

Dr Warehouse

User Guide



Version	Action	Date	Author
v1	<i>Writing</i>	<i>June 2, 2017</i>	<i>Nicolas Garcelon</i>
v2	<i>Editing</i>	<i>September 5, 2017</i>	<i>Nicolas Garcelon</i>

Contents

CONTENTS	1
TABLES OF FIGURES:	2
1. ACCESS	4
2. GENERAL SEARCH ENGINE PHILOSOPHY	4
3. HOME PAGE AND MENU BAR	5
4. SEARCH PATIENTS	6
 4.1. THE SEARCH ENGINE	6
4.1.1. THE "FULL TEXT" FILTER	6
4.1.1.1. EXACT PHRASE	6
4.1.1.2. USING LOGICAL OPERATORS:	6
4.1.1.3. USE OF WILDCARDS:	7
4.1.1.4. USE OF PROXIMITY IN THE TEXT:	7
4.1.1.5. THE ATOMIC QUERY SEARCHES THE TERMS THAT CO-OCCUR IN THE SAME RECORD:	7
4.1.1.6. PRECALCULATION OF THE NUMBER OF PATIENTS:	8
4.1.1.7. EXTEND THE SEARCH TO SYNONYMS AND HYPOONYMS:	8
4.1.1.8. NEGATION AND FAMILY HISTORY:	8
4.1.2. THE "STRUCTURED" FILTER	9
4.1.3. ADVANCED SEARCH FOR FULL TEXT AND STRUCTURED FILTERS	10
4.1.4. TEMPORAL CONSTRAINTS	11
4.1.5. THE PATIENT FILTER	12
 4.2. THE SEARCH ENGINE RESULT	13
4.2.1. RESULT TAB	13
4.2.1.1. PATIENT LIST	13
4.2.1.2. POPULATE A COHORT	14
4.2.1.3. SAVE THE QUERY AND AUTOMATIC QUERY	17
4.2.1.4. SEARCH IN RESULT	17
4.2.1.5. EXPORT PATIENTS TO EXCEL	18
4.2.1.6. FILTER THE RESULT	19
4.2.1.7. SHARE THE QUERY	19
4.2.2. STAT DATA TAB	20
4.2.3. CONCEPTS TAB	20
4.2.4. BIOLOGY TAB	22
4.2.5. MAP TAB	23
4.2.6. CLUSTERING TAB	23
4.2.7. EXPORT STRUCTURED DATA TAB (UNDER DEVELOPMENT)	24
5. A PATIENT'S RECORD	25
 5.1. DOCUMENTS	26
5.1.1. VIEW THE RECORDS	26

5.1.2. SEARCH THE RECORDS	27
5.1.3. EXTRACT DATA FROM RECORDS: WEIGHT, SIZE, ETC.	28
5.2. BIOLOGICAL RESULTS	29
5.3. TIMELINE	31
5.4. THE PATIENT'S CARE PATHWAY	31
5.5. COHORT	32
5.6. EXTRACTED CONCEPTS	33
5.7. SIMILARITY CALCULATION	34
6. MY COHORTS	37
6.1. VIEWING A COHORT:	37
6.2. DESCRIPTION TAB	37
6.3. TABS "PATIENTS INCLUDED", "PATIENTS EXCLUDED", "PATIENTS QUERIED"	37
6.4. "PATIENTS IMPORTED" TAB OR IMPORT PATIENTS	38
6.5. "PATIENT COMMENTS" TAB	39
6.6. "CONCEPTS" TAB	39
6.7. "SIMILAR PATIENTS" TAB	39
7. MY QUERIES	40
8. NOTIFICATIONS	41
9. MY PROFILE	41
10. CONTACT	41
11. LOGS	42
12. ETL	42
13. CREDITS	42

Tables of figures:

Figure 1: The Dr. Warehouse home page	5
Figure 2: The search engine - default display	6
Figure 3: Example of 3 full text atomic searches	8
Figure 4: Adding a structured atomic query	9
Figure 5: Search for the "creatinemia" test in the biology thesaurus	9
Figure 6: Example of structured atomic query on biological results	10
Figure 7: Advanced search on each atomic query	10
Figure 8: Adding a temporal constraint between 2 atomic queries	11
Figure 9: Patient filter	12
Figure 10: The search engine's general result	13
Figure 11: Open a document in the search engine result	14

<i>Figure 12: Select a cohort or create a cohort and include patients in it</i>	15
<i>Figure 13: Selection of a cohort: New links and buttons available</i>	15
<i>Figure 14: Include, exclude, query and comment on a patient for a cohort</i>	16
<i>Figure 15: Save a query and automatically rerun it every month</i>	17
<i>Figure 16: Refine the search of a previous search</i>	18
<i>Figure 17: Export the patients found by the search engine to Excel</i>	18
<i>Figure 18: Quick search on the result of a search.</i>	19
<i>Figure 19: Share a query with one or several collaborators</i>	19
<i>Figure 20: display the descriptive statistics of patients found by the search engine</i>	20
<i>Figure 21: Automatic extraction of concepts contained in the records - sort by relevance</i>	22
<i>Figure 22: Biological results of patients found by the search engine</i>	23
<i>Figure 23: Perform a clustering from the search results</i>	24
<i>Figure 24: Quick search in the menu bar from the surname and start of the forename or vice versa</i>	25
<i>Figure 25: a patient record - overview</i>	26
<i>Figure 26: patient record - view the records</i>	27
<i>Figure 27: search for the term "biopsy" in all the records of a patient</i>	28
<i>Figure 28: patient record - automatic extraction of the patient's weight over time in the records</i>	29
<i>Figure 29: patient record - biological results as a document list</i>	30
<i>Figure 30: patient record - biological results in table form</i>	30
<i>Figure 31: TimeLine of a patient</i>	31
<i>Figure 32: a patient's care pathway, at hospital department level</i>	32
<i>Figure 33: patient record - list of cohorts in which the patient appears</i>	33
<i>Figure 34: patient record - list of concepts extracted from records</i>	34
<i>Figure 35 - patient record - similarity calculation - the most similar patients</i>	35
<i>Figure 36: patient record - similarity calculation - table of possible diagnoses</i>	36
<i>Figure 37: Viewing a cohort - description tab</i>	37
<i>Figure 38: Cohort - patients included tab</i>	38
<i>Figure 39: Cohort - Import patients in the cohort from an external list</i>	39
<i>Figure 40: My queries - viewing a saved query and patients found each month</i>	40
<i>Figure 41: My profile - my rights, my stats, my department</i>	41
<i>Figure 42: Usage log of the data warehouse</i>	42

1. Access

Link: <http://drwh.nck.aphp.fr>

Use your APHP ID and password to login.

Dr. Warehouse works on all browsers. However, Mozilla Firefox or Chrome are preferred.

2. General search engine philosophy

Dr. Warehouse is a data search tool focused on medical records. It has been developed to allow you to search for your patients from text contained in records, similar to if you were using Google on your patient records.

However, the text search can generate false positives or false negatives. This is why you should not hesitate to refine your queries by excluding certain expressions: diabetes not gestational diabetes etc. The full text search engine allows quite complex queries to be created in order to be specific enough or on the contrary more sensitive.

However, if the engine returns false positive patients, the interface is perfectly adapted to quickly detect them.

It also allows encoded data associated with records to be searched. For example, in the case of laboratory records containing patients' biological results, you can free text search patients whose records contain a specific term, or do a structured search for a biological value higher than the upper limit.

For structured data which is not usually associated with a record, we have reconstructed an *ad hoc* record in order to retain our basic premise. For example, the ICD10 diagnosis codes are grouped together within the same document for a patient, a date and a department. You can then search for patients through a free text search with the ICD10 diagnosis code or title, or using a structured search on the ICD10 code.

3. Home page and menu bar

The home page allows the situation of the warehouse and your data to be summarized (

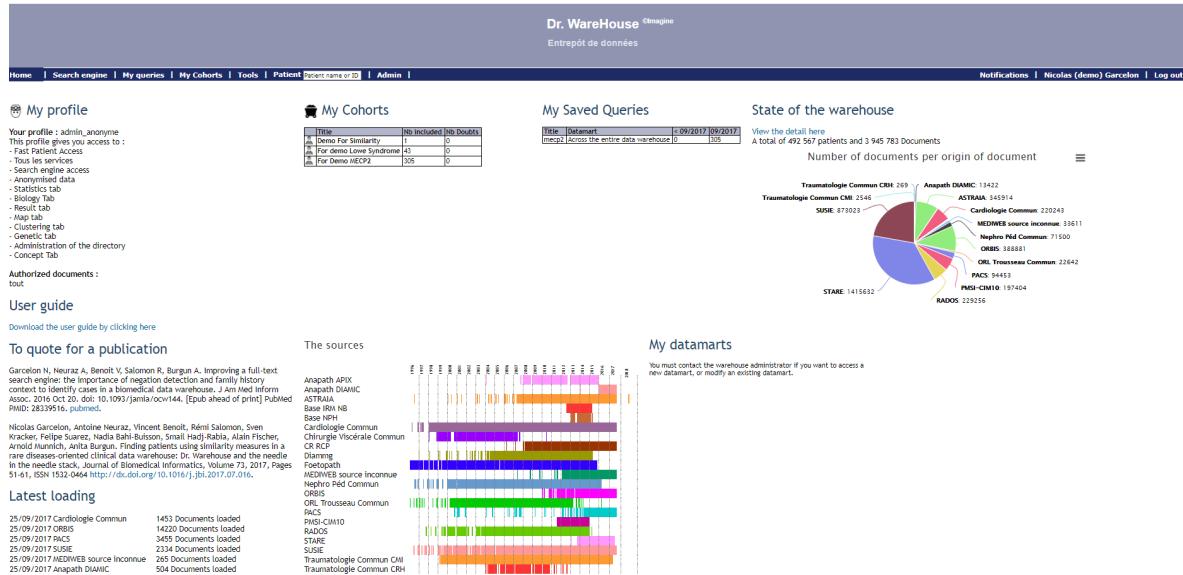


Figure 1).

