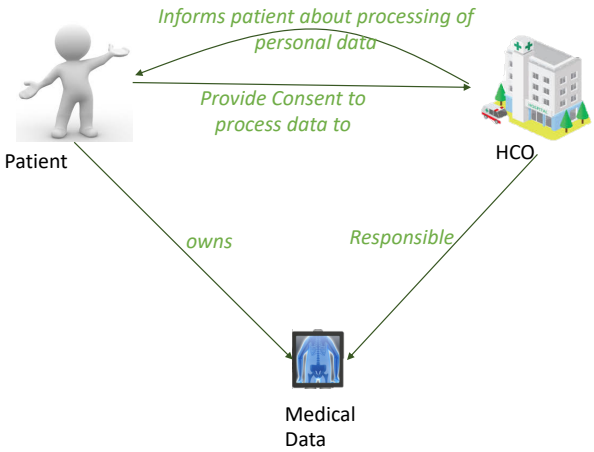

Figure 8-1 - Key Roles in the Medical Information Sharing Process

Figure 8-1 highlights the main entities involved when dealing with information sharing of medical data and their respective roles and relationships. In more details, we can identify the following entities:

- *Medical data*: they represent the pieces of information pertaining to an individual’s health that need to be shared among parties and that must be protected;
- *Patient*: is the person that is concerned by the medical data and that owns them (in GDPR terminology he/she is the data subject);

- *Health Care Organization (HCO)*: is the entity that produces, accesses, store and/or manipulate medical data and that is responsible for the data protection.

Let us note that every HCO is a composite entity as data protection responsibilities are spread across multiple individuals inside and outside the HCO.

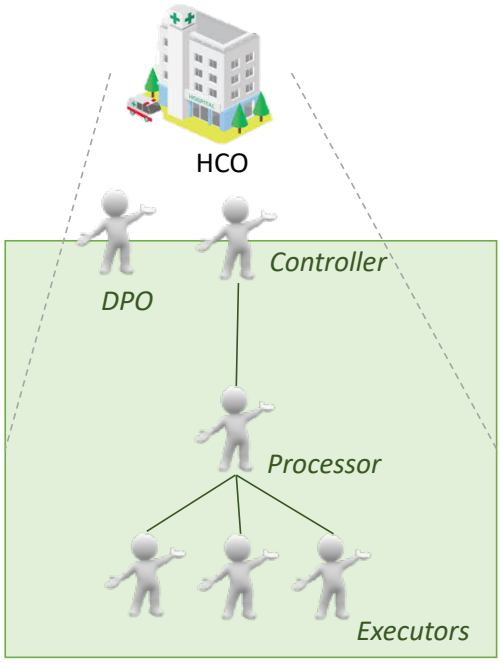

Figure 8-2 - Highlight on HCO roles responsible of the Medical Data Protection

Figure 8-2 reports the main roles involved in the data protection required by GDPR, which are:

- **Controller**: the entity (natural or legal person) which determines the purposes and means of the processing of personal data. The controller is responsible for ensuring and being able to demonstrate that processing is performed in compliance with GDPR. The controller is also responsible for maintaining records of processing activities. Note that the controller may also act jointly with other (joint) controllers within the same organization.
- **Processor**: the entity (natural or legal person) which processes personal data on behalf of the controller (typically under the constraints dictated by a contract). Note that a controller may authorize multiple processors (to the processing of the same data), and a processor may engage other processors (with explicit authorization of the controller).
- **Executors**: the persons, typically afferent to the technical staff, that are authorized by the processor to process personal data (under a statutory obligation of confidentiality). They are the actual executors of the processing of data.
- **Data Protection Officer** (**DPO**): a person designated by the controller and the processor, which has at least the following duties: (i) inform and advise the controller, the processor and the staff involved in processing of personal data, on the correct application of the provisions of the GDPR (including DPIA), (ii) monitoring compliance to GDPR, (iii) act as the point of contact and cooperate with the supervisory authority. Note that a DPO is not required to be afferent to the organization he/she provides his/her service to, and that the same person may be designated as DPO by multiple organizations.

Concerning the possible data processing and in particular the sharing of personal data (and particularly of medical data), let us recall that, according to GDPR, the following constraints should be satisfied:

1. The patient is and remains the owner of every medical data.
2. Except for specific categories of data, the patient must provide his/her consent to an HCO before it can process (including the sharing) data. Let us note that GDPR also specifies key roles that must be appointed inside every organization in order to process data properly.
3. The patient can modify his/her consent to data processing at any time.
4. Concerning the sharing, GDPR explicitly states that it is allowed for the minimum time needed.
5. The patient has his/her right to be “forgotten” i.e. the right to request the erasure of his/her personal data. This can be done at any time.

8.1.1 Information Sharing Flow in the Health Care Domain

In this section, we will introduce a simple scenario that will allow us to discuss the general “as is” situation and point out the directions towards the “how it could be”.

Sharing Scenario

Bob is a patient that did some specialized medical exam in the hospital HCO_1. The results of Bob’s exam are produced and stored by HCO_1. After some time, Bob needs to take other exams at hospital HCO_2 and he would need to have shared the results of his previous exam between HCO_1 and HCO_2.

Information Sharing Flow: “as is” case

Now, we will briefly describe the interaction pattern that is currently in place when two HCOs need to share medical data of a patient as in the following scenario.

When Bob arrives at HCO_2, HCO_2 needs to take the following steps in order to get the data shared:

1. Bob should bring the data himself from HCO_1 to HCO_2. If this is not the case, then
2. HCO_2 asks (typically with a physical interaction) to HCO_1 to share Bob’s exam results
3. HCO_1 checks if it is allowed, according with Bob’s data processing consent, to share Bob’s data with HCO_2. Also in this case, this action is performed by manually checking the papers filled in and signed by Bob when he did exams at HCO_1.
4. Two cases may happen:
    a. Bob signed a general sharing consent that allows to HCO_1 to transmit the data. In this case, HCO_1 is authorized to share the data and can send them to HCO_2. Currently, this option is rare and when it happens, data are typically transferred manually (by courier or trough Bob himself).
    b. Bob did not provide the consent yet. In this case, HCO_1 cannot share the data.

This sharing pattern is depicted in Figure 8-4 - Futuristic sharing pattern for medical data

.

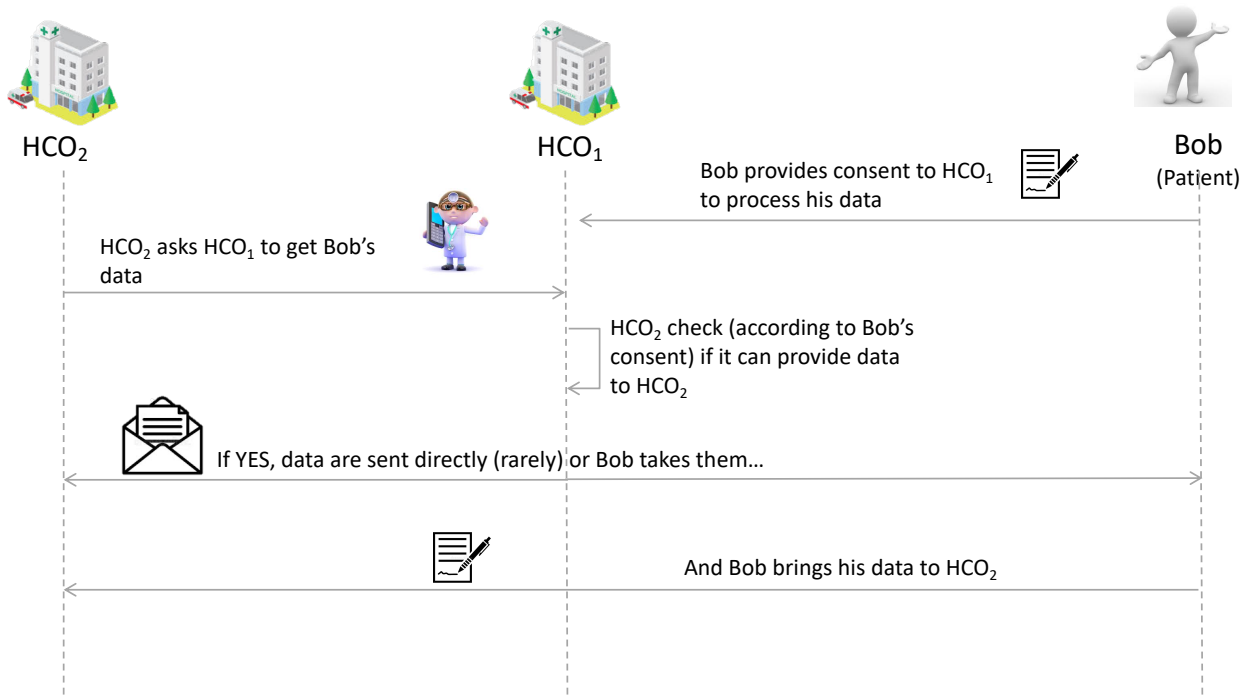

Figure 8-3 Actual sharing pattern for medical data

The main limitation of the current sharing pattern can be summarized in:

1. Sharing medical data may require time and possibly multiple interactions, also involving the patient in the loop.
2. Sharing medical data is currently a physical point-to-point interaction. If the same set of data need to be shared with multiple parties, it would require multiple sharing patterns to be in place.
3. Sharing is currently asymmetric. HCO_1 may have the consent to share Bob’s data with HCO_2 but the vice versa may not be true.
4. The sharing is not supported by a structured and well-defined digital process. In addition, typically data are transferred using unsecure communications (e.g. e-mails).

