Transformation of Microbiology Data into a Standardised Data Representation using OpenEHR

Antje Wulff^{1*#}, Claas Baier^{2*#}, Sarah Ballout¹, Erik Tute¹, Kim Katrin Sommer¹, Martin Kaase³, Anneka Sargeant⁴, Cora Drenkhahn⁵, Infection Control Study Group⁶, Dirk Schlüter², Michael Marschollek^{1#}, Simone Scheithauer^{3#}

¹Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany

²Institute for Medical Microbiology and Hospital Epidemiology, Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany

³Institute of Infection Control and Infectious Diseases, University Medical Center Göttingen (UMG), Georg-August University Göttingen, Robert-Koch-Str. 40, 37075 Göttingen, Germany.

⁴Institute of Medical Informatics, University Medical Center Göttingen (UMG), Georg-August University Göttingen Robert-Koch-Str. 40, 37075 Göttingen, Germany

⁵IT Center for Clinical Research (ITCR-L) and Institute of Medical Informatics (IMI), University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany

⁶List of study group members can be found in the main manuscript.

*these authors contributed equally to the work

*Corresponding author at:

Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany. E-Mail address: antje.wulff@plri.de (A.Wulff). URL: http://www.plri.de (A. Wulff)

*Corresponding author at:

Institute for Medical Microbiology and Hospital Epidemiology, Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany. E-Mail address: baier.claas@mh-hannover.de (C. Baier)

Supplementary Information

Supplementary Data 1: Exemplary comparison sheet for microbiology reports	. 3
Supplementary Data 2: Overview of the value lists defined and mapped onto terminologies* 1	10
Supplementary Data 3: OpenMibi application 1	12
Supplementary Data 4: Query parameters for a selected encounter1	13
Supplementary Data 5: Retrieving Data from the OpenEHR-based Data Repository by AQL 1	14
Supplementary Data 6: Retrieving Data from the OpenEHR-based Data Repository by openMibi 2	20
Supplementary Data 7: Retrieving Data from the OpenEHR-based Data Repository by openCQA 2	23
Supplementary Data 8: Results of the data quality check with pre-defined tasks and an extract of thre	e
exemplary result sets with different complexity levels of the openCQA tool2	28
Supplementary Data 9: Automated data quality check: openCQA result (Distribution of microbiology	
reports)	30
Supplementary Data 10: Automated data quality check: openCQA result (Distribution of bacterial	
counts)	31

Supplementary Data 1: Exemplary comparison sheet for microbiology reports

Exemplary detailed comparison for the retrospective review of microbiology reports (evaluation)

Microbiology Report for Patient #12345*				
Criteria OpenEHR Representation Gold Standard Dev				
Result report available?	Yes 🖂 No 🗆	Yes ⊠ No □		
Report state	Preliminary	Final		
Report time	24.05.2018	24.05.2018		
Receiving laboratory location	ceiving laboratory location Ward 81 Ward 81			
Total number of specimens in the entire report 1 1		1		
Specimen #1				
Specimen type	Urine (permanent catheter)	Urine (permanent catheter)		
Total number of the pathogens in the specimen22		2		
Culture	Aerobic and anaerobic culture	Aerobic and anaerobic culture		
Specimen identifier	123456789	123456789		
Date/time of sampling	22.05.2018 20:54	22.05.2018 20:54		
Body site	n/a	left		
Pathogen #1				
Pathogen name	Escherichia coli	Escherichia coli		
Pathogen quantity10^5/ml10^5/ml		10^5/ml		
Pathogen #2				
Pathogen name	Enterococcus faecalis	Enterococcus faecalis		
Pathogen quantity	10^5/ml	10^5/ml		

Antibiogram for Pathogen #1			
Which germ relates to the Antibiogram?	Specimen #1 Urine (permanent catheter) Pathogen #1 (Eschrichia coli)	Specimen #1 Urine (permanent catheter) Pathogen #1 (Eschrichia coli)	
Antibiotic name	Ampicillin	Ampicillin	
MIC value	>= 32	>=32	
Susceptibility	R 🛛 I 🗆 S 🗆		
Antibiotic name	Ampicillin/Sulbactam	Ampicillin-Sulbactam	
MIC value	>= 32	>=32	
Susceptibility	R 🛛 I 🗆 S 🗆		
Antibiotic name	Piperacillin/Sulbactam	Piperacillin Sulbactam	
MIC value	-	-	
Susceptibility	R 🛛 I 🗆 S 🗆	R 🛛 I 🗆 S 🗆	
Antibiotic name	Piperacillin/Tazobactam	Piperacillin-Tazobactam	
MIC value	>= 128	>=128	
Susceptibility	R 🛛 I 🗆 S 🗆	R 🛛 I 🗆 S 🗆	
Antibiotic name	Cefuroxim	Cefuroxim	
MIC value	= 16	16	
Susceptibility	R 🛛 I 🗆 S 🗆	R 🛛 I 🗆 S 🗆	
Antibiotic name	Cefuroxime-Axetil	Cefuroxime-Axetil	
MIC value	16	16	
Susceptibility	R 🛛 I 🗆 S 🗆	R 🛛 I 🗆 S 🗆	
Antibiotic name	Cefpodoxim	Cefpodoxim	
MIC value	= 2	2	
Susceptibility	R 🛛 I 🗆 S 🗆	R 🛛 I 🗆 S 🗆	

