Cloud Computing in Europe

Appendix 8

Healthcare and

Human Health Research

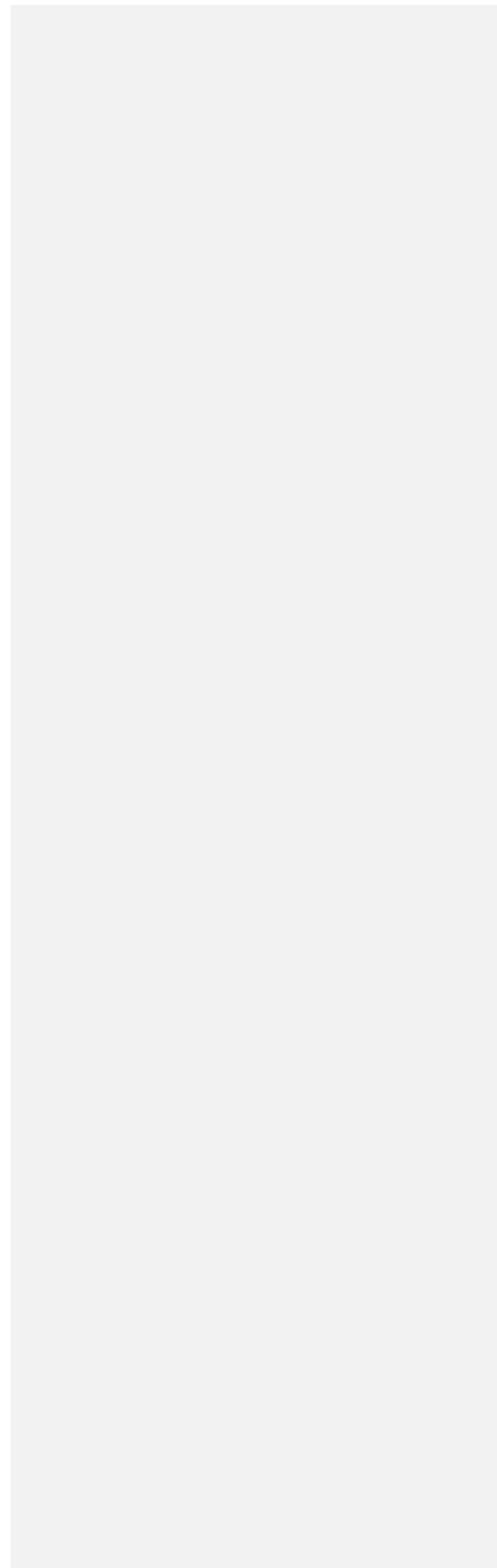15 July 2020

h-cloud.eu

Title	Cloud Computing in Europe. Appendix 8: Healthcare And Human Health Research
Lead Editor	Mark Dietrich
Contributors (in alphabetical order)	Mark Dietrich, Silvia Piai
Version	0.7
Date	15 July 2020
Confidentiality Notice	**Confidential** - The information contained in this document and any attachments are confidential. It is governed according to the terms of the project consortium agreement

Important Notice: Working Document

This briefing is an annex to the Green Paper v 0.7. The Green Paper is an outcome of the H-CLOUD project, a Coordination and Support Action that has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement n. 871920.
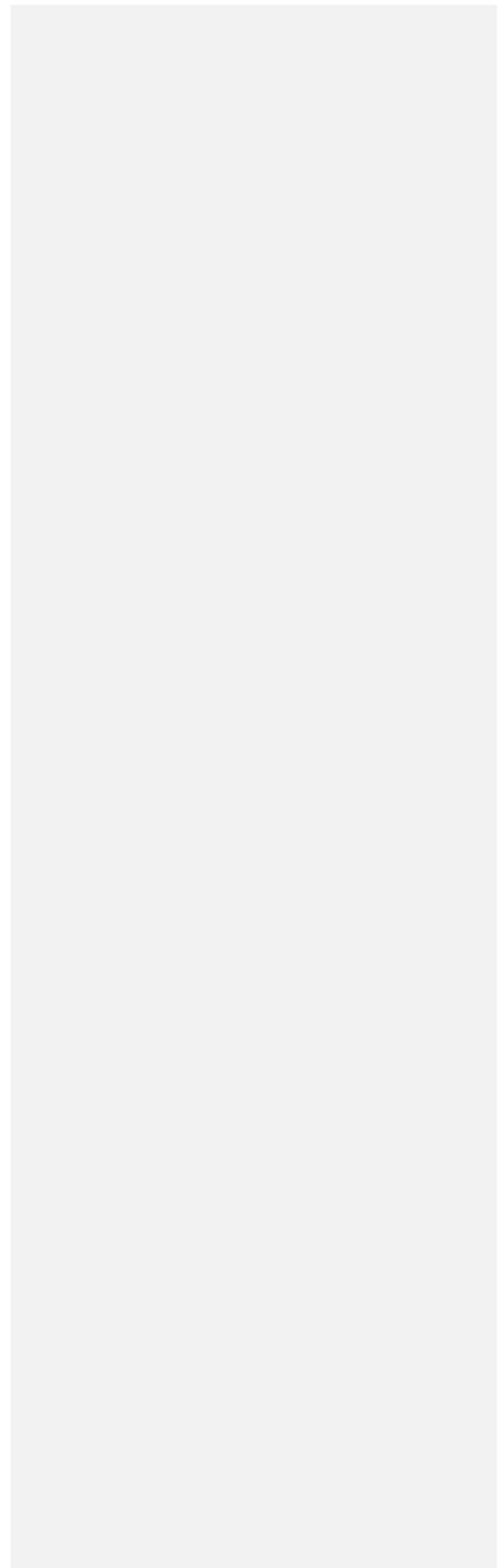
Copyright notice: © 2020-2022 H-CLOUD Consortium