Information Sharing Flow: “as it could be” case

Ideally, a secure information sharing platform supporting the exchange of medical data should be perceived as an on-line facility automating the flow described in the previous section. Thus, in order to speed up the data sharing and access, ideally, HCO_2 could retrieve shared data simply by accessing a digital platform, namely *Innovative Secure Information Sharing Platform* (InSISP), having the role of “mediator” between entities willing to contribute to the information sharing and supporting a “proactive” sharing by creating a shared repository where HCOs can look for needed data.

In particular, when HCO_2 needs to access Bob’s data, it could try to get them directly from HCO_1, trough InSISP, provided that Bob gave his consent to data sharing with HCO_2 and HCO_1 proactively pushed such data to InSISP.

In order to get the data shared, HCO_2 needs to take the following steps:

1. HCO_2 asks to InSISP if there is an HCO having Bob’s exam results and try to get them.
2. InSISP mediates the request with HCO_1 by allowing it to check if it is allowed, according to Bob’s data processing consent, to share Bob’s data with HCO_2.
3. Two cases may happen:

a. Bob already provided the consent to data sharing with HCO_2. In this case, HCO_1 is authorized to share the data and can send them to HCO_2.
b. Bob did not provide the consent yet. In this case, HCO_1 cannot share the data but it can ask Bob if he is willing to change his consent to data processing to allow the data sharing. If Bob provides his consent now, data can be shared; otherwise, the sharing request is simply dropped.

The sharing pattern is depicted in Figure 8-4 - Futuristic sharing pattern for medical data

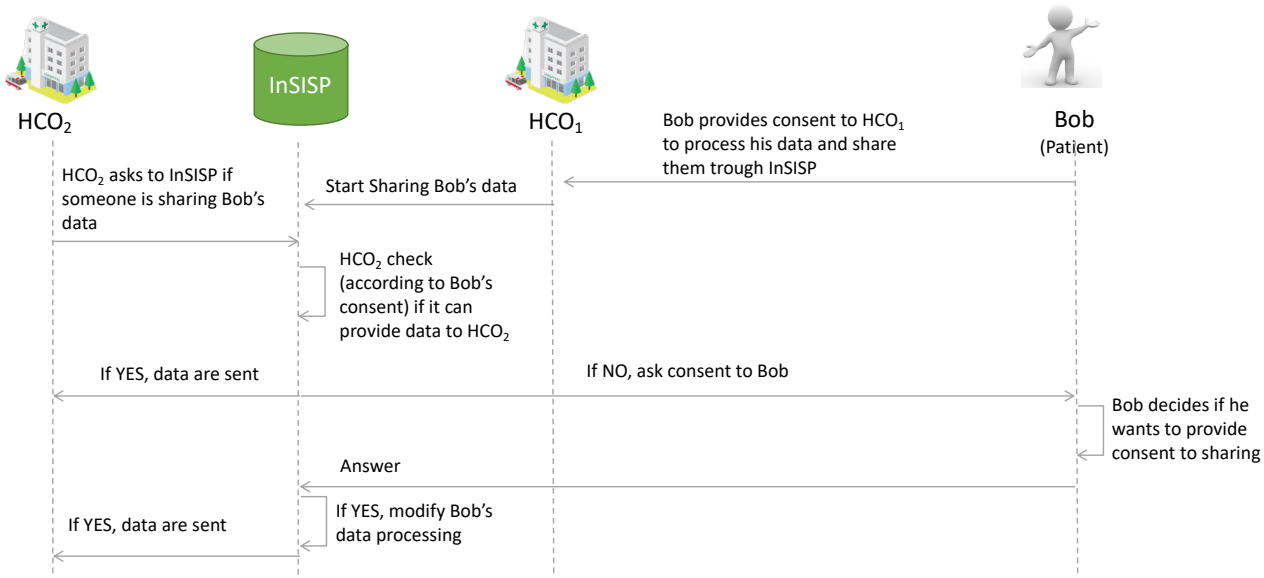

Figure 8-4 - Futuristic sharing pattern for medical data

Let us note that, in order to support this data sharing scenario, every HCO must keep track and store all consents to data treatment provided by every patient and behaves accordingly. In addition, they should be connected in some way and they should be able to support sharing request automatically, in compliance with data processing consents and preserving CIA properties.

Let us recall that, currently, each HCO works generally independently from the others and there not exists any general sharing network among them. In addition, the register maintaining data process consent is managed manually by physically collecting and managing signed paper where the patient declares his/her willingness to share and the parties involved in the sharing.

8.2 A novel view on Information sharing

From an innovative perspective, we can construct a platform that support sharing of medical data among its clients (i.e., HCOs) trying to overcome limitations pointed out in the previous section.

The basic idea is to design a platform that is able to support a fast and efficient medical information sharing both at national and cross-national level considering sharing constraints, included those imposed by GDPR.

From an abstract point of view, the Innovative Secure Information Sharing Platform (InSISP) can be seen as a data repository that can be accessed by HCOs (i.e., clients) to store and retrieve shared data by using a common interface and format.

Figure 8-5 The InSISP high level view

In order to guarantee dependability and security of the InSISP Platform we need to answer the following questions:

1. Who are the authorized entities allowed to interact with the platform (i.e., who are the clients)? Is it a publicly open service or can it be accessed only by registered users?
2. Who is managing data shared? Is it a centralized system or a distributed one?
3. How is the sharing implanted? Message exchanges or shared memory?

Concerning the first question and considering that authorization (especially those related to data sharing) are granted to specific HCOs we can conclude that the platform cannot be completely open, but accesses must be regulated and participating entities need to be identified univocally.

As a consequence, a federation of collaborating entities must be created and tools must be designed in order to manage and support the federation evolution.

Once the federation is established, participating entities can start sharing data according to the data processing consent provided by patients. To this aim, we can distinguish two main functionalities: (i) Data sharing and (ii) Data Processing Consent Management.

Data sharing is the core functionality of InSISP and, as the name suggests, it is aiming at supporting the actual sharing of medical data between entities participating in the federation. This functionality has the main objective of taking care of data that need to be shared and to maintain their CIA properties during all the sharing lifetime.

The Data Processing Consent Management functionality has the aim of managing the registry of Data Processing Activities where patients’ consents are stored and accessed to support properly and in compliance with GDPR requirements, the data sharing process.

Let us note that all the functionalities described so far require the internal maintenance of some table to store them. Thus, we can assume that each of them internally needs to maintain a data storage component (obviously storing different information).

Figure 8-6 summarizes a possible decomposition of the InSISP and highlights the three storage components.

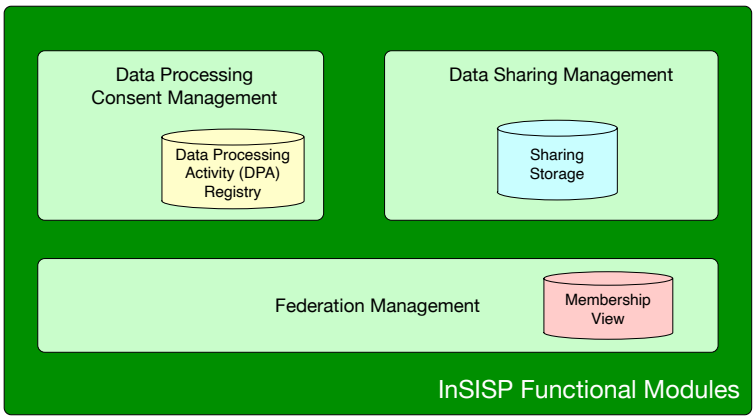

Figure 8-6 - InSISP Functional decomposition

In the following sections, we will detail the interfaces of each functional block and we will discuss possible deployment options for its internal storage component.

8.2.1 Federation Management

Let us recall that the federation management functionality has the aim to manage the federation lifecycle and in particular it should allow new HCOs to join and HCOs no more interested in participating into the federation to leave.

