

# 1 A data management system for precision medicine

2

3 John J. L. Jacobs,<sup>1,\*</sup> Inés Beekers,<sup>1</sup> Inge Verkouter,<sup>1</sup> Levi B. Richards,<sup>1</sup> Alexandra Vegelien<sup>1,2</sup>, Lizan D.  
4 Bloemsma,<sup>3</sup> Vera A. M. C. Bongaerts,<sup>4</sup> Jacqueline Cloos,<sup>5</sup> Frederik Erkens,<sup>6</sup> Patrycja Gradowska,<sup>7</sup>  
5 Simon Hort,<sup>6</sup> Michael Hudecek,<sup>8</sup> Manel Juan,<sup>9,10,11</sup> Anke H. Maitland-van der Zee,<sup>3</sup> Sergio Navarro  
6 Velázquez,<sup>9,10</sup> Lok Lam Ngai,<sup>5</sup> Qasim A Rafiq,<sup>12</sup> Carmen Sanges,<sup>8</sup> Jesse Tettero,<sup>5</sup> Hendrikus J. A. van  
7 Os,<sup>13</sup> Rimke C. Vos,<sup>14</sup> Yolanda de Wit,<sup>5</sup> Steven van Dijk.<sup>1</sup>

8

9 <sup>1</sup> Clinical Care & Research, ORTEC B.V., Zoetermeer, The Netherlands.

10 <sup>2</sup> Faculty of Mathematics, VU, Amsterdam, The Netherlands.

11 <sup>3</sup> Department of Pulmonary Medicine, Amsterdam UMC, The Netherlands.

12 <sup>4</sup> Department of Public Health and Primary Care, LUMC, Leiden, The Netherlands.

13 <sup>5</sup> Department of Haematology, Amsterdam UMC, The Netherlands.

14 <sup>6</sup> Fraunhofer IPT, Aachen, Germany.

15 <sup>7</sup> Dutch-Belgian Hemato-Oncology Cooperative Group Data Center–Erasmus MC Cancer Institute,  
16 Rotterdam, The Netherlands

17 <sup>8</sup> University Clinic Würzburg, Germany.

18 <sup>9</sup> Foundation for Biomedical Research, Barcelona, Spain.

19 <sup>10</sup> Immunology department, Hospital Clinic of Barcelona, Barcelona, Spain.

20 <sup>11</sup> HSJD-Clinic Immunotherapy platform, Barcelona, Spain

21 <sup>12</sup> Advanced Centre for Biochemical Engineering, University College London, UK.

22 <sup>13</sup> National eHealth Living Lab, LUMC, Leiden, The Netherlands.

23 <sup>14</sup> Public Health & Primary Care, Leiden University Medical Center

24 \* Corresponding author. [John.Jacobs@ortec.com](mailto:John.Jacobs@ortec.com)

25 **NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

26 **MESH:** precision medicine, personalized medicine, personalised medicine, P4 medicine, medical data  
27 management system, LogiqSuite.

28

## 29 Abstract

### 30 Introduction

31 Precision, or personalised medicine has advanced requirements for medical data management  
32 systems (MedDMSs). MedDMS for precision medicine should be able to process hundreds of  
33 parameters from multiple sites, be adaptable while remaining in sync at multiple locations, real-time  
34 syncing to analytics and be compliant with international privacy legislation. This paper describes the  
35 LogiqSuite software solution, aimed to support a precision medicine solution at the patient care  
36 (LogiqCare), research (LogiqScience) and data science (LogiqAnalytics) level. LogiqSuite is certified  
37 and compliant with international medical data and privacy legislations.

### 38 Method

39 This paper evaluates a MedDMS in five types of use cases for precision medicine, ranging from data  
40 collection to algorithm development and from implementation to integration with real-world data.  
41 The MedDMS is evaluated in seven precision medicine data science projects in prehospital triage,  
42 cardiovascular disease, pulmonology, and oncology.

43

### 44 Results

45 The P4O2 consortium uses the MedDMS as an electronic case report form (eCRF) that allows real-  
46 time data management and analytics in long covid and pulmonary diseases. In an acute myeloid  
47 leukaemia study data from different sources were integrated to facilitate easy descriptive analytics  
48 for various research questions. In the AIDPATH project, LogiqCare is used to process patient data,  
49 while LogiqScience is used for pseudonymous CAR-T cell production for cancer treatment. In both  
50 these oncological projects the data in LogiqAnalytics is also used to facilitate machine learning to  
51 develop new prediction models for clinical-decision support (CDS). The MedDMS is also evaluated

52 for real-time recording of CDS data from U-Prevent for cardiovascular risk management and from  
53 the Stroke Triage App for prehospital triage.

54

## 55 Discussion

56 The MedDMS is discussed in relation to other solutions for privacy-by-design, integrated data  
57 stewardship and real-time data analytics in precision medicine.

58

## 59 Conclusion

60 LogiqSuite is used for multi-centre research study data registrations and monitoring, data analytics  
61 in interdisciplinary consortia, design of new machine learning / artificial intelligence (AI) algorithms,  
62 development of new or updated prediction models, integration of care with advanced therapy  
63 production, and real-world data monitoring in using CDS tools. The integrated MedDMS application  
64 supports data management for care and research in precision medicine.

65

## 66 Introduction

67 Standardisation of medical practice has yielded great progress for modern medicine, but the number  
68 of new drugs approved per billion US dollars spent on research and development has halved roughly  
69 every nine years since 1950 in inflation-adjusted terms, implying that the efficacy of medical  
70 progress has gone down by a factor 80 (1-4). Animal models of disease lack human disease variations  
71 (5,6). Precision, or personalised medicine (7) promises significant therapeutic improvements by  
72 prediction, prevention, personalisation or stratification, and participation of patients (8,9). It  
73 depends on distinguishing different disease mechanisms at a detail level where animal models have  
74 limited predictive value.

75

## 76 Medical data management for precision medicine

77 Precision medicine studies clinically relevant i.e., large effects, allowing limited clinical studies (2,10)  
78 and patients stratification in small groups (11,12). Medical data science projects increase in  
79 complexity due to the evaluation of different biomarkers, either separately or in numerous  
80 combinations (13,14). While clinical trials in conventional medicine often limit their criteria by  
81 exclusion of potential comorbidities, precision medicine aims to include real-world data as relevant  
82 cases of human disease (15-18). Statistical analyses combining collections of real-world data (19,20)  
83 and clinical trial recordings (21) challenges the prerequisites for medical data management systems  
84 (medical DMSs; MedDMSs). Appropriate tools, like MedDMSs for medical data science on stratified  
85 interventions are needed for precision medicine (22,23) and lack of these lack slows down the  
86 transition to precision medicine (24). The MedDMSs are also crucial to in data collection for the  
87 mathematical development of artificial intelligence (AI) algorithms (25).

88 Precision medicine defines new requirements for MedDMSs. In traditional medicine, a MedDMS has  
89 dozens of parameters from a single site. MedDMSs for precision medicine should be able to process

90 hundreds or thousands of parameters to allow distinguishing and discrimination between various  
91 forms of disease with different mechanisms and/or requiring different interventions. It is unlikely  
92 that a single site will have relevant numbers for all stratified subgroups, thus the MedDMS should be  
93 able to integrate data from multiple sites, while maintaining compliant with privacy legislation, e.g.  
94 the EU General Data Protection Regulation (GDPR). The integration should include data validation  
95 and coordination for progressive insights.

96 These requirements together point towards a centrally governed cloud solution that integrates data  
97 from multiple sites, while maintaining the appropriate privacy level and access on a need-to-know  
98 basis for each subset of data.

99

## 100 [Requirements for a MedDMS for precision medicine](#)

101 Privacy protection legislations, like the GDPR, call for privacy-by-design and fine-tuned data access  
102 (26). Distinct medical conditions and research areas need different data models. Data models should  
103 be plastic for progressive insights and robust for long term data storage and management. Medical  
104 data are inherently complex with static data (e.g. sexes), data with a defined start and/or end date  
105 (e.g. diagnosis), repeated lab test results (e.g. haemoglobin), different kinds of sequence variations  
106 (e.g. mutations, indels, translocations), immunological cytokine profiles (27), and systems biology  
107 (28,29).

108 We are at the eve of implementation of precision medicine (30), but enhanced data management  
109 systems (DMSs) are needed for precision medicine (31,32). Real-world data are important both for  
110 initial model development, but also for real-world feedback in the process of continuous learning for  
111 continual improvement of the models. Most drugs and other interventions lack data for real-world  
112 evidence, beyond the controlled clinical trials, as they are hard to capture in current medical practice  
113 (19). Clinical studies are mostly performed under ideal circumstances with patients that fit the

114 inclusion and exclusion criteria, which often implying a single disease without comorbidities, while  
115 real patients mostly have multiple comorbidities. Real-world data without tight inclusion and  
116 exclusion criteria increase the complexity of data.

117 When data become more complex, statistical rules request more data, making data sharing crucial  
118 for precision medicine. The GDPR allows the use of anonymous data in research, and pseudonymous  
119 data when patients give their consent. Medical data should be sharable between studies, implying  
120 that data should be findable, accessible, interoperable, and reusable (FAIR) (33,34). Data dictionaries  
121 allow FAIR data dictionaries to added to the data in LogiqSuite. In some European countries, patients  
122 should give informed consent to every scientific study that is performed with their data. Based on  
123 either opt-in or opt-out, it should be possible to include or exclude data from analytics.

124 Data quality is the heart of data science (35). A developed mathematical prediction model is unlikely  
125 to be of better quality than its data input. Handling big and complex medical data are challenges for  
126 MedDMSs and thus for implementation of precision medicine (36-39). Data quality starts with input  
127 validation (40). The MedDMS should also have dynamic forms to minimise redundant question (41).  
128 It should have reports for user feedback, be integrated in care settings for real-world data, as well as  
129 research settings for trial data. Moreover continuous improvement requires data integration with  
130 clinical-decision support (CDS) tools for precision medicine.

131 In medical practice and research settings, responsible physicians, and investigators, respectively are  
132 assigned to cases, but their departments are also involved. Medical specialists often have authorised  
133 crosstalk between different cases of their patients, with the regulated rights for data viewing  
134 authorisation at the case level to avoid unnecessary complexity.

