DataTools4Heart

A European Health Data Toolbox for Enhancing Cardiology Data Interoperability, Reusability and Privacy

Deliverable D8.1:
Dissemination Plan

Reference	**D8.1_ DataTools4Heart_LYN_27092023**
Lead Beneficiary	LYN
Author(s)	Beatrice Bressan
Dissemination level	Public
Type	Report
Official Delivery Date	27/09/2023
Date of validation of the WP leader	26/09/2023
Date of validation by the Project Coordinator	27/09/2023
Project Coordinator Signature	

DataTools4Heart is funded by the European Union's Horizon Europe Framework Under Grant Agreement No. 101057849.

Version log

Issue Date	Version	Involved	Comments
31/08/2023	V1.0	B. Bressan (LYN)	1st draft (PM)
05/09/2023	V1.1	B. Bressan, C. Capelli, E. Morley-Fletcher (LYN)	2nd draft (integration LYN comments)
12/09/2023	V1.2	X. Puig (UB)	Comments and format changes
15/09/2023	V1.3	B. Bressan (LYN)	3rd draft (integration UB comments)
27/09/2023	Final	X. Puig, P. Gkontra, K. Lekadir (UB)	Revised, corrected final version
15/09/2024	Final v2	B. Bressan (LYN), X. Puig (UB)	Revised and final version 2

Executive Summary

This deliverable illustrates the DataTools4Heart *Dissemination Plan* with the relevant activities carried out in the first year of the project implementation. After an introduction of the *WP8 – Coordination, Management, Dissemination, Communication and Exploitation,* its objectives, tasks and deliverables, the *Dissemination Plan* has been articulated into a diversified range of activities, such as web communications, webinars, dissemination events, publications, and multimedia. This also includes internal in-depth interactions with the DataTools4Heart partners and engagements, feedback gathering, collaboration and initiatives with relevant stakeholders within the Consortium partners' networks. Furthermore, a large seminar (attended by over 100 people) was organised at M9, to discuss both legal and technological issues related to the health data reuse together with other EU-funded projects. This seminar has represented a useful experiment, gathering not only projects from the same cluster of DataTools4Heart, but also other EU-funded projects signalled by the European Commission, who actively supported the seminar successful outcome. This experiment has shown that the dialogue-based approaches to dissemination can provide an interesting path to be pursued for raising attention on critical issues of common interest. Finally, DataTools4Heart has leveraged a variety of communication channels (website and social media as well as events and publications), the support of Consortium partners, and other partner projects channels. This document also presents the activities carried out in M1-M12, which include the preparation of the DataTools4Heart's branding (logo, icons, infographics, templates, etc.), the dissemination materials, the website setup, the social media channels, and the future action items.

Table of Contents

Acronyms

Artificial Intelligence	AI
Cardiovascular Disease	CVD
Code of Conduct	CoC
Electronic Health Record	HER
European Commission	EC
European Union	EU
General Data Protection Regulation	GDPR
Grant Agreement	GA
Horizon Results Booster	HRB
Innovative Tools for Electronic Health Records and patient registries	InToEHR
Intellectual Property Rights	IPR
Natural Language Processing	NLP
Named Entity Recognition	NER
Open Access	OA
Project Coordinator	PC
Strengths, Weaknesses, Opportunities, Threats	SWOT
Virtual Assistant	VA

List of figures

List of tables

1 Introduction

To address unmet Cardiovascular Diseases (CVDs) clinical needs, a major shift towards integrative data-driven approaches to develop personalised cardiovascular medicine is needed. To fulfil such needs, DataTools4Heart is co-creating and developing a comprehensive, federated, privacy-preserving cardiology data toolbox including standardised data ingestion and harmonisation tools, multilingual Natural Language Processing (NLP), federated machine learning and data synthesis methods, as well as Virtual Assistants (VAs) to help scientists and clinicians navigate through large-scale multi-source cardiology data, while complying with European regulations and data standards. These innovative tools, which will enable Electronic Health Records (EHR) data interoperability, quality, and reusability in cardiology, are built on different previous H2020 projects' achievements.

Healthcare data re-use in Europe faces ethical and legal issues, a high diversity in data formats and languages, and a lack of technical and clinical interoperability. Such issues can be properly addressed only on condition that a wider and deeper standardisation is ensured in the European Union as to what data anonymisation means in practice (and its difference from data pseudonymisation) and that homogeneous conditions are laid down for undertaking a secondary processing of health data for delivering better and personalised medicine and nurture clinical research. In this sense, feedbacks and guidance from regulatory authorities are crucial, also with a view to overseeing and approving methodologies and standards, to build a medical community common basis.

This document presents the DataTools4Heart *Dissemination Plan*, which has been structured in strict synergy with the project implementation, conceiving appropriate measures in accordance with expected results and relevant coordination, management, dissemination, communication, and exploitation needs. The plan integrates different activities, including web-based communications, multi-purpose events, ground-level and specialised publications, specific initiatives for the liaison with regulatory authorities in the field, as well as a series of cross-fertilisation activities directed towards relevant research initiatives. To carry out such a plan, the Consortium relies on specific dissemination and communication channels and tools, as well as contact networks and close cooperation among its partners in their different areas of expertise.

2 Work Package 8: Coordination, Management, Dissemination, Communication and Exploitation

2.1 Key aim and objectives

As specified in the Grant Agreement (GA), DataTools4Heart's Work Package 8 (WP8) is led by University of Barcelona (UB), acting as Project Coordinator (PC) pursuing the main project objectives of ensuring an effective action planning and coordination, as well as appropriate monitoring, reporting and quality control, risk management and mitigation, administrative and financial management, while providing prompt interface with the European Commission (EC). All Consortium partners jointly promote the project, also raising awareness about its objectives within what it will endeavour to foster, continuously guaranteeing a strong external dissemination and internal communication and reaching out to a wider scientific community interested in these objectives. One WP8 key element is to establish a dissemination plan to ensure the highest level of public outreach, the proper engagement of stakeholders in co-designing and validating the DataTools4Heart value proposition and long-term vision, while aligning with policy makers and regulators at the European Union (EU) level and cooperating with the EC.

2.2 Tasks and deliverables

WP8 - *Coordination, Management, Dissemination, Communication and Exploitation* (Table 1) is led by UB, with the Dissemination task led by Lynkeus (LYN), and the contribution from the whole Consortium in terms of stakeholders' engagement and results' dissemination within their networks.

UB is the leader of *T8.1* (*Scientific Coordination)*, which coordinates the DataTools4Heart's financial aspects, flagging possible conflicts to the appropriate bodies and handling risks assessment with an appropriate contingency plan, and of T8.2 *(Project Management)*, which ensures timely reporting to the EC as per the GA, prepares relevant documentation, and sets up internal reviews and processes for the deliverables preparation carried out with all WP leaders.

LYN, as leader of *T8.3 (Dissemination),* prepares the appropriate materials and events in cooperation with all partners, which will deploy the *Dissemination Plan* via their websites and attend several conferences as part of their dissemination activities.
Under *T8.4 (Communication)* and *T8.5 (Exploitation)*, UB sets up cross-project working groups and establishes a close relationship with other partners and EU projects under this call, while LYN regularly curates and updates the web-based communication channels content and produces an *Exploitation Plan*, organising dedicated workshop with relevant external

stakeholders, to gather additional feedback on exploitation potential, and finally an *IPR Management Policy*.

Table 1: WP8 tasks and deliverables

Task	Title	Deliverable	Title
T8.1	Scientific Coordination (UB, LYN: M1-M48)	D8.1	Dissemination plan (LYN: M12)
T8.2	Project Management (UB, ALL: M1-M48)	D8.2	Exploitation Plan, IPR Management Policy (LYN: M40)
T8.3	Dissemination (LYN, UB, ESC, BSC, TRANS, ALL: M1-M48)		
T8.4	Communication (UB, LYN, ALL: M18-M48)		
T8.5	Exploitation (UB, ALL: M12-M 48)		

2.3 WP activities: M1-M12 and beyond

Focusing specifically on the dissemination activities, the first 12 months of DataTools4Heart were dedicated to a series of tasks, such as:

- *M1-M4: branding, communication setup and presentation:*
  - Creation of the logo, branding proposals, PowerPoint (PPT) presentation templates
  - Preparation of the banner for the EU funding acknowledgment
  - Launch of the first news for the kick-off meeting of the project inception

- *M5-M8: social media channels and website realisation:*
  - Release of the logo, branding, QR code, PPT presentation template (Annex I)
  - Conception, design, and publication of the website (Annex II)
  - Setup of social media Twitter and LinkedIn accounts (Annex III)

- *M8-M12: social media and dissemination plan:*
  - Elaboration of the social media plan
  - Release of the brochure and booklet (Annex IV)
  - Drafting of the *Dissemination Plan*

At M12, the Consortium completed the production of the main dissemination and promotion materials. At the same time the internal communication infrastructures is currently up and running (with the internal communication channels activation via mailing lists, regular meetings, and teleconferences).

Between M6 and M12, the Consortium has enlarged its stakeholders list and continues engaging relevant entities through DataTools4Heart's website and social media, producing dedicated multimedia (e.g., teaser videos and some interviews to partners' representatives).

2.4 WPs dependencies

WP8 activity is linked to all the other WPs (Figure 1) in collaboration with DataTools4Heart's Consortium partners (Figure 2).

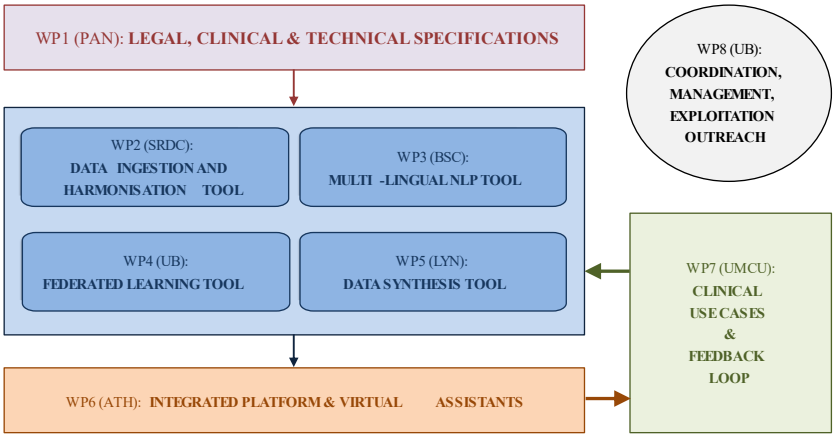


Figure 1: Interactions of WP8 with the other WPs

Figure 2: DataTools4Heart Consortium Partners

3 Dissemination Plan

3.1 Identification of stakeholder groups and target audiences

All dissemination and communication activities are conceived to provide a clear understanding of DataTools4Heart's vision, mission, focus and objectives, and are ultimately aimed at informing key stakeholders (i.e., primary audiences, public) about the project results, making them widely available for further research and longer-term development purposes, in the form of publications, and research data and tools. Given the importance of stakeholders' engagement and involvement in a co-design and co-creation process, DataTools4Heart's dissemination activities are primarily focused on supporting the DataTools4Heart's events, maximising outreach and engagement. In this sense, dissemination is aimed at engaging

relevant communities to spread interest and gather feedback all along the project for providing bases for its longer-term usage and sustainability.