In particular, this functionality supports the following operations:

1. *Join the federation*: HCOs should be able to become part of the federation at any time.
2. *Leave the federation*: HCOs may decide to leave the federation at their will and the InSISP should support the removal of the entity from the membership and should notify the end of the sharing to connected entities
3. *Get the Federation membership*: allows a HCO participating in the federation to get the current membership and know the set of HCOs potentially involved in the sharing. The current membership is expressed through a *view*.

A view is a set of records, one for each current member of the federation. For each member of the federation, its record must provide the following information:

- HCO identifier in the federation
- Public keys that can be used to verify HCO data integrity
- The categories of data that are currently available for the sharing.

A HCO appears in a view as soon as its join is completed and should appear in any view until it starts the leave operation.

Discussion about possible design and deployment options

Let us note that the Federation Management intrinsically relies on the execution of a distributed protocol running. Thus, the two main options to implement a federation membership service are: (i) *client/server* or (ii) *peer-to-peer*.

In the client/server case, the current membership of the system is maintained by a Trusted Third Party (TTP). When a new HCO wants to join the federation it simply needs to contact the TTP, identify and authenticate itself with the TTP that will proceed by adding it to the current view. Similarly, when a HCO in the federation

wants to leave, it simply needs to notify the TTP that will remove it from the current view. The current membership can be obtained again by querying the TTP.

The main advantage of this option is that all the complexity of the membership management is delegated to the TTP. However, this also implies that the TTP is clearly a single point of failure for the system as well as its main bottleneck.

In the peer-to-peer approach HCOs collaborate to maintain a consistent view of the system by exchanging messages and trying to reach a consensus on the sequence of views generated to include new members and to remove old ones.

[Chockler01] deeply discussed the formalization and the specification of the group membership service while more recently, [Aguilera11] considered the problem of building a reconfiguration service to support the development of a distributed shared storage.

Let us note that in all these cases the emphasis is on how to provide a consistent view to all the members. To the best of our knowledge, there not exists any approach investigating the cost of realizing a membership service using Blockchain technologies.

The main advantage of the peer-to-peer approach is its intrinsic resilience. In addition, in peer-to-peer settings it is also possible to consider a blockchain-based approach to construct the sequence of consistent views providing the view auditability property for free.

The main drawback is the cost imposed by the management of consistent views as it requires to run coordination and synchronization protocols among all the participants that would bring poor scalability in case of a highly dynamic federation.

Let us note that, when considering a membership service, all the members are assumed to be uniquely identified. So, from the perspective of the InSISP development, there should be an external service that is able to support the identification of members and to provide them with a digital identity, So, the identification task is delegated to an external service and is out of scope of the Federation Management[26].

8.2.2 Data Processing Consent Management

The Data Processing Consent Management function has the aim to support the development of a digital Data Processing Activity Registry (DPA) to store and access patients’ data processing consents.

Let us recall that data processing consent is granted by a patient for a specific set of data, to a specific set of entities and for a specific purpose and period of time. In order to support the automatic verification of patients consents, such information must be stored and managed by the HCO.

Patient	Document/Data Set	Sharing Beneficiary Entities	Sharing Expiration Time
Bob	Blood exam id 1234	HCO_2	March 31st 2020
Bob	Blood exam id 1234	HCO_3	April 30st 2020
Bob	X ray Image 0987	HCO_2	March 31st 2020

Table 6 - Example of Data Processing Activity Table at HCO_1.

[26] In the PANACEA project this is taken into account and realized by designing and implementing an Identity Management Platform (IMP) that support the secure information Sharing Platform (SISP).

Table 6Table 6 shows an example of a possible data structure that can be used by an HCO to store and manage data processing consents.

Let us note that every patient has the right to modify his/her consents and has the right to be forgotten i.e., at any point in time he/she may ask to revoke all his/her previous consents (while also erasing any data identifying him/her).

To this aim, the Data Processing Consent Management functionality should offer the following operations:

1. *Provide new Consent*: it allows to add a new entry to the table and to specify the beneficiary entities and to set the expiration time of the consent
2. *Update Consent*: it allows to modify and existing consent by allowing the sharing with a new beneficiary or by removing a beneficiary.
3. *Remove (all) Consent*: it basically implements the right to be forgotten by deleting all the consent previously provided by a specific patient.

Discussion about possible design and deployment options

Let us note that the Data Processing Consent Management function support every HCO in managing its own Data Processing Activity Registry. Thus, from this point of view, we can say that it is local to every HCO.

As a consequence, the most appropriate choice is to design it as a local data store managed and accessed only by one HCO. Of course, in order to increase the resiliency and security of the storage it can be also replicated but all the replicas will be still managed by the same HCO.

A distributed design raises a privacy issue in patients’ information. Indeed, even if data in the registry are not sensible by themselves, they could be easily correlated to infer sensitive information about patients and would result in a privacy violation e.g., by looking to the list of HCOs where Bob did his analysis you may infer that Bob is affected by a specific disease. To solve this issue, it is necessary to employ anonymization scheme generation an extra cost without any particular advantage in term of reliability or security.

8.2.3 Data Sharing Management

Data Sharing is the core functionality of the InSISP as it manages the real transfer of medical data between parties. It offers just one main operation i.e., the Get Data that is used to retrieve a specific piece of data for a specific patient and transfer it according with the patient consent.

We can consider three main options to design and deploy this functionality:

1. Client/server
2. Peer-to-peer with message exchanges
3. Peer-to-peer with shared memories

In the client/server case, data available for the sharing are copied and pushed toward a centralized Trusted Third Party that will take care of satisfying the sharing request. In order to be GDPR compliant, a specific consent to move data to the TTP must be signed as well as the consent to share data with all the federation members[27]. Figure 8-7 shows an overview of a possible client/server design.

27 Let us note that this second set of consents could be removed if every HCO provide to the TTP a copy of its Data Processing Activity Registry. However, as anticipated in the previous section, this would add the complexity of finding a good anonymization scheme that allows to preserve patients’ privacy still allowing to the TTP to check consents.

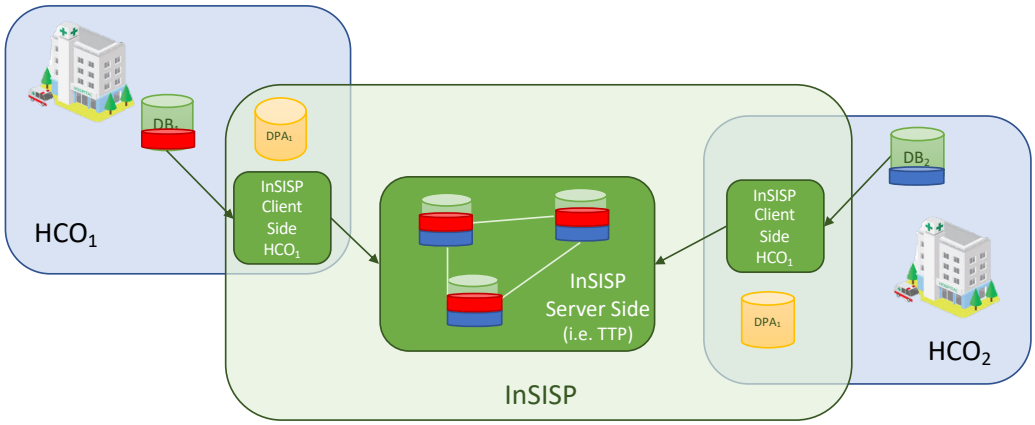

Figure 8-7 - Overview of the Client/Server Design with a TTP

In the peer-to-peer case the idea is that the sharing is realized by letting HCOs in the federation cooperate among them. We can distinguish two cases: cooperation realized through message exchange and cooperation realized through shared memory.

Let us consider the message exchange case before. In this case, we can implement exactly the sharing flow described in Section 8.1.1 where the sharing is realized using an ad-hoc request-reply communication pattern.

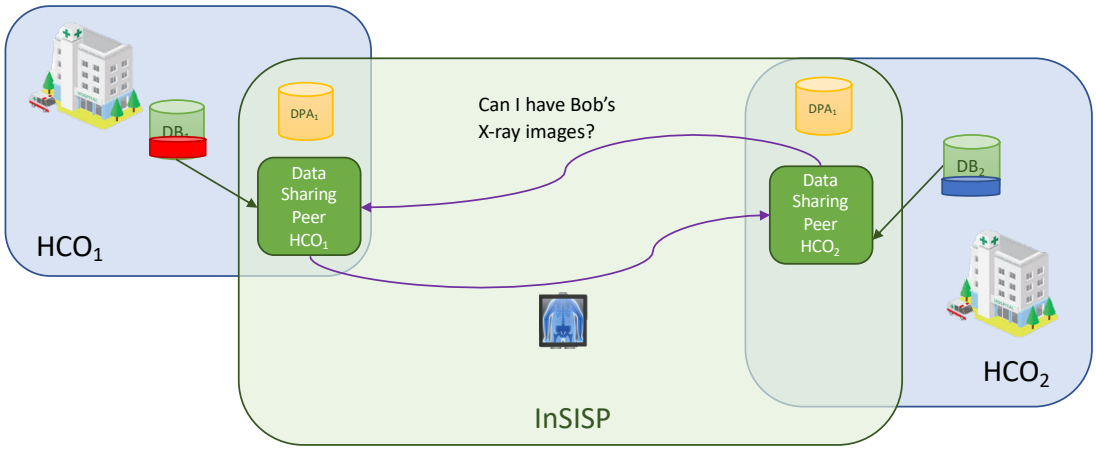

Figure 8-8 Overview of the Peer-to-peer with message exchanges Design

Figure 8-8-8 shows an overview of a possible Peer-to-peer with message exchanges Design.