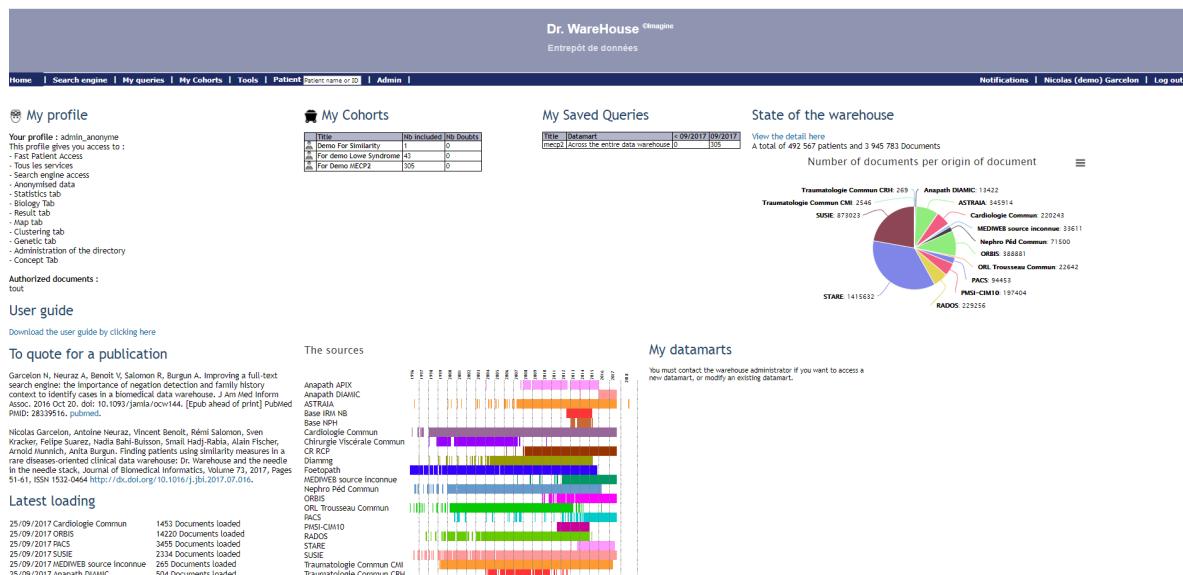


Figure 1: The Dr. Warehouse home page

You can view:

- your user profile and the scope of your data
- the list of your cohorts
- your saved queries and new patients found each month.
- the publication to cite in case of publications accomplished with the warehouse
- the warehouse situation: number of documents per source, with the passing of time in years and the last loading dates for each source.

The menu bar allows you to access (one chapter is dedicated to each item):

- the search engine
- your saved queries

- your cohorts
- tools (under development)
- quick access to patients by their full name or patient hospital ID
- notifications
- your profile and statistics

A link bar at the bottom allows access to:

- logs: warehouse usage statistics
- contact: contact details in case of problems
- ETL: details of loaded data
- credits: credits of image rights and JavaScript libraries

4. Search patients

Click on "search engine" in the menu bar at the top.

As specified in the introduction, you can perform free and structured text searches (Figure 2). To do this you can add multiple queries called "atomics" or sub queries. Each atomic query will give you the number of patients found. So when you carry out the search, the engine displays the intersection of patients who meet all atomic queries.

Search for patients

Across the entire data warehouse

1

Extend to synonyms :

+Advanced - Rewrite the query

+ Add a full text filter

+ Add a structured filter

+ Time constraints

+ Patient filter

START A SEARCH ?

Figure 2: The search engine - default display

4.1. The search engine

4.1.1. The "full text" filter

When you click on "search engine" in the menu bar, a free text search field is displayed; it is numbered "1" in a red circle.

You can enter the terms that you want to search in the records directly.

4.1.1.1. Exact phrase

The engine will search the phrase that you have entered exactly: "tetralogy of Fallot" is different to "tetralogy Fallot".

However, capital letters do not change the result:

"tetralogy of Fallot" = "tetralogy of fallot"

4.1.1.2. Using logical operators:

The search engine allows you to use logical operators to specify your query:

AND, OR, NOT

for example: diabetes and (tetralogy of fallot OR tetralogy fallot)

The use of brackets allows you to specify to the terms for which the operators apply.

NOT is placed before the expression to exclude. You cannot do an atomic query only containing a "NOT diabetes". For this reason we will see later in the advanced search that you can stipulate the atomic search as an exclusion criterion.

4.1.1.3. Use of wildcards:

You can use a wildcard character if you want to make your query more sensitive:

for example: diabet%

for example: diabet% and (tetralogy of fallot OR tetralogy fallot)

This will search for diabetic, diabetes etc.

4.1.1.4. Use of proximity in the text:

In some queries it may be useful to be able to specify a maximum distance between two words to get the most accurate results. To do this, just use the command **NEAR**.

Syntax: NEAR ((Word 1, Word 2 Word k), maximum number of words separating them)

For example: diabet% and NEAR ((tetralogy, fallot), 2)

You can specify if the word order has to be respected (by default, it is not): TRUE or FALSE:

For example: diabet% and NEAR ((tetralogy, fallot), 2, TRUE)

4.1.1.5. The atomic query searches the terms that co-occur in the same record:

In the example: eczema and infection% and thrombocytopenia

The search engine will search for all patients who have at least one record containing all the terms of the atomic query.

If a record contains only "eczema" and another contains only "infection" and "thrombocytopenia", the patient will not be found.

If you want to search for all patients who have at least one record with "eczema", one record with "infection" and one record with "thrombocytopenia", just add 3 full text filters, and state the search term in each filter (Figure 3). The engine will return the intersection of patients with all records found.

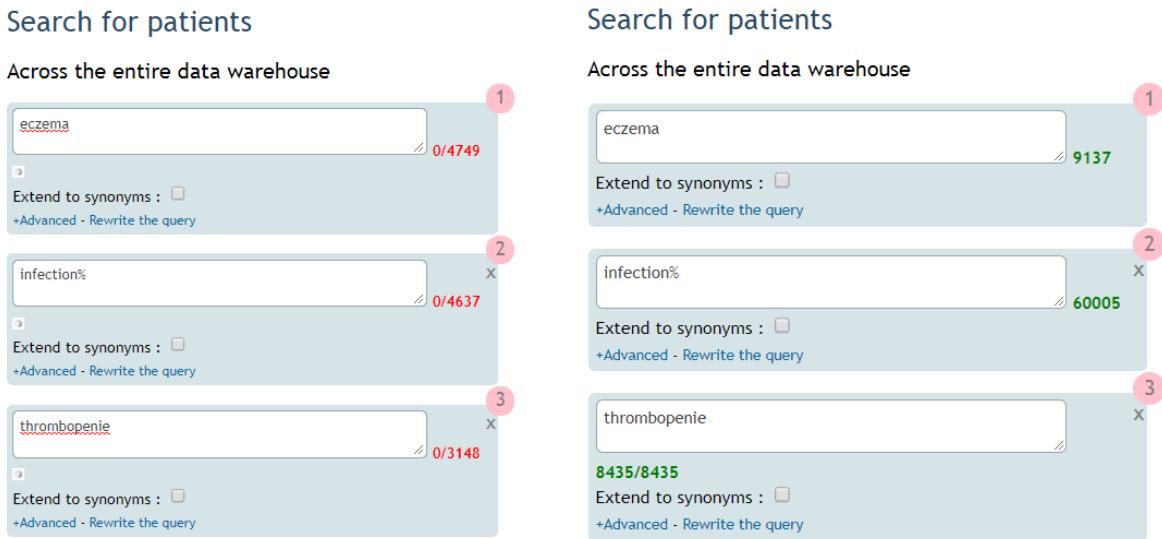


Figure 3: Example of 3 full text atomic searches

4.1.1.6. Precalculation of the number of patients:

When your cursor is in a "full text" atomic query, by pressing the keyboard enter key, the engine will precalculate the number of patients found for the atomic query. You can also launch the precalculation by clicking "?", or on the numbers displayed as "150/200" in green to the right of the free text.

This allows you to quickly test your query before running it to display the patients.

Once the precalculation has finished, the numbers displayed give you: the number of patients found who came into your department at least once / the number of patients found in the entire hospital.

4.1.1.7. Extend the search to synonyms and hyponyms:

Under the text field, you have the option to check a box "Extend to synonyms". This allows you to find records containing a synonym or a hyponym of the search term.

For example, if you search for "endocrine disorders", this will also find "diabetes".

If you search for "SCID", the system will also find "Omenn syndrome".

4.1.1.8. Negation and family history:

By default, the system excludes negative turns of phrase (sometimes unsuccessfully): "absence of diabetes", "has not allowed the conclusion of lupus" etc. This considerably reduces the number of false positives in results. But once again, errors may appear.

By default, the system also excludes phrases regarding the patient's relatives. Errors may also exist here.

The advanced search allows the inclusion or exclusion of these phrases and expressions that are excluded from results by default. Or you can also search specifically on these phrases.

The advanced search is available for each atomic query. It will be described in detail in section 4.1.3.

For family history: Simply specify in the drop-down menu "context": "Patient", "Family history" or "full text". By default, this is "Patient".

For negation, just specify in the drop-down menu "negation": "exclude" or "include". By default, negation is excluded.

4.1.2. The "structured" filter

You have the option to add as many "structured filters" as you like.

The structured filter will allow the data stored in encoded form in the data warehouse to be searched (Figure 4).