135

## 136 Validation of the MedDMS for precision medicine

137 From these requirements, we set up LogiqSuite, a MedDMS to facilitate precision medicine in multi-  
138 centre research studies, data analytics, AI development of prediction models, integration of care and  
139 research, and real-world data monitoring in using CDS tools. We describe five different use cases of  
140 LogiqSuite in seven topics of oncology, cardiovascular medicine, pulmonology, and pre-hospital  
141 triage. The use cases included fusing datasets, study monitoring, integrating care and research,  
142 development, and the implementation of a MedDMS. In different use cases, the usability of  
143 LogiqSuite as a MedDMS for precision medicine is evaluated.

144



## 145 Materials & methods

### 146 Concept

147 Our MedDMS is an integrated solution dubbed LogiqSuite, consisting of LOGIQ applications built on  
148 logic to LOG data Intelligently and Quantitatively. The concept is to record medical case data at the  
149 source to enable medical data science for precision medicine at the levels of patients, subjects, and  
150 data analytics. LogiqCare uses directly identifiable patient data to avoid patients being mix-up in  
151 clinical care. LogiqScience uses pseudonymised subject data for scientific research, laboratory, and  
152 other activities, where sample identity is crucial but direct identification of patients is undesirable.  
153 LogiqSuite allows seamless integration of LogiqCare and LogiqScience to aid collaboration between  
154 physicians and researchers. The data of LogiqCare and LogiqScience are synced to LogiqAnalytics for  
155 data analytics, in real-time, while the data are deidentified and depersonalised by controlling and  
156 minimising identifiable information. Separating LogiqAnalytics from the LogiqCare and LogiqScience  
157 database allows to filter out personal identifiers as well as to filter data not that should not be used  
158 for analytics (e.g., when records with unreliable or unconsented data) or data that do not conform to  
159 the need-to-know basis for analytics.

160

### 161 Generic Design

162 All data handling and storage is encrypted in LogiqSuite according to the GDPR, ISO27001, and Dutch  
163 healthcare regulatory guidelines i.e., NEN7510 (42). LogiqSuite is a robust cloud-native solution in  
164 Microsoft Azure cloud. Data are restricted to be handled and stored in the European Economic Area.  
165 The solutions have a web interface accessible with current versions of Google Chrome, Mozilla  
166 Firefox, Microsoft Edge, and (limited) on Safari for iOS and macOS. LogiqSuite is built using modern  
167 software development best practices (continuous integration and delivery, using a test, pilot, and  
168 production deployment, automated tests, various checks, and balances on quality). New

169 developments are delivered under a feature toggle which allows for quickly providing new  
170 functionality or security fixes.

171 Data isolation can be achieved by running in a multitenant environment (logical separation) or by  
172 having a dedicated environment per project (physical isolation). Backups are done automatically for  
173 both point-in-time restore (seven days into past) and long-term full backups (weekly are retained for  
174 30 days, monthly for 365 days). Scalability is possible in the compute, data, and web layers to cope  
175 with any workload. Specific details can be found in the technical documentation (43).

176

## 177 [Technical details](#)

178 LogiqSuite uses standard protocols: REST API for querying data and data ingress (migration),  
179 CSV/JSON serialisation for import/export of data, TLS 1.2 for encryption of communication channels,  
180 and OIDC or SAML for integrations with Identity Providers (e.g. Active Directory) (Figure 1). Full  
181 details can be found at the technical paper (44).

182 Data are stored in LogiqSuite as a stream of immutable events, which allows historical  
183 reconstruction. These events are communicated over a shared message bus that any internal service  
184 can subscribe to. For example, the service that is responsible for the web site maps the events to a  
185 database that is optimised for querying data. A service from LogiqAnalytics can take the same events  
186 and map them to a database specially tailored for a specific use case. Notably, sensitive personal  
187 identifiable information (PII) can be removed during this mapping, ensuring that research data are  
188 always anonymous. All data streams in LogiqSuite are evaluated in real-time and maintain  
189 continually in sync.

190 LogiqSuite is developed according to privacy-by-design principles and data sharing on a need-to-  
191 know basis, using the cloud-native capabilities of Azure (App Services, Azure SQL, Cosmos DB, Vault,  
192 Service Bus, etc). LogiqSuite relies on Auth0 for authentication-as-a-service (45) i.e., to establish

193 connections with customer identity providers in a reliable and secure manner. Customers can also  
194 connect using just their own account, without the need for an enterprise connection. Multifactor  
195 authentication (MFA) is enforced for individual users.

196 LogiqAnalytics leverages the Microsoft Power BI platform to allow users to collaborate on datasets  
197 and create insightful dashboards. Users of the same project can share dashboards using the  
198 Microsoft Teams environment.

199 Projects that use the Sciencrew platform for publishing content enjoy seamless integration in  
200 LogiqSuite. LogiqSuite is continuously improved (46). LogiqCare & LogiqScience provide a set of REST  
201 APIs for integration with third party systems for data accessibility (47).

202

203 **Figure 1. Data flows in LogiqSuite.** In brief, data are stored in read-, and write-model data bases, where the  
204 write-model is the single source of true, and the read a fast cache of aggregated data.

205

## 206 Case-centred data in LogiqCare and LogiqScience

207 LogiqCare and LogiqScience provide different access to case data of patients and subjects,  
208 respectively. Depending on the users' roles, directly identifiable or pseudonymous data are hidden,  
209 viewable, or editable. Depending on the implementation use case, patients, and subjects can be  
210 coupled, e.g. for data science collaboration between care and research projects. The central concept  
211 is the *case*, which can be fully customised for a certain diagnosis. Users will have worklist for the  
212 tasks to be done at their group or department.

213 Generally, writing and viewing access of data is mostly authorised at the *case* level, analogous to  
214 medical practice. The *patient* or the *subject* is the identifier for the person to which the *case* belongs.  
215 Organising data in *cases* allows collaborations between medical departments, like pulmonology and  
216 oncology for cases with comorbidity of asthma and oncology, respectively. *Cases* may contain

217 *consultations* to structure data in anamnesis, direct measurements, and questionnaires send  
218 (in)directly to the patient. *Cases* may also contain *test cascades* to collaborate between  
219 departments, like a pulmonology department requesting blood tests from haematology. These test  
220 cascades will orchestrate the appropriate rights to view and edit, as well as the workload list of  
221 various departments. *Test requests* are generated by the appealing departments and the *test results*  
222 and/or *test conclusions* are entered by the performing departments. Data management can be  
223 organised over the boundaries of any single organisation, as LogiqSuite is fully cloud based and can  
224 allow multiple organisations to collaborate under precise restrictions. This collaboration could be  
225 sharing data at the *case* level or within *test cascades*.

226 Descriptions in the user interface and data fields can be translated to the user's preferred language,  
227 e.g. to avoid translation of physician-patient interactions, and patient's language, e.g. in the case of  
228 questionnaires sent directly to patients. When entering choices of options, the user interface is  
229 switch to the user language (e.g. Dutch), while syncing the appropriate English translation to  
230 LogiqAnalytics for data science purposes.

231

## 232 [Solution design](#)

233 The LogiqSuite application is open to external interactions to other applications, like clinical decision  
234 support tools (e.g., U-Prevent, which is also built by Ortec), and has open API for third party apps  
235 (Figure 2). Data communication can be done using *patient*, *subject*, *case*, or *test* identifiers.

236

237 **Figure 2. LogiqSuite landscape.** LogiqSuite (blue) with interactions to other solutions from ORTEC (orange) and  
238 external sources (green). In brief, LogiqCare allows recording of (a) Patient data in LogiqCare, and (b) Subject  
239 data in LogiqScience, with (c) Cases as a central database. Case data are (d) synced to LogiqAnalytics for real-  
240 time descriptive statistics, and (e) ready-to-use for medical math, using AI to develop new prediction models.

241 Prediction models can be (f) productised for risk predictions for clinical-decision support, that are (g) coupled  
242 with the central cases. Central cases also allow (h) open API communication with third party apps.

243

#### 244 [Data transfer access](#)

245 Data from other databases can be mapped and imported into LogiqSuite by extract – transform –  
246 load (ETL) procedures. The ETL process allows verifiable, reliable, and repeatable data import from  
247 different sources into LogiqSuite. During the Transform step of ETL, data curation was also  
248 performed on the data from the various sources.

249

#### 250 [Data science in LogiqAnalytics](#)

251 LogiqAnalytics consists of a SQL database with numerical data and predefined choice options. All  
252 relevant data are real-time synced to LogiqAnalytics. Some data might enhance personal  
253 identifiability, like dates of birth and visits, and free text. In analytics the dates are converted to  
254 numerical data like age at diagnosis or duration of disease. Free texts are not suitable for analytics,  
255 so data intended to be analysed should be grouped into option prior data entry. The database  
256 structure in LogiqAnalytics can be configured to the needs of the project using SQL view tables, e.g.  
257 mapped into a relevant data model structure. This table structure also allows different parties to see  
258 different domains of data.

259 Additionally, it supports proper metadata documentation by an integrated data dictionary. The  
260 deidentified data are available in real-time for descriptive analytics in generated reports with  
261 Microsoft Power BI, Excel, SPSS, and other applications. The data can also be used for advanced  
262 analytics and AI-powered model development.

## 263 Results

264 Although biomedical research has some generic principles, it also has a sheer infinite number of  
265 putative solutions. Implementation in different projects is preceded by customization and regulation  
266 of user access.

267

### 268 Customisation process

269 DMSs for Personalized medicine should be flexible to allow the rapid creation of complex databases  
270 for collaboration of many scientific groups. Database design should be plastic and robust for  
271 progressive insights and historical consistency. Upon saving data, templates are used to build  
272 entities, that warrant data integrity independent of later database changes. The databases are  
273 configured using data templates for *cases, consultations, and test requests, test results, and test*  
274 *conclusions*. These are fully designable using standard building blocks in LogiqCare and LogiqScience.