Dissemination activities are pursued via a variety of channels, such as peer-reviewed and scientific journals and generalist publications; project events; print-based (e.g., posters, brochures) and multimedia-based (e.g., videos communication) materials; web channels (website, social media). All materials and channels have been produced and activated initially within the unrolling of the first activities and will be further developed in connection with the Consortium partners. The paramount objective of DataTools4Heart dissemination effort is to make the knowledge acquired, available not only to the EC, but also to the widest audience possible, through public dissemination of the project results, involving in the process an increasing number of stakeholders.

Thus, DataTools4Heart is using different channels, contents, and communication tools to address a variety of specific target groups of interest addressing the academic, industrial, clinical, and societal level. Tables 2 and 3 provide an overview of the target audiences, the relevant key messages, and desired outcomes of the interaction, together with the selected communication tools, based on their effectiveness in engaging specific stakeholders' groups.

Table 2: Dissemination approaches grouped according to different targets

<table>
<tr><th>Target</th><th colspan="2">Channel</th><th>Content</th><th>Approach</th></tr>
<tr><td>• Partners
• Researchers
• Public</td><td colspan="2">Website
(ongoing)</td><td>• Project info
• News
• Contact</td><td>• Description
• Visualisation</td></tr>
<tr><td>• Public
• Patients
• Cardiologists</td><td rowspan="2">Social
(weekly)</td><td>Twitter</td><td>• Stories
• Events
• Activities</td><td>• Storytelling
• Visual</td></tr>
<tr><td>• Experts
• Professionals
• Businesses</td><td>LinkedIn</td><td>• Information
• Events
• Collaborations</td><td>• Practices
• Guidelines</td></tr>
<tr><td>• Researchers
• Collaborators
• Technicians</td><td colspan="2">Open Access
(regularly)</td><td>• Novelties
• Results
• Networks</td><td>• Papers
• Proceedings</td></tr>
</table>

Table 3: Dissemination messages, channels grouped according to different stakeholders

Stakeholder	Message	Channel
• Patients • Citizens	• CDV data re-used securely • Personal data right respected • Research for better healthcare • Reduced costs for society • Patients co-shaping healthcare future	• Magazines • Conferences • Websites • Forums • Partners

• Public • Agencies	• Cross-border data re-use for better care, diagnostic, CDV cost-effectiveness • CoC for accountability, transparency, trust enhancement • Multilingual NER standardisation for EU international position improvement	• Meetings • Advocacies • Agencies
• Industry • Researchers	• Federated platform, standardisation tools, data models for complexity elimination among heterogeneous multi-site data • AI model testing/training sets from local to multi-site data repositories for precision/accuracy improvement • NLP tools extraction/organisation info for data increasement • AI VAs for interpretation improvement and navigation across complex multi-source data • Synthetic data as a novel solution for overcoming big data sharing issues	• Journals • Workshops • Digital hubs • Organisations • Meetings
• Providers	• Tools/systems for improved diagnostic solutions' development, transferable to clinical practice and research trials • Big data-driven digital technologies for clinician decision making support and healthcare personalisation improvement • AI VAs diagnostic tools' usability and adoption facilitation	• Journals • Summits • Associations

DataTools4Heart Consortium partners, in collaboration with academics and industry focus on organising matchmaking events and dedicated seminars to target various sectors. Thus:

- Dedicated materials (i.e., templates, posters, banners, brochures with DataTools4Heart's logo/infographics and visual identity) are disseminated via institutional channels.
- Dissemination events are advertised through the website and social media, leveraging the stakeholders' network (social media followers) and partners' communication channels.
- Different activities are promoted to spread out the project mission and objectives, attracting stakeholders from different areas to contribute to the validation of the project value proposition.

DataTools4Heart's dissemination relies also on the extensive network created by the University of Barcelona, as well as on medical, academic, and business networks of the Consortium partners. Besides stakeholder communities, affiliated societies and patient associations, Consortium partners are contributing to the DataTools4Heart dissemination through their institutional communication channels.

3.2 Implementation phases

The *Dissemination Plan* will follow the different phases of the DataTools4Heart implementation illustrated below together with its timeline and milestones (Figure 3).

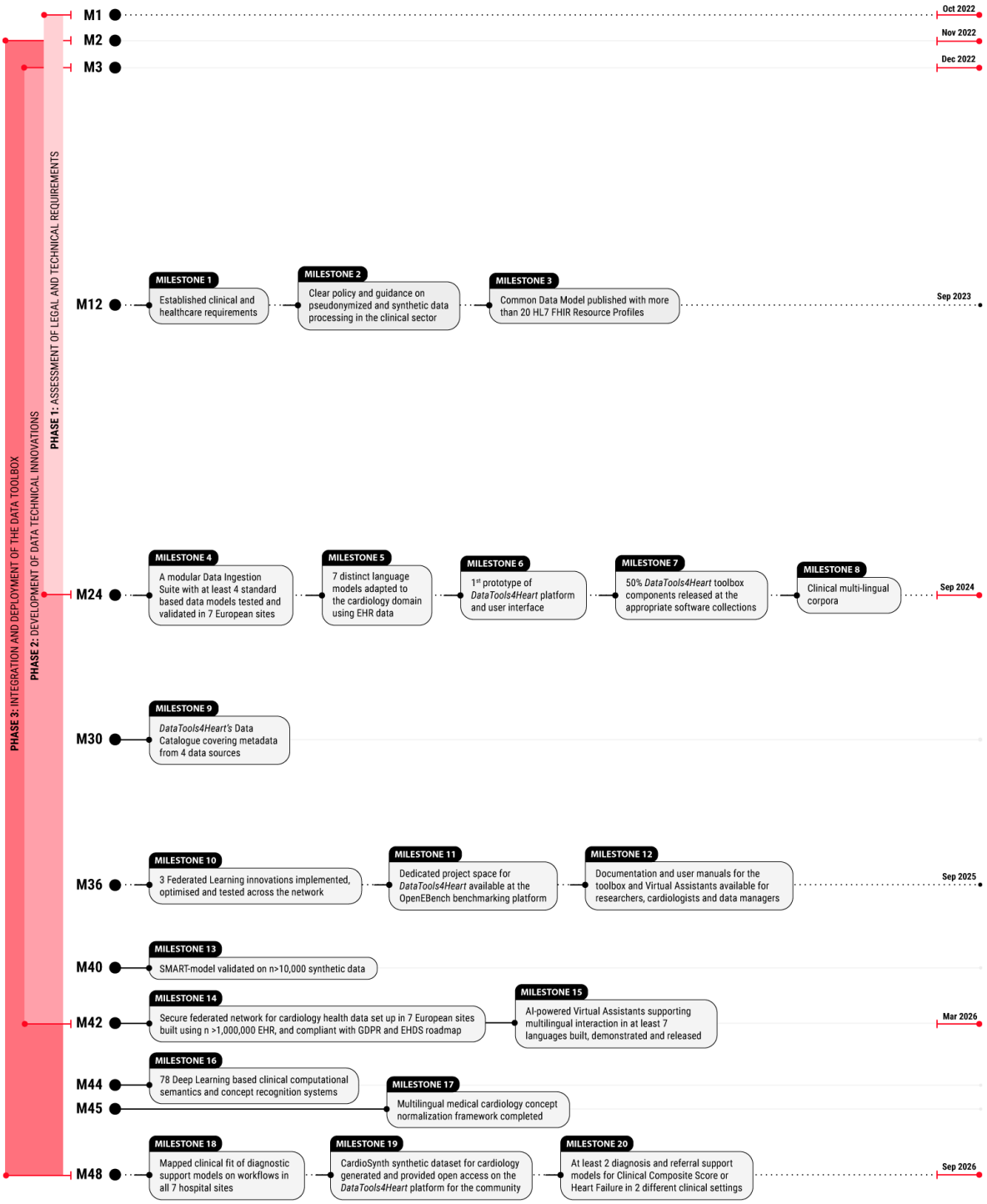


Figure 3: DataTools4Heart implementation timeline with relevant milestones

This *Dissemination Plan* has had a preparatory phase, dedicated to laying out the foundation of the actual activities by formulating an overall approach based on an analysis of the stakeholders and relevant audiences as well as factoring in the Consortium stakeholder

networks (section 4.3), the creation of a specific project branding and web-based material (Annexes I, II, III, IV), communication channels and tools (section 3.3), and the initial community building and communication (all this section).

After the completion of the initial tasks, the subsequent phase of the *Dissemination Plan* has been devoted to the actual dissemination of results through publications and events, impact assessment-based communications, together with stakeholders' engagement and audience expansion. Thus, facilitating feedback gathering on the developed solutions and vision and carrying out the exploitation opportunities' analysis. In the final phase of the project, the *Dissemination Plan* will entail the gathering of the conclusion from stakeholders on the project solutions, finalisation and realisation of its exploitation, management, and sustainability.

3.3 Channels and tools

Beyond leveraging its partners' institutional communication channels and networking tools (websites, social media, press releases, mailing lists, etc.), the Consortium is leveraging dissemination tools, conceived to reach a wide array of audiences via a variety of both mainstream (e.g., website, social media) and specialised (e.g., publications, events) channels, including:

- *Web channels: website, social media (Twitter, LinkedIn)* and *a publication web repository* (website), to be populated on a regular basis with news and achievements.
- *Print-based materials: brochures, booklets*, accompanying the project implementation, for general communication purposes to be distributed during specific events.
- *Multimedia: photos* (e.g., at dissemination events, project meetings), *videos* (e.g., interviews, teaser clips, demos) to be shared through the website and social media.
- *Dissemination events: conferences*, attended by Consortium partners; *workshops* for demoing the project's achievements; *webinars* for informing the whole community.
- *Publications: peer-review articles* (e.g., journals or proceedings), *generalists' publications* (e.g., newspaper or magazine news, press releases).