1. CONTENTS

2. HEALTHCARE AND HUMAN HEALTH RESEARCH

This demand scenario looks broadly at ICT use cases arising from patient care and human health research. It considers many healthcare sectors:

- fundamental research (such as genomics),
- clinical trials and advanced clinical practice (e.g. using the latest results in genomics to improve bedside care),
- hospital, primary and community care (care delivered outside hospitals through the health system and through at-home social services)
- Public health to prevent and react to health crises that affect the population as a whole.

Increasingly the healthcare and health research sectors are merging into a continuum of care-provision, with corresponding demands to integrate the underlying ICT ecosystems to improve effectiveness (better outcomes) and efficiency (lower costs to society).

D-H Challenge 1: Healthcare ICT functions need to be coordinated across multiple healthcare organizations. This will require both sharing of sensitive data, and coordination of activities related to that data, as well as collaborative IT functions, such as coordinated enterprise resource planning (ERP) applications.

D-H Recommendation 1: Federation of healthcare ICT functions. Healthcare ICT could benefit from a federated solution, since it has several attributes required for successful "federation", namely:

- the need to combine disparate activities, as well as geographically separate activities, into the cohesive improvement of human health,
- the presence of multiple healthcare providers that prefer to operate as peers
- the fact that many healthcare providers are publicly funded and resist the idea that external services to be integrated into their own services should be "purchased" rather than provided as a public good.

Deleted: may

Deleted: regional
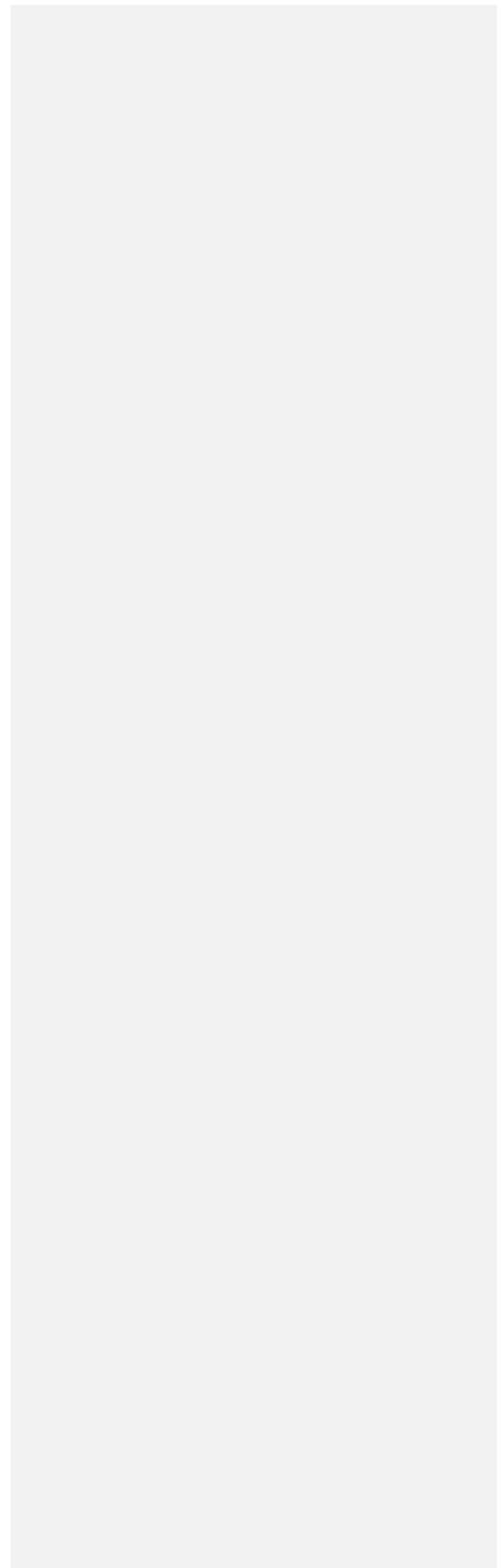
1   PERSONAL HEALTH INFORMATION

The sensitivity of personal health information ("PHI") has long been recognized. Even before GDPR, European countries had in place a distributed and complex legal regime for health data protection influencing the cybersecurity and infrastructure strategies of their healthcare organizations.