Search code : Thesaurus ▼ Go ? X
+Advanced

Figure 4: Adding a structured atomic query

To add a structured constraint:

- 1- Select a thesaurus in the drop-down menu "thesaurus", for example "Biology".
- 2- Then search for the code by typing the terms describing it, for example: "blood creat" then click on "Go" (Figure 5)

Search code : Biologie ▼ Go ? X
+Advanced

- CHIMIE DU SANG (5981548)
 - Créatinine (µmol/l)Sérum/Plasma Enzymologie (404718)
 - Créatinine (µmol/l)Sérum/Plasma (4)
 - 1 / créatininémie (µmol/l) (l/µmol) (36660)
 - Schwartz Lyon (taille*k)/creat (ml/min/1.73m2) (140)

Figure 5: Search for the "creatinemia" test in the biology thesaurus

- 3- Open up the thesaurus which is displayed below the search field and select the laboratory test that you are interested in. The number of existing tests in the warehouse is displayed in brackets, which allows you to see which laboratory test is most used. Choose "Creatinemia (µmol/l) Serum/Plasma enzymology (381569)".
- 4- You now have to specify the desired value range. For this you have 5 different options (Figure 6):
 - a. Value higher than the upper limit, lower than the lower limit or out of bounds
 - b. Value (<, >, =) X
 - c. Value between X and Y
 - d. Value higher than X times the upper limit
 - e. Value lower than X times the lower limit

Search code : Biologie ▼ creat sang Go X 3

Créatinine ($\mu\text{mol/l}$) Sérum/Plasma Enzymologie

Valeur ▼

Or Value at

Or Value between and

Or Value greater than times the upper bound

Or Value less than times the lower bound

- CHIMIE DU SANG (5981548)

- Crétaténine ($\mu\text{mol/l}$)Sérum/Plasma Enzymologie (404718)
- Crétaténine ($\mu\text{mol/l}$)Sérum/Plasma (4)
- 1 / créatininémie ($\mu\text{mol/l}$) (l/ μmol) (36660)
- Schwartz Lyon (taille*k)/creat (ml/min/1.73m²) (140)

+Advanced

Figure 6: Example of structured atomic query on biological results

4.1.3. Advanced search for full text and structured filters

You have the option of adding criteria to atomic queries by opening up the "advanced search" form for each query (Figure 7).

Search for patients

Across the entire data warehouse

1

eczema 9137

Extend to synonyms :

-Advanced - Rewrite the query

Exclude :

Age of patient at date of document : of years at years
of month at month

Origine du document : Select one or more sources

Hospital department : Choose one or more hospital departments

Date of document : from to

Information available over a period of at least : years

Context : ▼

negation Excluded ▼

2

infection% 60005

Extend to synonyms :

+Advanced - Rewrite the query

Figure 7: Advanced search on each atomic query

The additional filters are:

- Exclude: You can define the atomic query as exclusion criterion. So all patients found by this query will be excluded from the result.

- Patient's age on the date of the document: you can define an age range in years or in months on the date of the document searched for.
- Origin of the document: you can specify the document source (imaging, Laboratory Information Management System etc.)
- Department: the department that produced the document
- Date of the document: you can specify a range of dates within which the document was produced
- Information present over a period of at least: you can specify that you want the patients found to have at least 2 documents containing information and that they are spaced at least XX years apart. For example, you want patients who have lupus in at least 2 records, and that there is at least 5 years between these two records. You can put a decimal number if you want a time range of 1 month or a few days (1 day = 0.0027 years - 1 month = 0.08 years)
- As seen previously, for the free text search you can specify the context: patient (by default), family history, full text
- As seen previously, for the free text search, you can specify if you want to include negation or not (by default, it is excluded).

4.1.4. Temporal constraints

You can add temporal constraints between the atomic queries (Figure 8).

- Time constraints

Select 2 sub-queries and the constraint :

Filtered A

Filtered B

Constraints :

Simultaneous

Before after

Period

Figure 8: Adding a temporal constraint between 2 atomic queries

For example, you want to find patients who had severe eczema followed at least one year later by a severe asthma attack:

- Adding a full text filter type No. 1 atomic query: severe eczema
- Adding a full text filter type No. 2 atomic query: severe asthma
- Adding a temporal constraint:
 - Choose the 2 atomic queries to add the temporal constraint to:
 - Filter A: Filter No. 1 (filter numbers correspond to the number in the red circles of the atomic queries)
 - Filter B: Filter No. 2
 - Select the constraint type and if applicable specify the duration:
 - Simultaneous: the two documents found are dated on the same day
 - Before / After: The documents found by filter A precede the documents found by filter B. You have to specify if it is a maximum, minimum, or strict minimum (i.e. there have been no documents

found between the 2) limit and you have to specify the number of days, months or years.

- Period: the documents found by A and B are at a temporal distance without taking the order into account
- Click on "Add" so that it is added as a new atomic query.

The calculation time can be particularly long. This aspect is being optimized.

4.1.5. The patient filter

The patient filter is specific because it filters the result by the patient's condition today (Figure 9).

- Patient filter

The screenshot shows a form titled 'Patient filter' with the following fields:

- Sex : dropdown menu
- Age of patient today (years) : of at
- Patient / Deceased : living deceased
- Age of patient at death (years) : of at
- Period of 1st Coming: from to
- Follow up of minimum : years
- Exclude patients from these cohorts :
- Choisissez une ou plusieurs cohortes (button)

Figure 9: Patient filter

You can specify:

- The patient's sex
- The patient's age today by an age range in years
- If the patient is living or deceased
- If deceased, the age at which the patient died
- The period during which the patient came to the hospital for the first time
- A limit to patients who have been monitored by the hospital for a minimum number of years
- An exclusion of patients already present in cohorts. You can select several of them

These filters are not considered in the precalculation at the level of each atomic query. They will be taken into account at the level of the overall result, after the search is launched.

4.2. The search engine result

Once the query has been run, it is archived and can be found under the search engine in "Query history". Just click on a query so that it pre-fills the search engine.

The search engine result appears on several tabs (Figure 10): Result, Stat Data, Concepts, Biology, Map, and Clustering. We will present each of the tabs.

The screenshot shows the Dr. WareHouse search engine interface. At the top, there's a header with the logo 'Dr. WareHouse @Imagine' and 'Entrepôt de données'. Below the header, a navigation bar includes links for Home, Search engine, My queries, My Cohorts, Tools, Patient (with 'Patient name or ID'), Admin, Notifications (Nicolas (demo) Garcelon), and Log out.

The main area is titled 'Search for patients' and 'Across the entire data warehouse'. It displays two search boxes: one for 'eczema' (9137 results) and another for 'thrombopenie' (8435/8435 results). Below these boxes are buttons for 'Extend to synonyms' and 'Advanced - Rewrite the query'.

On the left, there are filters: 'Add a full text filter', 'Add a structured filter', 'Time constraints', and 'Patient filter'. At the bottom left are 'START A SEARCH' and help buttons.

The central part of the interface shows a list of patients. Each patient entry includes a thumbnail, name, age, and a list of documents associated with them. For example, 'Patient M, 18 years' has 461 documents, and 'Patient M, 58 years' has 3137 documents. Each document entry includes a link to 'View other documents'.

At the bottom, there are buttons for 'FEED A COHORT', 'SAVE QUERY', 'SEARCH ON RESULT', 'EXPORT PATIENTS TO EXCEL', 'FILTER THE RESULT BELOW', and 'SHARE YOUR QUERY'.

Figure 10: The search engine's general result

4.2.1. Result Tab

The header shows the number of patients displayed out of the total number of patients found. The patients displayed are only those who have gone into the user's department at least once.

It then shows the number of documents found for authorized patients out of the total number of documents found.

It then displays the user query in French. This includes atomic queries in free and structured text, temporal constraints and patients' filters.

It then displays buttons enabling certain actions to be performed, which will be detailed below:

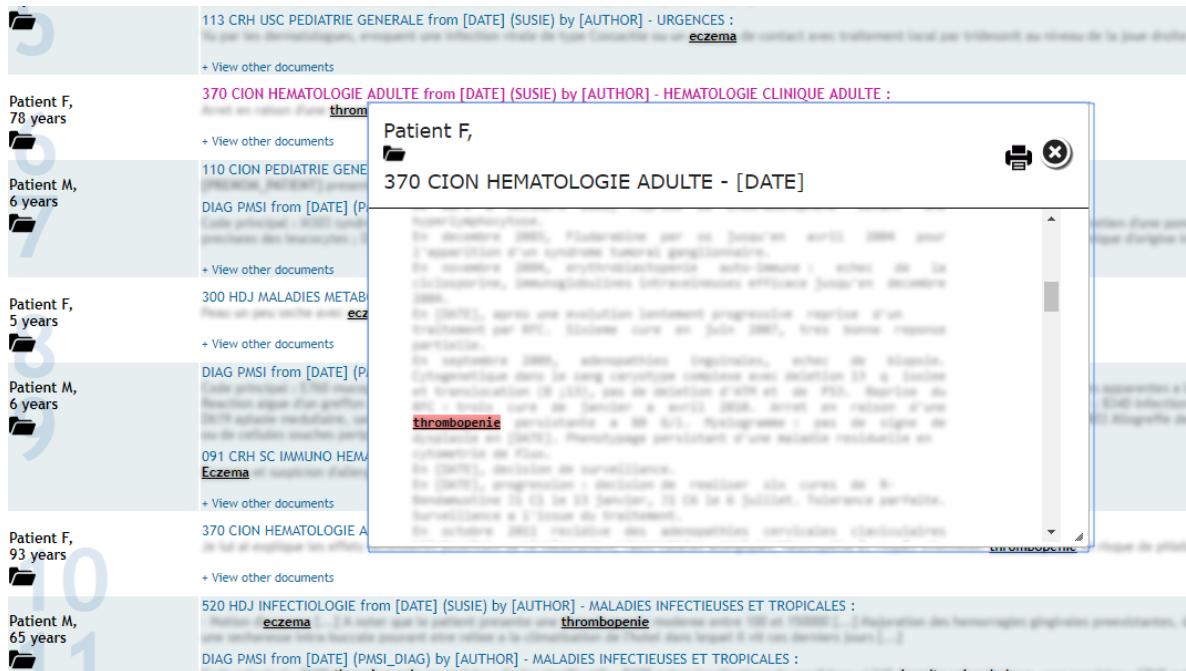
- Save query,
- Search in result,
- Export patients to Excel
- Filter result,
- Share query,
- Populate a cohort

And finally it displays all patients found and authorized for the user.