4 External Dissemination

4.1 DataTools4Heart website

DataTools4Heart's website (DataTools4Hearth.eu), finalised and released at the beginning of M6, is structured with the following six sections:

- *Homepage:* an overview of the DataTools4Heart *objectives*, as well as *news*, acknowledgments of EU funding, *contacts*, and the *Privacy Policy.*

- *Project:* a section providing a more in-depth description of DataTools4Heart's *vision*, *mission*, *focus*, and *objectives*, as well as its *implementation* timeline and milestones (Figure 3 above), and a description of the *expected results*. This page includes also the *ethical and legal* aspects containing the *guidance on pseudonymised and synthetic data* and the *Code of Conduct for health data re-use*, drafted for providing guidance on policy and programme measures to be incorporated within the DataTools4Heart project to facilitate the effective application of the GDPR.
- *Outreach:* a list of publications (*scientific papers* and *conference proceedings*). The *scientific papers* page includes publications based on DataTools4Heart's research, results, and fields, while *conference proceedings* include articles edited by partners for their participation to conferences.
- *Dissemination:* a collection of all *materials* (brochures, booklets, etc.) and *news* (achievements, meetings, public events, press releases, etc.), produced during the project for promotional purposes.
- *Consortium:* the geographical localisation and description of all DataTools4Heart's *partners*, their role, and the contact person as well as all job *vacancy* announcements.
- *Contact: contact form* for request for information or other inquiries from website users.

4.2 DataTools4Heart social media

DataTools4Heart's LinkedIn (https://www.linkedin.com/company/datatools4heart/) and Twitter (https://twitter.com/DataTools4Heart) social media accounts have been activated in M1 and are currently establishing continuous connections with organisations and individuals to create wide followers' audiences.

During the first year, DataTools4Heart LinkedIn page has counted 97 followers, 550 views and 230 unique visitors, compiling the following visitors' metric statistics (Figure 4), while its Twitter page has been followed by 11 projects and/or organisations: euCanSHare, EarlyCause, SIMCor, Lynkeus, UCLClinicalCardioEng, ESFRI_eu, ILSP – Athena RC, aidava_project, EHRA-PATHS, @realm_eu, and Retention Project.

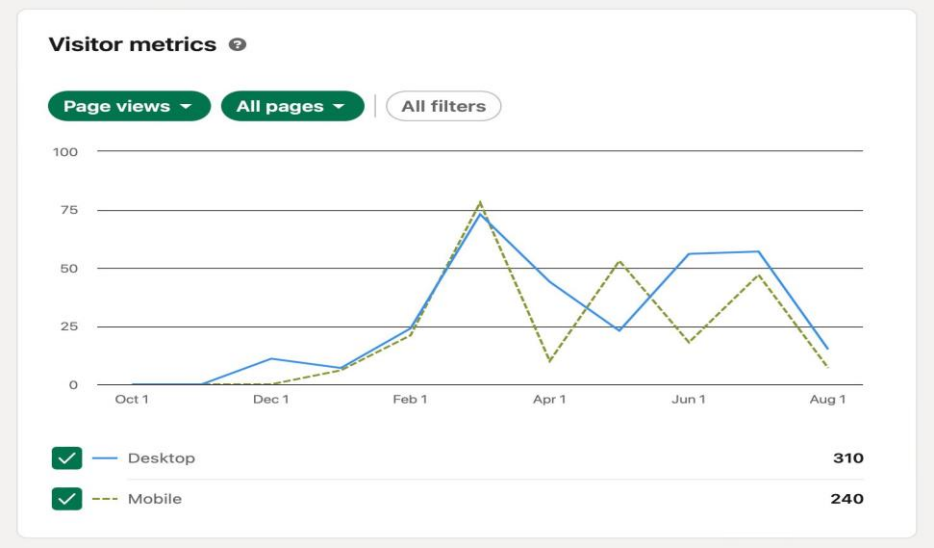


Figure 4: Visitor metric of the DataTools4Heart LinkedIn page during the first year

To maximise the social media dissemination, a dedicated approach, whose key elements are illustrated in Table 4, has been put in place (and continuously optimised via the analysis of the interaction with the project partners), to produce the most effective communication for each identified set of target audience, leveraging different medias and channels.

Table 4: DataTools4Heart's social media approach key elements

Who is the audience?	Where is the audience?	When is it communicated?	What are the messages?	Why is it innovative?	How is it executed?

• Industry
• Research
• Policy makers
• Patients
• Healthcare agencies
• Citizens
	• Czech Republic
	• France
	• Greece
	• Italy
	• Netherlands
	• Romania
	• Spain
	• Sweden
	• Turkey
	• UK
	• *…beyond*
		• LinkedIn (weekly)
		• Twitter (weekly)
			• Meetings
			• Vision
			• Mission
			• Objectives
			• Results
			• Articles
				• Personalised care
				• Centred medicine
				• Fast access
				• Higher efficacy
				• Legal, ethical
					• Social media
					• Editorial plan:
					  ○ written
					  ○ visual
					• Data analysis:
					  ○ visitors
					  ○ updates
					  ○ followers

On the base of the DataTools4Heart's social media approach, posts are published on Twitter and LinkedIn on a weekly basis and continuously during the DataTools4Heart's events. Such posts are regularly shared also by all DataTools4Heart partners. Furthermore, multimedia content, such as videos and interviews to the members of the Consortium are in preparation and will be disseminated via dedicated accounts and/or channels. Such videos are regularly recorded during face-to-face Consortium meeting.

4.3 Consortium institutional channels

Besides official channels and tools, DataTools4Heart also relies on the support of its Consortium partners using their channels listed in Table 5 (Website, Twitter, Facebook, LinkedIn, Instagram, YouTube, Newsletter, Press Release, Mailing List) to further share and circulate project-related contents and materials.

Table 5: Consortium partners institutional channels for the Dissemination Plan leverage by Website: site, Twitter: X, Facebook: FB, LinkedIn: in, Instagram: IG, YouTube: YT, Newsletter: NL, Press Release: PR, Mailing List: ML

Partner	Site	X	FB	in	IG	YT	NL	PR	ML
University of Barcelona	✓	✓	✓	✓	✓	✓		✓	✓
Amsterdam University Medical Centre	✓	✓	✓	✓	✓				✓
ATHENA Research Center	✓	✓	✓	✓			✓	✓	✓
Barcelona Supercomputing Center	✓	✓	✓	✓	✓	✓	✓	✓	✓
Bucharest Emergency Clinical Hospital	✓								✓
European Society of Cardiology	✓	✓	✓	✓	✓	✓		✓	✓
Fondazione Policlinico Universitario A. Gemelli	✓		✓	✓	✓				✓
International Clinical Research Center	✓	✓	✓	✓				✓	✓
LYNKEUS	✓	✓	✓	✓					✓
PANETTA	✓	✓		✓		✓			✓
Region Stockholm, Karolinska University Hospital	✓	✓	✓	✓	✓	✓		✓	✓
SIEMENS	✓	✓	✓	✓	✓	✓	✓	✓	✓
Software Research & Development Consultancy	✓	✓	✓						✓
TRANSLATED	✓	✓	✓	✓					✓
University College London	✓	✓	✓	✓	✓	✓	✓	✓	✓
University Medical Center Utrecht	✓								✓
Vall D'Hebron Hospital Research Institute	✓	✓	✓	✓	✓	✓			✓

4.4 External Advisory Boards

Throughout DataTools4Heart's action, its stakeholders and relevant international partners are continuously consulted via the dedicated external advisory boards for reviews of the project key activities and outcomes.

The external advisory experts have been selected at the beginning of the project and will serve for the project duration to ensure that the needs and requirements of European innovators can be considered to deliver possible products and services for exploitation that benefit citizens, patients, and society at large. External experts are envisaged to meet periodically for discussions with internal experts from the DataTools4Heart Consortium to possibly review the project developments and services, to provide input and feedback, and to support in reviewing the project value proposition. The external advisory experts represent institutions such as: St George's University of London, Imperial College London, Uppsala Clinical Research Centre, Heidelberg University, and University College London.

4.5 Presentations on DataTools4Heart

Crucial dissemination on project information is conveyed also through conferences' attendance. Table 6 shows all the presentations on DataTools4Heart carried out during the first year.

Table 6: Presentations on DataTools4Heart held on 2023

Type	Topic	Audience	Title	Date	Place
Summit	Cardiology and Cardio policy	• Cardiologists	ESC Spring Summit	09-10 Mar	Nice, France
Congress	Cardiology	• Cardiologists • Health professionals	ESC Acute Cardiovascular Care	24-26 Mar	Marseille, France
Congress	Preventive Cardiology	• Cardiologists • Health professionals	ESC Preventive Cardiology	13-15 Apr	Málaga, Spain
Congress	Cardiology and Heart rhythm	• Scientists • Healthcare professionals • Arrhythmia experts	EHRA Congress	16-18 Apr	Barcelona, Spain
Congress	Cardiovascular Imaging	• Cardiologists • Radiologists • Nuclear Medicine specialists	EACVI	10-12 May	Barcelona, Spain
Congress	Courses on interventional cardiovascular medicine	• Healthcare professionals	EuroPCR Course	16-19 May	Paris, France
Congress	Heart Failure	• Cardiologists	Heart Failure	20-23 May	Prague, Czechia
Congress	Cardiovascular nursing	• Healthcare professionals	ACNAP	20-23 May	Edinburg, UK

Congress	Transforming the digital healthcare: digital health polemics: ○ *AI,* ○ *Cyber innovation* ○ *Mental health* ○ *Climate health* ○ *Active aging* ○ *Telehealth* ○ *Digital maturity*	• Digital health stakeholders: ○ *Policy makers* ○ *Payers* ○ *Tech developers* • Life sciences groups • Patient groups	Radical Health Festival	12-15 Jun	Helsinki, Finland
Online Seminar	Reuse of data in healthcare	• EC • Digital health stakeholders • EU funded consortia	Privacy and tech scenarios on reuse of health data: EU projects' roundtable	19 Jun	Online
Congress	Cardiology and cardiovascular-related diseases	• Cardiologists	ESC Congress	25-28 Aug	Amsterdam, Netherlands

In particular, a seminar titled "*Privacy and Tech Scenarios on Reuse of Health Data: EU Projects' Roundtable*", organised and chaired by LYNKEUS and PANETTA and inaugurated by the European Commission, took place online on the 19th of June 2023. Scope of this event was to analyse together with other 13 EU-funded projects (listed in section 6) the issues relating to the reuse of personal and health data for research purposes, both from a legal and ethical as well as a technological standpoint.

4.6 Upcoming events

Table 7 shows some events in 2023 and 2024 where Consortium partners' representatives are likely to attend as participants or speakers, for networking, disseminating, and promoting the DataTools4Heart's results and developments.

Table 7: Future events for possible partners' attendance to present DataTools4Heart

Type	Topic	Audience	Title	Date	Place
Congress	Cross-disciplinary course to be up to date on 3 valves: ○ Aortic ○ Mitral ○ Tricuspid	• Cardiologists	PCR London Valves	19-21 Nov 2023	London, UK
Conference	Machine Learning for	• AI scientists	NeurIPS	10-16	New

	healthcare			Dec 2023	Orleans, USA
Congress	Acute Cardiovascular care	• Cardiologists	ESC Acute Cardiovascular Care	08-10 Mar 2024	Athens, Greece
Congress	Cardiology, Heart rhythm	• Healthcare professionals	EHRA	07-09 Apr 2024	Berlin, Germany
Congress	Cardiology and prevention	• Healthcare professionals	ESC Preventive Cardiology	25-27 Apr 2024	Athens, Greece
Conference	Nuclear Cardiology and Cardiac CT	• Clinicians • scientists	ICNC-CT	19-21 May 2023	Seville, Spain

4.7 Open Access publications

Project publications will include peer-review publications (e.g., journal or conference proceedings articles) and generalists' publications (e.g., magazine articles, press releases).