This environment, along with a historic lack of interoperability of healthcare business solutions, has led to the creation of siloed local IT infrastructure protected by robust security technologies. These silos have been centred around hospitals, physician practices and diagnostic/lab services, with limited capabilities for data sharing and collaboration. National health systems have worked to support varying degrees of electronic communication and data sharing with these "silos", while working to manage the risks associated with these initiatives.

The advent of GDPR introduced a definition of health data to be used within the EU and a common set of rules for processing this data. However, differences of GDPR implementation persist across member states. According to GDPR, PHI can be processed in some circumstances that require additional safeguard measures. These circumstances include:

- provision of direct health
- social care
- treatment or the management of health
- social care systems
- protecting subjects' vital interests, and
- based on explicit consent.

Processing can be also lawful for reasons of public interest in the area of public health, for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes. All processing should be based on European Union or member state law, guarantee suitable and specific measures to safeguard the fundamental rights of the data subjects and be proportionate to the aim pursued.

With these provisions, the GDPR gives broad boundaries to countries to confirm the details of the lawful basis for processing of health data, further derogations, and safeguards through national laws or statutes.

Overall, this has lead to misalignment in the interpretation of healthcare data protection rules between national data protection authorities (DPAs). For example, the positions of French and Italian DPAs related to the secondary use of healthcare data for Covid-19 pandemic management are stricter than those of the UK, Denmark, Norway and Ireland, which have taken a more pragmatic and a less literal approach. National differences have therefore translated into specific requirements for how and where PHI should be stored on ICT infrastructure -- sometimes requiring physical storage within those jurisdictions.

Deleted: rather
Deleted: disseminated
Deleted: al
Deleted: ,
Deleted: . This means there are
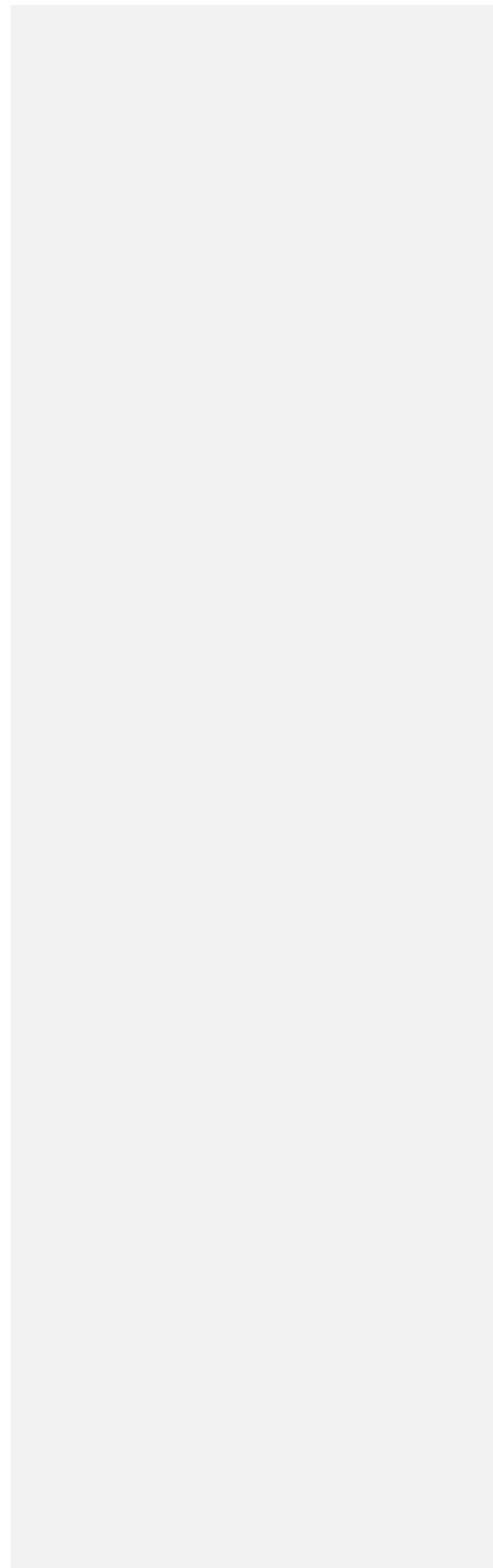Deleted: ,
Deleted: ,
Deleted: is
Deleted: ing
Deleted: a
Deleted: ,
Deleted: ly
Deleted: located

2 THE INTEGRATION OF SECTORS HAS AFFECTED PERSONAL HEALTH INFORMATION

The integration of different healthcare sectors has also merged requirements around how PHI is handled. The storage of most PHI is governed by a variety of laws and regulation, limiting the extent to which this data can be transferred, at the same time as innovations in healthcare are driving increased demands for access and/or transfer of this data.

The technologies of edge computing, wearable devices, internet of things (as well as robotics and artificial intelligence) are growing in importance for remote monitoring and telemedicine, and for enabling mobile point of care and "anywhere" healthcare -- creating new challenges for securely collecting, storing, transmitting and processing PHI

Increased access to PHI is essential for effective research, development of therapies, timely clinical intervention, and long-term chronic disease management, as well as management of health care facilities and the healthcare system overall.