275 Users lead the customisation process by defining their templates in an Excel file, supported by a  
276 medical data scientist of the LogiqSuite team (Figure 3). The medical data scientists provide crucial  
277 input for data stewardship before data collection to facilitate real-time data analytics. After  
278 verification by a medical data scientist, the Excel template is converted to a Logiq template on the  
279 pilot environment. User interface of templates could be translated to any language e.g., for  
280 international studies. The templates function as data-entry screens in the pilot environment. Next,  
281 template validation is done by an appointed user. After customer approval, the template is made  
282 available on the production environment. Changes could be implemented rapidly, e.g. if needed  
283 within a day, while maintaining the careful multistep process.

284 In the LogiqAnalytics database data could be regrouped if desired for certain analytics, but in general  
285 the structure of the LogiqCare and LogiqScience template is used. Data can be enriched with a user-  
286 defined data dictionary, to give meaning to values for analysis, and to facilitate FAIR data exchange.

287

288 **Figure 3. Template customisation process.** In brief, (a) medical data scientist (MDS) provides a template to  
289 designated (b) medical researcher (MRes) who will design the desired data flows with MDS advice. When  
290 finalised, (c) the MDS will run the python script to upload the template in the Pilot environment, where (d) the  
291 medical researcher will validate the template, for (e) improvements or (f) to instruct the MDS to (g) transfer  
292 the template to the production environment.

293

### 294 Regulation of user access

295 Access control is supervised using a documented four-eyes principle i.e., the person granting access  
296 is not the same person as executing access. Project leaders assign the persons to decide on role-  
297 (RBAC) and attribute-based access control (ABAC), reading and writing rights for a maximum of one  
298 year per request. The requests are logged automatically for traceability and executed manually by  
299 ORTEC to avoid abuse.

300 LogiqSuite provides its own Azure AD as Identity Provider (IdP) to which members of another  
301 organisation can be invited using federation identity. This implies that company policies for data  
302 access apply. Authenticating to the LogiqSuite Azure AD requires an additional MFA, that might be  
303 merged into the external Azure AD's authentication policy. Single sign-on is supported. For individual  
304 users, LogiqSuite can provide its own Azure AD as IdP, equipped with MFA that can be enforced.  
305 User access is granted by the project lead and executed by ORTEC to obey the four-eyes principle.  
306 Only specific ORTEC users have roles with exclusive rights, like Admin, to manage accounts,  
307 departments, and groups, and Configurator, to edit studies and templates for data, e-mails, and  
308 reports. Roles for customer users are constrained to User and Viewer, limited by attributes for  
309 accounts, departments, and Care and/or Science. To avoid unintended interactions between  
310 functions, distinct RBAC and ABAC are separated for each task. After login, LogiqSuite users with  
311 multiple functions can switch between their active functions with distinct RBAC and ABAC, to avoid

312 unintended control due to combining functions, e.g. an editing right in function A should not change  
313 a viewing right in function B.

314

315 [Evaluation of precision medicine use cases.](#)

316 LogiqSuite has three different privacy levels i.e., directly identifiable, pseudonymous, and (shear)  
317 anonymous for care, research (science), and analytics, respectively. This allows users to select their  
318 desired balance between protection of patient and data safety.

319 In seven different biomedical projects we could identify five different use-cases for precision  
320 medicine: (I) reorganisation and integration of data analytics for descriptive statistics in oncology, (II)  
321 real-time monitoring of a multicentre clinical study for pulmonary disease, (III) integration of clinical  
322 patient data with GMP data in oncology, (IV) AI development of for clinical decision support (CDS)  
323 models in oncology, and (V) integration of real-world data with CDS in cardiovascular risk  
324 management and prehospital triage.

325

326 [I. Integration of databases for analyses](#)

327 Different databases contained parts of the clinical follow-up data on minimal residual disease (MRD)  
328 from Acute Myeloid Leukaemia (AML) (48). In this research project, data was available in different  
329 databases from haemato-oncology of Adults in the Netherlands (HOVON) (49) and local databases  
330 built by the Amsterdam UMC, which had data on patient-survival follow-up and follow-up  
331 leukaemia-aberrant immune phenotypes (LAIPs), respectively (Figure 4). A common data model was  
332 crafted combining the structure of these databases in a subject-case structure with consults for  
333 direct analysis, and test cascades with test-requests for samples, test-results for direct analyses, and  
334 test-conclusions for LAIP-specific data.



335 A research MedDMS was built in LogiqScience, and the data were synced to LogiqAnalytics. The data  
336 consist of subject, the genetic characterisation of their AML, the characterisation, classification, and  
337 quantification of LAIPs, and therapy and survival data. Data was loaded using ETL from a csv file out  
338 of a query from the HOVON database and multiple local data management systems in Amsterdam  
339 UMC and GIMEMA in Rome.

340

341 **Figure 4. Extract-Transform-Load integration for AML.** In brief, (a) data was extracted from different  
342 databases, which were coupled with unique identifiers, (b) data transformation consisted of mapping to the  
343 LogiqSuite's data structure and values, that were translated and curated before loading into LogiqSuite, (c)  
344 which are written and stored as F-sharp code, to allow tracking and reusing of the procedure, and (d) data are  
345 loaded in LogiqSuite and organized in Patient or Subject – Case – Consults & Test cascade structures.

346

## 347 II. Real-time monitoring of a multilocation clinical study

348 Clinical studies in precision medicine are often performed at multiple locations in different  
349 languages, while they ideally should be monitored, orchestrated, and managed at a central location  
350 in real time. The results are real-time synced for data analytics. The LogiqScience orchestrates the  
351 clinical study as a central electronic case report form (eCRF) in LogiqScience. Different attributes  
352 were used for ABAC to limit data access on a need-to-know basis, e.g. by only providing access to  
353 subjects of a user's centre. Two additional institutes contributed data on lifestyle intervention and  
354 air pollution exposure, which had different roles limiting which subset of the Subject's data could be  
355 accessed. Additionally, Test cascades allowed cooperation between departments while properly  
356 controlling data access on a need-to-know basis.

357 The Long Covid use case (50) belongs to the precision medicine for lung diseases (P4O2) research  
358 project, a multi-site clinical study for lung diseases for precision medicine, including a Long Covid  
359 study. Data was entered by five different inclusion centres directly into LogiqScience as an eCRF. The

360 eCRF templates used by the researchers were designed in English, but the questionnaire templates  
361 for the patients were available in Dutch. The multilingual properties of LogiqCare and LogiqScience  
362 offer these possibilities, but LogiqAnalytics is synced in English only. Choices, dates, and numerical  
363 values were synced to LogiqAnalytics, allowing real-time monitoring of the progress in subject  
364 inclusion and data completeness by Power BI dashboards (Figure 5). Data on air pollution by  
365 ultrasonic personal aerosol sampler was transferred into the database by ETL. Researchers could  
366 monitor the data availability dashboard and provide real-time feedback to users, such that data  
367 could be completed in time. As a result of the real-time sync to LogiqAnalytics, as soon as the data of  
368 the last subject was entered into the eCRF, the analytics database was ready for analysis.

369

370 **Figure 5. Early study dashboard P4O2 Covid showing trial progress state in April 2022.** In brief, data are  
371 synced in analytics, showing real-time study overview. Arrows point to missing data, which are indicated in  
372 grey. This information was used for real-time identification of missing data, and early corrective actions.

373

### 374 [III. Integration of care and precision medicine production.](#)

375 The AIDPATH project aims to set up AI-powered decentralised production of advanced (CAR-T cell)  
376 therapies in hospitals all over Europe. Production of precision medicine or advanced therapeutic  
377 medicinal products, like the production of autologous CAR-T cells, requires integration of clinical  
378 care and good manufacturing practice (GMP) production data for advanced therapies. Using  
379 LogiqSuite, it is possible to integrate a patient's clinical care data in LogiqCare with pseudonymised  
380 CAR-T cell production data in LogiqScience (Figure 6).

381 The AIDPATH project consortium aims to have AI-guided production of autologous chimeric antigen  
382 receptor T-cells (CAR-T cells) distributed over various UMCs (51). In the AIDPATH project, two flows  
383 are combined, a clinical with patient data with a GMP production with subject data, allowing direct

384 patient identification where needed, and preserving patient privacy where possible. Specific  
385 pretreatment patient data are crucial for CAR-T cell production and thus integrated in shared cases  
386 between patients and subjects. Some details are only needed for care or manufacturing purposes, so  
387 case data might be shared, but underlying consultations and test cascades could be unshared. This  
388 allows seamless collaboration between care and manufacturing, ensures the pseudonymised  
389 production data are unambiguously coupled to the patient, without sacrificing patient privacy (52).

390 The decentralised AIDPATH production system uses laboratory devices to automatically process the  
391 CAR-T cells. The devices are digitally connected to the COPE (Control Optimise Plan Execute). COPE  
392 acts as a manufacturing execution system and connects to the devices using their respective  
393 software interfaces. This enables detailed live data logging, process control as well as online  
394 production parameter modification. The COPE Software acts as a user interface for the automation  
395 environment showing the pseudonymised patient data from the LogiqSuite platform. Using the  
396 LogiqSuite platform, data from all decentralised production sites will be registered in a standardised  
397 manner. The resulting LogiqAnalytics database will uniformly combine data from all sites and be  
398 crucial for continuous learning and model development.

399

400 **Figure 6. Integration of care and GMP processes in AIDPATH.** In brief, UMC 1 has (a) care and (b) CAR-T cell  
401 productions of (c) autologous cells, coupling the products and (d) their quality control to the (e) patient  
402 treatment. Care will be (f) recorded in LogiqCare and (g) ATMP / GMP CAR-T cell production and (h) quality  
403 control in LogiqScience, which are (i) interconnected. If a clinical-decision tool is developed this can be (j) used  
404 to determine a desired CAR-T cell product from personal clinical data, and (k) provide this information to adapt  
405 the production process. Data within a case (l) contains consults and test cascades (Figure 4), (m) data from  
406 other institutes (UMC 2) is added to LogiqAnalytics and all data are gathered for (n) continual learning to  
407 improve the clinical-decision support tool.