4.2.1.1. Patient List

Patients are displayed in no particular order. The system only displays the first 50 patients. To display more, click "Show more" at the bottom of the page.

For each patient, you can view the records found by the search engine. An extract of the document containing the search terms is displayed. You can view a document in its entirety by clicking on its title (Figure 11). The terms searched are colored to quickly identify the context of the term. You can open multiple documents and move them by moving the mouse with the left button held down on the header of a document. You can also print a document by clicking 



Patient F, 78 years
Patient M, 6 years
Patient F, 5 years
Patient M, 6 years
Patient F, 93 years
Patient M, 65 years

370 CION HEMATOLOGIE ADULTE from [DATE] (SUSIE) by [AUTHOR] - HEMATOLOGIE CLINIQUE ADULTE :

Patient F,
Patient F,
Patient F,
Patient F,
Patient F,
Patient M,

370 CION HEMATOLOGIE ADULTE - [DATE]

eczema
thrombopenie
dermitis seborrhcea

Figure 11: Open a document in the search engine result

For each patient, the system only displays the first three documents found. To display all documents just click on "+ display other documents"

To access the patient's data in its entirety, just click on the record  next to the name. A new window opens to display the patient's record.

4.2.1.2. Populate a cohort

You can populate an existing cohort or create a new one from a search record.

Create or select a cohort

When the "populate a cohort" button is clicked, a form is displayed allowing you to choose an existing cohort or create a new one (Figure 12).

FEED A COHORT

- Create a cohort

Title :

Share with : Choose users

Description :

ajouter

Or Select an existing cohort to include new patients :

Choose a Cohort

Figure 12: Select a cohort or create a cohort and include patients in it

To create the cohort, click on "create a cohort", specify the title, the people to share it with (if relevant), and a description.

To select an existing cohort, select it in the list of your cohorts.

Include patients in the cohort

Once a cohort is selected, new links and buttons appear (Figure 13).

Result **Cohort : Demo For Similarity** x Stats data Concept Biology Genes Map Clustering

461 patients
3137 Documents

Filtered 1 : Documents containing 'eczema', Excluding negations
Filtered 2 : Documents containing 'thrombopenie', Excluding negations

FEED A COHORT SAVE QUERY SEARCH ON RESULT EXPORT PATIENTS TO EXCEL FILTER THE RESULT BELOW SHARE YOUR QUERY

- Create a cohort

Title :
Share with : Choose users
Description :

ajouter

Or Select an existing cohort to include new patients :

Demo For Similarity x

Opened cohort : Demo For Similarity x

Include all of the following patients in the cohort
 Exclude all patients below from the cohort
 In doubt all patients below
 Import all the patients below into the cohort
 Do not display patients already included or excluded

Patient M,
18 years
090 HDJ IMMUNO HEMATOLOGIE from [DATE] (SUSIE) by [AUTHOR] - IMMUNO-HEMATOLOGIE PEDIATRIQUE :
thrombopenie
+ View other documents

Patient M,
370 CRH HOP SEM HEMATOLOGIE from [DATE] (SUSIE) by [AUTHOR] - HEMATOLOGIE CLINIQUE ADULTE :

Figure 13: Selection of a cohort: New links and buttons available

Access to patients already included / excluded:

A new "cohort" tab appears between "result" and "stat data". This tab will display the patients included, excluded, queried and pending.

Block processing of all patients found:

In the result tab, between the buttons and the list of patients found you can:

- Include in the cohort all patients found.
- Exclude from the cohort all patients found.
- Query all patients found (pending clarification)
- Import all patients found to include them later

Clicking on one of these links will apply the action to all patients found. This will modify patients already included, excluded, queried or imported in the cohort and must therefore be used with caution.

Manual processing of patients found:

For each patient found, 4 new buttons are displayed (Figure 14):

- : to include the patient in the cohort. The line turns green.
- : to exclude the patient from the cohort. The line turns red.
- : to query the patient. The line turns gray.
- : to add a comment about the patient. If a patient is not included, excluded or queried, the patient is included in the "imported" patients for this cohort. You can delete your comments. They will be visible to collaborators with whom you share the cohort.

The screenshot shows a list of patients with their names, ages, and medical records. Each record includes a set of four icons for cohort management. Below the list is a section titled 'Comments in the cohort' containing a text input field and a 'Add' button.

Patient M, 18 years	090 HDJ IMMUNO HEMATOLOGIE from [DATE] (SUSIE) by [AUTHOR] - IMMUNO-HEMATOLOGIE PEDIATRIQUE : <u>thrombopenie</u>
Patient M, 58 years	370 CRH HOP SEM HEMATOLOGIE from [DATE] (SUSIE) by [AUTHOR] - HEMATOLOGIE CLINIQUE ADULTE : <u>thrombopenie</u>
Patient M, 36 years	520 HDJ INFECTIOLOGIE from [DATE] (SUSIE) by [AUTHOR] - MALADIES INFECTIEUSES ET TROPICALES : <u>thrombopenie</u>
Patient F, 7 years	091 CRH SC IMMUNO HEMATOLOGIE from [DATE] (SUSIE) by [AUTHOR] - IMMUNO-HEMATOLOGIE PEDIATRIQUE : <u>Thrombopenie</u>
Patient M, 7 years	DIAG PMSI from [DATE] (PMSI_DIAG) by [AUTHOR] - PEDIATRIE GENERALE : <u>thrombopenie</u>
Comments in the cohort : <input type="text"/> <input type="button" value="Add"/>	

Figure 14: Include, exclude, query and comment on a patient for a cohort

On selecting a cohort, the line background color is automatically updated with the patients already imported or excluded in the cohort.

If you repeat a search and the cohort is already selected, the results will automatically be colored with the information of the selected cohort.

Manage and export the patients from the "cohort": XXX" tab

You can quickly access patients you have included, excluded, queried or imported by clicking on the "cohort: XXX" tab.

By clicking on you can then export patients to Excel.

By clicking on you can copy and paste to import these patients directly into Gecko® for example.

To leave a cohort, just close the "cohort XXX" tab.

Manage all your cohorts

In the menu bar, you have the option to access all your cohorts. This page is explained in detail in Chapter 6.

4.2.1.3. Save the query and automatic query

You have the option to save a query and for it to be run automatically every month.

Just specify a title and check the option "Automatically run this query every month" (Figure 15). It will then appear in "my queries" as well as on the home page (

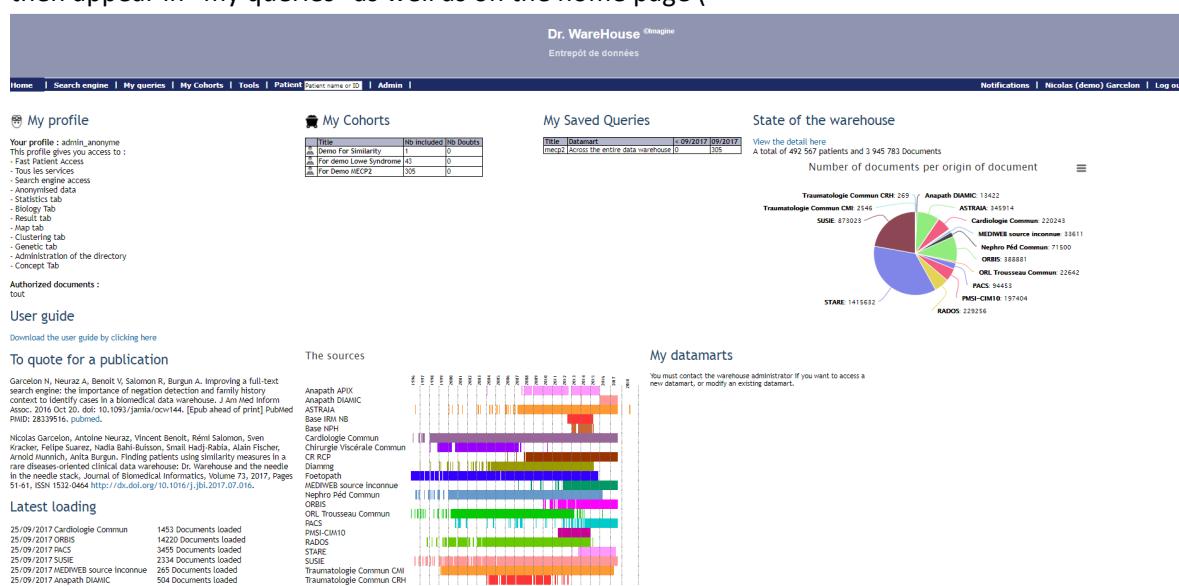


Figure 1).

FEED A COHORT **SAVE QUERY** **SEARCH ON RESULT** **EXPORT PATIENTS TO EXCEL** **FILTER THE RESULT BELOW**

SHARE YOUR QUERY

Save query

Title :

Execute this query automatically every month:

Save

Figure 15: Save a query and automatically rerun it every month

4.2.1.4. Search in result

You have the option of running a full search on the result of a search.

Clicking on the button "search in result" opens a new search engine in a new window. You then have the option to refine your search (Figure 16).

Instead of having "On the patients of your departments" at the top of the search engine, you have "Refine the previous result".

By default, all patients from the previous search are displayed, with all tabs.

You can then use all the features of the search engine solely on the results of the previous query.

Dr. WareHouse ©Imagine
Entrepôt de données

Home | Search engine | My queries | My Cohorts | Tools | Patient Patient name or ID | Admin |

Search for patients

Refine the previous result

Extend to synonyms :
+Advanced - Rewrite the query

1
461 patients
57132 Documents

Filtered 1 :

Result Stats data Concept Biology Genes

+ Add a full text filter
+ Add a structured filter
+ Time constraints
+ Patient filter

START A SEARCH ?