Antibiotic name	Ceftriaxon	Ceftriaxon
MIC value	-	-
Susceptibility	R I I S 🛛	R I I S 🛛
Antibiotic name	Cefotaxim	Cefotaxim
MIC value	<= 1	<=1
Susceptibility	R I I S 🛛	R I I S 🛛
Antibiotic name	Ceftazidim	Ceftazidim
MIC value	<= 1	<=1
Susceptibility	R 🗆 I 🗆 S 🖂	R 🗆 I 🗆 S 🛛
Antibiotic name	Gentamicin	Gentamicin
MIC value	<= 1	=1
Susceptibility	R 🗆 I 🗆 S 🖂	R 🗆 I 🗆 S 🛛
Antibiotic name	Tobramycin	Tobramycin
MIC value	-	-
Susceptibility	R □ I □ S ⊠	R 🗆 I 🗆 S 🖂
Antibiotic name	Fosfomycin	Fosfomycin
MIC value	<= 16	<=16
Susceptibility		R I I S 🛛
Antibiotic name	Nitrofurantoin	Nitrofurantoin
MIC value	<= 16	<=16
Susceptibility		R I I S 🛛
Antibiotic name	Levofloxacin	Levofloxacin
MIC value	<= 0.12	<=0.12
Susceptibility	R I I S 🛛	R I I S 🛛
Antibiotic name	Ciprofloxacin	Ciprofloxacin
MIC value	<= 0.25	<=0.25

Susceptibility		
Antibiotic name	Moxifloxacin	Moxifloxacin
MIC value	-	-
Susceptibility	R 🗆 I 🗆 S 🖂	R 🗆 I 🗆 S 🖂
Antibiotic name	Meropenem	Meropenem
MIC value	<= 0.25	<=0.25
Susceptibility	R 🗆 I 🗆 S 🛛	
Antibiotic name	Sulfamethoxazol/Trimethoprim (Cotrimoxazol)	Cotrimoxazol
MIC value	>= 320	>=320
Susceptibility	R 🛛 I 🗆 S 🗆	R 🛛 I 🗆 S 🗆
Antibiogram 2		
Which germ relates to the Antibiogram?	Specimen #1 Urine (permanent catheter) Pathogen #2 (Enterococcus faecalis)	Specimen #1 Urine (permanent catheter) Pathogen #2 (Enterococcus faecalis)
Antibiotic name	Ampicillin	Ampicillin
MIC value	<= 2	<=2
Suscentibility		
Susceptibility		
Antibiotic name	Ampicillin/Sulbactam	R □ I □ S ⊠ Ampicillin-Sulbactam
Antibiotic name MIC value	R L I L S 🛛 Ampicillin/Sulbactam <= 2	R □ I □ S ⊠ Ampicillin-Sulbactam <=2
Antibiotic name MIC value Susceptibility	R \Box I \Box S \boxtimes Ampicillin/Sulbactam <= 2 R \Box I \Box S \boxtimes	R I I S 🛛 Ampicillin-Sulbactam <=2 R I S 🖾
Antibiotic name MIC value Susceptibility Antibiotic name	R I I S X Ampicillin/Sulbactam <= 2 R I S X Gentamicin High-Level	R □ I □ S ⊠ Ampicillin-Sulbactam <=2 R □ I □ S ⊠ Gentamicin (Hochresistenz)
Antibiotic name MIC value Susceptibility Antibiotic name MIC value	R □ I □ S ⊠ Ampicillin/Sulbactam <= 2 R □ I □ S ⊠ Gentamicin High-Level -	R □ I □ S ⊠ Ampicillin-Sulbactam <=2 R □ I □ S ⊠ Gentamicin (Hochresistenz) -
Antibiotic name MIC value Susceptibility Antibiotic name MIC value Susceptibility	R I S × Ampicillin/Sulbactam <= 2 R I S × Gentamicin High-Level - R I S ×	R I S 🛛 Ampicillin-Sulbactam <=2 R I S 🖾 Gentamicin (Hochresistenz) - R I S 🗆
Antibiotic name MIC value Susceptibility Antibiotic name MIC value Susceptibility Antibiotic name	R \square I \square S \boxtimes Ampicillin/Sulbactam <= 2 R \square I \square S \boxtimes Gentamicin High-Level - R \square I \square S \boxtimes Doxycyclin	R □ I □ S ⊠ Ampicillin-Sulbactam <=2 R □ I □ S ⊠ Gentamicin (Hochresistenz) - R □ I □ S □ Doxycycline
Antibiotic name MIC value Susceptibility Antibiotic name MIC value Susceptibility Antibiotic name MIC value MIC value	R I S ⊠ Ampicillin/Sulbactam <= 2 R I S ⊠ Gentamicin High-Level - R I S ⊠ Doxycyclin -	R I I S X Ampicillin-Sulbactam <=2 R I I S X Gentamicin (Hochresistenz) - R I I S I Doxycycline -
Antibiotic name MIC value Susceptibility	R I S ⊠ Ampicillin/Sulbactam <= 2 R I S ⊠ Gentamicin High-Level - R I S ⊠ Doxycyclin - R I S □	R I S ⊠ Ampicillin-Sulbactam <=2 R I S ⊠ Gentamicin (Hochresistenz) - R I S □ Doxycycline - R I S □