In the Peer-to-peer with shared memories design, each HCO creates a shared memory space where it stores all the pieces of data that can be shared according to the Data Processing Activity Registry. As an example, let us consider the case where Bob provided the consent to HCO_1 to share his x-ray with HCO_2. This means that HCO_1 and HCO_2 create locally a shared space where they will store a copy of all Bob’s x-ray images (the red slice in Figure 8-9-9).

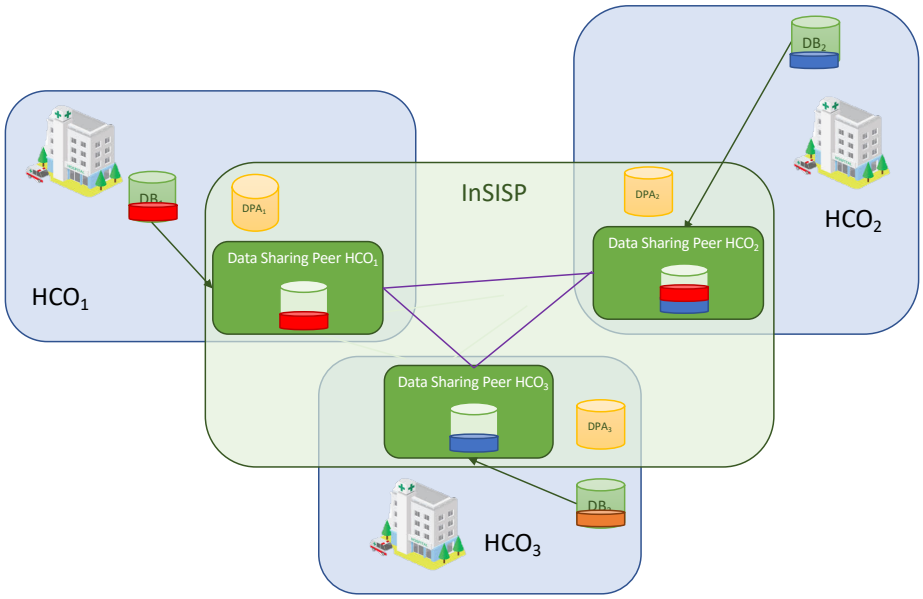

Figure 8-9 - Overview of the Peer-to-peer with shared memories Design

Discussion about possible design and deployment options

Technically, all the three considered designs are feasible. However, as anticipated in the previous section, the client/server scenario poses several challenges from the point of view of the consent needed in order to make it compliant and, in particular, the main issue is the large set of consent that are necessary and that patients may be reluctant to provide.

Contrarily, the two point-to-point designs solve this issue as they exploit locally the consent information and move data only toward authorized HCOs.

8.2.4 Evaluation and recommendations

According to the considerations done in the previous sections, Table 7 summarized the viable options for the design of each functionality of the InSISP.

		Is it a valid Design and Deployment Option?	
		Yes	No
DPA Registry supporting the Data Processing Consent Management	Local	X	
	Centralized		X
	Distributed		X
Sharing Storage supporting the Data Sharing	Local	X (peer-to-peer with message exchange)	
	Centralized		X
	Distributed	X	
Membership View supporting the Federation Management	Local		NA
	Centralized	X	
	Distributed	X	

Table 7 – Summary of Design options and recommendations

9. Evaluation of emerging technical solutions: Blockchain

In this chapter we will discuss about the pro and cons of adopting blockchain technologies in the PANACEA project. In particular, we first provide background notion on blockchain, then we will present and discuss the most relevant existing blockchain technologies and finally we will assess advantages and disadvantages of using it to develop a secure information sharing platform to exchange medical data.

9.1 Background

A blockchain is a *decentralized*, *distributed* data structure used to store *transactions* (aggregated in blocks) across many computers.

Each block is linked to the previous one trough a cryptographic hash and it is a data structure which allows to store a list of transactions. In the blockchain, a transaction abstracts and allows to keep track of an exchange or interaction between two entities.

Transactions are created and exchanged by peers of the blockchain network and modify the state of the blockchain data structure.

Blockchain can be classified according to two different perspectives: Data access and Data management.

Concerning Data access, we can have:

- *Public Blockchain*: there are no restrictions on reading blockchain data and submitting transactions for inclusion into the blockchain;
- *Private Blockchain*: direct access to blockchain data and submitting transactions is limited to a predefined list of entities.

Concerning Data management, we can have:

- *Permissioned Blockchain*: transaction processing is performed by a predefined list of peers with known identities;
- *Permissionless Blockchain*: no restrictions on identities of transaction processors (i.e., blocks creators).

Combining the two perspective, we can have four categories as depicted in **Error! Reference source not found.**.

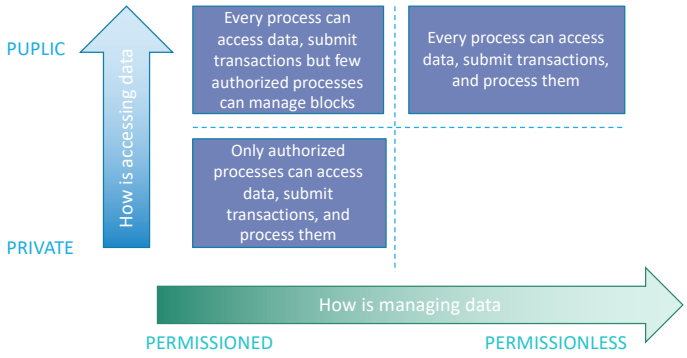

Figure 9-1 - Blockchain Classification Overview

Let us note that private permission less architectures do not make too much sense as they would require that the set of processes managing the blockchain (i.e., those creating and attaching blocks) is open while the set or readers and writer is closed. Thus, we will restrict our attention only on the other three categories.

In addition to CIA properties, blockchain also supports auditability and transparency as any reader is able to verify the correctness of the state of the system. Indeed, storing all the transactions, it is possible to re-play (starting from a correct checkpoint) the entire history and check that the current state is consistent with the set of recorded transactions.

Let us note that, when using a blockchain to store data, there exists an inherent trade-off between transparency and privacy. Indeed, if the primary requirement is to have a fully transparent system, we need to accept that anyone is allowed to see any piece of information (sacrificing privacy). Conversely, if the primary requirement is to have a private system, it will not provide transparency. A trade-off between transparency and privacy is however possible but it will come at the cost of efficiency, as it would require employing complex cryptographic primitives.

9.2 Evaluation of existing technologies for private permissioned blockchain

In this section we will present the main existing technologies able to support the design and development of blockchains.

For each technology considered, we provide a brief description and a table that highlights the following aspects

System model	This field specifies the type of blockchain that can be implemented with the considered technology (i.e., private vs public, permissioned vs permissionless) and the assumptions about the network compositions, roles, etc.
Transaction data model	This field explains how transactions are modelled and implemented by the current technology.
Consensus	This field provides information about the consensus procedure followed to generate and attach blocks to the chain.
Internal token	This field specifies if the technology employs some token to regulate the block creation process.
Governance	This field provides information about the entities/research groups etc that are currently involved in the development of the specific technology. It basically refers to the degree to which decision-making power is distributed in the blockchain community. It tries to answer the question of who can make what decisions on a blockchain platform.

9.2.1 Ethereum

Ethereum [Ethereum2014] was born in 2015 and immediately represented a big step ahead in the world of blockchain platforms, as it attempts to generalize the monetary transactions support already provided by Bitcoin, by allowing to add a custom, programmable logic to transfers. Thanks to its built-in fully-fledged Turing-complete programming language, it enacts the first public framework for developing *smart contracts* for blockchains, i.e., programs that can be used to encode arbitrary types of decentralized blockchain applications, including but not limited to cryptocurrencies. Ethereum has its internal cryptocurrency called Ether, which is currently also used for the internal PoW mining system.

The basic component of Ethereum is the account. The Ethereum blockchain tracks the state of every account. All state transitions on the Ethereum blockchain represent value and/or information transfers between accounts.

A *per-address* transaction model is used (opposed to Bitcoin’s *per-output* transaction model).