D-H Challenge 2: Different healthcare sectors are integrating, which changes how PHI is handled. Healthcare innovations are driving increased need for access and/or transfer of that data across organizational boundaries. These changes are happening alongside the existing regulations for the storage, transfer and use of PHI. [Policy, Deployment]

D-H Recommendation 2: Create a common distributed data management solution, compliant with the GDPR, and particularly built with "Privacy by Design[1]", in order to enable increased capabilities in the healthcare sector.

---

[1] Cavoukian, Ann. "Privacy by Design in Law, Policy and Practice – A White Paper for Regulators, Decision-makers and Policy-makers"
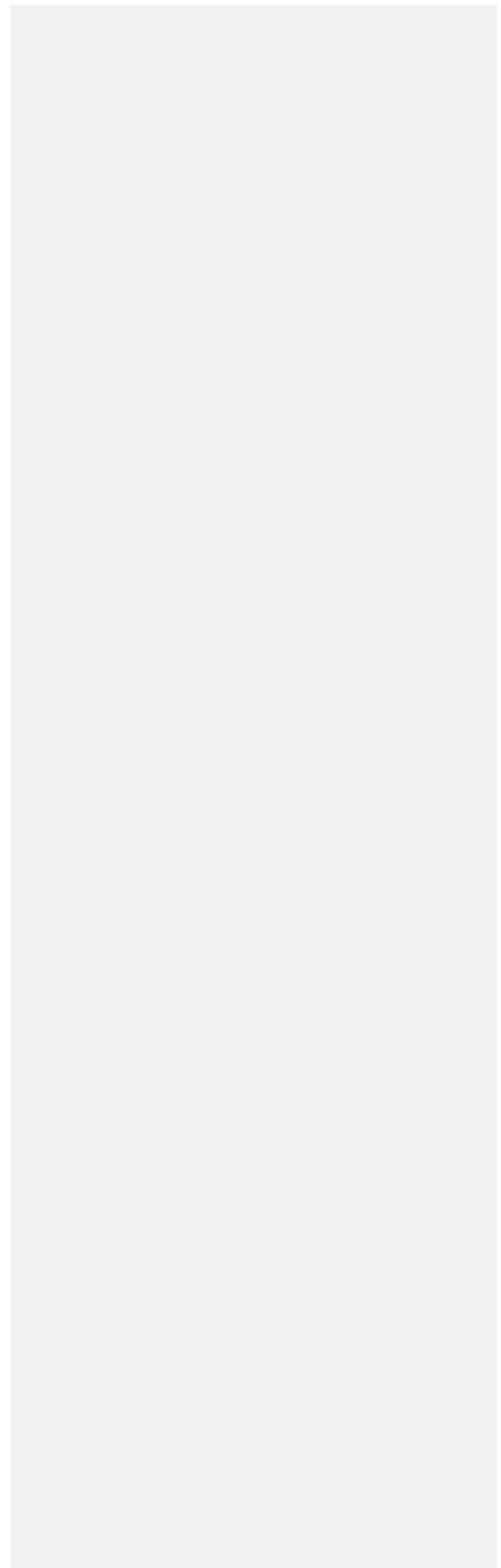
3 AGGREGATION OF PERSONAL HEALTHCARE INFORMATION

Human health research has struggled to balance the value of aggregating large cohorts of patient/sample data against the challenges of protecting the personal health information incorporated in that data. The early presumption was that anonymized data could be shared in a research context, and research enterprises created their own "firewalls" for anonymizing data, which could then be aggregated and shared without concern for privacy. However de-identification and anonymization of genomic data have been shown to offer weak protection against re-identification. Re-identification is easier when genetic data is combined with phenotype data (describing the individual) and clinical data (describing the illness).

D-H Challenge 3: Even anonymized human health research data must be treated as PHI, since it can be re-associated with other data to recreate PHI. [Deployment]