MIC value	>= 16	>=16
Susceptibility	R 🛛 I 🗆 S 🗆	
Antibiotic name	Tigecyclin	Tigecyclin
MIC value	<=0.12	<0.12
Susceptibility	R I I S 🛛	
Antibiotic name	Nitrofurantoin	Nitrofurantoin
MIC value	<= 16	<=16
Susceptibility	R I I S 🛛	
Antibiotic name	Levofloxacin	Levofloxacin
MIC value	= 1	1
Susceptibility	R I I S 🛛	
Antibiotic name	Imipenem	Imipenem
MIC value	<= 1	<=1
Susceptibility	R I I S 🛛	
Antibiotic name	Vancomycin	Vancomycin
MIC value	= 2	2
Susceptibility	R I I S 🛛	R 🗆 I 🗆 S 🛛
Antibiotic name	Teicoplanin	Teicoplanin
MIC value	<= 0.5	<0.5
Susceptibility	R I I S 🛛	
Antibiotic name	Sulfamethoxazol/Trimethoprim (Cotrimoxazol)	Cotrimoxazol
MIC value	<= 10	<=10
Susceptibility		
Antibiotic name	Quinupristin/Dalfopristin	Quinu-/Dalfopristin
MIC value	= 8	8
Susceptibility		

Antibiotic name	Linezolid	Linezolid
MIC value	= 2	2
Susceptibility	R □ I □ S ⊠	R 🗆 I 🗆 S 🛛

Table 1: Exemplary comparison sheet for microbiology reports

Supplementary Data 2: Overview of the value lists defined and mapped onto terminologies*

Archetype	ltem	Consented value list	Terminology
Result report	State	Preliminary report, final report	Local terms
Laboratory test	Test name	Microbiology examination [fixed value]	Local terms
Specimen	Specimen type	Blood culture aerobic, blood culture anaerobic, blood culture, swab + free text for additional values	SNOMED CT
Anatomical location	Body site name	Anus, nose, nasopharyngeal, throat, rectum, arterial, venous + free text for additional values	SNOMED CT
Laboratory test analyte	Analyte	Presence, Absence	Local terms
Laboratory test analyte	Analyte result (identified pathogen)	List available in CKM: https://ckm.highmed.org/ckm/document?cid=1246. 159.96	SNOMED CT
Pathogen details	Frequency	List available in CKM: https://ckm.highmed.org/ckm/document?cid=1246. 159.101	SNOMED CT
Pathogen details – slot antibiogram	Antibiotic name	List available in CKM: https://ckm.highmed.org/ckm/document?cid=1246. 159.98	LOINC
Pathogen details – slot antibiogram	Susceptibili ty	Resistant, Susceptible (standard dosing regimen), Susceptible (increased exposure)	e.g. EUCAST
Pathogen details	MDRO class	e.g. MRSA, VRE	Local terms, based on national definitions

		List available in CKM:	
Laboratory	Order	https://ckm.highmed.org/ckm/document?cid=1246.	
test	request	<u>159.97</u>	LOING
		+ free text for additional values	

Table 2: Overview of archetypes used for representing microbiology data

*To outsource the complexity of terminology interaction from within applications, the HiGHmed consortium decided to license a terminology server: *Ontoserver* by CSIRO (*doi: 10.1186/s13326-018-0191-z*)



Supplementary Data 3: OpenMibi application

Figure 1: Full view of the openEHR-based application for microbiology data (openMibi)

Supplementary Data 4: Query parameters for a selected encounter

Number	Section	Items displayed	AQL parameter
(1)	Microbiology Report	Encounter identifier, status, date/time, order-ID, recipient location, order request	EHR-ID Encounter-ID
(2)	Specimens	Specimen type, specimen identifier, date/time collected, date/time received, anatomical location	EHR-ID Report-ID
<mark>(</mark> 3)	Microoganism	Name, result, date/time of result, germ number	EHR-ID Report-ID Specimen-ID
(4)	Antibiograms	Date/time generated, antibiotic name, resistance, minimum inhibitory concentration	EHR-ID Report-ID Specimen-ID Pathogen-ID

Query parameters for a selected encounter (openMibi)

AQL, Archetype Query Language

Table 3: Query parameters for a selected encounter in the openMibi application

Supplementary Data 5: Retrieving Data from the OpenEHR-based Data Repository by AQL

Example:

Find all patients with a blood culture sample carrying more than one species

Based on the standardized data repository many different data queries can be tailored and specified by the user. Therefore, the archetype query language (AQL)¹ can be used. AQL is a query language that acts directly upon the openEHR-based data model (called *archetype*) and not the database level meaning that each AQL query is executable in any openEHR based data repository as long as the same archetypes are used.

Step 0: Check availability of data in openEHR based format and access rights and credentials

Data must be integrated into and available from an openEHR based data repository (e. g. *Better Platform* by Better d.o.o. in Ljubljana², SL or *Ehrbase* (open source), commercial support by Vitasystems in Mannheim, DE³) at each institution, according to the consented and standardized data models. The user who would like to access data needs to gain access rights, credentials and connection parameters (e.g. the link to the query editor of the chosen repository or the link to a specific REST API).

<u>Step 1: Identification of the openEHR template representing the data required</u> In our case, we are looking for blood culture samples and microbiology results, so we need to search data in the data model *microbiology report*⁴.

¹ <u>http://www.openehr.org/releases/QUERY/latest/docs/AQL/AQL.html</u>

² <u>https://platform.better.care/</u>

³ <u>https://ehrbase.org/</u>

⁴ <u>http://ckm.highmed.org/ckm/templates/1246.169.69</u>



Figure 2: Overview of the hierarchy structure of the openEHR template for microbiology results

Step 2: Identification of the data items needed

According to our example, we need the (1) patient identifiers as the result of our query. With these identifiers, we can further look for more details of each patient (which is not included in this example).