There are two types of accounts: (i) *externally owned accounts* (EOA), which are controlled by a private key, and (ii) *contract accounts*, which enable to run contracts. A contract account is controlled by its contract code and can only be activated by an EOA.

System model	Ethereum support the development of both public and private blockchains. The blockchain, is supported by a peer-to-peer network whose nodes are divided in full and light nodes. Full nodes (validator) are in charge of maintaining the ledger, while light nodes are in charge of verifying transactions. Being a public and open network, anyone can join the network as light node or as fully node, pseudo-anonymous. Furthermore, transactions are publicly available using online block explorers.
Transaction data model	Ethereum models transactions through smart contract. Smart contracts are written using Solidity or Serpent high level language. In order to develop and deploy an Ethereum application, each customer must have gas[28] available to interact with the application. All modifications to contracts data must be performed by its code (similarly to *enforced* stored procedure in classical relational data bases). To modify contracts data, blockchain users performs modification *requests*, and the contract determines whether and how to fulfil those requests. For example, the smart contract for a financial ledger performs the same three tasks as the administrator of a centralized database: checking for sufficient funds, deducting from one account, and adding to another. The *Ethereum Virtual Machine (EVM),* is the runtime environment that enable running the smart contracts, compiling them into byte code (using an EVM compiler), and finally storing them on the blockchain using an Ethereum client. Every node participating in the network runs the EVM as part of the block verification protocol. Contract executions, triggered by messages or transactions, are purposefully redundant in the sense that every instruction is executed on *every node of the network*. This enables to reach consensus on the system state without the need of trusted third parties or oracles. The main drawback is the cost: for every executed transaction, senders have to pay a specified cost, expressed in a number of units of *gas*. The main purpose of gas is preventing deliberate attacks and abuse of the Ethereum network, such as endless running transactions.
Consensus	The consensus protocol works at Ledger level. Since transactions have a specific business logic that is expressed by smart contracts, each miner adheres the following procedure: 1. *Transactions are collected in a mempool*: miners receive transactions which are bouncing in the network and previously propagated by other peers.

[28] Gas is the internal pricing for running a transaction or contract in Ethereum.

	2. *Transactions are executed*: when the block is complete, miner executes the code on the local EVM that is attached in each transaction 3. Block and header are created. 4. Consensus. Miners are incentivized to collaborate in this procedure as they are rewarded with quantity of Ether. Currently, step 4 is implemented using a consensus protocol based on Proof of Work (PoW), yet, it features an algorithm (called ETHash) which differs from the one used in Bitcoin since its computation requires both memory and bandwidth. Since the frequency of block creation is regulated by the difficulty factor, and since in Ethereum each block is created every 15 seconds, we have that the difficulty of the hash puzzle is much lower than the one used in Bitcoin, but this also leads to an high increment of forks in the chain because more than one miner may solve the block. To solve this issue, Ethereum uses the GHOST protocol, where the main chain is identified by following the chain with the highest computation efforts. This coincide with the longest chain because more blocks are present in a chain, more efforts miners spent on that path to mine it; so the longest chain will become the final or canonical path and hence the current state of the global EVM. Many software implementing the client-side protocol of Ethereum have been developed. Parity[29], for instance, provides access to the basic Ether wallet functions, as well an Ethereum GUI browser that provides access to all the features of the Ethereum network including dApps. It is a *full node* wallet, which means that clients store the blockchain on their computer. Other examples are Mist or Ethereum-wallet.
Internal token	The Ether cryptocurrency is used to reward the miners.
Governance	Ethereum has a (informal) business process coordinating improvements to the Ethereum core protocol specifications, client APIs, and contract standards. People can propose their ideas as an issue or pull request to the EIPs (as per Ethereum Improvement Proposals) repository. After basic filtering, the proposal will receive a number and is published in draft form. For an EIP to become Active it will require the mutual consent of the community. Those proposing changes should consider that ultimately, as for Bitcoin, consent may rest with the consensus of the Ethereum users.

9.2.2 Hyperledger Fabric

Hyperledger [HyperledgerFabricProj] is a *consortium* (composed of enterprises and associations) hosted by the Linux Foundation which aims to develop a set of frameworks for realizing various blockchain implementations. Fabric is one of these frameworks that enables building a private, permissioned ad-hoc blockchain, with several functionalities specifically designed for enterprise networks [HyperledgerFabricArch].

System model	In private permissioned blockchains, nodes are known and identified parties that share a common goal but do not necessarily trust each other. As a consequence, the network is not peer-to-peer and various roles are distributed among nodes with different permissions and tasks.

[29] https://www.parity.io/

	In particular, we can distinguish: • *Clients* (also called light nodes): they simply submit transactions to the network and represent the end-users. • *Orderers*: they act as a sort of ordering service since they provide a communication channel between clients and peers over which messages (containing transactions) can be broadcasted. They play an important role especially in the agreement phase since they need to route transactions through special paths. • *Peers*: they maintain the ledger. In order to do that, they receive ordered update messages (in the form of blocks) from orderers and commit new transactions to the ledger. *Endorsers* are a special type of peer. Their task is to endorse or validate a transaction by checking whether it fulfils necessary and sufficient conditions (e.g., the provision of required signatures). Basically, the role of peers is the analogue of validators, sealers or miners of the previous technologies.
Transaction data model	Fabric uses a transaction model similar to smart contracts called *Chaincode* and it is the first blockchain system that runs distributed applications written in general-purpose programming languages (such as Go, Java, Javascript). This is a significant advancement to existing blockchain platforms, which require code to be written in a domain-specific language and thus requires specific training (such as Solidity or Serpent). Furthermore, Chaincode programs do not depend on a native cryptocurrency, allowing thus to encode arbitrary *assets* and the transaction instructions (business logic) for modifying them.
Consensus	Consensus is achieved at Transaction level. Fabric is a modular architecture i.e., few elements are pluggable and represent a sort of template for several enterprise applications. As an example, it provides several consensus modules to allow the adaptation of the technology to a specific application. In Hyperledger Fabric, consensus is defined as the verification of the correctness of a set of transactions included in a block. Since this depends on the specific application, consensus is ultimately achieved when the order and results of a block transactions have met the explicit policy criteria that take place during a transaction’s journey from proposal to commitment. Differently from traditional blockchain platforms (where transactions and smart contract are typically executed after the consensus and all the parties execute all contracts), Fabric uses an approach where transactions are endorsed and executed (and data is returned) before ordering them and before the blockchain reaches consensus in the chain. The consensus occurs therefore at *transaction-level*, where orderers provide channels where only nodes involved in a transaction receive it and validate it. A channel is a private blockchain overlay which allows for data isolation and confidentiality. A channel-specific ledger is shared across the peers in the channel, and transacting parties must be properly authenticated to a channel in order to interact with it. The consensus algorithm is concerned of the ordering service, where a cluster of orderers runs an algorithm for ordering transactions and agree on their order. The application use-cases and its fault tolerance model should determine which

	ordering plugin to use among the available ones BFT Smart, Simplified Byzantine Fault Tolerance (SBFT), HoneyBadger. The performance of Hyperledger Fabric blockchains has not been extensively tested. However, it is expected that they can process transactions at a rate greater than 10k transactions per second using a BFT consensus model.
Internal token	No internal token is present as the system is made for private and permissioned uses where no monetary incentives are needed.
Governance	Since the system is closed and permissioned the governance is internally managed by the Linux Foundation.

9.2.3 MultiChain

MultiChain [Multichain15] is an open-source platform for the creation and deployment of private, permissioned blockchains, either within or between organizations. It originates from a fork of Bitcoin Core and shares a similar protocol and transaction format. A node set up through MultiChain can also act as a node on the Bitcoin.

Although being initially envisioned for use-cases in the institutional financial sector, MultiChain can be implied in very different settings, where custom assets can be created and whose full lifecycle can be handled on-blockchain (issuance, transfer, exchange, reissuance, escrow, redemption, destruction). It also natively supports use cases with focus on general data retrieval, timestamping and archiving, like publish-subscribe systems rather than the transfer of assets between participants. Streams are append-only collections and can be used to implement key-value and time series database, as well as identity-driven databases where entries are classified according to their author. Creators of a Multichain network determines the initial set of administrators, as well as whether anyone can connect to the network without restriction. Administrators can also dynamically control permissions to the blockchain for specific users while the blockchain is active. Such permissions include the ability to send, receive, or create assets, and the ability to create blocks. Decisions to alter permissions are made via consensus among administrators.

Creators control how fast the network moves and the size of the data in the blockchain, the average time to add a block, the maximum block and transaction sizes, and the maximum size of metadata output. They can also change the mining protocol (optionally choosing a PoW mechanism and its characteristics). Regardless of these settings, the blockchain will use a randomized round-robin system to add blocks.