Following EC's guidelines for dissemination of results, all publications produced within DataTools4Heart will be made available in Open Access (OA) upon publication or after an embargo period, if any. Thus, Consortium partners will publish primarily on OA journals, or elsewhere, making the full-text available (as published version, pre-print, or accepted manuscript) in scientific archives, and possibly in other sources as well (e.g., institutional archives, etc.).

Specialists' publications will mostly include journal or conference proceedings articles in relevant fields, available for OA. Publications of particular interest will also be disseminated through DataTools4Heart's website. Table 8 lists all articles that have been published by DataTools4Heart Consortium members since the beginning of the project and their contributions with the project tasks.

Table 8: List of DataTools4Heart publications in 2023

Type	Title	Author	Publisher	Date	Contribution
Journal article	*Explainable Artificial Intelligence and Cardiac Imaging: Toward More Interpretable Models*	A. Salih, I. Boscolo Galazzo, P. Gkontra, A. M. Lee, K. Lekadir, Z. Raisi-Estabragh, S. E. Petersen	Lippincott Williams and Wilkins Ltd.	12 Apr	Deep learning and machine learning models, and convolutional neural networks to improve interpretability of AI. Work linked to the

					development of Federated Learning Innovations in WP4 (T4.2-T4.4)
Journal article	*Druggable proteins influencing cardiac structure and function: Implications for heart failure therapies and cancer cardiotoxicity*	A. F. Schmidt, M. Bourfiss, A. Alasiri, E. Puyol-Anton, S. Chopade, M. Van Vugt, S. W. Van der Laan, C. Gross, C. Clarkson, A. Henry, T. R. Lumbers, O. Van der Harst, N. Franceschini, J. C. Bis, B. K. Velthuis, A. S. J. M. Te Riele, A. D. Hingorani, B. Ruijsink, F. W. Asselberts, J. Van Setten, C. Finan	American Association for the Advancement of Science	26 Apr	Research on Cardiovascular Magnetic Resonance variables. This impacts on treatment optimization for heart failure patients by identifying novel targets for drugs. Work linked to the heart failure clinical use case in WP7 (T7.1)

5 Internal Dissemination

An important part of the dissemination effort within DataTools4Heart is devoted to optimising the cooperation and alignment among the Consortium partners. For this reason, several activities, including the organisation of various online and face-to-face meetings, have been put in place to consolidate the internal partnership within the project. During the first year of the project the following events have been organised:

- *Kick-off meeting* held in Barcelona (University of Barcelona, 6th-7th October 2022) with all Consortium partners.
    - *Outcome:* All Consortium partners established the necessary steps to carry out all the project achievements, including the technical and general meetings schedule.
- *First Consortium meeting* held in Barcelona (University of Barcelona, 2nd-3rd March 2023) with participants from all Consortium partners.
    - *Outcome:* The Consortium partners introduced the use-cases, the objectives of the project, and dissemination approach, including the structure of the Data Catalogue.
- *Governing Board e-meeting* on a regular monthly basis with all WP leaders with participants from all Consortium partners.
    - *Outcome:* All WP leaders discussed about WPs' achievements, issues, and actions to take per each WP to establish best directions to move forward.

- *Dissemination group e-meetings* on a regular basis with Project Coordinator (UB), Dissemination Task Manager (LYN) and Partner (ESC) members to exchange ideas.
    - *Outcome:* Dissemination group members coordinated and planned joint dissemination initiatives and activities, including web-based channels' presence.
- *An online session* (June 2023) with questionnaires and specifications (Annex V) to align DataTools4Heart's vision and objectives for setting up a preliminary SWOT analysis.
    - *Outcome:* Engagement with new stakeholders and researchers (section 6).

To ensure a swift and efficient internal dissemination within the project, minutes of these meetings have been provided when needed and, when technically possible, the meetings have been recorded.

Moreover, a dedicated Google Drive folder has been created to facilitate document sharing, and a Slack channel has been also activated to facilitate direct interaction among partners in near real-time.

After summer 2023, on-site and online meetings together with other public events will be organised, communicated, and disseminated through DataTools4Heart social media, website, and other communication channels from all partner institutions.

6 Engagement with stakeholders and researchers

A crucial activity of DataTools4Heart is the engagement with stakeholders and researchers working on data reuse. Over the past year, two main activities have been carried out.

First, DataTools4Heart joined the "*Innovative Tools for Electronic Health Records and patient registries*" (InToEHR) cluster facilitated by EC and composed of the five following projects or research consortia (in alphabetical order of their acronyms):

- AIDAVA (GA 101057062): AI powered Data Curation & Publishing Virtual Assistant.
- DataTools4Heart (GA 101057849): A European Health Data Toolbox for Enhancing Cardiology Data Interoperability, Reusability and Privacy.
- eCREAM (GA 101057726): enabling Clinical Research in Emergency and Acute care Medicine through automated data extraction.
- IDEA4RC (GA 101057048): Intelligent Ecosystem to improve the governance, the sharing and the re-use of health Data for Rare Cancers.
- RES-Q PLUS (GA 101057603): Comprehensive solutions of healthcare improvement based on the global Registry of Stroke Care Quality.

Participation to InToEHR will facilitate dissemination of progresses and findings and contact with the Horizon Results Booster (HRB) platform when appropriate. HRB provides support to boost the potential exploitation of the research results of Horizon projects, as well as aid in dissemination and commercial goals for marketable project outputs.

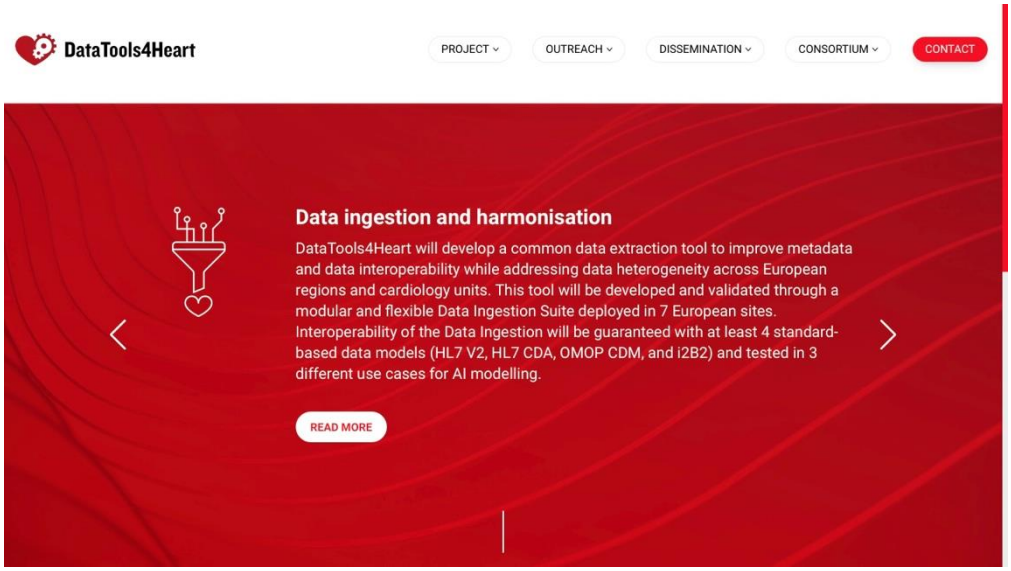
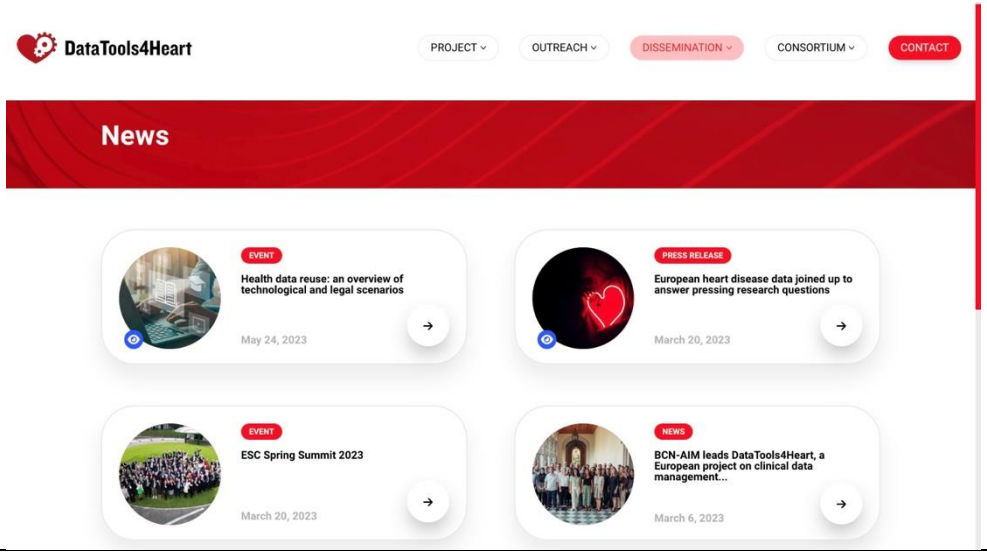
Second, DataTools4Heart organised the online seminar titled "*Privacy and Tech Scenarios on Reuse of Health Data: EU Projects' Roundtable*". The seminar took place online on 19th of June 2023, and was organised and chaired by LYNKEUS and PANETTA. The event was inaugurated by the European Commission and included presentations from 14 EU funded projects which shared insights, and perspectives shared by the presenters regarding the secondary processing of health data for research purposes. The online seminar was attended by over 120 participants representing EC, academic and research institutions and representatives from industry. Together with the seminar, an online survey was conducted amongst participants from 19 projects pre-identified by the European Commission. The survey aimed to map the main technological solutions adopted in the field of reuse of health data and capture other Consortiums' and experts' opinions regarding opportunities and barriers. The survey was made of 16 questions with a mix of multiple choice and Likert scale questions. Such activities contribute to compile a SWOT analysis (Annex V) to provide valuable insights into the landscape in which DataTools4Heart aims to develop and deploy its solutions for the effective reuse of health data in cardiac research. The findings demonstrate that EU research is at the forefront of developing technological solutions for data processing and data protection within the existing legislative framework that encourages research. The innovative and promising technologies proposed by DataTools4Heart align well with this research landscape. Such dissemination activity is crucial to gain visibility for the project and lead the change in the reuse of health data.