However, we would like to find patients with specific microbiology results. Consequently, we also need data items referring to (2) an existing blood culture sample (data item: *specimen type* of the archetype specimen⁵), and (3) a culture result (data item: *analyte result* [identified pathogen] of the archetype laboratory analyte result⁶). To find out whether the blood culture sample delivers one or more results, we are also interested in the (4) *specimen identifier* (data item: laboratory specimen identifier of the archetype specimen).

Step 3: Reproduce template hierarchy in query

All relevant data items are included in the above-mentioned template. For the query, we need to reproduce the template hierarchy (Figure 2) in the query as needed for the search request.

To look for microbiology reports in the repository independent of a specific patient, we need to specify in the query that a so called 'composition' with the name

⁵ https://ckm.highmed.org/ckm/archetypes/1246.145.748

⁶ https://ckm.highmed.org/ckm/archetypes/1246.145.396

'microbiology report' must be available in any electronic health record (*contains composition microbiology report*).

Since we are searching for samples with microorganisms found, we need to specify that the report contains a sample (*contains cluster specimen*) and a finding (*contains cluster culture result*). As you can see in Figure 2, the specimen and the culture results are nested into another element called 'laboratory test result' which is why we need to include this above those two 'contains' statements as follows:

```
[...]
EHR e
CONTAINS COMPOSITION c[openEHR-EHR-COMPOSITION.report-
result.v1,'Microbiology report']
CONTAINS (
CLUSTER d[openEHR-EHR-CLUSTER.specimen.v1]
AND
OBSERVATION o[openEHR-EHR-OBSERVATION.laboratory_test_result.v1]
CONTAINS CLUSTER w[openEHR-EHR-CLUSTER.laboratory_test_analyte.v1]
)
```

Step 4: Find the correct technical path pointing to the data items

A patient can be identified by the so-called *ehr-id* representing an unique identifier of every patient. The corresponding so-called path is *e/ehr_id/value*.

Each archetype item is identified by a unique path that can be accessed to retrieve data from this archetype. The corresponding paths can be looked up for instance in the CKM (German) ⁷. Clicking on 'Show Paths' visualizes an unique path to the specific data item.

The required paths for this search request are:

- (1) e/ehr_id/value as patient_id
- (2) d/items[at0001]/value/id as specimen_id
- (3) d/items[at0029]/value/value as specimen_type
- (4) w/items[at0001]/value/value as pathogen_identified

The first letter points to the corresponding archetype carrying this data item (see above, e.g. specimen comes with the variable 'd' which is why the path to *specimen_id* stats with d/).

⁷ http://ckm.highmed.org/ckm/templates/1246.169.69

Step 4: Create SELECT FROM statement

It is now possible to create the query using the SELECT FROM statement as known

from other query languages.

'DISTINCT' can be used to assure that now duplicates are retrieved.

```
SELECT DISTINCT
e/ehr_id/value as patient_id,
d/items[at0001]/value/id as specimen_id,
d/items[at0001]/value/value as specimen_type,
w/items[at0001]/value/value as pathogen_identified
FROM EHR e
CONTAINS COMPOSITION c[openEHR-EHR-COMPOSITION.report-
result.v1,'Microbiology report']
CONTAINS (
CLUSTER d[openEHR-EHR-CLUSTER.specimen.v1]
AND
OBSERVATION o[openEHR-EHR-OBSERVATION.laboratory_test_result.v1]
CONTAINS CLUSTER w[openEHR-EHR-CLUSTER.EHR-CLUSTER.specimen.v1]
)
```

Step 5: Limit results

We are only interested in blood culture samples and in microbiology results delivering one or more microorganisms found which is why we can add a 'WHERE' statement in which such limiting requirements can be defined as follows:

```
WHERE
exists w/items[at0001]/value/value and
d/items[at0029]/value matches {'blood culture aerobic', 'blood culture
anaerobic', 'blood culture'}
```

Step 6: Sort results

The results can be ordered by the specimen id and the specimen type to identify whether two or more results were found in a specific specimen.

```
ORDER BY order by d/items[at0001]/value/id, d/items[at0029]/value/value asc
```

Step 7: Finalize query and send to repository

The final query for instance can look like this:

```
SELECT DISTINCT
e/ehr_id/value as patient_id,
d/items[at0001]/value/id as specimen id,
d/items[at0029]/value/value as specimen type,
w/items[at0001]/value/value as pathogen identified
FROM EHR e
CONTAINS COMPOSITION c[openEHR-EHR-COMPOSITION.report-
result.v1,'Microbiology report']
      CONTAINS (
          CLUSTER d[openEHR-EHR-CLUSTER.specimen.v1]
          AND
          OBSERVATION o[openEHR-EHR-OBSERVATION.laboratory test result.v1]
          CONTAINS CLUSTER w[openEHR-EHR-
CLUSTER.laboratory_test_analyte.v1]
          )
WHERE
exists w/items[at0001]/value/value and
d/items[at0029]/value matches {'blood culture aerobic', 'blood culture
anaerobic', 'blood culture'}
ORDER BY
order by d/items[at0001]/value/id, d/items[at0029]/value/value asc
There are different ways to send this query to the data repository according to the
```

repository chosen and the connection approaches (see step 0). Since these are technical issues, we abstain from presenting these.