System model	The system is a permissioned peer-to-peer network, where transactions between peers are mediated by a handshaking process that achieves proof of identity and security using public key cryptography. In MultiChain, when two nodes connect: 1. Each node presents its identity as a public address on the permitted list. 2. Each node verifies that the other’s address is on its own version of the permitted list. 3. Each node sends a challenge message to the other party. 4. Each node sends back a signature of the challenge message, proving their ownership of the private key corresponding to the public address they presented. If either node is not satisfied with the results, it aborts the peer-to-peer connection.
Transaction data model	It uses bitcoin’s protocol, transaction and blockchain architecture, with changes only to the handshaking. All other features are implemented using metadata and modifications to the validation rules for transactions and blocks. The main difference with respect to smart contract-based chains is that in Multichain data is embedded immutably in a blockchain, but the *code* for interpreting that data is in the node or application layer. This is deliberately different from the smart

	contracts, in which code is embedded in the blockchain and runs in a virtual machine. Furthermore, Multichain supports freely customizable, indexed, append-only streams, where one or more *publishers* digitally sign each item. Streams integrate within MultiChain permissions in that, each item in a stream is represented by a blockchain transaction, in the same way that assets can only be issued by certain addresses.
Consensus	MultiChain uses a randomized round-robin system for block-adders and a concept of *mining diversity* which is a configurable parameter, that governs how long a block-adder has to wait, before the other nodes will accept another block from him. With a configured strictness of zero, any block-adder can add any block meaning it’s very tolerant but also increases the risk that a single block-adder or small group of block-adders can spoof the system. At the other extreme of the scale, (strictness of 1) once a miner adds a block, he/she has to let every other block-adder add a block before you can add again (i.e., rotating leader). This stops single or groups of block-adders from creating forks, but if a node goes offline then at some point no further blocks will be added while the network waits for the offline node to add the next block. Strictness lets adjust the balance between security and technical malfunction risk. This may significantly increase the Byzantine fault tolerance (BFT) of MultiChain blockchains in comparison to some other blockchains. Block time can be as low as 2 seconds.
Internal token	No native coin. In a MultiChain blockchain you can create coins or digital tokens which have an immutable ownership history, which all nodes track and verify. However, by default there is no underlying coin to pay transaction fees and to reward block-adders, although it can be optionally configured to do so if needed.
Governance	Being an open source product for implementing private blockchains, although Multichain Developers release new versions of the code, the governance is left to the enterprise implementing the blockchain.

9.3 Is it worth to use Blockchain technologies inside InSISP? And which ones?

In this section, we will use the methodology proposed in [Staderini18] to evaluate if blockchain is a valid option to support the InSISP deployment in the Healthcare domain.

This methodology considers the following three steps:

1. Requirement analysis to assess blockchain benefit (cfr. Figure 9-2);
2. Evaluation of the most appropriate blockchain solution where the designer is guided on the choice of the most suitable blockchain category, based on blockchain-specific criteria depending on who are readers and writers of the data and who is allowed to generate data.
3. Blockchain configuration selection which assists the designer throughout the decision-making process for the configuration of the blockchain compliantly with the chosen category and the given project requirements

Let us remark that blockchain is intrinsically a distributed system and it makes no sense trying to use it when a distributed setting is not appropriate.

According to the conclusion derived in Section 8.2.4, a distributed design scenario can be considered for the storages supporting respectively the Federation Membership function and for the Data Sharing function. Let us analyze them individually.

9.3.1 Step 1: Blockchain yes or no?

This step starts from the analysis of requirements related to the component under analysis and has the final objective to understand the benefit of adopting a blockchain-based solution for its low-level design and development.

The factors that are considered in this step are listed in the following:

- Data or State Storage. The first element to consider is to check if the module under analysis needs to store data or system state. If no information needs to be stored, clearly no blockchain is needed.
- Immutability and Data Integrity. With immutability we refer to the property of a data to be never changed in time (i.e., a constant value that is never updated). If immutability is a requirement, then blockchain is certainly an option, as this is probably the most distinctive property of any blockchain. Integrity is strictly related to immutability and this is why they are analysed together, also considering that they are both closely related to cryptography. If a component requires data protection from unauthorized modifications, then this requirement can be met with a blockchain.
- Non-repudiation. Non-repudiation means that the author of some message/data cannot deny that it produced the message. This is another fundamental property which can be easily satisfied using blockchains.
- Multiple writers. This criterion considers the multiplicity of entities in charge of writing data in the storage. If only one entity is a writer, thus a common database is probably most appropriate than a blockchain especially from the performance perspective i.e., in terms of throughput and latency.
- Trusted Third Party always online. A Trusted Third Party (TTP) is an entity which facilitates interactions between mutually mistrusting entities. If in the system a TTP is required and it is planned to be always online, entities can delegate to it write operations as transactions or state changes. Therefore, the TTP plays the role of a trusted deliverer and verifier. In this case, a blockchain, known for being a trust less technology, becomes useless, and the methodology brings to the related output. Otherwise, it can happen that the involvement of a TTP is planned but not for being always online: in this case it could play the role of an authority giving authorizations for permissioned blockchains. Alternatively, a TTP may not exist at all. In the latter situations it is not possible to exclude the recommendation of using a blockchain.
- Writers are known and trusted. If all the entities interested in writing, know and mutually trust each other, a blockchain is superfluous, then not recommended (again mainly for performance issues).

The flow chart of the first step is shown in Figure 9-2.

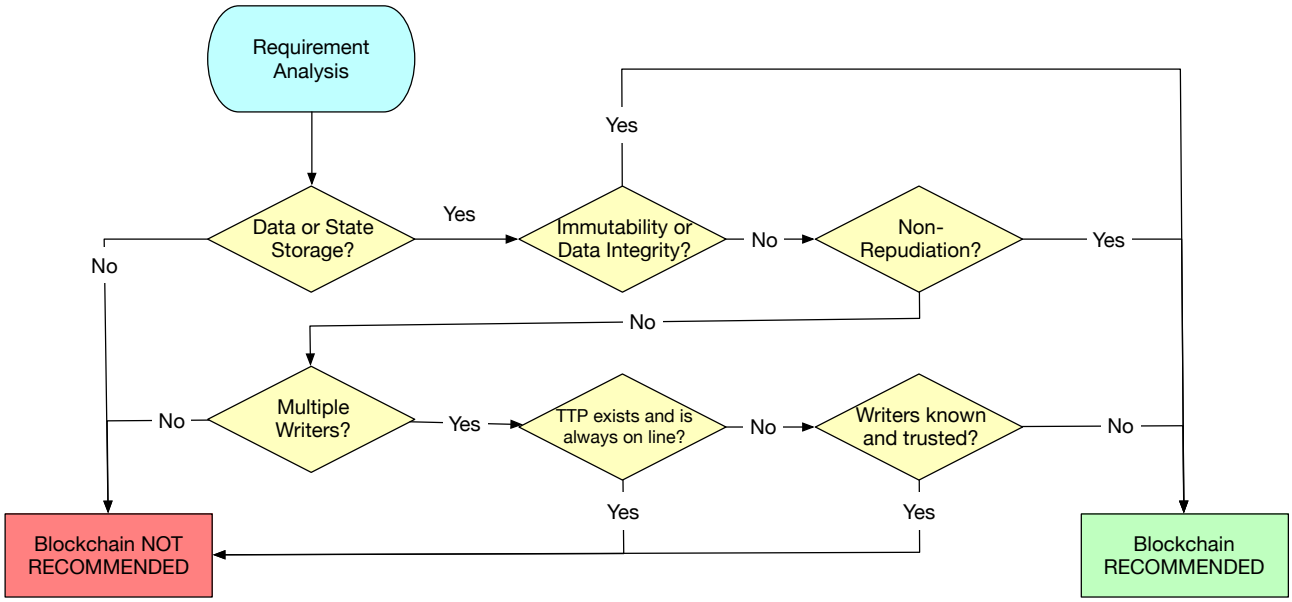

Figure 9-2 – Step 1 Flow Chart: Blockchain yes or not?

Federation Management Block Assessment

Let us now apply the methodology to the Federation management block. Let us note that, in order to manage the federation, the module needs to store at least current members’ identifiers and possibly other related information used to identify and authorize other members. Thus, in the following, when discussing about data stored by the Federation management, we will refer to this set of identifiers.

Table 8 highlights the considerations done for each selected criterion.