Query history

Search:

Date	Queries
No data available in table	

Patient M,
18 years
Patient M,
58 years
Patient M,
36 years

Base IRM from [DATE] (Base IRM) by [AUTHOR] - RADIOLOGIE PEDIATRIQUE :
350 CION NEUROCHIRURGIE from [DATE] (SUSIE) by [AUTHOR] - NEURO-CHIRURGI
Scanner cérébral avec IV from [DATE] (RADOS) by [AUTHOR] - RADIOLOGIE PEDIAT
CR de consultation from [DATE] (ORBIS) by [AUTHOR] - HEMATOLOGIE CLINIQUE A
370 CION HEMATOLOGIE ADULTE from [DATE] (SUSIE) by [AUTHOR] - HEMATOLOGI
Scanner thoraco-abdominopelvien avec injection from [DATE] (PACS) by [AUTHOR
TDM Thoracique sans IV from [DATE] (RADOS) by [AUTHOR] - RADIOLOGIE ADULTE
DIAG PMSI from [DATE] (PMSI_DIAG) by [AUTHOR] - MALADIES INFECTIEUSES ET T
HEMATOLOGIE ADULTES SCID from [DATE] (KDOS) by [AUTHOR] - HEMATOLOGIE CL
Compte-rendu texte joint HEMATO from [DATE] (STARE) by [AUTHOR] - MALADIES
CR HDJ from [DATE] (ORBIS) by [AUTHOR] :
Scanner thoracique sans injection from [DATE] (PACS) by [AUTHOR] - RADIOLOGIE
CR de radiologie/imagerie from [DATE] (MEDIWEB) by [AUTHOR] - CARDIOLOGIE :
371 CRH ALLO GREFFE SI HEMATOLOGIE from [DATE] (SUSIE) by [AUTHOR] - HEMA

Figure 16: Refine the search of a previous search

4.2.1.5. Export patients to Excel

In the result tab (Figure 17), the "Export patients to Excel" button can be used to export the name, date of birth, Patient hospital ID, and telephone number of the patients found to an Excel spreadsheet.

The screenshot shows a search results interface with various tabs at the top: Result, Stats data, Concept, Biology, Genes, Map, and Clustering. Below the tabs, it displays '348 patients' and '5312 Documents'. A message box lists three filters: 'Filtered 1 : Documents containing 'eczema' , Excluding negations', 'Filtered 2 : Documents containing 'infection%' , Excluding negations', and 'Filtered 3 : Documents containing 'thrombopenie' , Excluding negations'. At the bottom, there are buttons for 'FEED A COHORT', 'SAVE QUERY', 'SEARCH ON RESULT', 'EXPORT PATIENTS TO EXCEL' (which is highlighted with a red box), 'FILTER THE RESULT BELOW', and 'SHARE YOUR QUERY'. Below these buttons, there is a 'Save query' section with fields for 'Title' and 'Execute this query automatically every month'. The main content area shows patient results for 'Patient M, 18 years' and 'Patient M, 58 years', each with a list of documents related to 'INFECTION', 'eczema', and 'infection'.

Figure 17: Export the patients found by the search engine to Excel

4.2.1.6. Filter the result

If you want to do a simple free text search on the patients found, click on "filter the result below". You can enter a few terms (by using AND, OR, NOT and %) to filter the list of patients (Figure 18).

The screenshot shows the same search results interface as Figure 17. The 'FILTER THE RESULT BELOW' button is highlighted with a red box. Below it, there is a 'Filter the result' section with a 'Full text query' input field containing '[eczema infection]'. The main content area shows patient results for 'Patient M, 18 years' and 'Patient M, 58 years', each with a list of documents related to 'INFECTION', 'eczema', and 'infection'.

Figure 18: Quick search on the result of a search.

4.2.1.7. Share the query

If you want to share your query with collaborators so that they can run it for their data (Figure 19). Click on "Share the query". Give a name to this query as well as the recipients.

They will receive a notification. Clicking this pre-fills the search engine. They simply have to run the search.

This feature can also prove useful in the context of a telephone hotline. The hotline can thus create a query and share it directly with the person who asks for help.

The screenshot shows a user interface for sharing a query. At the top, there are several buttons: 'FEED A COHORT', 'SAVE QUERY', 'SEARCH ON RESULT', 'EXPORT PATIENTS TO EXCEL', 'FILTER THE RESULT BELOW', and 'SHARE YOUR QUERY'. The 'SHARE YOUR QUERY' button is highlighted with a red border. Below these buttons, the text 'Share your query' is displayed. Underneath, there is a field labeled 'Name your query:' containing the text 'eczema infection thrombopenie'. Another field labeled 'Select people to share:' contains two names: 'Garcelon Nicolas' and 'Neuraz Antoine', each with a small 'X' icon to the right. At the bottom left of the form is a 'Share' button.

Figure 19: Share a query with one or several collaborators

4.2.2. Stat data tab

The "stat data" tab shows the demographic statistics of the population found (Figure 20): age pyramid on the date of the 1st record found, age pyramid of patients today. These two pyramids are reproduced for children from 0 to 20 years.

The distribution of the number of patients found per year, differentiating patients already seen in previous years.

The distribution of the number of documents found per data source.

The distribution of the number of documents found per department, differentiating patients already seen in another department at an earlier date.

At the top of the age pyramid, there are two links allowing you to view the care pathway of the patients found.



Figure 20: display of the descriptive statistics of patients found by the search engine

4.2.3. Concepts Tab

This tab allows you to view the record's extracted concepts.

By default, it displays the extracted concepts of the records containing the search criteria. If you want to display the extracted concepts from all patients' records, simply click "on all documents of patients found". The concepts are refreshed automatically.

There are 3 parts to the display (Figure 21):

- A concept cloud, the size of which depends on the frequency of the concept
- A table listing all concepts and calculating different scores:
 - o Number of patients
 - o % of patients with regard to the results
 - o The TF-IDF score of concepts based on the frequency of the concept in the results over the frequency of documents in the data warehouse containing this concept
 - o The PSS (Phenotypic Specificity Score) which gives the number of patients with the concept in the result over the total number of patients in the warehouse who have this concept.
 - o The Case Weighted PSS, which modulates the PSS according to the number of patients with this concept.

The table can be sorted by clicking on the column headings. It can also be filtered using a free text search field.

- A graph using the 3 scores from the table

Search for patients

Across the entire data warehouse

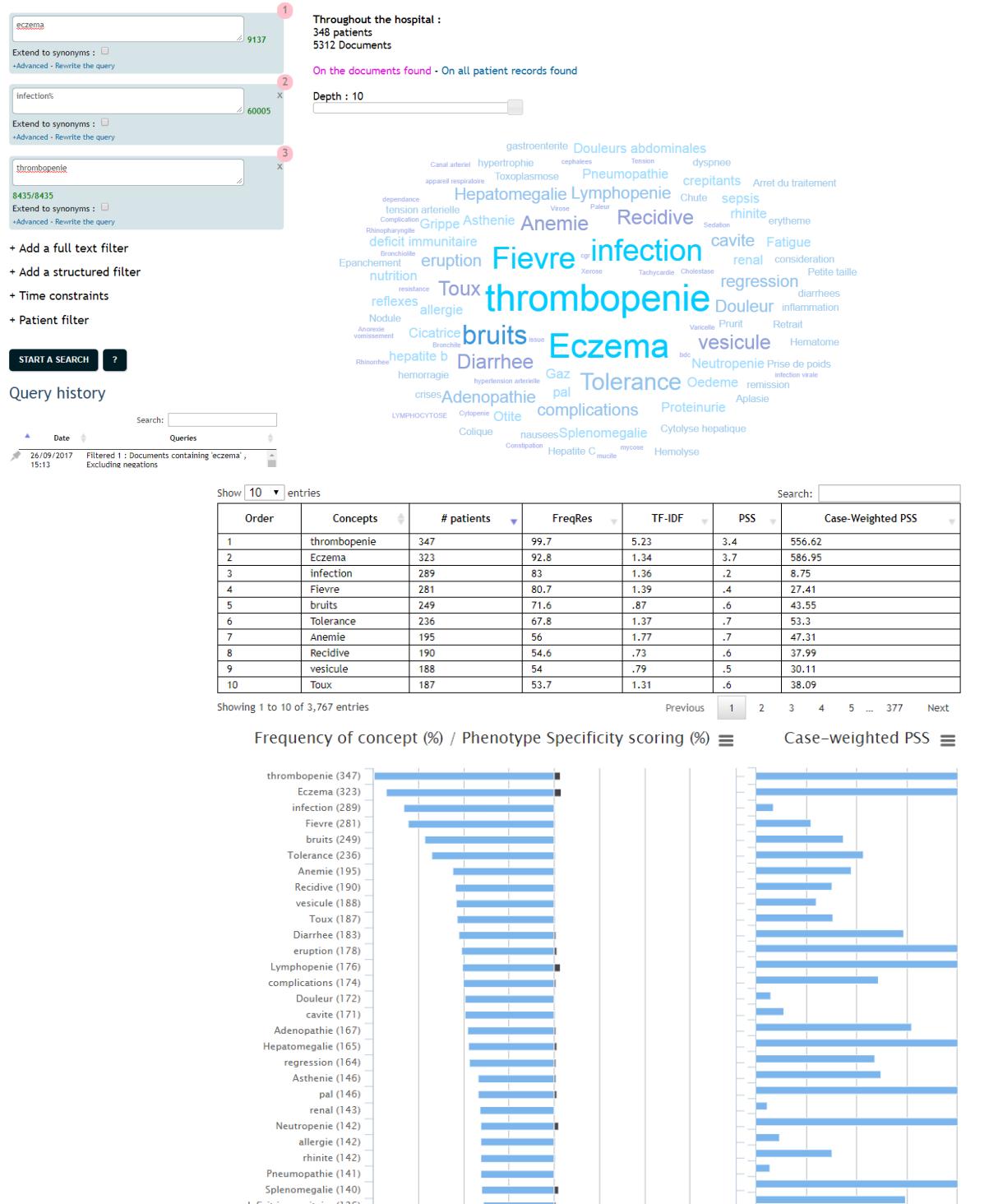


Figure 21: Automatic extraction of concepts contained in the records - sort by relevance

4.2.4. Biology tab

This tab allows you to view the aggregated biological results of all patients found.