patient_id	specimen_id	specimen_type	pathogen_identified
12345	00001	blood culture aerobic	Staphylococcus epidermidis
12345	00002	blood culture anaerobic	Staphylococcus epidermidis
678910	00003	blood culture aerobic	Micrococcus luteus
111213	00004	blood culture aerobic	Staphylococcus epidermidis
111213	00005	blood culture anaerobic	Staphylococcus epidermidis
141516	00006	blood culture aerobic	Staphylococcus capitis
141516	00007	blood culture anaerobic	Staphylococcus capitis
171819	80000	blood culture aerobic	Streptococcus anginosus
171819	00008	blood culture aerobic	Pediococcus pentosaceus
171819	00009	blood culture anaerobic	Staphylococcus epidermidis

A result might look like this (in a table format):

Table 4: AQL result (table format)

By looking at these structured results, it is easy to identify all patients with a blood culture sample carrying more than one species. However, there are a lot more possibilities to define queries, e.g. you can also work with so-called 'GROUP BY' and 'COUNT' statements. By this, it is also possible to identify all patients with a blood culture sample carrying more than one species fully automatically. These queries

come with an increased complexity, which is why we decided to present the abovegiven simplified query instead.

Supplementary Data 6: Retrieving Data from the OpenEHR-based Data Repository by openMibi

Example:

Find all Klebsiella pneumoniae isolates resistant to Ciprofloxacin

Specific applications can be developed to interact with the data repository. An example is our self-developed tool called openMibi. We included a simple query builder to illustrate the capabilities of such tools for retrieving data as specified by a user through such application.

Step 0: Check availability of data in openEHR based format and access rights and credentials for openMibi

Data must be integrated into and available from an openEHR based data repository (e. g. *Better Platform* by Better d.o.o. in Ljubljana, SL⁸ or *Ehrbase* (open source), commercial support by Vitasystems in Mannheim, DE⁹) at each institution, according to the consented and standardized data models. The user who would like to access data needs to gain access rights, credentials and connection parameters (e.g. the link to the openMibi tool).

Step 1: Open tool

OpenMibi is a web-based tool that can be accessed via a specific link when installed at the participating institution.

HiGHmed openMibi		Logout
	Patientensuche Suchen nach: C Ehr-Id DI-Zahl	
	Datensatzsuche Erreger, z.B. Klebsiella pneumoniae Antibiotikum, z.B. Ciprofloxacin Resistenz, z.B. R	
	Suchen	



⁸ <u>https://platform.better.care/</u>

⁹ https://ehrbase.org/

Step 2: Configure parameters

The user can either retrieve all microbiology reports by searching for a specific patient (as explained in Supplementary Information xx) or by using the newly added exemplary query builder. The user can define the desired microorganisms of interest, the antibiotic name and the resistance.

Step 3: Creating query

In the background, the application now creates a query based on the archetype query language (AQL) and sends it to the repository.

Step 4: Visualisation of results

The openMibi tool now is able of visualising all microbiology reports by their unique identifiers according to this request.



Figure 4: OpenMibi Query Builder with first results

Step 5: Browsing through reports

By clicking on a specific report identifier, the full report will be visualised as presented in Supplementary Data 3 and can be browsed.

Microbiology report						
Case Identifier: 123456 Report state: Final Date/time: 16/03/2017 Order Identifier: 111222 Recipient location: 73 Order request: aerobe/snaerobe culture	Case identifier: 123456 Report state: Final Outer/ine: 13/03/2017 Order identifier: 222383 Recipient location: 73 Order request: aerobe/anaerobe cultur	Case identifier: 123456 Report Mate: Final Date/time: 07/004/2017 Order identifier: 333444 Recipient hontion: 27 Order request: MIBA				
Specimens		Microorganisms	Antibiogram			
Specimen type: Blood culture aerobic		Nama: Klabriella oneumonine	Date/time generated:	15/03/2017		
Specimen identifier: 9999999999999		Result: Evidence	Antibiotic name	Resistance	MIC	MLAB name
Date/time collected: 11/03/2017, 9:45 pm		Date/time of result: 14/03/2017	Ampicillin	R	>32	Ampicillin
Date/time received: 12/08/2017, 1:54 pm		Germ number: 1	Ampicillin/Sulbactam	R	>32	Ampicillin-Sulbactam
Anatomical location: Other MLAB name: Blood culture			Aztreonam	R	>16	Aztreonam
		Name: Klebsiella pneumoniae	Cefotaxim	R	>8	Cefotaxim
		Result: Evidence	Ceftazidim	R	>32	Ceftazidim
Specimen type : Blood culture aerobic		Date/time of result: 14/03/2017	Cefuroxim	R	>16	Cefuroxim
Specimen identifier : 8888888888888		Germinumber: 2	Ciprofloxacin	R	>4	Ciprofloxacin
Date/time collected : 11/03/2017, 9:45 pm			Colistin	R	=4	Colistin
Anatomical location : Other			Ertapenem	R	34	Ertapenem
MLAB name: Blood culture			Gentamicin	R	>8	Gentamicin
			Meropenem	R	>16	Meropenem
			Monifloxacin	R	>2	Moxifloxacin
			Piperacillin/Tazobactar	m R	>32	Piperacillin-Tazobactam
			Sulfamethoxazol/Trime (Cotrimoxazol)	ethoprim R	>16	Cotrimoxazol
			Sulfamethoxazol/Trime (Cotrimoxazol) Tetracyclin	ethoprim R R	>16 =8	Cotrimoxazol Tetracyclin
			Sulfamethoxazol/Trime (Cotrimoxazol) Tetracyclin Tigecyclin	ethoprim R R S	>16 =8 =0.5	Cotrimoxazol Tetracyclin Tigecyclin

Figure 5: OpenMibi – visualisation of microbiology report

Supplementary Data 7: Retrieving Data from the OpenEHR-based Data Repository by openCQA

Example:

MIC distribution of Ciprofloxacin and Cefotaxim in *Klebsiella pneumoniae* isolates from patients in 01/2017- 06/2017

Another example on how users can interact with the standardized data repository through an application can be visualized with our self-developed openCQA tool¹⁰. Here, the user needs to specify a query that delivers all potential relevant data sets. After that, the user can apply different statistical calculations that are automatically proposed by the tool. In the end, a specific analysis, such as the distribution of MIC for specific antibiotics and microorganisms (e.g. *MIC distribution of Ciprofloxacin and Cefotaxim in Klebsiella pneumoniae isolates from patients in 01/2017-06/2017*) are visualized.