7 Annexes

7.1 AI: Preview of some DataTools4Heart branding materials

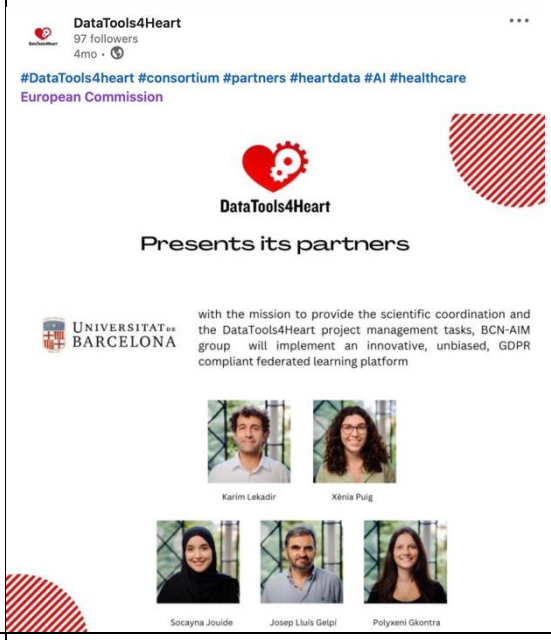
Logo	DataTools4Heart	PPT Template	

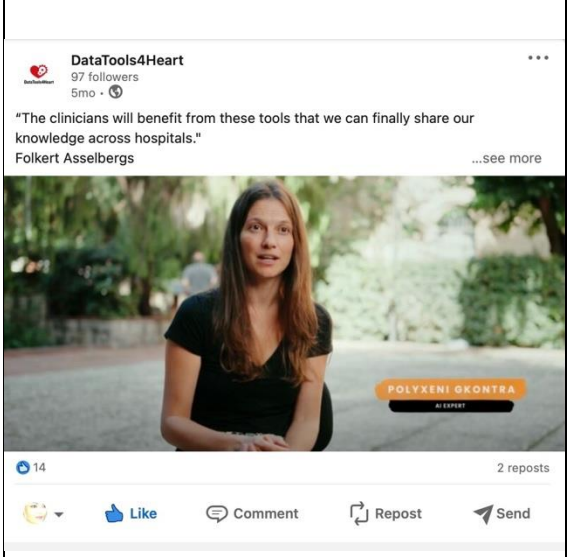
7.2 AII: Preview of some DataTools4Heart website sections

Homepage	
News	

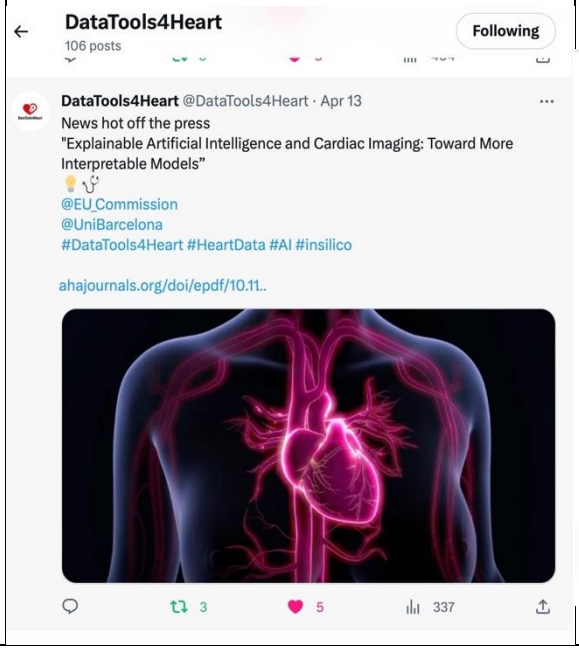
7.3 AIII: Preview of some DataTools4Heart social media posts

LinkedIn





Twitter

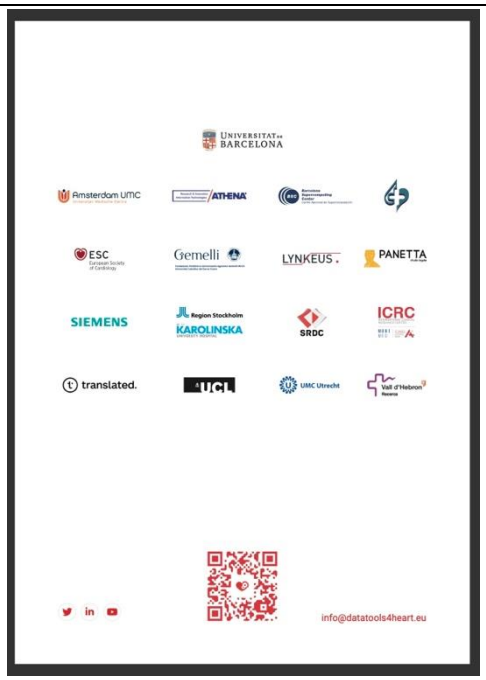


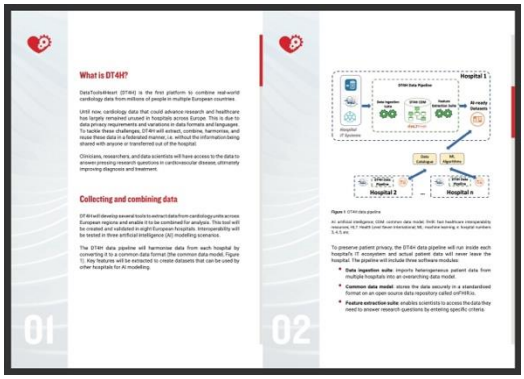
7.4 AIV: Preview of some DataTools4Heart dissemination material

Brochure

Booklet


DataTools4Heart
DataTools4Heart is funded by the European Union's Horizon Europe Framework Under Grant Agreement No. 101057849


UNIVERSITAT DE BARCELONA
Amsterdam UMC
ATHENA
ESC
Gemelli
LYNKEUS
PANETTA
SIEMENS
Region Stockholm
KAROLINSKA
SRDC
ICRC
translated.
UCL
UMC Utrecht
Vall d'Hebron
info@datatools4heart.eu


What is DT4H?
Collecting and combining data
01
02


Converting the data to common clinical codes
Training the model to recognise clinical information
Clinician-in-the-loop
03
04


Preserving patient privacy
DT4H aims to securely combine clinical care data by employing innovative techniques such as federated learning (Figure 4) and data synthesis. Instead of centralising data collection from hospitals in one place to train the AI model for addressing specific research questions, the model itself will be sent to each hospital for onsite training using local data. Additionally, synthetic data will be generated to mimic the target population while ensuring that no individual information is retained. The synthetic data generation will be rigorously assessed over the course of the project to ensure its quality. The resultant models will then be aggregated at a central location. The updated model will be returned to each site to continue training and the process will be repeated until the model is fully trained. An open-source privacy-conscious synthetic dataset, CardioSynth, will be an important legacy of DT4H.
Figure 4: Training a model using federated learning
n: hospital numbers 3, 4, 5, etc.
Virtual assistants will help users navigate the platform
DT4H will develop an integrated platform incorporating all elements of the system, from collecting data, converting it to common clinical codes, and combining it anonymously to preserve patient privacy (Figure 5).
Users will access the system via the "DT4H entry point". They will be able to input information into the data catalogue and extract information using search criteria such as "type 2 diabetes".
AI-powered virtual assistants operating in seven languages will help researchers and clinicians to navigate the platform and access the information they need from the large-scale cardiology datasets. Ease of use will be examined in pilot studies conducted in eight European hospitals using real-world data.
Figure 5: DT4H integrated platform
ML: machine learning; NLP: natural language processing; SMPC: secure multi-party computation.
05
06

7.5 AV: DataTools4Heart: Mapping technological solutions for health data use/re-use[1]

SYNOPSIS

Cardiovascular disease (CVD) is the main cause of mortality worldwide, representing about a third of annual deaths, and patients' CVD medical care produces a very large amount of data. Re-use of health data is a great opportunity for research but poses also challenges to protect the rights of individuals. DataTools4Heart is aimed at designing methods to reuse such data to facilitate research and improve the conditions of patients affected by CVD. An unprecedented generation of innovative emerging data and technologies to be used in healthcare is now available worldwide. DataTools4Heart, together with other projects, has been funded during such an exciting and yet rapidly evolving period, making it crucial to stay up to date with the landscape of novelties in terms of legislation and technological solutions. Thus, joining forces with other related projects has been deemed essential in view of exerting a positive impact.

This report attempts to map the main technologies developed to ensure a safe re-use of health data, highlighting both opportunities and barriers related to their implementations. Such assessment was conducted using direct (a survey) and indirect methods (analysing methods and proposals of other related projects). The results are thus presented herein, in form of a SWOT analysis, also based on the activities presented on 19th of June 2023 during the aforementioned online seminar "*Privacy and tech scenarios on reuse of health data: EU projects' roundtable*".

Introduction

DataTools4Heart aims to co-create, develop, and demonstrate a comprehensive, federated, privacy-preserving cardiology data toolbox including standardised data ingestion and harmonisation tools, multilingual natural language processing, federated machine learning and data synthesis methods. Thus, virtual assistants to help scientists and clinicians navigate through large-scale multi-source cardiology data, while complying with European regulations and data standards.

Throughout the project, DataTools4Heart will be "*developing robust novel solutions compliant with legal requirements*" and consider requirements stipulated in the relevant European regulations, such as the General Data Protection Regulation (GDPR), the Medical Device Regulation, the Data Governance Act, and the recently published EC proposal for AI regulations, as well as current plans for a European Health Data Space (EHDS).

The Project has been funded during an exciting and yet rapidly evolving period which aims to foster and nurture the re-use of health data for research purposes within the EU. An unprecedented generation of data and technologies is now available across the globe. It is therefore crucial to capture such moment to stay up to date with the landscape of novelties in terms of legislation and technological solutions. In addition, the cooperation with other projects is key to maximise the impact of DataTools4Heart and join forces to have a real impact on current developments in the medical, technological and AI industries.

This document reports the study which was carried out to map and assess the main projects aiming to enable the re-use of health data and provides a SWOT analysis of technological solutions and frameworks in Horizon-funded projects in the context of the latest novelties of the EU legislation. Such SWOT analysis will help finding common topics among cluster of projects and facilitate the interactions within the research community.

Investigation

This analysis was based on two phases:

1) An online survey was conducted amongst participants from 19 projects pre-identified by the European Commission (Table a). The survey aimed to map the main technological solutions adopted in the field of reuse of health data and capture other Consortiums' and experts' opinions regarding opportunities and barriers. The survey was made of 16 questions with a mix of multiple choice and Likert scale questions.
2) Thirteen presentations (as shown in Table b) were assessed from the EU Projects' Roundtable already described above". The purpose of assessing these presentations was

[1] Beatrice Bressan, Claudio Capelli, Edwin Morley-Fletcher (Lynkeus), Lorenzo Cristofaro (Panetta), Draft 1.0, 19th July 2023

to analyse and evaluate the content, insights, and perspectives shared by the presenters regarding the secondary processing of health data for research purposes. The presentations covered various topics related to privacy, technology, and the challenges and opportunities associated with health data reuse.