By default, it displays all the aggregated tests for all patients found (Figure 22): number of patients, laboratory test count, mean, median, standard deviation, number of values above the upper limit, %

of values above the upper limit, number of values below the lower limit, % of values below the lower limit.

You can select a laboratory test and display the list of patients with the aggregated data for each patient: laboratory test count, mean, median, standard deviation, number of values above the upper limit, % of values above the upper limit, number of values below the lower limit, % of values below the lower limit.

The tables can be sorted by clicking on the column headings.

Exams	Nb patients	Nb exams	min	max	Median	Average	Standard deviation	Nb val > upper bound	% Upper bound	Nb val < lower bound	% lower bound	Trend
1 / creatinémie (μmol/l) / μmol	83	1548	0	.13	.02	0	0					
1,25 dihydroxychalciferol (ng/ml) Serum Radiimmunoanalyse	73	188	6	230	55.5	56.4	35.4	16	9	12	6	
25 hydroxycholecalciferol (ng/ml) Serum/Plasma Radiimmunoanalyse	125	431	4	210	29	32.7	20.1	33	8	219	51	
3 hydroxybutylique acide (aceto-perchlorate) mmol/l Sang	4	7	.07	.36	.12	.2	.1					
Acide urique Plasma	33	72	1.1	151.6	18.49	27.6	29.6	9	13	21	29	
Acéto-perchlorique acide mmol/l Sang	6	13	.04	.24	.07	.1	.1					
ADN natif anticorps Test de Fair UI/ml Sérum Radiimmunoanalyse	48	75	1.3	77.4	3.8	6.6	10.6					
Adénovirus ADN Titrage Copier/mg Sang PCR quantitative/Argene	5	38	.677	25690000	841000	3286861.1	5525299.4	38	100			
Adénovirus ADN Titrage log copier/ml Sang PCR quantitative/Argene	5	38	2.8	7.4	5.95	5.4	1.5	38	100			
AIE 75 anticorps cpn Serum Radiglobin assay	6	14	.21	3895	194	920.9	1424.6	7	50			
Alanine aminotransférase UI Serum/Plasma	250	8237	0	4051	34	72.6	180.2	3192	39	60	1	
Albumine /g Sérum Electrophorèse 2	6	7	46.1	59.3	56.7	7						
Albumine /g Sérum Electrophorèse 2	63	253	5.8	59.6	36.3	36	6.8	5	2	100	40	
Albumine /g Sérum Electrophorèse 2	6	7	30.9	53.9	41.1	41.8	7.2					
Albumine /g Sérum/Plasma	187	496	6.7	55.2	32.8	32.7	6.7	87	2	2496	50	
Albumine/protéines totales % Sérum/Plasma Electrophorèse	110	569	21.1	52.3	39.3	38.2	5.2	12	2	299	53	
Albumine/protéines totales % Sérum/Plasma Electrophorèse	110	577	32.2	89.6	59.1	58.3	7.3	53	9	171	30	
Alentzuromyg /μg Sérum Immunoenzymologie	6	31	11.2	48.5	18.9	21.2	8.6					
Alpha 1 globulines /g Sérum Electrophorèse 2	6	7	21.2	2.5	2.6	.5						
Alpha 1 globulines /g Sérum Electrophorèse 2	62	253	.8	44.4	3	2.1	.6					
Alpha 1 globulines /g Sérum Electrophorèse 2	6	7	1.3	2.5	1.9	1.9	.5					
Alpha 1 globulines /g Sérum/Plasma Electrophorèse	110	568	.9	8	3.2	3.3	.9	202	36	14	2	
Alpha 1 globulines /protéines totales % Sérum/Plasma Electrophorèse	111	569	2.2	15.9	4.8	5.2	1.8	253	44	19	3	
Alpha 2 globulines /protéines totales % Sérum Electrophorèse 2	6	7	8.5	14.3	10.7	10.9	2.3	4	57			

Figure 22: Biological results of patients found by the search engine

4.2.5. Map Tab

This tab allows you to view patients on a map of the world. They are geolocated by their zip code only. There are two displays available: Google Map and a slimmed down map.

4.2.6. Clustering Tab

This tab allows you to create sub-groups of patients from concepts extracted from their records. The user can specify the minimum similarity distance between two patients so that they are connected.

The result is represented in 2 forms: a network and a table.

For the network graphic representation (Figure 23), the nodes are patients and the links between nodes represent the similarity between 2 patients. The thickness of the link is calculated according to the similarity distance between two patients. Nodes can be clicked on to access the patient record. Links can be clicked on to view the list of concepts in common between two patients. The size of the concept represents the weight of the concept in the similarity calculation (TF-IDF score).

Under the network, the system displays a table representing patient groups and concepts in common for each group.

It is up to the user to vary the minimum distance (in %) to get a relevant representation allowing new associations to be detected

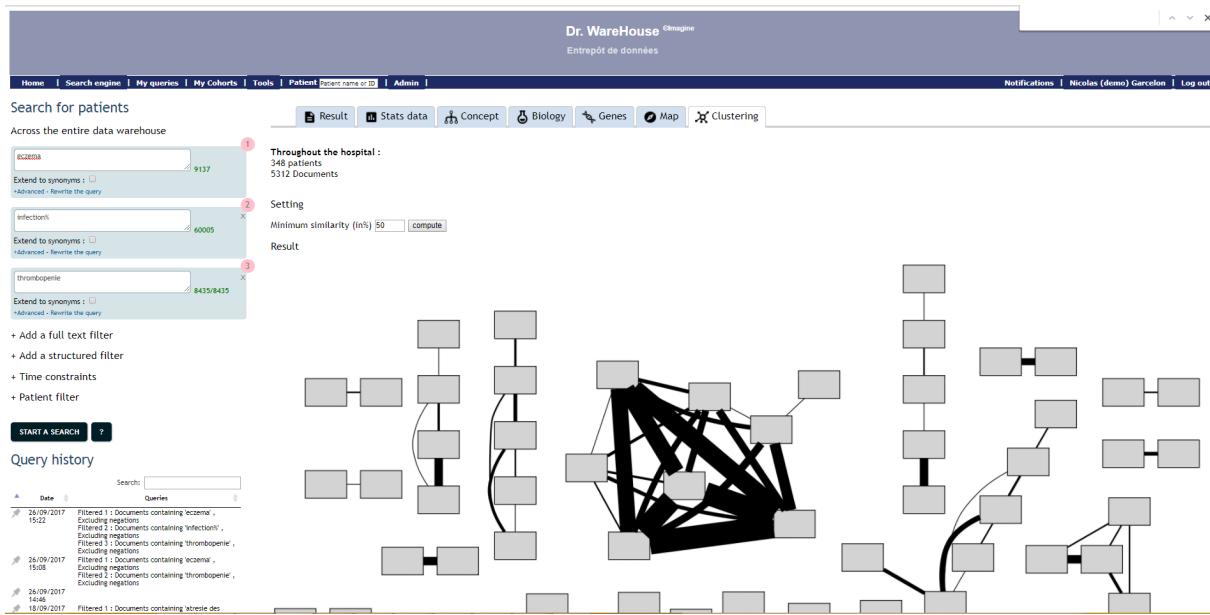


Figure 23: Perform a clustering from the search results

4.2.7. Export structured data tab (under development)

Under construction

From patients found, the user will be able to select the structured data of interest. The export will have 11 columns:

Patient hospital ID, name, date of birth, laboratory test code, laboratory test title, numerical result of the laboratory test, textual result of the examination, lower limit, upper limit, test date.

5. A patient's record

A patient's record is accessible in several different ways:

- Directly from the menu bar: when the user starts typing a patient's surname and first name or vice versa, the system offers a patient list from which the patient can be selected using the arrows on the keyboard or the mouse (Figure 24). If no patient matches, the system will search for patient names with a similar spelling. You can also use the patient's hospital ID to display this.

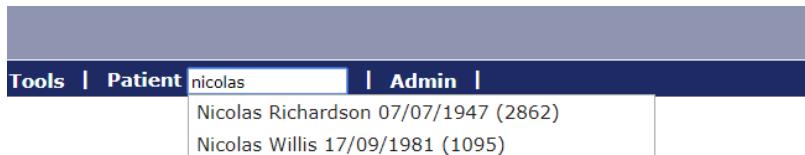


Figure 24: Quick search in the menu bar from the surname and start of the first name or vice versa

- By clicking on this icon
- By clicking on the network graph nodes that represent patients.

A patient record is represented with tabs (Figure 25):

- Documents
- Biological results
- Timeline
- Care pathway
- Extracted concepts
- Similarity

Richardson Nicolas 07/07/1947 M (2862)
70 years

Documents

Biology

TimeLine

Journey

Cohort

Concepts

Similarity

SEARCH

23 Documents found 

pubmed American journal of physiology. Renal physiology Dr Gattone Vincent H	21/04/2014
pubmed The Journal of clinical investigation Dr Fedele Sorin V	05/04/2011
pubmed Clinical nephrology Dr Ling Galina	13/11/2009
pubmed PloS one Dr Eisenberger Tobias	29/07/2007
pubmed American journal of physiology. Renal physiology Dr Hopp Katharina	18/01/2007
pubmed Kidney international Dr Tao Shixin	17/11/2006
pubmed Journal of pediatric gastroenterology and nutrition Dr Maditz Kaitlin H	04/11/2006
pubmed Veterinary pathology Dr Herder V	06/07/2005
pubmed Molecular and cellular probes Dr Herbst S M	22/12/2003
pubmed Gene Dr Tavira Beatriz	11/10/2002
pubmed Biochemistry and cell biology = Biochimie et biologie cellulaire Dr Hu Qiaolin	16/09/2002
pubmed Pathology, research and practice Dr LÃ³pez JosÃ© I	05/01/2001

Figure 25: a patient record - overview

5.1. Documents

The document tab allows you to view all the patient's documents, other than laboratory records. The documents are sorted by descending order.