408

#### 409 IV. AI development from a database

410 In two oncological projects, AML-MRD and AIDPATH, data from LogiqAnalytics is used for machine  
411 learning algorithms to develop predictive models for survival analysis with competing risks, that can  
412 be used as CDS tools (53). The data was imported from other software solutions (e.g., Access-, SPSS-,  
413 and Excel-databases) using ETL procedures and integrated in LogiqScience. The data from AML-MRD  
414 has been described above (e.g. Figure 4). The data for the AIDPATH use case 5 was derived from  
415 multicentre clinical trial ARI-0001 (54,55) and ARI-0002h (56).

416 In the AML-MRD case the competing risks were graft versus host disease (GvHD) and toxicity  
417 (chemotherapy related risk), and tumour relapse. In the CAR-T cell case the competing risks were  
418 cytokine-release syndrome, immune suppression, and tumour relapse.

419 Data from LogiqScience was synced to LogiqAnalytics. Available and putative relevant biomedical  
420 parameters were selected, based on their potential mechanistical relevance and their timely  
421 availability for prognosis in practice. Machine learning was employed directly on data in  
422 LogiqAnalytics (57). All these activities can be performed directly on the real-time synced data in  
423 LogiqAnalytics, allowing the CDS prediction model to be updated without having to rebuild the  
424 machine learning flow. Currently, we are working on the parameter selection and data analytics in  
425 these projects to develop a prediction model.

426

#### 427 V. Real-world data integration with clinical-decision support tools

428 ORTEC has developed and connected two CDS tools, U-Prevent and the Stroke Triage App, to  
429 LogiqScience to collect data directly after CDS use. Since LogiqAnalytics is fed real-time from  
430 LogiqCare and LogiqScience, this allows continuous learning of the CDS prediction model. After  
431 expert evaluation for MDR this results in continual learning to improve the AI CDS tools.

432 U-Prevent is an MDD-certified clinical decision support solution with multiple prediction models for  
433 cardiovascular risk management. Practical use is greatly enhanced by using data directly from  
434 electronic health records (58). U-Prevent allows users to enter missing data or to use imputation  
435 (59). In the recent PROSPERA project (unpublished results), we have coupled LogiqSuite to U-Prevent  
436 (see Figure 2), allowing it to prefill data in U-Prevent from the primary care electronic health records  
437 in U-Prevent. After data revision and updates by the physician, the U-Prevent CDS tool can be used  
438 in patient care to support shared decision making and allocating appropriate care. At the end of the  
439 U-Prevent session, the updated data and CDS outcomes are stored in LogiqScience. Hereby,  
440 LogiqSuite facilitates efficient use of a clinical decision support tool and automatic registration of  
441 follow-up data, in this case in a pseudonym way due to local interpretations of the GDPR. Moreover,  
442 the analytics environment yields real-time dashboard with included patients, genders, risks profiles,  
443 prevention strategies, and more, that can easily be selected per primary care centre. These key-  
444 performance indicators (KPIs) facilitate benchmarking of various practitioners.

445 For prehospital triage of patients with suspected acute stroke we developed the Stroke Triage App, a  
446 MDR class I CDS tool (unpublished results), based on the risk predictions and time tables of PRESTO-  
447 2 implementation study (60,61) and disease history and routing information. In brief, the Stroke  
448 Triage App supports routing decisions for emergency care by advising on the adequate hospital level  
449 for stroke care in relation to the patient's prehospital triage assessment. Pseudonymous patient  
450 identifiers, data from prehospital triage, GPS location, routing information, and advised outcome are  
451 stored in LogiqScience. During the study the data can be enriched with clinical follow-up data. This  
452 will allow the continual evaluation of the Stroke Triage App and facilitate continuous improvement  
453 of the underlying CDS prediction model for prognosis used in routing decisions.

## 454 Discussion

455 Precision medicine stratifies diagnosis-treatment combinations between similar diseases, using large  
456 and complex datasets. This implies the need for good quality large databases of complex medical  
457 data that should be GDPR compliant, collect real-world data, integrate care and research, allow FAIR  
458 data exchange with multiple parties, communicate with clinical-decision support tools, easy to adapt  
459 for progressive insight, facilitate real-time analysis, and provide ready-to-use data for AI methods to  
460 develop new prediction models for prognostic, diagnostic and therapeutic purposes. In different use-  
461 cases we have described a MedDMS system that fulfils the basic criteria. The different approach of  
462 the integrated solution does not end but opens important subjects of discussion.

463

## 464 Data protection

465 MedDMSs for precision medicine goes beyond data sharing in the cloud and demands a higher level  
466 of traceable procedures. This includes documentation of implementation, data stewardship for the  
467 traceability of design modification, data validation steps, and privacy protection by a fine-grained  
468 authorisation network. The automatic recorded and verified process documentation becomes more  
469 important where research complexity grows in number of parameters and inclusion centres. The  
470 plan for data management is crucial, so a MedDMS cannot seen independent from its procedures to  
471 maintain confidentiality of data and consistency of the database.

472 LogiqSuite allows integration of medical care and clinical research with real-time available  
473 descriptive statistics, and machine learning in one solution, with different and appropriate privacy  
474 levels for each use case. LogiqSuite distinguishes itself from other MedDMS by its interactions with  
475 CDS prediction models for precision medicine as shown in the two double use-cases. LogiqSuite has  
476 prepares real-world data in real-time for development. Data from LogiqSuite is directly available in  
477 CDS tools, facilitating their easy use. This is linked to the feedback of the (modified) input and results

478 into LogiqSuite. Syncing these data in real-time to LogiqAnalytics facilitates the cycle of continual  
479 learning for precision medicine.

480 Sharing data from different study sites applies standard and interoperable solutions for  
481 implementing and managing medical registries to each site (62). LogiqSuite allows this with the same  
482 data templates, but ABAC restricts data access to those authorised to the relevant attributes e.g.,  
483 institutes. Input validation is an important asset of data stewardship (40). The medical data scientists  
484 guide the use of these and other features in the MedDMS implementation.

485

## 486 [Comparing MedDMSs](#)

487 LogiqSuite is not the first cloud MedDMS. REDCap exists since 2009 (63) and is used for multilocation  
488 eCRFs (64). Castor EDC (65) and FIMED (66) also support storage of medical research data. Castor  
489 EDC has a signable integrated informed consent for research purposes, which should technically be a  
490 trusted third party (TTP) i.e., with separate access authentication. LogiqSuite distinguishes from  
491 other MedDMS using four-eyes principles to guard data structure, quality, and access. Since this  
492 process loops in the medical data scientists, they can also give advice for dynamic templates and live  
493 analytics, which are new for many medical researchers. LogiqSuite also has added template  
494 translations, sending of questionnaires to patients or subjects, connectivity with open API, and  
495 support with data transfer from other sources by ETL. LogiqSuite users are charmed by the real-time  
496 study monitoring, which is suitable for monitoring study completeness and exerting needed  
497 corrective actions.

498

## 499 Integrated versus federated MedDMS.

500 Not all involved in precision medicine choose to integrate MedDMSs in a single solution. Some  
501 parties propose blockchain with communication tools between physically separate MedDMSs (67).  
502 Separate MedDMSs offers some easy advantages for privacy protection, but ABAC and RBAC  
503 strategies on a need-to-know basis could also do this. Federated learning methods in healthcare are  
504 developed to allow machine learning to be exerted. The major challenge in machine learning is the  
505 poor data quality yielding poor CDS models. While federated learning is pioneered in various  
506 projects, methods for data stewardship and curation in federated learning still needs to be explored.  
507 Separate MedDMSs will have different interpretations of parameters and different data quality. Poor  
508 quality data could be curated, which is a critical step in machine learning (68). Input validation and  
509 data curation techniques might be suboptimal for federated approaches (69,70). Options are to  
510 perform this either centrally (71) or by general rules (e.g. x times the standard deviation) (72). In our  
511 experience, these methods have limited efficacy for curing complex, large, divergent medical data  
512 with missing properties. Other methods for data curation are needed for efficient use of federated  
513 learning techniques. Especially, systematic shifts due to confounders between data subgroups  
514 should be detected and understood. Ideally data stewardship assesses an anonymised data set  
515 completely, since this facilitates the application of the complex rules for data stewardship more  
516 appropriately.

517

## 518 Beyond the MedDMS.

519 Even if many medical research groups collaborate in single database, there will always be medical  
520 data beyond the current collaborations, which also needs to be integrated for scientific research. So  
521 MedDMSs should be prepared for FAIR data exchange. Medical data are complex by their nature,  
522 consisting of various domains (e.g. anamnesis, follow-up, disease classifications, lab tests,  
523 medication) and includes various types of data (e.g. dates, values, repeated measurements), which



524 use various methods for data classifications. After early and continued initiatives for Observational  
525 Medical Outcomes Partnership (OMOP) common data model (73,74), the scientists expanded this to  
526 collaborate in different settings for FAIR data exchange (75). Common data models are important for  
527 various purposes and crucial for AI (76). Systems are being developed to add FAIR standards de novo  
528 to MedDMSs (77). Multiple initiatives exist to draw data details of FAIR sharing (78), many in  
529 national or regional settings. Combining complexity of medical data and the regional approach, we  
530 foresee that there will be multiple relevant common data models and corresponding data  
531 dictionaries. The LogiqAnalytics data can be filled with all relevant predefined common data model  
532 structures, using SQL views. In LogiqSuite multiple data dictionaries can be integrated for FAIR data  
533 exchange.

## 534 Conclusions

535 LogiqSuite is a MedDMS that integrates directly identifiable care data, pseudonymous science data  
536 and minimal identifiable to anonymous analytic data. LogiqSuite was evaluated in five different use  
537 cases for precision medicine, including data research, multilocation study monitoring, integration of  
538 research and production data with care data, real-time use of data for prediction model  
539 development, and input and registration of data from CDS tools. The value of this MedDMS was  
540 shown in different biomedical fields, including oncology, cardiovascular risk management,  
541 pulmonology, and prehospital triage.