Step 0: Check availability of data in openEHR based format and access rights and credentials for openCQA

Data must be integrated into and available from an openEHR based data repository (e. g. *Better Platform* by Better d.o.o. in Ljubljana¹¹, SL or *Ehrbase* (open source), commercial support by Vitasystems in Mannheim, DE¹²) at each institution, according to the consented and standardized data models. The user who would like to access data needs to gain access rights, credentials and connection parameters (e.g. the link to the openCQA tool).

Step 1: Open tool

OpenCQA is a web-based tool that can be accessed via a specific link when installed at the participating institution.

¹⁰ <u>https://pubmed.ncbi.nlm.nih.gov/33750371/</u>

¹¹ <u>https://platform.better.care/</u>

¹² https://ehrbase.org/

≡ Derive by type	Aggregations	Dimensions	AQL Paths	Export	User
barplots_enabling_SMICS	^ Nr.	Tags	Execution	DomainPaths/Results	
HDQF	Click to add MMs fro	om this knowledge-base. Clear filter		Run	all MMs Delete MMs
quantitative_analyse_per_k	eimz				openCQA



Step 2: Create query

By clicking on 'AQL', the user can define an AQL query to load data into the tool.

≡ !	Derive by type Aggregations Dimensions AQI Paths Export	User
	X	ults
barplots	SELECT distinct w/items[at0026]/value/id as specimen_id,	
HDQF	<pre>w/items[at0027]/value as isolate_id, w/items[at0001]/value/value as pathogen_identified, g/items[at0004]/value as pathogen_identified,</pre>	MMs Clea
	q/items[at0001]/value/magnitude as MIC_value, g/items[at0001]/value/magnitude at MIC_value, g/items[at0001]/value/magnitude status as MIC magnitudestatus.	nts")>_1) {
quantitat	d/items[at0015]/value as collection_time FROM EHR e	auntallia d
Outcom	CONTAINS COMPOSITION C CONTAINS (ounts)>-
	CLUSTER d(openEHR-EHR-CLUSTER.specimen.v1) AND OPEENMETION c(openEHR EHR OPEENMETION) charactery foot normalis w1)	nts")>-1) 8
ML_Kno	CONTAINS (CLUSTER w[openEHR-EHR-CLUSTER.laboratory_test_analyte.v1] CONTAINS (CLUSTER w[openEHR-EHR-CLUSTER.laboratory_test_analyte.v1] CONTAINS (CLUSTER z[openEHR-EHR-CLUSTER.erregredetails.v1] and CLUSTER g[openEHR-EHR-	
Bericht z	CLUSTER.laboratory_test_analyte.v1]))	
Diaman		
Diagnos		
Impfstat		
Kennzei		
SARS-C	Max. page size: 10000	
Mikrobio	logischer Befund (CIM)	openCQA

Figure 7: OpenCQA - AQL builder

Step 3: Choose pre-defined methods

By clicking on 'derive by type', a set of pre-defined statistical methods are loaded and can be chosen (e.g. count, min, max, median).

Step 4: Edit pre-defined methods

By clicking on 'edit', the pre-defined methods can be edited and customized. For instance, in 'group', a grouping rule can be specified to reach a 'count per group' as needed in our example. Furthermore, the 'characterization' input field can be filled with any R code to define new functions or methods. In our case, we wrote a new method to visualise the results as a customized bar plot.