At the top of the list of documents, the  icon allows you to open all records one below the other and print them if necessary.

5.1.1. View the records

When a document title in the list on the left is clicked, the document in question opens up to the right. The  icon next to the document title allows you to print the individual document (Figure 26).

Dr. WareHouse ©Imagine
Entrepôt de données

[Home](#) | [Search engine](#) | [My queries](#) | [My Cohorts](#) | [Tools](#) | [Patient](#) | [Admin](#) |

Richardson Nicolas 07/07/1947 M (2862)
70 years

[Documents](#) [Biology](#) [TimeLine](#) [Journey](#) [Cohort](#) [Concepts](#) [Similarity](#)

SEARCH

23 Documents found

pubmed American journal of physiology. Renal physiology Dr Gattone Vincent H	21/04/2014	Veterinary pathology, by Dr Herder V, the 06/07/2005
pubmed The Journal of clinical investigation Dr Fedele Sorin V	05/04/2011	Polycystic kidneys and GM2 gangliosidosis-like disease in neonatal springboks (<i>Antidorcas marsupialis</i>). Clinical, gross, histopathologic, electron microscopic findings and enzymatic analysis of 4 captive, juvenile spr
pubmed Clinical nephrology Dr Ling Galina	13/11/2009	34; 12%) were affected by either one or both disorders in a German zoo within a period of 5 years (2008-2013). Macr
pubmed PLoS one Dr Eisenberger Tobias	29/07/2007	brachygnathism in 2 animals. Histopathologically, kidneys of 4 animals displayed cystic dilation of the renal tubul
pubmed American journal of physiology. Renal physiology Dr Hopp Katharina	18/01/2007	peripheral nervous system, hepatocytes, thyroid follicular epithelial cells, pancreatic islets of Langerhans and r
pubmed Kidney international Dr Tao Shixin	17/11/2006	electron-lucent vacuoles, up to 4 μ m in diameter, were present in neurons. Enzymatic analysis of liver and kidney
pubmed Journal of pediatric gastroenterology and nutrition Dr Maditz Kathrin H	04/11/2006	activity at the same level of total Hex, suggesting a hexosaminidase defect. Pedigree analysis suggested a monogen
pubmed Veterinary pathology Dr Herder V	06/07/2005	polycystic kidney and a GM2 gangliosidosis similar to the human Sandhoff disease. Whether the simultaneous occurre
pubmed Molecular and cellular probes Dr Herbst S M	22/12/2003	
pubmed Gene Dr Tavira Beatriz	11/10/2002	
pubmed Biochemistry and cell biology = Biochimie et biologie cellulaire Dr Hu Qiaolin	16/09/2002	
pubmed Pathology, research and practice Dr López José I	05/01/2001	
pubmed Case reports in urology Dr Hazra Anup	12/07/2000	
pubmed PLoS one Dr DrAfgemÁviller Michaela	19/01/2000	
pubmed Experimental and toxicologic pathology : official journal of the Gesellschaft für Toxikologische Pathologie Dr Shoieb Ahmed	24/11/1999	
pubmed Current opinion in pediatrics Dr Hoyer Peter F	08/11/1999	
pubmed Nephron. Clinical practice Dr Riella Cristian	30/03/1999	
pubmed Journal of pediatric genetics Dr Guay-Woodford Lisa M	21/03/1999	
pubmed Pediatric research Dr Pavlov Tengis S	12/11/1997	
pubmed Nephron Dr Le Corre StÅspanhe	11/11/1997	
pubmed Archives of gynecology and obstetrics Dr Banks Nicole	12/02/1997	
Liver transplantation : official publication of the American		

Figure 26: patient record - view the records

5.1.2. Search the records

You can search through the patient's documents by using the search engine at the top of the list of patients. The search engine is the same as for the general search engine, with the use of AND, OR and NOT. The documents are then filtered with an extract that contains the search terms (Figure 27). The search is automatically performed on text enriched by terminological enrichment.

The screenshot shows the Dr. WareHouse interface. At the top right, it says "Dr. WareHouse ©Imagine" and "Entrepôt de données". Below the header is a navigation bar with links: Home, Search engine, My queries, My Cohorts, Tools, Patient (with a sub-link "Patient name or ID"), Admin. The main content area displays patient information: "Richardson Nicolas 07/07/1947 M (2862)" and "70 years". Below this are several tabs: Documents, Biology, TimeLine, Journey, Cohort, Concepts, and Similarity. A search bar contains the term "neonatal" and a "SEARCH" button. Below the search bar, it says "3 Documents found" and shows a list of three results.

Journal of pediatric genetics, by Dr Guay-Woodford Lisa M, the 21/03/1999

Autosomal recessive polycystic kidney disease: the prototype of the hepato-renal fibrocystic diseases.

Autosomal recessive polycystic kidney disease (ARPKD) is a severe, typically early onset form of renal cystic disease, hypertension and progressive renal insufficiency. However, the disease has multisystem manifestations and a comprehensive approach is required. Clinical features include hepatomegaly, portal hypertension, and ascites. Renal manifestations include bilateral renal tubular dysgenesis, renal tubular acidosis, and progressive renal failure. Other manifestations include liver cirrhosis, portal hypertension, and ascites. Management includes dialysis, liver transplantation, and other interventions. The prognosis is generally poor, with most patients progressing to end-stage renal disease and death within 10-20 years of diagnosis.

Autosomal recessive polycystic kidney disease (ARPKD) is a severe, typically early onset form of renal cystic disease, hypertension and progressive renal insufficiency. However, the disease has multisystem manifestations and a comprehensive approach is required. Clinical features include hepatomegaly, portal hypertension, and ascites. Renal manifestations include bilateral renal tubular dysgenesis, renal tubular acidosis, and progressive renal failure. Other manifestations include liver cirrhosis, portal hypertension, and ascites. Management includes dialysis, liver transplantation, and other interventions. The prognosis is generally poor, with most patients progressing to end-stage renal disease and death within 10-20 years of diagnosis.

Autosomal recessive polycystic kidney disease (ARPKD) is a severe, typically early onset form of renal cystic disease, hypertension and progressive renal insufficiency. However, the disease has multisystem manifestations and a comprehensive approach is required. Clinical features include hepatomegaly, portal hypertension, and ascites. Renal manifestations include bilateral renal tubular dysgenesis, renal tubular acidosis, and progressive renal failure. Other manifestations include liver cirrhosis, portal hypertension, and ascites. Management includes dialysis, liver transplantation, and other interventions. The prognosis is generally poor, with most patients progressing to end-stage renal disease and death within 10-20 years of diagnosis.

Figure 27: search for the term "biopsy" in all the records of a patient

5.1.3. Extract data from records: weight, height, etc.

By using regular expressions, you can extract data which is written in the same way in the record. For example, weight will always be displayed as a series of figures + a period or a comma or nothing + the figures after the comma or none + one space or not + the unit in kg

These patterns can be represented as a regular expression:

We can therefore translate the weight of a patient by: [0-9]+[, .]?[0-9]* ?kg

For a detailed explanation of regular expressions, see:

https://fr.wikipedia.org/wiki/Expression_r%C3%A9gul%C3%A8re

By adding parenthesis in the regular expression, Dr. Warehouse will be able to specifically extract them (Figure 28):..... ([0-9]+[, .]?[0-9]*) ?kg.....

Then just copy and paste the table to Excel.

The same exercise can be done with height, blood pressure etc.

Patient M,
20 years

Documents Biology TimeLine Journey Cohort Concepts Similarity

..... ([0-9]+[,]?)?([0-9]*)>kg.....

SEARCH

Select table 16 Documents found 

[DATE]	Poids : 40.8 kg ; T°	40.8
[DATE]	Poids : 44 Kg. Abdo	44
[DATE]	poids de 5 kg . Les	5
[DATE]	poids = 45 kg. Tail	45
[DATE]	Poids : 37 kg Se	37
[DATE]	: poids 43,1 kg, tail	43,1
[DATE]	m, P : 43 kg.. Cus	43
[DATE]	: poids 41 kg, tail	41
[DATE]	Poids : 36 Kg 700.	36
[DATE]	, Poids 38kg (+ 1,	38
[DATE]	ns, O+, 30 kg, dece	30
[DATE]	et pese 38 kg, soit	38
[DATE]	, poids 31 kg ; tai	31
[DATE]	et pese 30.9 kg.	30.9
[DATE]	(+1) et pese 30kg... .	30
[DATE]	BEG P= 30kg (M)	30

Figure 28: patient record - automatic extraction of the patient's weight over time in the records

5.2. Biological results

This tab allows you to view the biological results as a record (Figure 29) or as a table containing all the results over time (Figure 30).

A search engine also allows you to search documents with free text queries.