542 LogiqSuite is unique in supporting real-time data analysis in care settings, which is also a beloved  
543 feature for scientific researchers. Moreover, LogiqSuite supports collaborative clinical care and  
544 research/laboratory workflows. The data can originate from manual entry in clinical care, eCRFs of  
545 research studies, clinical-decision support tools, import via ETL from third-party sources, and  
546 through FAIR-compliant open API. Available data can be monitored in real-time, providing tools for  
547 data monitoring and continuous feedback loop for data analytics and prediction model  
548 development. In clinical practice this facilitates continual learning with real-world data, since MDR  
549 legislation require a clinical evaluation for the validity of updated prediction models in every cycle.

550 Appropriate tools are crucial for KPI monitoring and progress in care and science. The LogiqSuite  
551 MedDMS application supports data management and science for advancement to precision  
552 medicine by data collection and collaboration, facilitating analytics and machine learning, and  
553 implementation of CDS tools.

554

555 **Acknowledgements:**

556 The authors would like to thank Bob Roozenbeek and Ruben van Wijdeven (Erasmus MC, Rotterdam,  
557 The Netherlands) for their invaluable contribution to the Stroke Triage App. The authors would like  
558 to thank Sandy Pratama, Thom Steenhuis, Marco Koeleman, Christian Hutter, Vlad Constantin Lipan  
559 and Alina Bratosin for their excellent technical assistance on the development of LogiqSuite.

560

561 **Conflicts of interest:** JJJ, IB, IV, LBR, & SD are employees of ORTEC B.V. which has a commercial  
562 interest in LogiqSuite.

563

564 **Author contributions:** Authors contributed on *medical data science* (JJJ, IB, IV, LBR), *LogiqSuite*  
565 *design* (JJJ, IB, IV, LBR, TS) and the use-cases *AML-MRD* (JC, LLN, JT, PG, JJJ, IB, IV, LBR), *P4O2*  
566 (AHMZ, LDB, IB, IV, JJJ, LBR), *AID PATH database* (JJJ, IV, IB, SH, FE, CS, SNV, MJ, MH), *AIDPATH AI*  
567 (JJJ, IV, AV, IB, SNV, MJ, CS, MH), *AML MRD AI* (AV, JJJ, IB, IV, JC, LLN, JT), *Stroke Triage App* (IB,  
568 JJJ), *Prospera* (VAMCB, HJAO, RCV, JJJ, IB, IV), and *Idea & initial draft writing* (JJJ). All authors  
569 contributed to the paper and agreed with its contents.

570

571

572 **Abbreviation list**

573	ABAC	Attribute-Based Access Control
574	AI	Artificial Intelligence
575	AML	Acute Myeloid Leukaemia
576	CAR-T	Chimeric Antigen Receptor T-Cells
577	CDS	Clinical Decision Support
578	DMS	Data Management System
579	eCRF	Electronic Case Report Form
580	ETL	Extract Transform Load
581	FAIR	Findable, Accessible, Interoperable, and Reusable
582	GDPR	General Data Protection Regulation
583	GMP	Good Manufacturing Practice
584	IdP	Identity Provider
585	LAIP	Leukaemia-Aberrant Immune Phenotype
586	MedDMS	Medical Data Management System
587	MDR	Medical Device Regulation
588	MFA	Multi-Factor Authentication
589	MRD	Minimal Residual Disease
590	OMOP	Observational Medical Outcomes Partnership
591	PII	Personal Identifiable Information
592	RBAC	Role-Based Access Control
593	TTP	Trusted Third Party

594

## 595   References

- 596   1.     DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug  
597         development costs. *J Health Econ.* 2003; 22: p. 151.
- 598   2.     Stewart DJ, Kruzrock R. Cancer: The Road to Amiens. *J Clin Oncol.* 2009; 27: p. 328-33.
- 599   3.     Scannell J, Blackley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D  
600         efficiency. *Nat Rev Drug Discov.* 2012; 11: p. 191-200.
- 601   4.     Lo AW. Can Financial Economics Cure Cancer? *Atl Econ J.* 2021; 49: p. 3-21.
- 602   5.     Den Otter W, Steerenberg PA, Van der Laan JW. Testing therapeutic potency of anticancer  
603         drugs in animal studies: a commentary. *Regul Toxicol Pharmacol.* 2002; 35: p. 266-272.
- 604   6.     Fosse V, Oldoni E, Bietrix F, Budillon S, Daskalopoulos EP, Fratelli M, et al. Recommendations  
605         for robust and reproducible preclinical research in personalised medicine. *BMC Medicine.*  
606         2023; 21: p. 14.
- 607   7.     Delpierre C, Lefèvre T. Precision and personalized medicine: What their current definition says  
608         and silences about the model of health they promote. Implication for the development of  
609         personalized health. *Front Sociol.* 2023; 8: p. 1112159.
- 610   8.     Chan IS, Ginsburg GS. Personalised medicine: progress and promise. *Ann Rev Genomics Hum*  
611         *Genet.* 2011; 12: p. 217-44.
- 612   9.     Snyderman R. Personalised health care: from theory to practice. *Biotechnol J.* 2012; 7: p. 973.
- 613   10.    Jacobs JLL, Characiejus D, Scheper RJ, Stewart RJE, Tan JFV, Tomova R, et al. The Amiens  
614         Strategy: small phase III trials for clinically relevant progress in the war against cancer. *J Clin*  
615         *Oncol.* 2009; 27: p. 3062-3.
- 616   11.    Shaban-Nejad A, Michalowski M, Buckeridge DL. Health intelligence: how artificial intelligence  
617         transforms population and personalised health. *NPJ Digit Med.* 2018; 1: p. 53.

- 618 12. Ollier W, Muir KR, Lophatananon A, Verma A, Yuille M. Risk biomarkers enable precision in  
619 public health. *Per Med*. 2018; 15: p. 329-342.
- 620 13. Kamb A. What's wrong with our cancer models? *Nat Rev Drug Discov*. 2005; 4: p. 161-5.
- 621 14. Characiejus D, Hodzic J, Jacobs JLL. "First do no harm" and the importance of prediction in  
622 oncology. *EPMA J*. 2010; 1: p. 369-375.
- 623 15. Ginsburg GS, Philips JA. Precision Medicine: From Science To Value. *Health Affairs*. 2018; 37: p.  
624 694-701.
- 625 16. Kalra D. The importance of real-world data to precision medicine. *Pers Med*. 2019; 16: p. 79-  
626 82.
- 627 17. Mc Cord KA. The personalized medicine challenge: shifting to population health through real-  
628 world data. *Int J Public Health*. 2019; 64: p. 1255-1256.
- 629 18. Christopoulos P, Schlenk R, Kazdal D, Blasi M, Lennerz J, Shah R, et al. Real-world data for  
630 precision cancer medicine-A European perspective. *Genes Chromosomes Cancer*. 2023; 62: p.  
631 557-563.
- 632 19. Arondekar B, Duh MS, Bhak RH, DerSarkissian M, Huynh L, Wang K, et al. Real-World Evidence  
633 in Support of Oncology Product Registration: A Systematic Review of New Drug Application  
634 and Biologics License Application Approvals from 2015-2020. *Clin Cancer Res*. 2022; 28: p. 27-  
635 35.
- 636 20. Douglas MP, Kumar A. Analyzing Precision Medicine Utilization with Real-World Data: A  
637 Scoping Review. *J Pers Med*. 2022; 12: p. 557.
- 638 21. Fountzilas E, Tsimberidou AM, Vo HH, Kurzrock R. Clinical trial design in the era of precision  
639 medicine. *Genome Med*. 2022; 14: p. 101.

- 640 22. Sheldon J, Ou W. The real informatics challenges of personalised medicine: not just about the  
641 number of central processing units. *Per Med.* 2013; 10: p. 639-45.
- 642 23. Müller H, Dagher G, Loibner M, Stumptner C, Kungl P, Zatloukal K. Biobanks for life sciences  
643 and personalised medicine: importance of standardization, biosafety, biosecurity, and data  
644 management. *Curr. Opin Biotechnol.* 2020; 64: p. 45-51.
- 645 24. Li C. Personalised medicine - the promised land: are we there yet? *Clin Genet.* 2011; 403-12: p.  
646 79.
- 647 25. Hallack JA, Azar DT. The AI Revolution and How to Prepare for It. *Transl Vis Sci Technol.* 2020;  
648 9: p. 16.
- 649 26. Vlahou A, Hallinan D, Apweiler R, Argiles A, Beige J, Benigni A, et al. Data Sharing Under the  
650 General Data Protection Regulation: Time to Harmonize Law and Research Ethics?  
651 *Hypertension.* 2021; 77: p. 1029-1035.
- 652 27. Delhalle S, Bode SFN, Balling R, Ollert M M, He FQ. A roadmap towards personalised  
653 immunology. *NPJ Syst Biol Appl.* 2018; 4: p. 9.
- 654 28. Loscalzo J. Systems Biology and Personalised Medicine A Network Approach to Human  
655 Disease. *Proc Am Thorac Soc.* 2011; 8: p. 196-8.
- 656 29. Samuels S, Balint B, von der Leyen H, Hupé P, de Koning L, Kamoun C, et al. Precision medicine  
657 in cancer: challenges and recommendations from an EU-funded cervical cancer biobanking  
658 study. *Br J Cancer.* 2016; 115: p. 1575-1583.
- 659 30. Yeatman TJ, Mule J, Dalton WS, Sullivan D. On the eve of personalised medicine in oncology.  
660 *Cancer Res.* 2008; 68: p. 7250–7252.
- 661 31. Agur Z. Biomathematics in the development of personalised medicine in oncology. *Future*  
662 *Oncol.* 2006; 2: p. 39-42.