D-H Recommendation 3: All human health research enterprises must integrate GDPR compliance into their data management tools and policies, even when dealing with data regarded as exempt since it was anonymized.
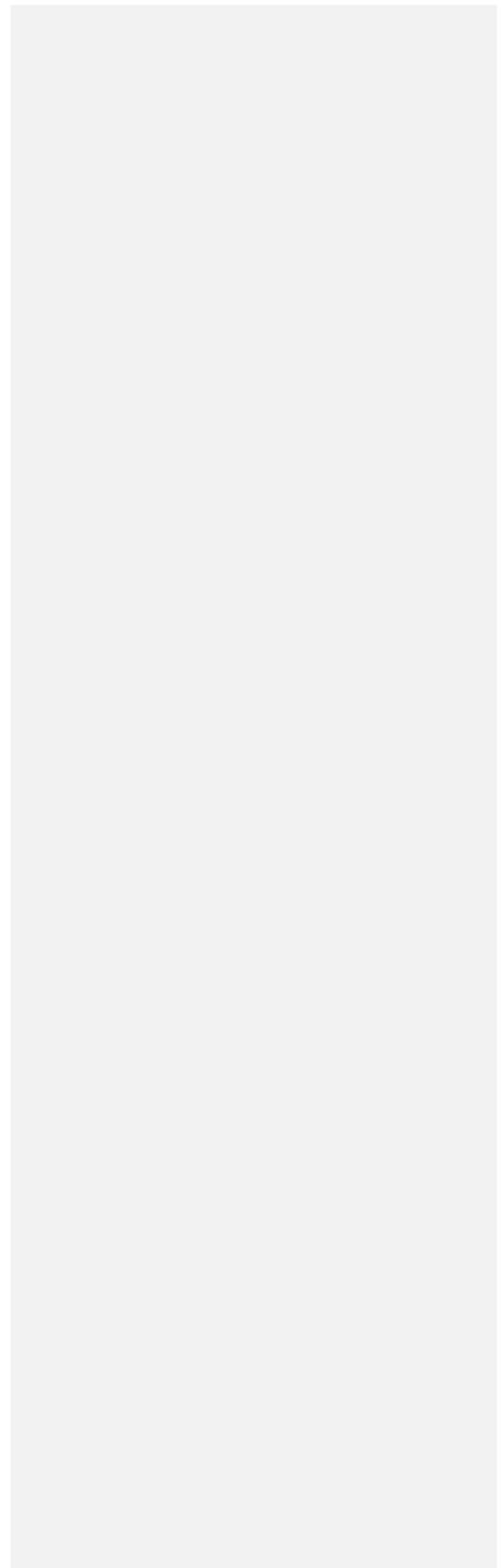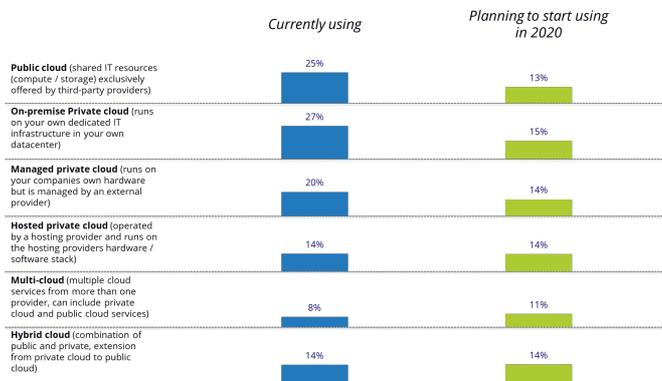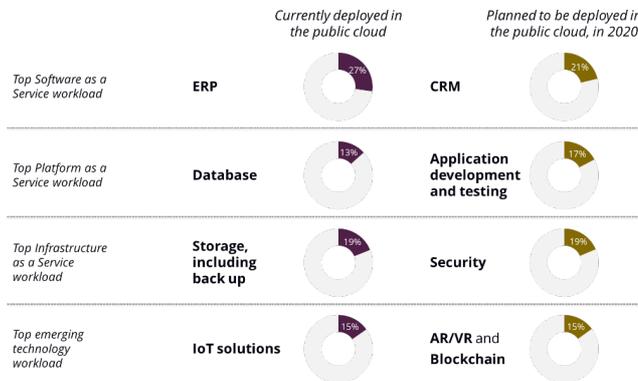
4 ADOPTION OF CLOUD TECHNOLOGIES BY HEALTHCARE PROVIDERS

Given the "siloed" nature of traditional healthcare ICT investments, and healthcare's naturally heightened concerns about data protection, security and privacy, cloud adoption in this sector mirrors adoption in the broader public administration sector, primarily investing in several flavours of private cloud, some IaaS public cloud, and a minority of organizations explicitly adopting multi- or hybrid cloud. Interestingly IDC research shows that European healthcare providers' key objectives when investing in cloud are improving their IT security posture, as well as their IT agility and efficiency. At the same time, however, key concerns on cloud adoption are also related to security, compliance, and IT governance (Including challenges related to defining standard services and SLAs).

Current and planned adoption by type of cloud

Currently using | *Planning to start using in 2020*

Public cloud (shared IT resources (compute / storage) exclusively offered by third-party providers) — 25% | 13%

On-premise Private cloud (runs on your own dedicated IT infrastructure in your own datacenter) — 27% | 15%

Managed private cloud (runs on your companies own hardware but is managed by an external provider) — 20% | 14%

Hosted private cloud (operated by a hosting provider and runs on the hosting providers hardware / software stack) — 14% | 14%

Multi-cloud (multiple cloud services from more than one provider, can include private cloud and public cloud services) — 8% | 11%

Hybrid cloud (combination of public and private, extension from private cloud to public cloud) — 14% | 14%

Source: *European Tech and Industry Pulse Survey*, 2019–2020 (healthcare sample n= 301 [WE = 243, CEE = 58])

Top solutions currently or planned to be deployed in the Public Cloud

Currently deployed in the public cloud | *Planned to be deployed in the public cloud, in 2020*

Top Software as a Service workload | **ERP** | 27% | **CRM** | 21%

Top Platform as a Service workload | **Database** | 13% | **Application development and testing** | 17%