The assessment process aimed to gain a comprehensive understanding of the technological advancements, opinions, and potential barriers surrounding this matter, taking into considerations the steps that the EU is taking to make the re-processing of health data and records easier in all member States. The results of the survey and the insights gathered from the presentations provide valuable information for further analysis and decision-making within the context of the European Commission's initiatives in this field.

Table a: Projects participating to the online flash survey

PROJECT	PROJECT TITLE	COORDINATOR
AIDAVA	AI powered Data Curation & Publishing Virtual Assistant	Universiteit Maastricht
DataTools4Heart	A European Health Data Toolbox for Enhancing Cardiology Data Interoperability, Reusability and Privacy	University of Barcelona
eCREAM	enabling Clinical Research in Emergency and Acute care Medicine through automated data extraction	Mario Negri Institute for Pharmacological Research
IDEA4RC	Intelligent Ecosystem to improve the governance, the sharing, and the re-use of health Data for Rare Cancers	Fondazione IRCCS Istituto nazionale dei tumori
RES-Q PLUS	Comprehensive solutions of healthcare improvement based on the global Registry of Stroke Care Quality	Masaryk University (Brno)
More-EUROPA	More Effectively Using Registries to suppOrt PAtient-centered Regulatory and HTA decision-making	Academisch ziekenhuis groningen
ONCOVALUE	Implementing value-based oncology care at European cancer hospitals: An AI-based framework for assessing real-life effectiveness of novel cancer therapies in real-time	Hus-yhtyma (Finland)
Real4Reg	Development, optimisation, and implementation of artificial intelligence methods for real world data analyses in regulatory decision-making and health technology assessment along the product lifecycle	Bundesinstitut fur arzneimittel und medizinprodukte
REALM	Real-world-data Enabled Assessment for heaLth regulatory decision-Making	Maastricht University
REDDIE	Real-world evidence for decisions in diabetes	Medical University of Graz
RETENTION	heaRt failurE paTient managEment and iNTerventIOns usiNg continuous patient monitoring outside hospitals and real-world data	ICCS (Greece)
TIMESPAN	Management of chronic cardiometabolic disease and treatment discontinuity in adult ADHD patients	Orebro (Sweden)
HEAP	Human Exposome Assessment Platform	Karolinska Institutet (Sweden)
WELLBASED	Improving health, wellbeing, and equality by evidenced-based urban policies for tackling energy poverty	Fundacion de la comunitat valenciana
ReCoDiD	Reconciliation of Cohort data in Infectious Diseases	Heidelberg University Hospital
ORCHESTRA	Connecting European Cohorts to Increase Common and Effective Response to SARS-CoV-2 Pandemic.	UniVR
UnCoVer	Unravelling Data for Rapid Evidence-Based Response to COVID-19)	Prins Leopold Instituut voor Tropische Geneeskunde (Belgium)

SYNCHROS	SYNergies for Cohorts in Health: integrating the ROle of all Stakeholders	Parc Sanitari Sant Joan de Déu
SIMCor	In-Silico testing and validation of Cardiovascular IMplantable devices	Charite (Berlin)

Table b: Projects and speakers presenting at the webinar

ORGANISATION/PROJECT	SPEAKER
EU COMMISSION	Jana Makedonska
EU COMMISSION	Christina Kyriakopoulou
DataTools4Heart	Polyxeni Gkontra
BBMRI-ERIC	Ilaria Colussi
REALM	Gokhan Ertaylan
PANETTA LAW FIRM	Lorenzo Cristofaro
ONCOVALUE	Jochem de Boer
EUCAIM / CHAIMALEON	Ricard Martinez
UnCoVer	José L. Peñalvo
AIDAVA	Isabelle de Zegher
eCREAM	Giulia Ghilardi
Real4Reg	Christoph Röthlein
RES-Q PLUS	Catalina Martinez Costa
SIMCor	Jan Brüning
RETENTION	Christina Nanou
RE-SAMPLE	Christos Kalloniatis
eMOTIONAL Cities	Antonio Cerciello

Results from the Survey

Responses were collected from 21 participants from a total of 14 projects. The results presented below show the mapping of the technological solutions for the reuse of health data and the opportunities and barriers according to opinions collected through a series of statements.

Mapping

The primary technological solutions commonly adopted in the mapped projects for reusing health data, include *Data Ingestion and Harmonization, Data Synthesis and AI models* (responses to **Question 2**, Figure a).

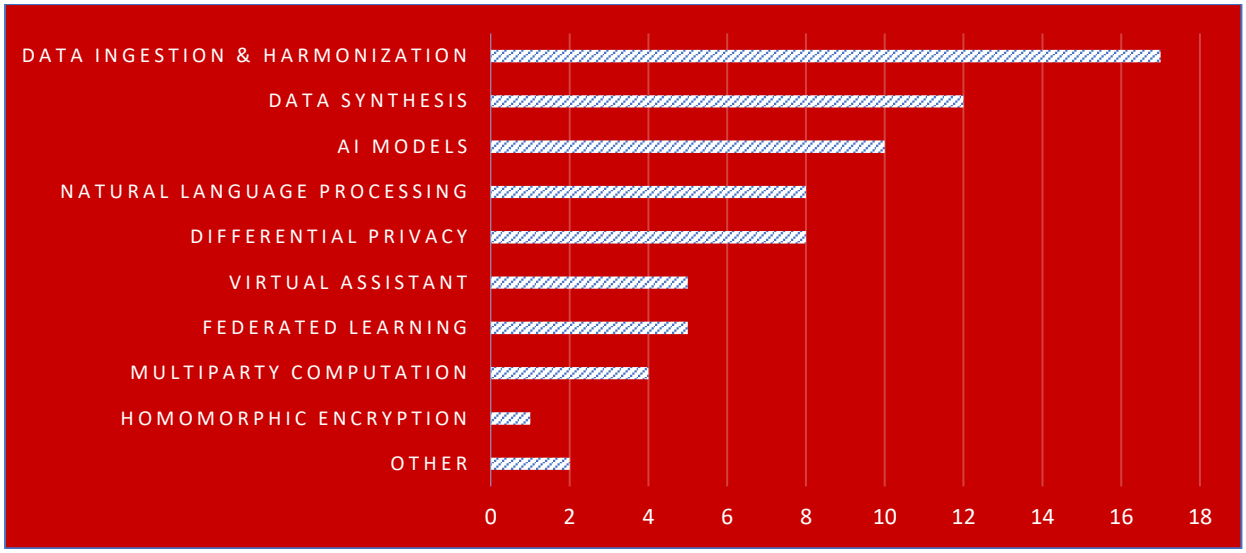


*Figure a: **Question 2**, Please indicate all the technological solutions adopted in your project on the re-use of health data (Please select all those apply)*

To ensure the security of the dataset and adhere to the principle of data minimisation, also by means of data anonymisation/pseudonymisation, data synthesis and differential privacy have been widely adopted to date (responses to **Question 9**, Figure b). These methods may play a crucial role in safeguarding data while minimising the amount of personally identifiable information present in the datasets.

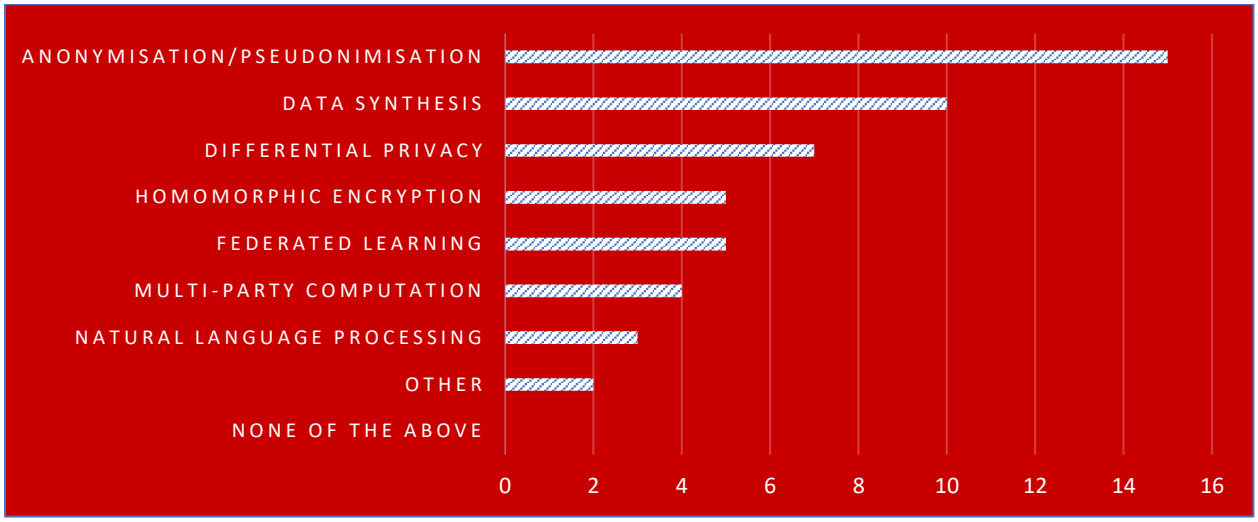


Figure b: **Question 9**, *Please select the technical solutions which have been adopted so far to ensure the security of the datasets and implement the principle of data minimisation (Please select all those apply)*

To date, the primary purposes for reusing health data have been focused on research endeavours supporting the development of predictive algorithms (responses to **Question 3**, Figure c). In the foreseeable future, applications are expected to expand to include decision support systems and delivery of personalised medicine. Consistently, researchers are considered the primary stakeholders to be affected by the emerging of such innovations according to the 55% of the respondents (responses to **Question 15**: *Which stakeholders are likely to be affected by the emerging of the technologies, such as Data Synthesis, Federated Learning, Natural Language Processing, Mult-Party Computation, Differential Privacy, Homomorphic Encryption, Anonymisation/Pseudonymisation*).