- 663 32. You K, Wang P, Ho D. N-of-1 Healthcare: Challenges and Prospects for the Future of  
664 Personalised Medicine. *Front. Digit. Health*. 2022; 4: p. 830656.
- 665 33. Tracz V, Lawrence R. Towards an open science publishing platform. *F1000Res*. 2016; 5: p. 130.
- 666 34. Noor AM, Holmberg L, Gillett C, Grigoriadis A. Big Data: the challenge for small research  
667 groups in the era of cancer genomics. *Br J Cancer*. 2015; 113: p. 1405–1412.
- 668 35. Modzelewska D, Sole-Navais P, Sandstrom A, Zhang G, Muglia LJ, Flatley C, et al. Changes in  
669 data management contribute to temporal variation in gestational duration distribution in the  
670 Swedish Medical Birth Registry. *PLoS One*. 2020; 15: p. e0241911.
- 671 36. Limaye N. Data management Redefined. *Perspect Clin Res*. 2010; 1: p. 110–112.
- 672 37. Yu KH, Hart SN, Goldfeder E, Zhang QC, Parker SCJ, Snyder M. Harnessing big data for precision  
673 medicine: infrastructures and applications. *Pac Symp Biocomput*. 2017; 22: p. 635-639.
- 674 38. Fahr P, Buchanan J, Wordsworth S. A Review of the Challenges of Using Biomedical Big Data for  
675 Economic Evaluations of Precision Medicine. *Appl Health Econ Health Policy*. 2019; 17: p. 443–  
676 452.
- 677 39. Viceconti M, Hunter P, Hose R. Big data, big knowledge: big data for personalised healthcare.  
678 *IEEE J Biomed Health Inform*. 2015; 19: p. 1209-15.
- 679 40. Data clinic. Online.; 2021. Available from: [https://elevatehealth.eu/courses/data-](https://elevatehealth.eu/courses/data-stewardship/)  
680 [stewardship/](https://elevatehealth.eu/courses/data-stewardship/).
- 681 41. Tran VA, Johnson N, Redline S, Zhang GQ. OnWARD: Ontology-driven Web-based Framework  
682 for Multi-center Clinical Studies. *J Biomed Inform*. 2011; 44: p. S48–S53.
- 683 42. Information security in healthcare - register NEN 7510. Online. Available from:  
684 <https://www.nen.nl/en/certificatie-en-keurmerken-nen-7510>.
- 685 43. ScienCrew. Online. Available from: <https://www.sciencrow.com/c/10446?title=LogiqSuite>.



- 686 44. Beekers I, van Dijk S, Pratama S, Jacobs JLL. ScienCrew. Online.; 2023. LogiqSuite. Architecture,  
687 Security & System requirement. Available from:  
688 <https://storage.imgzine.com/public/432/LogiqSuite+Architecture+and+Security+Documentati>  
689 [on.pdf](#).
- 690 45. Auth0. Online.; 2023. Available from: <https://auth0.com/>.
- 691 46. ScienCrew. Online.; 2023. Available from: <https://www.sciencrow.com/>.
- 692 47. Barker M, Chue Hong NP, Katz DS S, Lamprecht AL, Martinez-Ortiz C, Psomopoulos F, et al.  
693 Introducing the FAIR Principles for research software. Scientific Data. 2022; 9(Gruenpeter M,  
694 Martinez PA, Honeyman T.): p. 622.
- 695 48. Heuser M, Freeman SD, Ossenkoppele GJ, Buccisano F, Hourigan CS, Ngai LL, et al. 2021  
696 Update on MRD in acute myeloid leukemia: a consensus document from the European  
697 LeukemiaNet MRD Working Party. Blood. 2021; 2753-2767: p. 138.
- 698 49. Van Solinge TS, Zeijlemaker W, Ossenkoppele GJ, Cloos J, Schuurhuis GJ. The interference of  
699 genetic associations in establishing the prognostic value of the immunophenotype in acute  
700 myeloid leukemia. Cytometry B Clin Cytom. 2018; 94: p. 151-158.
- 701 50. Baalbaki N, Blankestijn JM, Abdel-Aziz MI, de Backer J, Bazdar S, Beekers I, et al. Precision  
702 Medicine for More Oxygen (P4O2)-Study Design and First Results of the Long COVID-19  
703 Extension. J. Pers. Med. 2023; 13: p. 1060.
- 704 51. Hort S, Herbst L, Bäckel N, Erkens F, Niessing B, Frye M, et al. Toward Rapid, Widely Available  
705 Autologous CAR-T Cell Therapy - Artificial Intelligence and Automation Enabling the Smart  
706 Manufacturing Hospital. Front Med (Lausanne). 2022; 9: p. 913287.
- 707 52. Bäckel N, Hort S, Kis T, Nettleton D, Egan J, Jacobs JLL, et al. The potential of Artificial  
708 Intelligence in automated CAR-T cell manufacturing. Front Mol Med Sec Cell therapy. 2023; 3:  
709 p. in press.

- 710 53. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. J.  
711 Am. Stat. Ass. 1999; 1446: p. 496-509.
- 712 54. Ortíz-Maldonado V, Rives S, Castellà M, Alonso-Saladrigues A, Benítez-Ribas D, Caballero-  
713 Baños M, et al. CART19-BE-01: A Multicenter Trial of ARI-0001 Cell Therapy in Patients with  
714 CD19+ Relapsed/Refractory Malignancies. Mol Ther. 2021; 29: p. 636-644.
- 715 55. Ortiz-Maldonado V, Alonso-Saladrigues A, Español-Rego M, Martínez-Cibrián N, Faura A,  
716 Magnano L, et al. Results of ARI-0001 CART19 cell therapy in patients with relapsed/refractory  
717 CD19-positive acute lymphoblastic leukemia with isolated extramedullary disease. Am J  
718 Hematol. 2022; 97: p. 731-739.
- 719 56. Oliver-Caldés A, González-Calle V, Cabañas V, Español-Rego M, Rodríguez-Otero P, Reguera JL,  
720 et al. Fractionated initial infusion and booster dose of ARI0002h, a humanised, BCMA-directed  
721 CAR T-cell therapy, for patients with relapsed or refractory multiple myeloma (CARTBCMA-  
722 HCB-01): a single-arm, multicentre, academic pilot study. Fractionated initial infusion and  
723 booster dose of ARI0002h, a humanised, BCMA-directed CAR T-cell therapy, for patients with  
724 relapsed or refractory multiple myeloma (CARTBCMA-HCB-01): a single-arm, multicentre,  
725 academic pilot study. 2023; 24: p. 913-924.
- 726 57. Verkouter I, Vegelien A, Beekers I, Navarro Velázquez S, Juan M, Sanges C, et al. Mathematical  
727 approach towards personalised prediction of the most efficient CAR-T cell product using  
728 survival analysis with competing risks. In 5th European CAR T-cell meeting, EHA-EBMT; 2023;  
729 Rotterdam, The Netheralnds.
- 730 58. Groenhof TKJ, Rittersma ZH, Bots ML, Brandjes M, Jacobs JJJ, Grobbee DE, et al. A  
731 computerised decision support system for cardiovascular risk management 'live' in the  
732 electronic health record environment: development, validation and implementation—the  
733 Utrecht Cardiovascular Cohort Initiative. Neth Heart J. 2019; 27: p. 435-442.

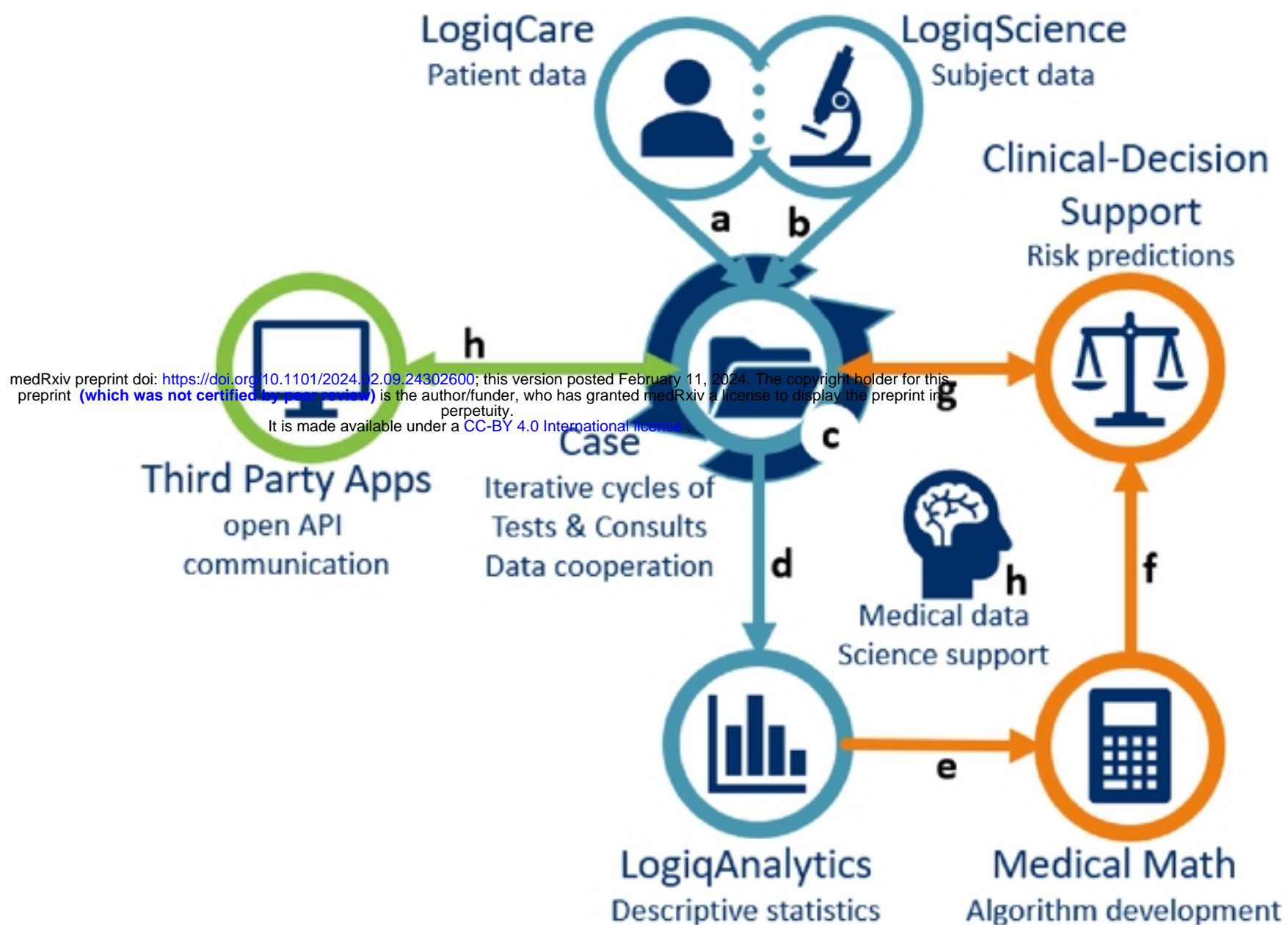
- 734 59. Nijman SWJ, Hoogland J, Groenhof TKJ, Brandjes M, Jacobs JIL, Bots ML, et al. Real-time  
735 imputation of missing predictor values in clinical practice. *Eur Heart J Digit Health*. 2020; 19: p.  
736 154-164.
- 737 60. Duvekot MHC, Venema E, Rozeman AD, Moudrous W, Vermeij FH, Biekart M, et al.  
738 Comparison of eight prehospital stroke scales to detect intracranial large-vessel occlusion in  
739 suspected stroke (PRESTO): a prospective observational study. *Lancet Neurol*. 2021; 20: p.  
740 213-221.
- 741 61. Nguyen TTM, van den Wijngaard IR, Bosch J, van Belle E, van Zwet EW, Dofferhoff-Vermeulen  
742 T, et al. Comparison of Prehospital Scales for Predicting Large Anterior Vessel Occlusion in the  
743 Ambulance Setting. *JAMA Neurol*. 2021; 78: p. 157-164.
- 744 62. Da Silva KR, Costa R, Crevelari ES, Lacerda MS, de Moraes Albertini CM, Filho MM, et al. Glocal  
745 Clinical Registries: Pacemaker Registry Design and Implementation for Global and Local  
746 Integration – Methodology and Case Study. *PLoS ONE*. 2013; 8: p. e71090.
- 747 63. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture  
748 (REDCap)--a metadata-driven methodology and workflow process for providing translational  
749 research informatics support. *J Biomed Inform*. 2009; 42: p. 377-81.
- 750 64. Obeid JS, McGraw CA, Minor BL, Conde JG, Pawluk R, Lin M, et al. Procurement of shared data  
751 instruments for Research Electronic Data Capture (REDCap). *J Biomed Inform*. 2013; 46: p.  
752 259-65.
- 753 65. Ottenhoff MC, Ramos LA, Potters W, Janssen MLF, Hubers D, Hu S, et al. Predicting mortality  
754 of individual patients with COVID-19: a multicentre Dutch cohort. *BMJ Open*. 2021; 11: p.  
755 e047347.
- 756 66. Hurtado S, García-Nieto J, Navas-Delgado I, Aldana-Montes JF. FIMED: Flexible management of  
757 biomedical data. *Comput Methods Programs Biomed*. 2021; 212: p. 106496.