barplots_enabling_SMICS HDOF quantitative_analyse_per_keimz Outcome_Measures ML_Knowledge Bericht zur Kontaktverfolgung (CIM) 1 edet Oder Group: Impfstatus (CIM) 2 edet 2 edet Group: function(r) (sun(is.na(r))) Franceichnung Erregernachwer SARS-CoV2 (CIM) 3 edet Beicht zur Kontaktverfolgung (CIM) 4 edet edet Bericht zur Kontaktverfolgung (CIM) 4 edet edet Bericht zur Kontaktverfolgung (CIM) 3 edet For (penahr=ehr-ci-uster:laboratory_test_analyte.v1/itemal (at0001/value/ mative.v1/itemal (at0001/value/magnitude_status: analyte.v1/items[at0001/value/magnitude_status: analyte.v1/items[at0001/value/magnitude_status: analyte.v1/items[at0001/value/magnitude_status: analyte.v1/items[at0001/value/magnitude_status: analyte.v1/items[at0001/value/magnitude_status: analyte.v1/items[at0001/value/magnitude_status: analyte.v1/items[at0001/value/magnitude_status: analyte.v1/items[at0001/value/magnitude_status: analyte.v1/items[at0001/value/magnitude_status: analyte.v1/items[at0001/value/magnitude_status: analyte.v1/items[at0001/value/magnitude_status: analyte.v1/items[at0001/value/magnitude_status: analyte.v1/items[at0001/value/magnitude_status: analyte.v1/items[at0001/value/magnitude_status: analyte.v1/items[at0001/value/magnitude_status: analyte.v1/item	\equiv Derive by type Aggr	egatior	ns D)imensions AQL	Paths	Export					Us	er î
HDQF quantitative_analyse_per_keimz	barplots_enabling_SMICS	^										
quantitative_analyse_per_keimz Value Tags:	HDQF											
Outcome_Measures Implication Tags: [wew_counts.Cprofloxed] DomainPaths: [/*domainPaths:/*dataset.cov/[operativ=hir-cluster.laboratory_test_analyte.v1]/isem/fite.v1]/	quantitative_analyse_per_keimz						o _lL	<=0,25	>0,25 to <=0,5	>0,5		
ML_Knowledge DomainPaths: Diagnose (CIM) 1 edit ["domainPath": "dataset-row/[openethr-ehr-cluster.laboratory_test_analyte.v1]/items[at0001]/value/magnitude","pathXtension":null_items": "ataset-row/[openethr-ehr-cluster.laboratory_test_analyte.v1]/items[at0001]/value/magnitude","pathXtension":null_items": "ataset-row/[openethr-ehr-cluster.laboratory_test_analyte.v1]/items[at0001]/value/magnitude","pathXtension":null_items": "ataset-row/[openethr-ehr-cluster.laboratory_test_analyte.v1]/items[at0001]/value/magnitude","pathXtension":null_items": "ataset-row/[openethr-ehr-cluster.laboratory_test_analyte.v1]/items[at0001]/value/magnitude","pathXtension":null_items": "ataset-row/[openethr-ehr-cluster.laboratory_test_analyte.v1]/items[at0001]/value/magnitude], >-1) && (iMM tags indexOf("Cefotaxim")>- Diagnose (CIM) 2 edit forup:	Outcome_Measures			Tags: level_counts,Ciprofloxaci								
Benchtzur Kontaktverfolgung (CIM) 1 1 det 1 det 1 det	ML_Knowledge			DomainPaths: [{"domainPath":"dat	aset-row	/[openehr-ehr-	cluster.laborat	ory_test_analyte.v!	1]/items[at0001]/value	1		
Diagnose (CIM) 2 edit Group: Characterization: p=1/3 ex (min.tags.index01(_Celulatin j=2)) Impfstatus (CIM) 2 edit Intervention (velucatin j=2) 3 edit function (v) (sum(is.na(v))) Mikrobiologischer Befund (CIM) 4 edit 4 edit edit 4 edit edit 6 edit	Bericht zur Kontaktverfolgung (CIM)	1	edit	/magnitude", "pathEx row/[openehr-ehr-clu ("domainPath": "data	tension" uster.la set-row/	<pre>inull,"item":' boratory_test fopenehr-ehr-</pre>	'item0", "datatyg analyte.v1]/ite :luster.laborato	e":"numeric","match ms[at0001]/value/ma rv test analvte.v1]	hedPath":"dataset- agnitude"}, v 1/items[at0001]/value	ts")>-1) && (iMM.tag	gs.indexOf("Cefotaxim")>-	
Impfstatus (CIM) 2 dene 2// (denole 250) (reduel Characterization: 3 edit done 3 edit done 4 edit done 4 edit done 1 1 4 edit done 1 1 4 edit done	Diagnose (CIM)	_	edit	Group:)1) aa (iiviivi.tags.		-
Kennzeichnung Eregernachwei SARS-CoV-2 (CIM) 3 done function (Y) (sum (15,na(Y))) nalyte v1/items[at0001]/value/magnitude_status analyte v1/items[at0001]/value/magnitude_status Mikrobiologischer Befund (CIM) 4 det det Image: Second Secon	Impfstatus (CIM)	2	clone edit	if (item0 < 2501) { retur Characterization:						analyte v1]/items[a	t00011/value/magnitude	-
Mikrobiologischer Befund (CIM) Patientenaufenthalt (CIM) 4 edit done Hevel_counts,Ciprofloxacin Hevel	Kennzeichnung Erregernachwei SARS-CoV-2 (CIM)	3	clone	function(v) {sum(!i	s.na(v))	}				nalyte.v1]/items[at0 analyte.v1]/items[a	001]/value/magnitude_statu t0001]/value/magnitude	ıs
Patientenaufenthalt (CIM) 4 4 6 6 6 6 6 7 6 7 7 7 7 7 7 7 7 7 7 7	Mikrobiologischer Befund (CIM)		edit	Save Abort Delete	мм	2021 1:1:1:1:47			- 13	nalyte.v1]/items[at0	001]/value/magnitude_statu	IS
	Patientenaufenthalt (CIM)	4	clone	level_counts,Ciprofloxad	cin	Execute	Dimension le <=0,25	Value Size 479 479 17 17				
Personendaten (CIM) >0,5 [197]197	Personendaten (CIM)			1			>0,5	197 197				
Personeninformation zur Infektionskontrolle (CIM) 5 edit level_counts, Cefotaxim Execute # tem0 (numeric) = dataset-row[openehr-ehr-cluster.laboratory_test_analyte v1]/items[at0001]/value/magnitude # item1 (string) = dataset-row[openehr-ehr-cluster.laboratory_test_analyte v1]/items[at0001]/value/magnitude status	Personeninformation zur Infektionskontrolle (CIM)	5	edit clone	level_counts,Cefotaxim		Execute	# item0 (numer # item1 (string)	ic) = dataset-row/[ope = dataset-row/[opene	enehr-ehr-cluster.laboratory_te ehr-ehr-cluster.laboratory_test	st_analyte.v1]/items[a _analyte.v1]/items[at0	it0001]/value/magnitude 001]/value/magnitude_statu	s
Prozedur (CIM)	Prozedur (CIM)										openCC	QA 🗸

Figure 8: OpenCQA – Edit pre-defined methods

Alternative step for 2-4

For reoccurring statistical tasks or calculations, all of the above-mentioned procedures can be stored in a so-called 'knowledge base'. They are directly loaded when starting openCQA.