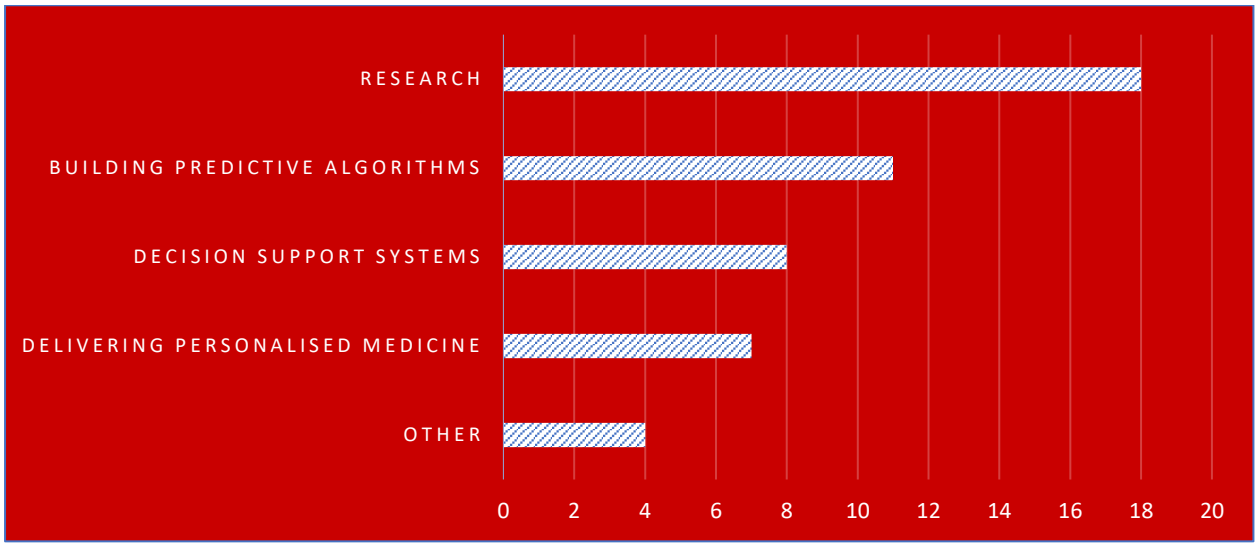


Figure c: **Question 3**, *Please specify the main intended use for the above selected technological solution(s) (Please select all those apply)*

Opportunities and Barriers

According to the opinions collected by the respondents (responses to **Question 9** and **Question 2**) the technologies regarded as most suitable for safeguarding the sensitive nature of health data are Differential Privacy (76.2% of participants agree/strongly agree on this statement), Federated Learning (66.7%) and Multi-Party Computation (66.6%) (responses to **Question 2** and to **Question 4**: *Please indicate your level of agreement with the following statements on the possibilities to protect the sensitive and confidential nature of health data*). These technologies offer robust methods to ensure the privacy and security of health data – and thus help achieving data minimisation – during analysis and collaborative learning processes.

Differential Privacy (76.2%) and Federated Learning (66.7%) have been also identified as the two technological solutions mostly suitable to facilitate data sharing (i.e., they allow a secure, privacy preserving methods for sharing health data; responses to **Question 5**: *Please indicate your level of agreement with the following statements on the use of novel technological solutions to facilitate the data sharing*). Interestingly, there is not such a consensus on the role that data synthesis and data harmonisation can play in helping to overcome the lack of common standards and protocols for exchanging and sharing health data.

According to the survey results, a significant majority of respondents (85.7%) agree/strongly agree that data harmonisation plays a crucial role in reducing errors and bias in research findings. Similarly, there is a consensus among respondents (52.4%) regarding the effectiveness of data ingestion tools in achieving this goal. However, opinions on virtual assistants are more mixed, with a relatively neutral stance expressed by 38.1% of the respondents (responses to **Question 6**: *Please indicate your level of agreement with the following statements about potential improvements of the quality of research findings*).

Data synthesis (61.9%) and homomorphic encryption (57.1%) have been identified as effective approaches to enhance patients' trust in the ethical and responsible use of their health data (responses to **Question 7**: *Please indicate your level of agreement with the following statements on the possibilities to increase patients trust*). By employing data synthesis techniques and implementing homomorphic encryption methods, healthcare providers and organisations may foster patients' confidence by ensuring privacy and maintaining the confidentiality of sensitive health information.

A federated approach to Machine Learning, which prevents data from being moved or extracted from local repositories and processing systems, has been identified as a means to comply with data minimisation and facilitate conditions for lawful reuse of data, as agreed upon by 71.4% of

the respondents (responses to **Question 8**: *Please indicate your level of agreement with the following statements on the local support to implement novel technological solutions;* and to **Question 10***: Please indicate your level of agreement with the following statement: a federated approach to Machine Learning, preventing data from being moved or pulled out from your local repositories and processing systems, helps complying with the conditions - including those applicable at national level in your country – for the lawful reuse of the data*).

Regarding the analysis conducted with privacy experts, such as Data Protection Officers, the majority of cases (14 out of 21 respondents) indicate that ongoing efforts have been made to assess the impacts of these technologies on personal data (responses to **Question 14**: *Have you analysed with privacy experts (e.g., Data Protection Officers) the impacts of the above technologies on personal data, to ensure that the rights vested to the individuals are sufficiently protected?*) This analysis aims to ensure that individuals' rights are adequately protected.

Furthermore, a significant majority (90% of the respondents) has the necessary expertise, resources, and means to apply anonymisation or pseudonymisation techniques to the datasets residing in their repositories (responses to **Question 12**: *Does your organisation/entity have, among others, the expertise, the resources, and the means to apply anonymisation, or pseudonymisation to the dataset residing in your repositories?*) This indicates a quite widespread – or supposed – capability to protect privacy while working with the data.

In terms of local governance at clinical collaborators, there appears to be a lack of support for implementing novel technologies (responses to **Question 8**).

Scaling up the implementation of such technologies in research and clinical setting appear to be a challenge (responses to **Question 13**: *Please indicate which challenges you foresee to guarantee the scalability of your technological solution for further expansion*). Below, a series of comments collected, which identify the main challenges grouped in three main reported categories:

Legal:

- Lack of clear definitions about of deidentification grade for anonymisation (this issue is also reported in TEHDAS's findings)
- Legal uncertainty, especially due to national fragmented scenarios
- Obstacles in crossing EU borders
- Legal challenges, especially relating to data anonymisation

Implementation:

- Organisational challenges, standardised and harmonised data collection
- Budget
- Capacity building
- Difficulty in explaining approach to ERCs, DPOs
- Research participants
- Interoperability with other medical repositories, ISO and MDR compliance
- Lack of IT systems or measurement devices' track, updates, and changes locally
- Lack of IT tools knowledge within hospitals

Technical:

- Manual interaction requirement for data extraction (API to clinical systems not available)
- Harmonisation blockages
- Attachment of metadata describing the process
- Patients' access to their data
- Federated Learning for training of AI models
- NLP tools' performance to extract relevant information from narratives
- Development of a user-driven system to realise k-anonymisation of databases
- Users able to decide which variables to intervene on to achieve the desired k-value

Finally, there is not a widespread knowledge on national provision applicable to establish specific conditions, including the implementation of the novel technological measures, for reusing health data for scientific research purposes (responses to **Question 11**: *Is there a national or industry*

code of conduct your organisation/entity already adheres to and applies in connection with the processing of health data and/or the reuse of data in the scientific and medical research area? And to **Question 16**: *To the best of your knowledge, is there any national provision applicable in your country which establishes specific conditions, including the implementation of the mentioned technological measures, for reusing health data for scientific research purposes?*)

Results from the Roundtable

The findings of the presentations and the collective discussions are here summarised according to the four categories of the SWOT analysis, which highlight strengths, weaknesses, opportunities, and threats of developing technological solutions for re-use of health data.

Strengths

The EU context: The European Commission has set up a competitive landscape which fosters ambitious projects to tackle the challenges of the reuse of health data. The upcoming European Health Data Space (EHDS) and the Artificial Intelligence Act will help the medical and research sectors progress faster and more efficiently. The timing is right for validating and standardising new technological pathways to foster the secondary processing of health data.

The legislation: The GDPR goal is to protect the individuals' rights and personal data, while still fostering economic growth of all industries by laying down conditions for lawful data processing and valorisation. This Regulation already includes ways to create an environment to facilitate research (e.g., derogation from the principle of limitation of the processing purposes with the presumption of compatibility; derogation from the storage limitation principle; exceptions to rights of the data subjects). A general principle is that transparency towards patients is key and a pre-condition for the lawfulness of any type of data processing. Although the interpretation of the GDPR might vary according to the context of use, a Code of Conduct (CoC) will be developed in DataTools4Heart to lay down elective rules which help all stakeholders of the research field achieve accountability and privacy-by-design. Other complimentary CoCs are going to help with supporting a fair and efficient access to health data.

Data availability: clinical and real-world data (RWD) are continuously increasing in quality and quantity. These can include different level of sensitive data (from less to highly sensitive data). Such availability fosters their use for decision-making process, regulatory and development of AI tools.

Technological solutions: the development of new tools and Privacy Enhancing Technologies is extremely advanced. The main strengths of the most proposed technologies are:

- Federated Learning: it avoids data sharing as the training of models occurs locally in compliance with hospitals' security perimeters (i.e., no real data travel outside the hospitals).
- Secure Multi-party Computation: it orchestrates privacy-preserving data.
- Semantic Data Harmonisation: it can create data resources catalogues of different formats.
- Natural Language Processing (NLP): it can be employed on multiple languages for Electronic Health Records (EHR) content interpretation and crucial information extraction; it is scalable with multi-lingual tools.
- Data Synthesis: it is used to generate virtual cohorts, which mimics real patients' characteristics with de-identified representation; useful to train models and evaluate software and device and share repositories with the community.
- Virtual Assistant: it facilitates the user to interrogate the database and support the decision-making process.
- Data Ingestion: it enables automatic extraction of structured outcomes from unstructured data (e.g., images, free text from EHR with NLP).

The implementation: all these projects result from strong consortia, which cover a large part of Europe, including UK and CH. This ensures access to large data sets. The average duration of such projects is 4 to 5 years and almost all the projects have just started. All the stakeholders are represented including technical partners, hospitals, clinical and academic institutes, and SMEs. Regulatory and legal partners are also sometimes included together with patients' groups. This ensures that high-level standard of GDPR compliance and AI ethics risk-based approach and

alignment with future regulations such EHDS and AI act are achieved. Some of the active projects build on previously developed platforms and technologies and include sustainability among their main goals.

Weaknesses

The EU context: The proposed projects cannot be considered only within a research context, but they cover a variety of domains. Thus, they often imply a striking difference among the needs of the different stakeholders, such as clinicians, researchers, and patients. The current health ecosystem is way more complex, highlighting widely unresolved clinical and organisational challenges. Sustainability of this type of research remains a major issue.