Criterion	Answer for InSISP in the healthcare domain
Do you need to store Data or State?	YES. The data to be stored are represented by the identifiers of federation members and their related information (e.g., keys used to verify the integrity of the messages)
Do you have Immutability or Data Integrity requirements?	Immutability NO. HCOs identifiers should not change. However, the set of identifiers may change in time as well as some of the additional information stored may need to be updated (e.g., public/private keys may need to be refreshed as well as digital certificates). Data Integrity YES. The federation membership should contain only the identifier of effective members. Identifiers should not be tampered or created.
Do you have non-repudiation requirements?	NO. Non-repudiation is desirable, but it is not currently a requirement.
Do you need to support multiple writers for the same data?	YES. Let us recall that the data stored is a set with all the identities of participating members. As a consequence, this set can be updated by any member that wants to leave or by a new member that is trying to join the federation.
Is there a TTP and is it always on-line?	NO. A TTP need to be assumed as bootstrap node to be accessed by members that wants to join the federation. However, this node is not guaranteed to be always on line.
Are writers known and trusted?	YES. In our context we are considering the creation of a federation among collaboration entities. Thus, participating members must be known and trusted.

Table 8 - Assessment for Federation Membership Data

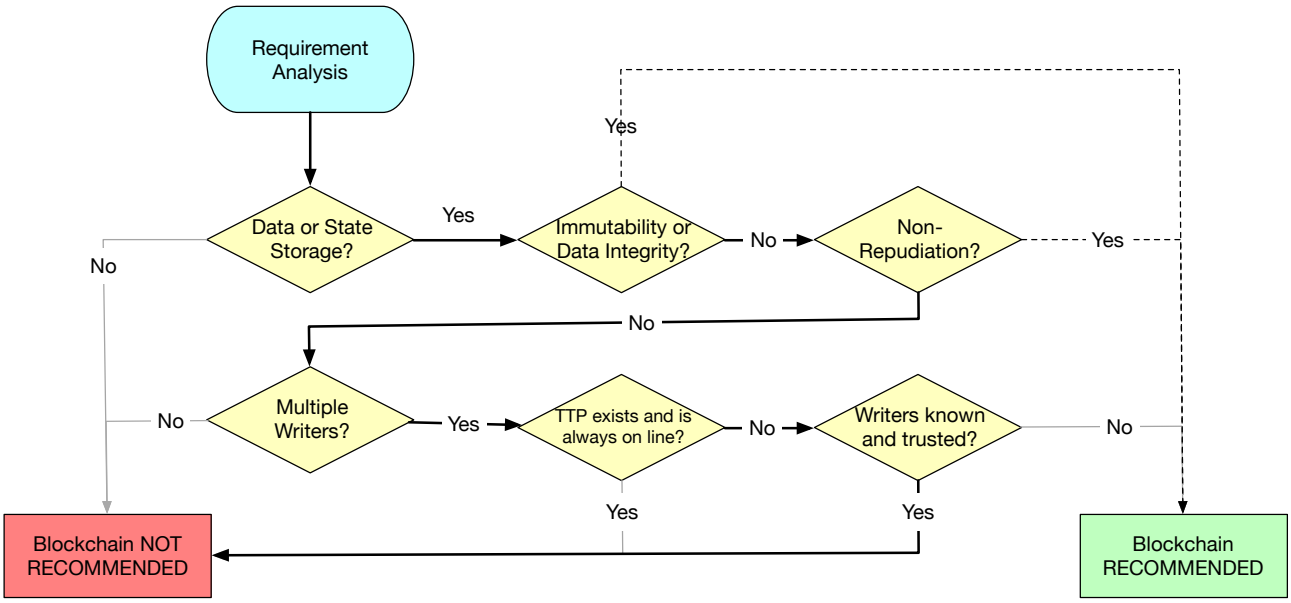

Figure 9-3 - Assessment for Federation Membership Data: Flow Chart

Figure 9-3 highlights on the flowchart the main path observed by the preliminary assessment on the Federation Membership block. At a first glance we get that Blockchain is not recommended to support the implementation of the federation membership mainly because there exists a basic level of trust between members of the federation. In addition, we are also assuming that the federation membership is also relying on a trusted external service providing digital and secure identities to participants.

However, we also highlighted the opportunity to consider a non-repudiation requirement and we considered the importance of preserving data integrity (i.e., to ensure that the current membership cannot be altered).

Thus, **if these two requirements become more relevant or if assumptions on the identity platform or about trustworthiness of participant cannot be met, then blockchain becomes immediately a viable solution for the Federation Management Block** (see dotted lines in Figure 9-3).

Data Sharing Block Assessment

Let us now apply the methodology to the Data Sharing block. Let us recall that this module is the core of the InSISP where medical data are actually shared. Thus, in the following, when discussing about data stored by the Data Sharing block will refer to medical data.

Criterion	Answer for InSISP in the healthcare domain
Do you need to store Data or State?	YES. As the name suggest, the Data Sharing Block is going to store medical data that need to be shared among the federation.
Do you have Immutability or Data Integrity requirements?	Immutability NO. Medical Data may need to be removed from the storage to guarantee the “*right to forgotten*” ruled out in GDPR or specific fields may need to be updated as stated in the “*right to rectification*” (contact information associated to medical data).

	Integrity YES. Medical Data has a strong integrity requirement (also coming from GDPR). In addition, immutability is also a highly desirable property.
Do you have non-repudiation requirements?	NO. Non-repudiation is desirable, but it is not currently a requirement.
Do you need to support multiple writers for the same data?	NO. When dealing with medical data, we are considering a type of data that are produced by a data producer and then becomes typically read-only (e.g., blood exam reports or x-ray images are produced and then they can just be accessed in read mode).
Is there a TTP and is it always on-line?	YES. Without loss of generality we can assume that the HCO that produced the medical is trusted and is always available.
Are writers known and trusted?	YES. Medical data can be produced only by HCO and they are assumed to be trusted and known in the federation.

Table 9 - Assessment for Medical Data

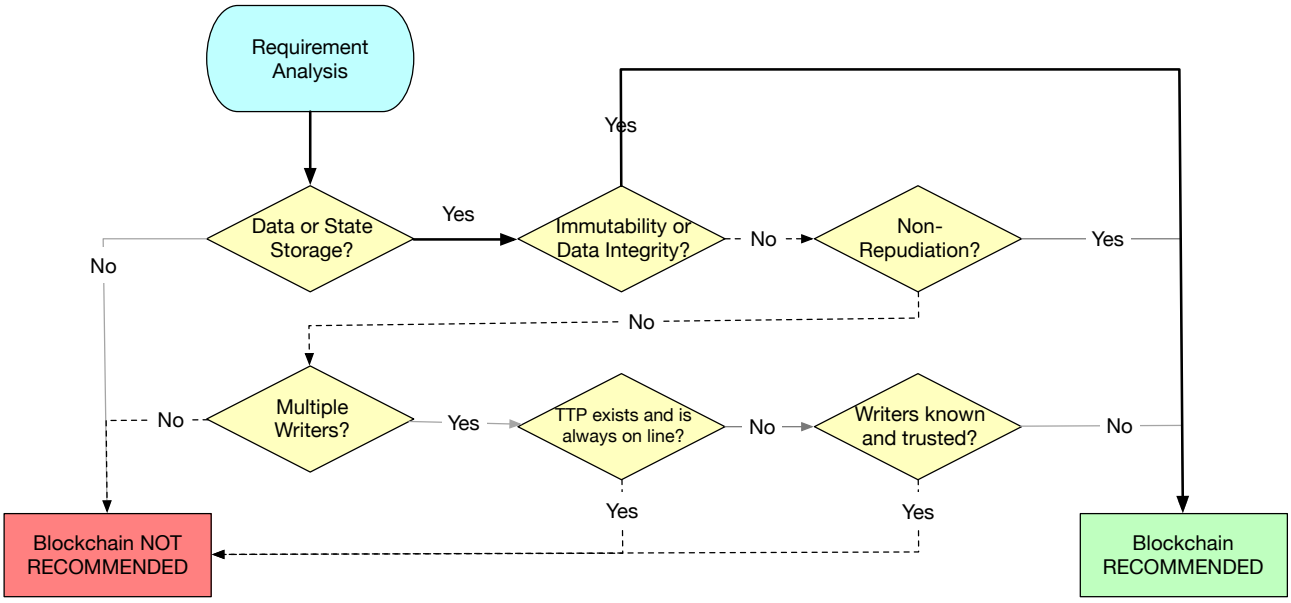

Figure 9-4 - Assessment for Medical Data: Flow Chart

Figure 9-4 highlights on the flowchart the main path observed by the preliminary assessment on the Data Sharing block. At a first glance we get that Blockchain is recommended to support the implementation of the data sharing mainly because it is supporting efficiently the data integrity requirement i.e., it allows to trace data accesses and verify their authorship and integrity.

Let us remark that medical data are basically read-only and, in our context, they are shared between trusted parties. Thus, the main benefit we can get by adopting a blockchain-based solution is the support for data integrity verification and data auditability.

However, this feature must be carefully balanced with the “right to be forgotten” requirement (i.e., a MUST requirement imposed by the GDPR regulation) and its implication on the adoption of a blockchain-based

solution. In order to support the implementation of the “right to be forgotten” we need to guarantee that data can be deleted from the blockchain when the data owner asks to do it.