Top Infrastructure as a Service workload | **Storage, including back up** | 19% | **Security** | 19%

Top emerging technology workload | **IoT solutions** | 15% | **AR/VR** and **Blockchain** | 15%

Source: *European Tech and Industry Pulse Survey*, 2019–2020 (healthcare sample n= 301 [WE = 243, CEE = 58])

D-H Challenge 4: Intrinsic concerns by healthcare organizations about implementing compliant data protection, security and privacy could represent a significant obstacle to cloud adoption by this sector. Since the healthcare sector is also fragmented, so significant numbers of healthcare providers do not have the scale to take on the challenge of data protection, security and privacy in the cloud, nor do citizens want publicly funded healthcare providers putting resources into duplicated solutions to this problem. They will hinder initiatives to combine the efforts of disparate healthcare providers into the cohesive improvement of human health. [Deployment]

D-H Recommendation 4: Technology solutions are required that simplify the adoption of cloud-based solutions, yet address the sector's underlying needs. For example, GDPR-compliant PaaS components robust enough to meet the needs of the healthcare sector. Indemnify their users against privacy breach fines when using these tools using best practices {applying reported patches/updating implementation-integration approaches}.
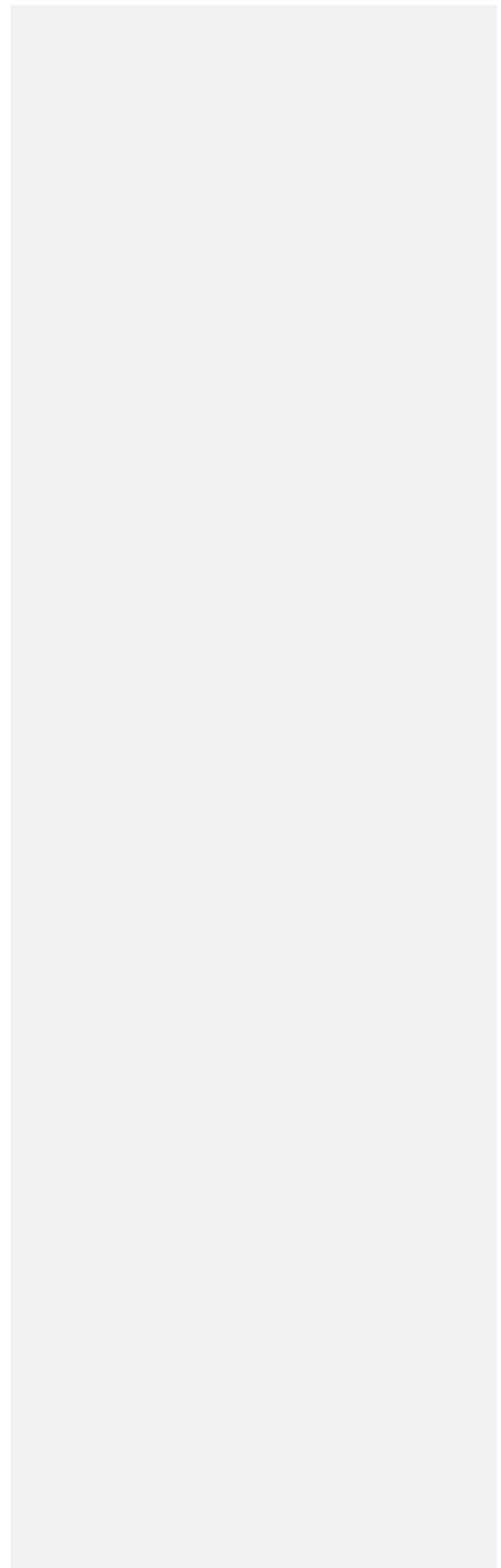
5 PHI DATA VOLUMES, DATA LAKES, OTHER APPROACHES TO DISTRIBUTED DATA ACCESS OR MANAGEMENT.

The Personal Health Information for one individual, in aggregate, does not typically represent a significant amount of data[2]:

● Electronic Health Records: a few MB, including narrative reports from health providers
● Personal genomic record: 200 GB (raw data from sequencers) or 2-3 GB (estimate for fully aligned genomic data).
● Diagnostic imaging: various sizes depending on technology and multiple "slices" per scan: CAT scan 20-30 GB.
● A critical mass of samples are needed, which require vast storage requirements, before cross population analysis can begin.

Even with significant numbers of diagnostic images, the total data for one individual could be less than, perhaps much less than, 1 TB based on current practices. Continuous monitoring of chronic conditions, as well as high frequency monitoring of patients in intensive care, could increase per person totals (on average) by roughly 10-20%.

Theoretically, for the ~500 million people who live in the EU, this 1 TB per person translates into 500 Exabytes. Even at 100 GB per person, this is 50 Exabytes for the EU. The total volume of EU PHI data, combined with the current restrictions on transferring that data, highlights the value of a distributed data management solution, capable of limiting data access to authorized users, while enabling the benefits of data aggregation and "big data" analysis to the extent possible.