The legislation: a broad margin of intervention of member States remains, when it comes to the processing and the reuse of health data, carrying together the issues of a wide and deep national fragmentation. Dealing with local laws and soft laws is challenging. The requirement of patient consent for the reuse of personal data is challenging. Consent is considered a legacy for the research area, but it is not always the adequate legal basis for both primary and secondary data processing. There is a well-recognised unbalance between patients and healthcare providers, which makes the consent process inappropriate to the day-by-day reality of the research sector, especially in the healthcare area. Anonymisation is quite clear for the GDPR perspective but not from a practical perspective. For example, there is still no common understanding of anonymisation and pseudonymisation processes. A recent ruling by the EU Court of Justice fosters a subjective (more precisely, a "data recipient's") perspective, instead of an objective perspective (as applied by many Data Protection Authorities so far), to evaluate whether a specific set of data should be considered anonymous rather than pseudonymous.

Data availability: most of the data lay unused and sparse at different centres and there is no easy solution to access them. Information is often highly unstructured. The data quality extraction is still challenging and there are difficulties in data privacy guarantees. There are many different data collections and representation ways.

Technological solutions: most of the technological solutions are at early stages of implementation or evolving fast, which make their implementation more complicated. The amount of training datasets required for developing some of the proposed tools is huge. In addition, these often require human supervision. Challenges remain in harmonising all different partners, data types, sources, privacy perspectives, work modalities, research, proposal scopes. In this sense, an adequate matrix to evaluate the privacy risks needs to be further consolidated, as well as the validation of synthetic data, with reference to Real World Data.

The implementation: overall, there is a tendency to focus on technical/practical points but not on ethical and legal aspects. Consent is considered a legacy for the research area, but it is not always the adequate solution to ensure the lawfulness of the processing of personal health data, including their reuse for research purposes. Legal, organisational, technical security and privacy requirements definition for patient consent, or other applicable legal grounds, may vary at member States' level and affect design and utility of the research. The involvement of clinical partners is obviously crucial, but this also poses challenges such as the interaction with complex clinical and organisational issues. Contradictions may emerge between the ability of doing research in emergency medicine and sustainability issues (e.g., lack of dedicated medical staff; lack of understanding the processes to access and handle data). There is a wide range of variability in guidelines, methods, knowledge not only across projects, but also across partners of the same project.

Opportunities

The EU context: the H2020 legacy aims to leave structural changes for data research. The implementation of the EHDS will hopefully bring novel and concrete opportunities for data reuse, which will ultimately bring benefits in the entire healthcare sector (i.e., better outcomes, reduced costs, better management, and improved quality of research). In the context of the new framework, innovative tools will be put forward such as *MyHealth@EU* and a platform to keep the dialogue open with researchers. The context is prone to generate novel creative ways to contribute to digital health. Cluster of EU funded projects facilitate the interactions of stakeholders involved in different ways and facing similar problems.

The legislation: harmonising the rules for all the member States is a priority. The GDPR and the recent proposals of AI Act and European Health Data Space go in that direction. The GDPR strikes a good balance between both the protection of personal data and the free movement of data, in view of enhancing valorisation of data and individual privacy. The AI Act lays down a ground-breaking set of rules to regulate and harmonise the design, go-to-market, and use stages for all AI systems, providing a classification leading to the application of layered obligations aimed at minimising the impact of high-risk systems and preventing the launch of prohibited applications. The EHDS pursues the complex purpose of ensuring the harmonisation of the regulatory framework applicable to the primary and secondary use of electronic health data, to foster interoperability across the EU, on one hand, and data reuse for research purposes, on the other hand.

Data availability: The culture of taking information to make data available to researchers is improving. A concept of data altruism is rising, considering user consent and a sort of personal data wallet. Communities can be created to identify data gaps and seek synergies. FAIR principles are established for databases for clinicians, researchers, health policy makers in the European and national legislations respect. Platforms for dissimilar data sources capable of streamlining ethical and legal aspects are needed. Data anonymisation is highly context dependent and far from being standardised, from a legal standpoint. Public biobanking are increasingly common. The data acquisition from different countries for common analysis ensure diversity of datasets.

Technological solutions: Novel technological solutions are being developed for guaranteeing data protection. Solutions such as federated learning and synthetic data can tackle challenges of data management and processing, ensuring data minimisation. Training costs can be reduced by the generation of synthetic data and by assessing their impact. This will reduce spending on drugs that yield little benefit and enable oncologists to promote value-based cancer care. Flexible data model for data integration will facilitate their conversion into multiple representations. In-silico modelling is very promising for verification and validation thanks to the possibility of creating larger cohorts to acquire using clinical trials. Unique insights can be attained combining (bio) markers, lifestyle characteristics, well-being parameters for complex chronic conditions exacerbations outset and the progression of holistic models' creation. Knowledge of exacerbations can improve the timely detection of preventive treatments. The same can apply to clinical endpoints definition for complex chronic conditions patients' monitoring and identification, allowing highly innovative secure handling of data and machine learning methodology by multi-party computation.

The implementation: There is a strong potential for cooperating to improve data health environment with a series of actions which facilitate the implementations of such projects, including: common data understanding insurance through terminology and ontology using an open platform for interoperability across EU countries; promotion of data use and reuse; design of innovative platform delivery for interventions based on patient monitoring outside hospitals, RWD analysis; a shared agenda to help public health decision making, once models are learned from data validation. In this context, the requests to implement such processes even further have been recorded together with the stimulus to go beyond a specific disease approach towards a more personalised approach. There is the opportunity to create usable standards for data quality, common data model, and analytical workflow.

Threats

The EU context: landscape is very complex also depending on the specific type of data (i.e., the challenge of being specific and general at the same time). This loss of connection and contextualisation is a problem for developing applications and for testing different applications for the same population. In addition, some member States still rely on individual consent to enable data reuse, while other allow the application of different legal grounds for undertaking the secondary processing of health data for research purposes. Competent Data Protection Authority are not aligned when it comes to the definition and standardisation of the measures and legal bases permitting data reuse. These divergent interpretations also give rise to conflicts between Data Protection Authorities and research ethics committees. It is important to be transparent about what is done with data and for what purpose the data are processed and reused, but to make this happen, consent is not the only and always best way.

The legislation: there is a lack of alignment of definitions and interpretation of some crucial concepts between different regulators. By going to anonymisation and/or pseudonymisation, there is a loss of context and connection with the subject, creating data applicable just for a specific setting. There is still fragmentation at national levels, which the EU legislator is attempting to address by means of the EHDS, challenges to be overcome in the GDPR in relation to the processing of health data and a long road to run across towards a CoC for health research. Finally, there is the issue of terminology and confusion of definitions between communities (i.e., for synthetic data) and regulatory (i.e., data holders in the Data Governance Act and the proposed Data Act). The GDPR is not aligned with some member States' laws, where national legislators exerted the delegation to establish further conditions and limitations for the processing of health data (Art. 9.4).

Data availability: the data are rarely shared and made available for the benefit of the research, due both to the intricate regulatory framework applicable to their sharing and reuse, which is further complicated by the differences existing at national level and which stokes the data controllers' fear of being sanctioned, and the lack of technical interoperability standards.

Technological solutions: safeguarding the privacy of data utilised in large language models, such as GPT-4 and ensuring respect for patients' privacy through effective anonymisation techniques or PETs is essential. It is necessary to handle the complexity associated with data that cannot be easily transferred between clinical sites and to address the intricacies of making datasets compatible and interoperable. Dealing with the issue of catastrophic forgetting, where machine learning models may forget previously acquired knowledge when trained on new data. It is necessary to navigate the obstacles presented by combining technologies like multi-party computation, federated learning and synthetic data that may support data privacy, and overcome the challenges of harmonising and integrating datasets with varying syntactic and semantic structures. Transforming raw data into curated datasets for analysis and modelling purposes and converting curated datasets from a population-level representation to an individual-level representation remains an issue.

The implementation: there are several challenges related to data management in healthcare settings. One issue is the lack of time in emergency departments to obtain consent from patients for the use of their data (when this is the applicable legal basis). Additionally, organising and maintaining data collection as part of standard clinical routines poses difficulties. Not all hospitals' databases are pseudonymised, and this raises concerns about patient privacy. Fulfilling the necessary requirements to share pseudonymised data for routine clinical use can be a lengthy process. Furthermore, there is often a lack of appropriate guarantees regarding access control, data protection and privacy, when sharing and reusing sensitive information, such as the risk of data breach. Managing the storage of sensitive health data prior to pseudonymisation, anonymisation, or deletion is another challenge. Some hospitals may face memory constraints in their databases for storing patients' data, and this can hinder data availability. The high operating costs associated with data management also pose a significant hurdle. Delays in database development for data sharing purposes further complicate the situation. In many cases, the failure by data controllers to comply with some specific requirement established by the legislation in force may prevent, or highly complicate, the sharing of personal and health data with the research community. The absence of contextual information, including the "5 W (Who, What, When, Where, and Why) and How" details, can jeopardise the interpretation of data. Synchronising and harmonising data for a comprehensive clinical data model may prove more than challenging. Companies also face reputational risks during the verification and validation of medical devices.

Conclusions

The analysis detailed in this document provides valuable insights into the landscape in which DataTools4Heart aims to develop and deploy its solutions for the effective reuse of health data in cardiac research. The availability of clinical and real-world data, ranging from less sensitive to highly sensitive information, offers significant opportunities for research purposes, including the development of diagnostic tools, support for decision-making, regulatory compliance, and AI tool development and training, among others. It is important to note that research efforts must not only harness the potential of data, but also ensure the protection of data subjects' rights.

The findings demonstrate that EU research is at the forefront of developing technological solutions for data processing and data protection within the existing legislative framework that encourages research. The innovative and promising technologies proposed by DataTools4Heart align well with this research landscape. However, their implementation still faces challenges, including issues related to data access, fragmented legislation across different EU countries, and the translation of these technologies into clinical centres operational practice.

Despite these challenges, the current landscape also presents novel opportunities. These opportunities can be realised through the development of advanced technologies such as large language models and by fostering collaboration with legal experts to establish codes of conduct that benefit the entire community. In this regard, the objectives of DataTools4Heart are aligned with the identified opportunities in this assessment.

It is important to recognise that the success of DataTools4Heart and other similar projects is threatened by factors such as the uncertainty of the legal framework and of organisational challenges. Therefore, a holistic approach involving all stakeholders is strongly recommended. This approach ensures that challenges are properly identified and addressed, leading to effective solutions that can drive progress in the field, while upholding legal and ethical considerations.

Acronyms

AI	Artificial Intelligence
API	Application Programming Interface
CoC	Code of Conduct
CVD	Cardiovascular Diseases
DPO	Data protection officer
EC	European Commission
EHDS	European Health Data Space
EHR	Electronic Health Records
ERC	European Research Council
EU	European Union
FAIR	Findability, Accessibility, Interoperability, and Reusability
GDPR	General Data Protection Regulation
GPT 4	Generative Pre-trained Transformer 4
HTA	Health Technology Assessment
ISO	International Organization for Standardization
MDR	Medical Device Regulation
NLP	Natural Language Processing
RWD	Real World data
SME	Small and Medium Enterprises
SWOT	Strengths, Weaknesses, Opportunities, and Threats