- 758 67. Pericàs-Gornals R, Mut-Puigserver M, Payeras-Capel MM. Highly private blockchain-based  
759 management system for digital COVID-19 certificates. *Int J Inf Secur.* 2022; 21: p. 1069-1090.
- 760 68. Alves VM, Auerbach SS, Kleinstreuer N, Rooney JP, Muratov EN, Rusyn I, et al. Curated Data In  
761 - Trustworthy In Silico Models Out: The Impact of Data Quality on the Reliability of Artificial  
762 Intelligence Models as Alternatives to Animal Testing. *Altern Lab Anim.* 2021; 49: p. 73-82.
- 763 69. Gu X, Sabrina F, Fan Z, Sohail S. A Review of Privacy Enhancement Methods for Federated  
764 Learning in Healthcare Systems. *Int J Environ Res Public Health.* 2023; 20: p. 6539.
- 765 70. Hirano T, Motohashi T, Okumura K, Takajo K, Kuroki T, Ichikawa D, et al. Data Validation and  
766 Verification Using Blockchain in a Clinical Trial for Breast Cancer: Regulatory Sandbox. *J Med  
767 Internet Res.* 2020; 22: p. e18938.
- 768 71. Tacconelli E, Gorska A, Carrara E, Davis RJ, Bonten M, Friedrich AW, et al. Challenges of data  
769 sharing in European Covid-19 projects: A learning opportunity for advancing pandemic  
770 preparedness and response. *Lancet Reg Health Eur.* 2022; 21: p. 100467.
- 771 72. Appelbaum L, Kaplan ID, Palchuk MD, Kundrot S, Winer-Jones JP, Rinard M. Development and  
772 Experience with Cancer Risk Prediction Models Using Federated Databases and Electronic  
773 Health Records. In Linwood SL, editor. *Digital Health.* Brisbane (AU): Exon Publications; 2022.  
774 p. Chapter 2.
- 775 73. OHDSI. Online.; 2023. Available from: <http://www.ohdsi.org/data-standardization/>.
- 776 74. Belenkaya R, Gurley MJ, Golozar A, Dymshyts D, Miller RT, Williams AE, et al. Extending the  
777 OMOP Common Data Model and Standardized Vocabularies to Support Observational Cancer  
778 Research. *JCO Clin Cancer Inform.* 2021; 5: p. 12-20.
- 779 75. Wilkinson MD, Dumontier M, Aalbersberg U, Appleton G, Axton M, Baak A, et al. The FAIR  
780 Guiding Principles for scientific data management and stewardship. *Sci Data.* 2016; 3: p.  
781 160018.

- 782 76. Huerta EA, Blaiszik B, Brinson LC, Bouchard KE, Diaz D, Doglioni C, et al. FAIR for AI: An  
783 interdisciplinary and international community building perspective. *Sci Data*. 2023; 10: p. 487.
- 784 77. Groenen KHJ, Jacobsen A, Kersloot MG, Dos Santos Vieira B, van Enckevort E, Kaliyaperumal R,  
785 et al. The de novo FAIRification process of a registry for vascular anomalies. *Orphanet J Rare*  
786 *Dis*. 2021; 16: p. 376.
- 787 78. Inau ET, Sack J, Waltemath D, Zeleke AA. Initiatives, Concepts, and Implementation Practices  
788 of the Findable, Accessible, Interoperable, and Reusable Data Principles in Health Data  
789 Stewardship: Scoping Review. *J Med Internet Res*. 2023; 25: p. e45013.
- 790



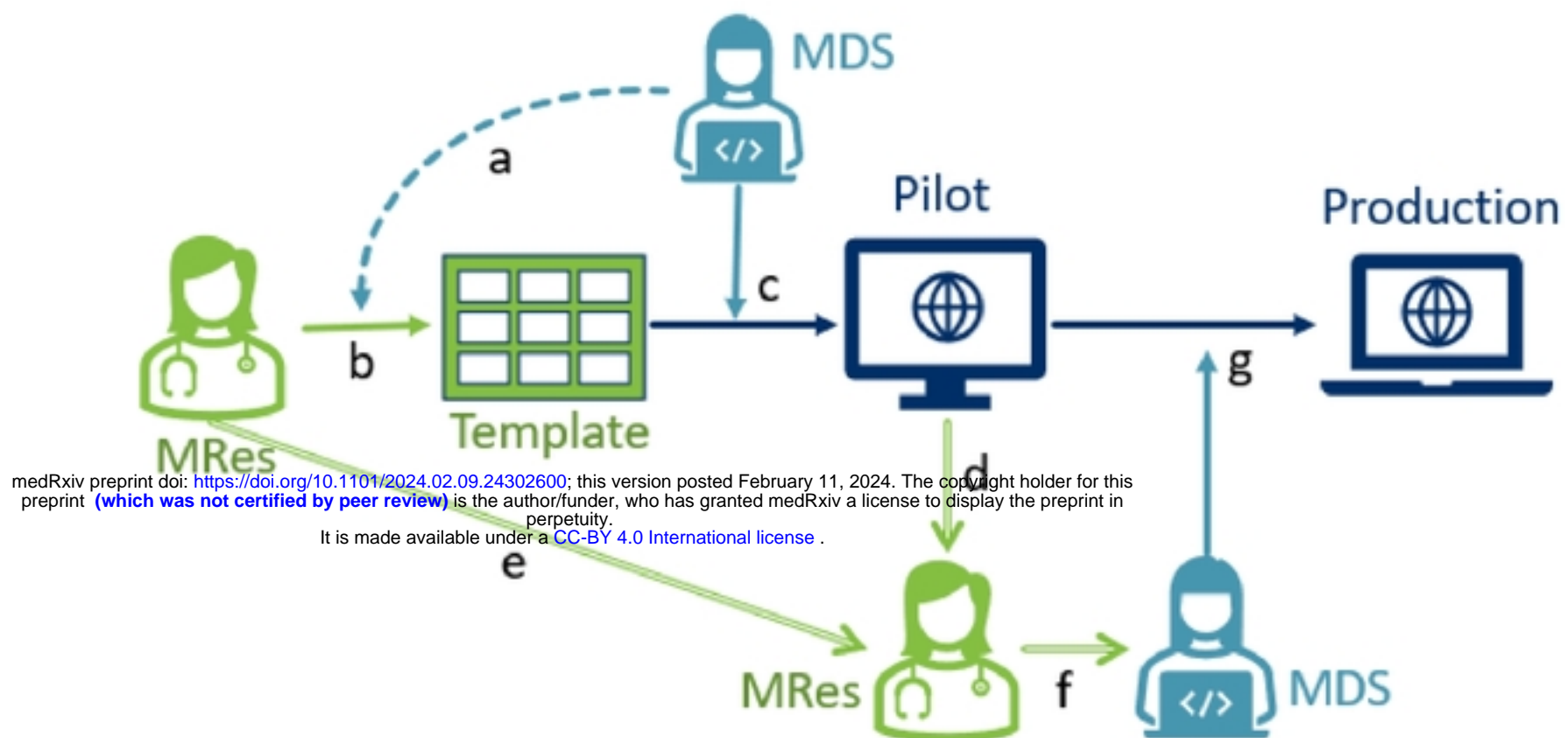
Figure 2



**Figure 2. LogiqSuite landscape.** LogiqSuite (blue) with interactions to other solutions from ORTEC (orange) and external sources (green). In brief, LogiqCare allows recording of (a) Patient data in LogiqCare, and (b) Subject data in LogiqScience, with (c) Cases as a central database. Case data are (d) synced to LogiqAnalytics for real-time descriptive statistics, and (e) ready-to-use for medical math, using AI to develop new prediction models. Prediction models can be (f) productised for risk predictions for clinical-decision support, that are (g) coupled with the central cases. Central cases also allow (h) open API communication with third party apps.