Step 5: Visualisation of results

By clicking on 'execute', the results for the chosen or edited methods are visualized.





	egation	is D	imensions AQL Paths	Export	User
barplots_enabling_SMICS	1	edit clone	barplot,Cefotaxim	Execute	# item0 (numeric) = other_data_input (iMM tags indexOf("level_counts")>-1) && (iMM tags indexOf("Cefotaxim")>-1) resultsValue # item0 (string) = other_data_input (iMM tags indexOf("level_counts")>-1) && (iMM tags indexOf("Cefotaxim")>-1) dimension_level
HDQF	2	edit clone	row_count	Execute	# item0 (string) = dataset-row/
quantitative_analyse_per_keimz	3	edit clone	level_counts,MIC	Execute	# item0 (numeric) = dataset-row/[openehr-ehr-cluster laboratory_test_analyte v1]/items[at0001]/value/magnitude # item1 (string) = dataset-row/[openehr-ehr-cluster laboratory_test_analyte v1]/items[at0001]/value/magnitude_status
Outcome_Measures					# item0 (numeric) = dataset-row[openehr-ehr-cluster laboratory_test_analyte v1]/items[at0001]/value/magnitude # item1 (stiring) = dataset-row[openehr-ehr-cluster laboratory_test_analyte v1]/items[at0001]/value/magnitude_status 14.4.2021 15:55.48 3
ML_Knowledge	4	edit clone	level_counts,Ciprofloxacin	14.4.2021, 15:55:47 Execute	Dimesion level Value Size
Bericht zur Kontaktverfolgung (CIM)					2025 to <=0,5 17 17 >0,5 197 197
Diagnose (CIM)	5	edit clone	level_counts,Cefotaxim	Execute	# item0 (numeric) = dataset-row/lopenehr-ehr-cluster laboratory_test_analyte v1}/items[at0001]/value/magnitude # item1 (string) = dataset-row/lopenehr-ehr-cluster laboratory_test_analyte v1}/items[at0001]/value/magnitude_status coercCQA

Figure 10: OpenCQA – Results (2)

The results for our example look like this:



Distribution of Klebsiella pneumoniae isolates according to MIC range (Ciprofloxacin)

Figure 11: Distribution of *Klebsiella pneumoniae* isolates according to MIC range (Ciprofloxacin)



Figure 12: Distribution of *Klebsiella pneumoniae* isolates according to MIC range (Cefoxtaxim)

Supplementary Data 8: Results of the data quality check with predefined tasks and an extract of three exemplary result sets with different complexity levels of the openCQA tool

Task	openCQA Result						
		Dimension		Count			
		Patients		51947			
		Microbiology reports		260.084			
		Specimens		303344			
		Species		754			
		Isolates		240594			
General information		Antibiograms		103742			
	Distribution of microbiology reports per patient (see Supplementary Data 9 for a box plot)						
	Mean			.006718			
	Median		2	2			
	Maximum		350	350			
			1				
	Percentage of MRSA on all Staphylococcus aureus isolates						
Proportion of MRSA	Number of Staphylococcus aureus isolates: 7916						
	Number of Methicillin-resistant Staphylococcus aureus isolates: 1724						
	= 1724/7916 = 0.218 = 21.8% (including copy strains)						
Proportion of Meropenem- resistant <i>Klebsiella</i>	Number of <i>Klebsiella pneumoniae</i> isolates: 3367 Number of <i>Meropenem-resistant Klebsiella pneumoniae</i> isolates: 164						
pneumoniae	= 104/3307 = 0.049 = 4.376 (including Copy Strains)						

	Distribution of bacterial counts in positive urine samples with							
	Escherichia coli as the sole microorganism (see Supplementary Data							
	10 for the diagram):							
	TO TOT the diagra	<u>am).</u>	1					
		Bacterial count	Frequency					
		not specified	242					
		10^2/ml	258					
		10^3/ml	477					
		10^4/ml	985					
		10^5/ml	3,638					
	Species		Number of isola	ites				
	Escherichia	a coli	11246					
	Staphyloco	occus aureus	7916					
	Pseudomo	nas aeruginosa	7775 4818					
	Enterococo	cus faecium						
found per species	Enterococo	cus faecalis	4483					
(with antibiogram)	Klebsiella p	oneumoniae	3367					
	Staphyloco	occus epidermidis	3274					
	Enterobact	er cloacae	2143					
	Klebsiella d	oxytoca	1740					
	Proteus mi	rabilis	1677					

Table 5: Results of the data quality check with pre-defined tasks and an extract of three exemplary result sets with different complexity levels of the openCQA tool

Supplementary Data 9: Automated data quality check: openCQA result (Distribution of microbiology reports)



Distribution of microbiology reports per patient



Supplementary Data 10: Automated data quality check: openCQA result (Distribution of bacterial counts)



Figure 14: OpenCQA result on distribution of bacterial counts in positive Urine samples with Escherichia coli as the sole bacteria