In practice, several national and regional healthcare systems are creating "data lakes" to support the various "big data" analyses of the healthcare data available across their jurisdictions[3]. Even under the current regime, these data lakes are problematic. Often these data lakes are not distributed. Rather data is being brought to a single facility to enable this analysis. Neither are they anonymized. Analyses are conducted under the auspices of the relevant healthcare systems, which continue to be responsible for the appropriate use of, and controlled access to, results derived from those analyses. These data lakes have been implemented on premises or as custom solutions on public IaaS providers (notably AWS, Microsoft etc) with support from consulting firms with experience working with healthcare providers and PHI.

In addition, the European research infrastructure ECRIN[4] has been created specifically to provide tools that support multi-national clinical trials, which face many of the same challenges handling PHI across regulatory boundaries. it is understood that ECRIN is still searching for a solution.

D-H Challenge 5: Healthcare data volumes, and current initiatives to aggregate and process significant healthcare datasets, highlight challenges in efficiently accessing, sharing and analysing multi-national, distributed healthcare data while maintaining the protection, security and privacy of that data and avoiding unnecessary data movement and duplication.

D-H Recommendation 5: ECRIN and custom regional/national data lake solutions should be examined to see if they contain the seeds of wider solutions to the problem of distributed personal health data management.

D-H Challenge 6: Variations in national regulation of PHI across the EU currently limit the feasibility of creating data lakes with data from multiple countries.

D-H Recommendation 6: The medical exploitation of aggregated sensitive data in the cloud is recognised as a problem at the EU parliament level. Efforts are being made to overcome the problems

---

[2] Data volume estimates compiled from multiple sources; available upon request.
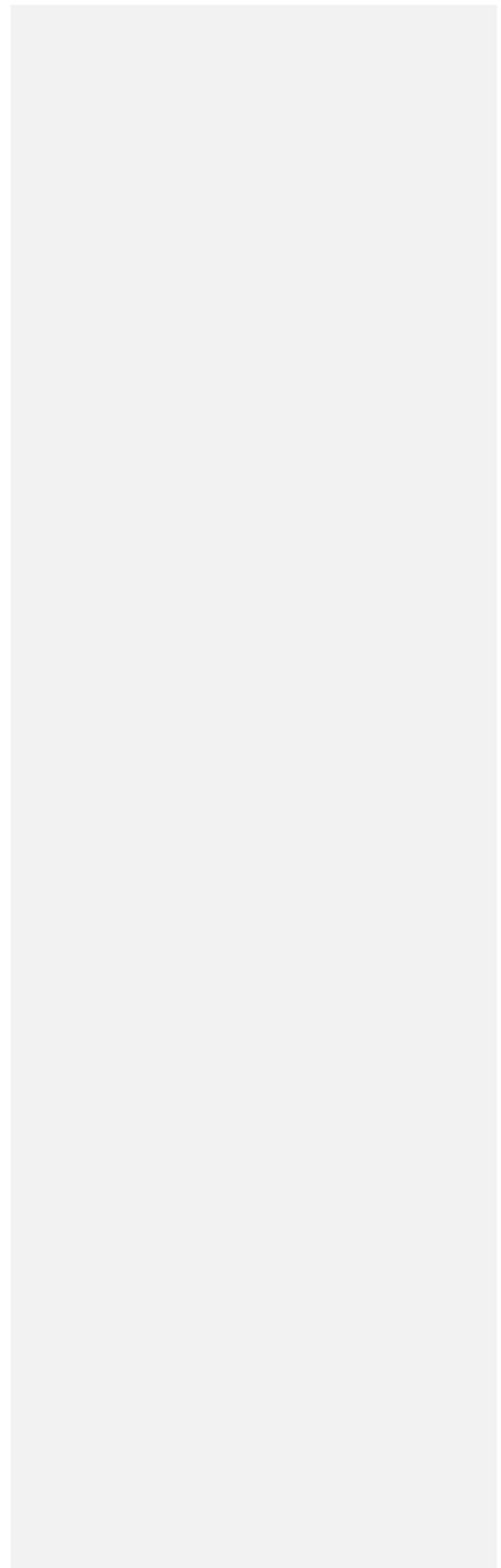
[3] The Secondary Use of Health Data and Data-driven Innovation in the European Healthcare Industry, IDC 2020 http://datalandscape.eu/data-driven-stories/story-6-secondary-use-health-data-and-data-driven-innovation-european-healthcare

[4] https://www.ecrin.org/

Formatted: Left

Deleted: http://datalandscape.eu/data-driven-stories/story-6-secondary-use-health-data-and-data-driven-innovation-european-healthcare

6 APIS PROVIDE A WAY TO AVOID DATA AGGREGATION

One way of developing insights based on larger sets of health data, without requiring aggregation of that data into a single data lake, is the idea of creating application programming interfaces (APIs) to allow remote users or software to interrogate local stores of health data about their contents. For example, a user could send an API request for statistics about the correlation of two genetic markers in a remote data set.