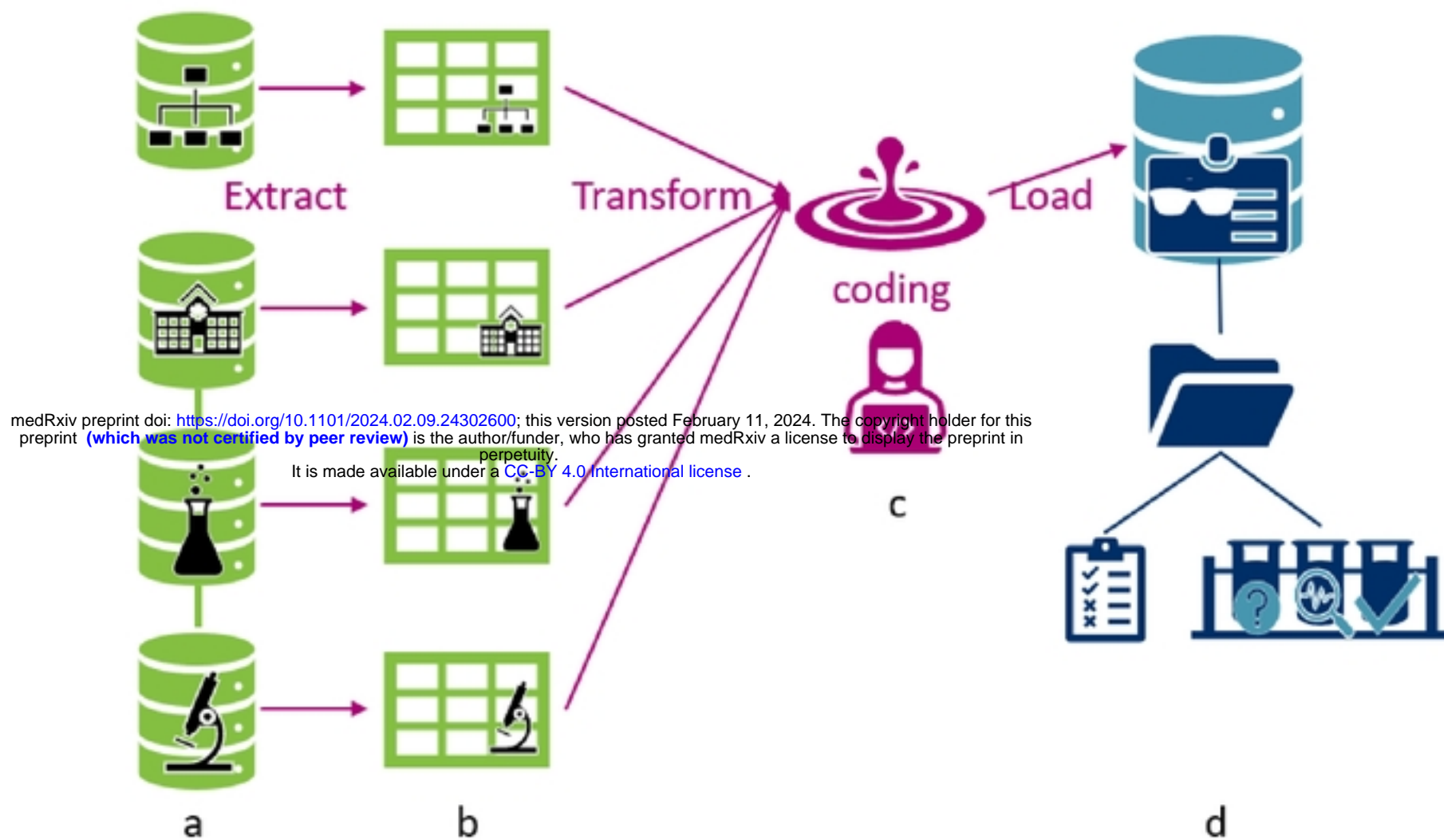
Figure 3



**Figure 3. Template customisation process.** In brief, (a) medical data scientist (MDS) provides a template to designated (b) medical researcher (MRes) who will design the desired data flows with MDS advice. When finalised, (c) the MDS will run the python script to upload the template in the Pilot environment, where (d) the medical researcher will validate the template, for (e) improvements or (f) to instruct the MDS to (g) transfer the template to the production environment.



Figure 4



**Figure 4. Extract-Transform-Load integration for AML.** In brief, (a) data was extracted from different databases, which were coupled with unique identifiers, (b) data transformation consisted of mapping to the LogiqSuite's data structure and values, that were translated and curated before loading into LogiqSuite, (c) which are written and stored as F-sharp code, to allow tracking and reusing of the procedure, and (d) data are loaded in LogiqSuite and organized in Patient or Subject – Case – Consults & Test cascade structures.

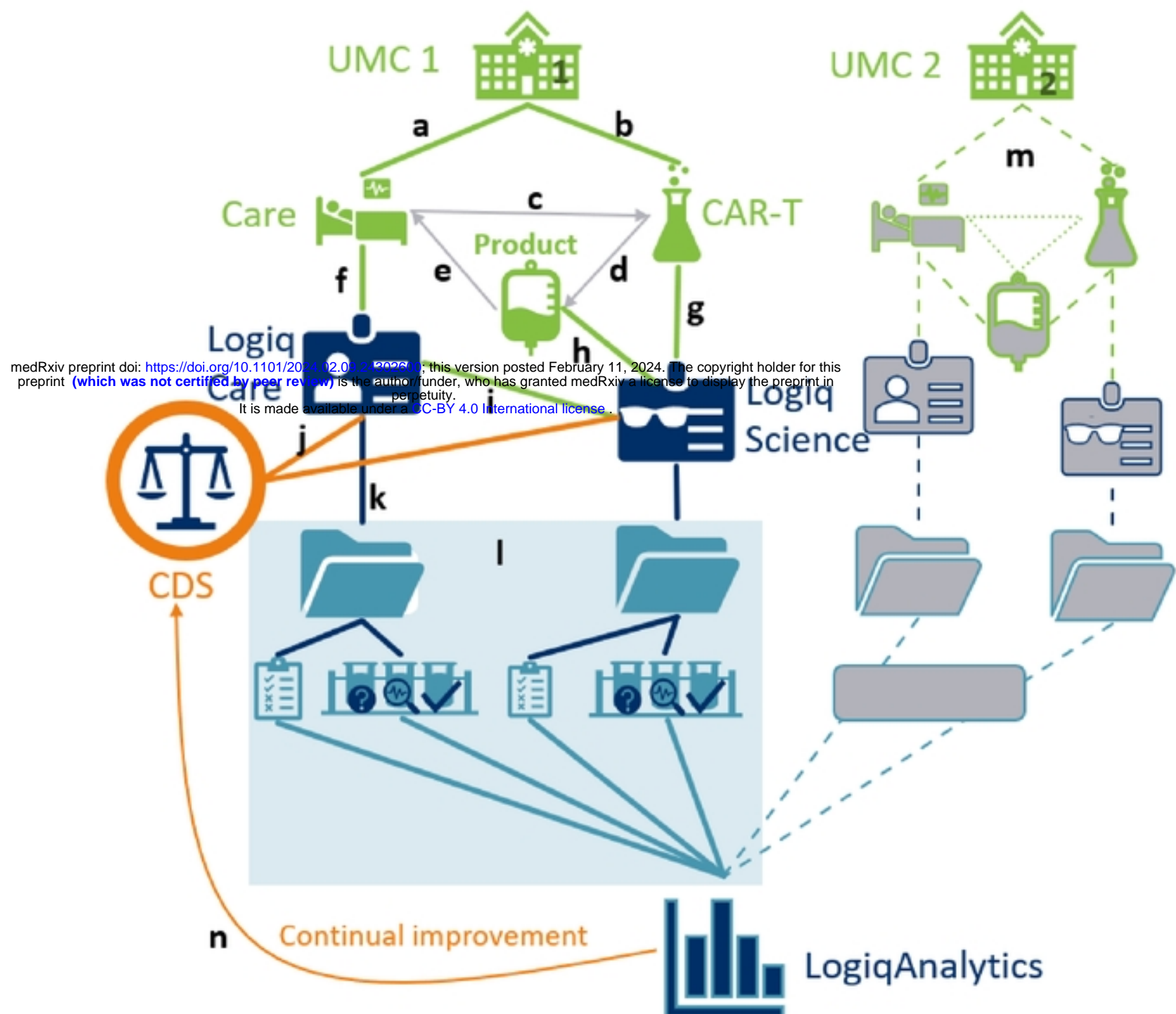
Figure 5



medRxiv preprint doi: <https://doi.org/10.1101/2024.02.09.24302600>; this version posted February 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

**Figure 5. Early study dashboard P402 Covid showing trial progress state in April 2022. In brief, data are synced in analytics, showing real-time study overview. Arrows point to missing data, which are indicated in grey. This information was used for real-time identification of missing data, and early corrective actions.**

Figure 6



**Figure 6. Integration of care and GMP processes in AIDPATH.** In brief, UMC 1 has (a) care and (b) CAR-T cell productions of (c) autologous cells, coupling the products and (d) their quality control to the (e) patient treatment. Care will be (f) recorded in LogiqCare and (g) ATMP / GMP CAR-T cell production and (h) quality control in LogiqScience, which are (i) interconnected. If a clinical-decision tool is developed this can be (j) used to determine a desired CAR-T cell product from personal clinical data, and (k) provide this information to adapt the production process. Data within a case (l) contains consults and test cascades (Figure ,m) Data from other institutes (UMC 2) is added to LogiqAnalytics and all data are gathered for (n) continual learning to improve the clinical-decision support tool.

medRxiv preprint doi: <https://doi.org/10.1101/2024.02.09.24302600>; this version posted February 11, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.  
It is made available under a [CC-BY 4.0 International license](#) .