Currently, deleting data efficiently from a blockchain is still an open research problem and the few existing solutions are currently based on the adoption of computationally expensive cryptographic techniques.

Furthermore, when dealing with medical data, there is also the additional complexity following by the huge heterogeneity of data to be considered (i.e., text, images, images+sounds, etc.). Blockchain technologies have been originally designed to deal with transactional data, of small size and in the form of numbers or strings. Currently it is not clear how to extend the paradigm to work with heterogeneous data. A possible solution to this issue could be to keep such heterogenous medical data stored locally in a classical data base and store in the blockchain only its hash. However, as it has been pointed out by Esposito et al. [Esposito18], it is still not clear how much privacy lawyers consider such meta-data as an expression of a personal data and thus the issue may remain.

As a conclusion, the application of blockchain technologies to InSISP is NOT suggested.

9.3.2 Step 2: Which Blockchain paradigm is good for me?

This step of the methodology is performed only for those blocks that ended up with blockchain as a viable solution. Thus, considering the outcome of the assessment performed in the previous section, **we will apply this step of the methodology only for the Federation Management block where blockchain may be an option** for storing the federation membership and its related information.

The Blockchain Choice step first considers reading and writing operations as main criteria for the choice. Reading is further decomposed into two criteria, namely:

1. Accessing the storage to get data (i.e., reading operation) and
2. Creating transactions to be stored in the ledger (i.e., creation operation).

The possible configurations for reading and creation operations are *open*, *restricted*, and *internal*, meaning respectively that the operation is permitted to any node, permitted to some privileged nodes only, and permitted to some privileged nodes belonging to a unique organization only.

Concerning writing, the criterion that must be taken into account is the mining criterion, which establishes the entities allowed to perform writing operations. Possible options for the mining are:

1. open mining, meaning that every peer node can be a miner,
2. consortium mining, where only some nodes are allowed to write, but they can be entities belonging to different organizations, or
3. internal mining, where the miners are just a subset of the nodes of the same organization.

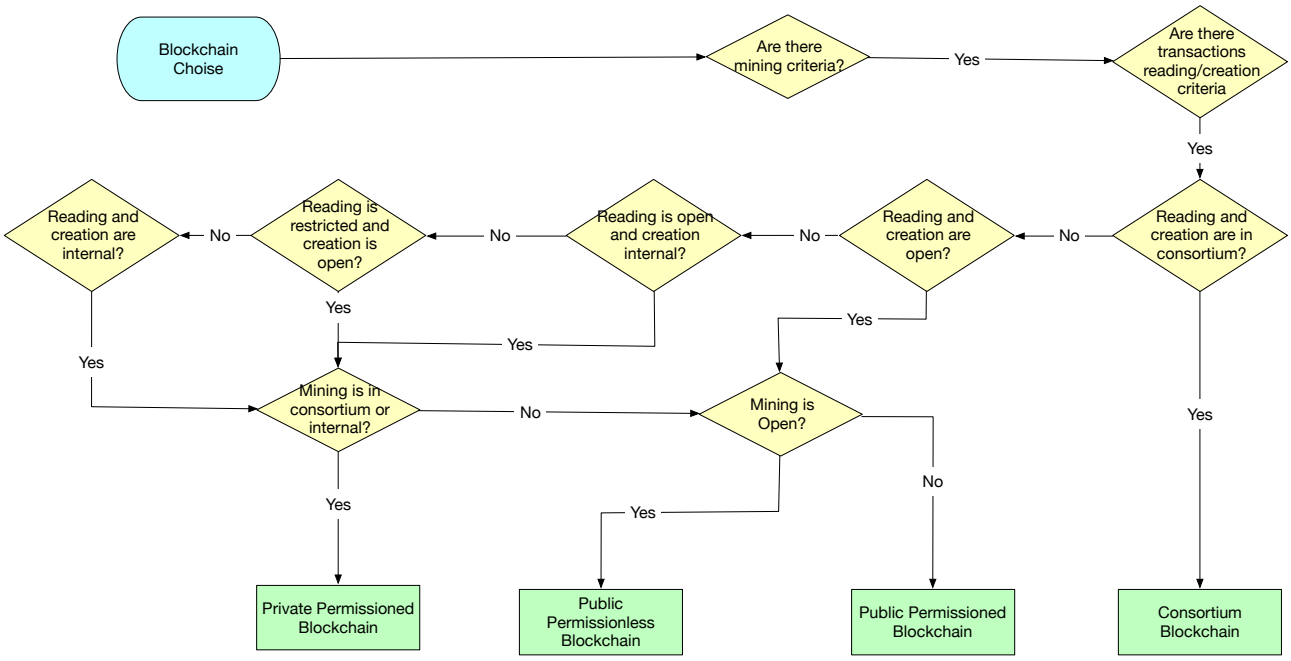

Figure 9-5 - Step 2 Flow Chart: Which Blockchain paradigm is good for me?

Figure 9-5 shows the flow chart guiding us in selecting the best blockchain option for the membership storage.

Concerning the write operation, let us recall that here we are considering who is allowed to append blocks in the blockchain. Thus, in our case, it is clear that we have mining requirements and that miners can be only entities belonging to the federation (i.e., we are considering a consortium mining). So the answer to the first check of the process is YES).

Concerning transaction reading and creation, we have requirements as well. In particular,

- creation is open as every HCO can try to join the federation and thus can generate a new transaction to be inserted in the current set of participants.
- Read is restricted as every HCO belonging to the federation is allows to get the current view.

It follows that, if blockchain is selected as technology, then the best option is a **consortium blockchain or a private permissioned blockchain**.

The selected technology should employ a lightweight distributed consensus protocol that assures low latency and high throughput, such as a rotating coordinator algorithm.

Among the considered existing implementations (described in Section 9.2), the best candidates matching the mentioned performance requirements and design principles seem to be Multichain or alternatively Hyperledger Fabric as they have been designed specifically for permissioned blockchains.

10. Conclusions

Information sharing in the health domain is a complex and challenging process, since there are many stakeholders involved, different and sometimes competing standards and solutions to choose from, and important security, ethics, and regulation related constraints for any proposed solution to comply with. Based on the information, content and analysis of this document, it is important to consider the following aspects when building a new platform for sharing medical information:

- Patient consent is of utmost importance and infrastructure should be in place for its registration, enforcement, and withdrawal
- Compliance with GDPR and national laws, and implementing solutions to address significant requirements, such as the “right to erasure”
- Produce Interoperable solutions by linking and interoperating with well-established standards such as document and data formats (e.g. CDA, DICOM) and metadata and value lists (e.g. SNOMED, LOINC) in order to support common understanding and integration
- Handle the whole security spectrum: authentication and authorization of users, data privacy, auditing for “post-mortem” analysis and non-repudiation, data integrity, and machine enforced trust among the sharing organizations
- Enable the unique identification of patients while at the same time exposing the minimal set of personal information in order to protect their privacy
- Performance, scalability, and availability of the whole platform should be high in order to support the health-related processes efficiently, and
- Be part of the healthcare ecosystem, which means allow easy integration with existing infrastructure by featuring interoperable “ports and adapters” interfaces.

Additionally, the IHE profiles should not be not neglected. In the 2015/1302 Commission Decision, after consulting the European multi-stakeholder platform on ICT standardization and sectoral experts, 27 IHE profiles have been identified for referencing in public procurement, such as XCA, XDS and XDS-I, PIX, PDQ, ATNA, BPPC, etc. [EU15]

For authentication, the use of biometrics is important but especially when combined with an additional factor that is not passwords. Therefore, the proposal for authentication of medical staff is to use a hardware token and one biometric characteristic. Face seems the most appropriate one, because it is compatible with wearing gloves, is evolving to be robust to medical masks, and because it is far easier to acquire than iris.

Cloud computing is now used everywhere and provides an important set of features, such as adaptive scalability, performance, and benefits from business perspective. Especially for the sharing of clinical information, cloud can be very advantageous especially in cases where central repositories or central coordination are needed. But organizations should also be wary about the data protection, privacy, and access control mechanisms that should be in place, either offered by the cloud provider or built in house, in order to properly handle sensitive data and comply with regulations such as GDPR.

Blockchain is a highly interesting technology that could be put in good use in information sharing, more specifically for supporting data integrity verification and data auditability. Nevertheless, there are some major issues to be resolved, such as the compliance with GDPR’s “right to be forgotten” requirement which, unless the blockchain implementation is adapted, requires the deletion of data from the blockchain when the data owner asks to do it and this is not feasible, by design, in the “traditional” Blockchain implementations. Furthermore, there is also the additional complexity following by the huge heterogeneity of medical data (i.e., text, images, etc.) that do not fit exactly to the original design of Blockchain. Due to these considerations, the usage of Blockchain technologies within PANACEA Secure Information Sharing Platform is NOT recommended.