- If the user had relatively little authorization (perhaps even anonymous), the API would provide general correlation statistics, perhaps responding with relative proportions of samples (i.e. % out of the total held in the remote data set) exhibiting different combinations of those genetic marker contents.
- A user with a middle level of authorization (perhaps a *bona fide* health researcher from the jurisdiction storing the remote data set) might receive specific sample counts, perhaps with some random noise added, and/or with small sample counts suppressed from the results.
- A user with a higher level of authorization (perhaps identified as covered by a data transfer agreement with the custodian of the remote data set) might receive specific sample counts, and might be able to further query the results for additional phenotyping information (e.g. ethnicity, presence of different disease diagnoses).

This approach was adopted by the Global Alliance for Genomics and Health[5] (GA4GH), after considering a number of alternatives. It has been widely adopted in international cancer research. The approach is very effective when each record of the "remote" data set holds a large amount of data on a single patient - - allowing fundamental correlations between items in that individual record to be assessed remotely and with reference to the individual record alone. The approach is less useful if information on single patients is spread across different data sets and linked only by their "keys", i.e. patient identifiers.

---

[5] https://www.ga4gh.org/
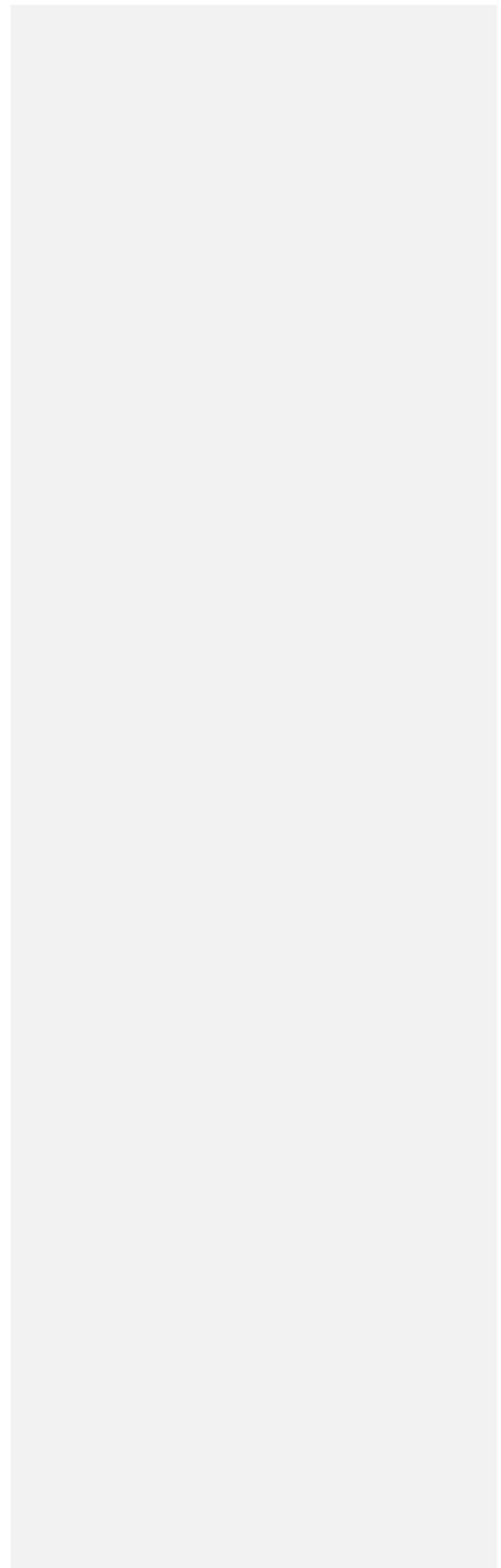
7   DEMAND SIDE CHALLENGES IN HEALTHCARE

Organisational challenge (Breadth)	Deployment sophistication		
	A: Relatively simple cloud deployments	**B: High Data protection and security needs**	**C: Sophisticated deployment of more advance technology**
Level 4: Cross sector coordination	Non-health Organisations already collecting health information from apps. Potential to integrate with health information.	Wider issue of access to healthcare data outside the sector. National regulations limit the ability to create data lakes across multiple countries.	
Level 3: Multiple orgs, same sector	Large amount of sharing across parts of the health care sectors, however much is still paper based. Large opportunity for efficiency improvements.	Privacy when storing and sharing personal health information (PHI) across the sector is key. Large volumes of data, means extensive distributed data management is needed. demands and moving compute to the source.	Exploitation of AI & advanced diagnostics relies on patient data (often anonymous) as basis for decision making models.
Level 2: Single larger org & supply chain	Many do not have the scale to deploy applications to the cloud.	Much personal data can be stored electronically. Personal Health data mainly kept in silos, and under strict policy guidelines to protect PHI from unauthorised disclosure and use.	Organisations tend to employ sophisticated but proven solutions from industry suppliers.
Level 1: Single small/med size org.	Siloed nature of healthcare applications means architecture hinders deployment and tends to hybrid private/public cloud solutions.		Smaller organisations trend to adopt industry wide solutions.