Dr. WareHouse ©Imagine
Entrepôt de données

[Home](#) | [Search engine](#) | [My queries](#) | [My Cohorts](#) | [Tools](#) | [Patient](#) Patient name or ID | [Admin](#) |

Patient M,
36 years

Documents Biology TimeLine Journey Cohort Concepts Similarity

Documents - Table - curves

SEARCH

280 Documents found

Document Title	Date
Compte-rendu texte joint HEMATO	19/09/2017 STARE
Compte-rendu texte joint VIROLOGIE	19/09/2017 STARE
Compte-rendu texte joint BIOCHA	19/09/2017 STARE
Compte-rendu texte joint VIROLOGIE	08/09/2017 STARE
Compte-rendu texte joint HEMATO	08/09/2017 STARE
Compte-rendu texte joint BIOCHA	08/09/2017 STARE
Compte-rendu texte joint BIOCHA	31/08/2017 STARE
Compte-rendu texte joint VIROLOGIE	31/08/2017 STARE
Compte-rendu texte joint BIOCHA	16/08/2017 STARE
Compte-rendu texte joint VIROLOGIE	26/07/2017 STARE
Compte-rendu texte joint EXPLO	26/07/2017 STARE
Compte-rendu texte joint IMMUNOA	26/07/2017 STARE
Compte-rendu texte joint BIOCHA	26/07/2017 STARE
Compte-rendu texte joint BIOCHA	26/07/2017 STARE
Compte-rendu texte joint BACTERIOLOGIE de 06H30 à 18H00	17/07/2017 STARE
Compte-rendu texte joint BIOCHA	17/07/2017 STARE
Compte-rendu texte joint BACTERIOLOGIE de 06H30 à 18H00	17/07/2017 STARE
Compte-rendu texte joint BACTERIOLOGIE de 06H30 à 18H00	03/07/2017 STARE
Compte-rendu texte joint HEMATO	03/07/2017 STARE
Compte-rendu texte joint BACTERIOLOGIE de 06H30 à 18H00	03/07/2017 STARE
Compte-rendu texte joint BACTERIOLOGIE de 06H30 à 18H00	03/07/2017 STARE

Compte-rendu texte joint VIROLOGIE , by [AUTHOR], the [DATE]

ASSISTANCE PUBLIQUE HOPITAUX DE PARIS Necker-Enfants malades

POLE DE BIOLOGIE ===== SERVICE : H01 INFECTIOLOGIE AD H3

DEMANDE N° : 1709M190167

Identite : [LASTNAME] [MAIDEN_NAME] [FIRSTNAME]
NIP : [HOSPITAL_PATIENT_ID] NDA : 611769497

Ne(e) le : [36 years]

Sexe : M

Edition du : [DATE] a 12:40

Analyse M190167 Val. Ref. Unites

10:30

MICROBIOLOGIE SEROLOGIE PARASITAIRE

Nature de prelevement : SERUM (Tube sec)

Recherche Aspergillaire

Depistage Antigene 0.45 <0.50 Index

Recherche par technique ELISA

Resultat Glucane *(s38A18.00) <80.00 pg/ml

Recherche d'antigene fongique par technique colorimetrique

Interpretation Presence d'un taux eleve d'antigene betaglucane dans le serum. La detection de beta glucanes circulants dans le serum peut-etre le temoin d'une infection fongique invasive. ATTENTION : Ce

Figure 29: patient record - biological results as a document list

Dr. WareHouse ©Imagine
Entrepôt de données

[Home](#) | [Search engine](#) | [My queries](#) | [My Cohorts](#) | [Tools](#) | [Patient](#) Patient name or ID | [Admin](#) | Notifications | Nicolas (demo) Garcelon | Log out

Patient M,
36 years

Documents Biology TimeLine Journey Cohort Concepts Similarity

Documents - Table - curves

	10-2013	11-2013	12-2013	01-2014	02-2014	03-2014	04-2014	06-2014	07-2014	08-2014	09-2014	10-2014	11-2014	12-2014	
Exams	31	21	12	09	30	18	13	03	24	10	02	22	19	10	25
Antécédent résistance (mg/l) Sérum/Plasma immunoanalyse															
Leucocytes (/ml) Urine Comptage automatique															
BACT HÉMOCULTURE															
Résultat des tests bactérien FA															
URINE															
Résultat des tests bactérien FA															
GLAIS LIPIQUE															
Cholestérol HDL (mmol/l) Sérum/Plasma	0,95	1	1,05	0,89	0,91	0,85	0,87	0,93	1,06	0,96	0,88	0,96			
Cholestérol HDL cholestérol LDL cholestérol VLDL Plasma Calcifiée															
Cholestérol total (mmol/l) Sérum/Plasma Calcifiée	2,31	2,15	2,32	1,77	1,93	1,73	2,29	1,88	2,51	1,93	1,93	1,77			
Triglycérides (mmol/l) Sérum/Plasma															
2,14															
GLAIS MARTIAL															
Par rapport au taux Plasma	7	8	7	12	10	7	6	8	7	7	7	6			
Par capacité de fixation (umol/l) Sérum/Plasma	61	64	64	67	68	62	66	72	68	61	64				
Ferritine (ug/l) Sérum	70	42	50	33	46	39	70	25	49	73	71				
Transferrine (g/l) Sérum/Plasma	2,44	2,57	2,57	2,84	2,73	2,46	2,66	2,88	2,72	2,45	2,55				
Transferrine (g/l) Sérum/Plasma taux de saturation Sérum/Plasma															
CHEMIE DES URINES															
Calcium (mmol/l) Urines	1,15		2,32												
Chlorures (mmol/l) Urines	72		241												
Creatinine estymatique (mmol/l)															
Glucosurie (mmol/l) Urines															
Magnétisme (mmol/l) Urines	11,42		3,8												
Osmosurie (mOsmol/kg) Urines Calcifiée	381		1089												
Proteines (g/l) Urines															
Phosphates (mmol/l) Urines															
Potassium (mmol/l) Urines	29		108												
Proteines (g/l) Urines															
Ratio albumine/globuline (g/L) cétoacéturie (mmol/l) (mg/mmol)															
Sodium (mmol/l) Urines	57		194												
Urine (immol/l) Urines	189		488												
Urea (immol/l) Urines			125												
Urea (mmol/L) Urines			124												
URINE DU SANG															

Figure 30: patient record - biological results in table form

5.3. Timeline

This tool allows you to view the hospitalizations and records of the patient over time (Figure 31). You can use the mouse thumbwheel to zoom in or out in order to have more or less detail over time.

The timeline is divided into 3 horizontal parts, allowing you to have a summarized and more macroscopic view of the life line. The lower areas can be used to move the life line from left to right with the mouse.

Hospitalizations and documents are clickable and allow you to view records in a new window.

The search engine allows you to highlight the areas containing the search terms. When you click on the part in the timeline, the document is opened and the term will be highlighted.

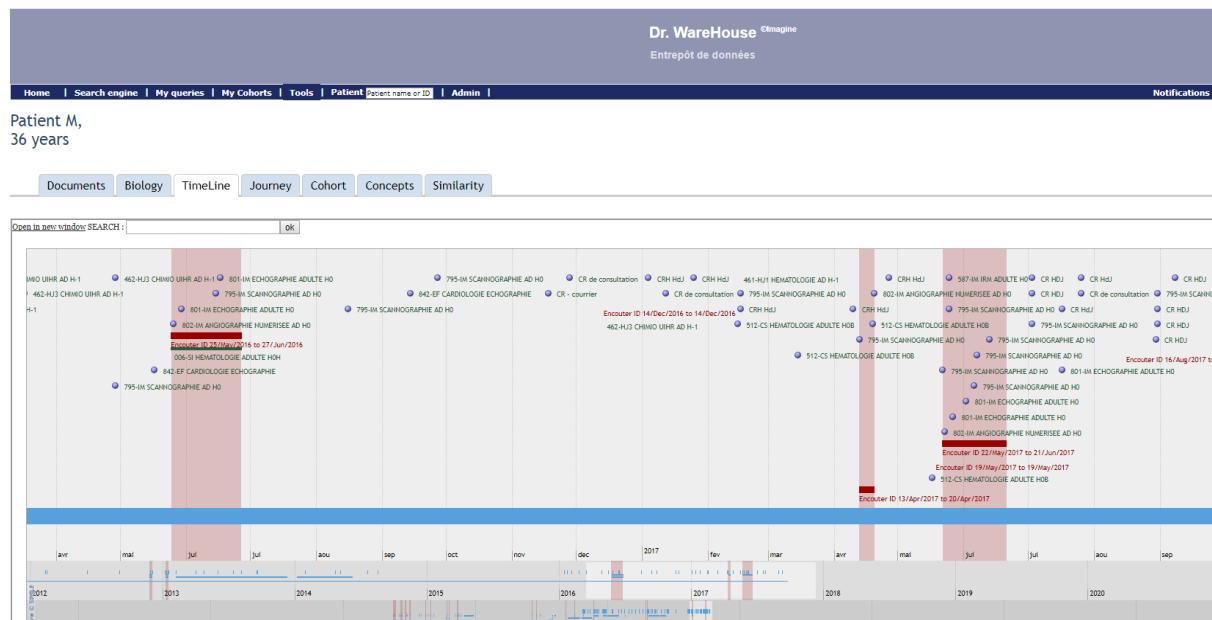
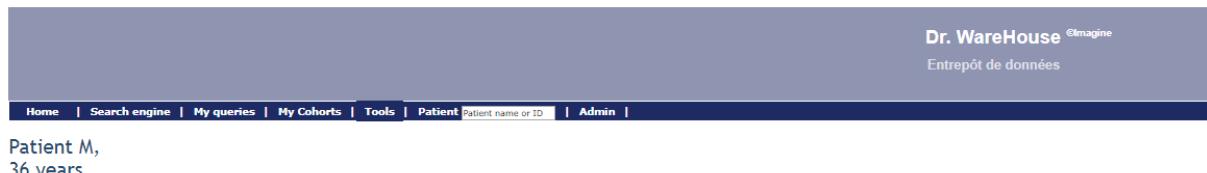


Figure 31: Timeline of a patient

5.4. Patient care pathway

This tab allows you to view all the patient's movements in the hospital (Figure 32). There are three levels of viewing:

- hospitalizations at department level
- hospitalizations at hospital unit level
- hospitalizations and consultations at hospital unit level



Journey

Show hospital journey per hospital department - Show hospital journey per unit - Show all movement (including consultation)

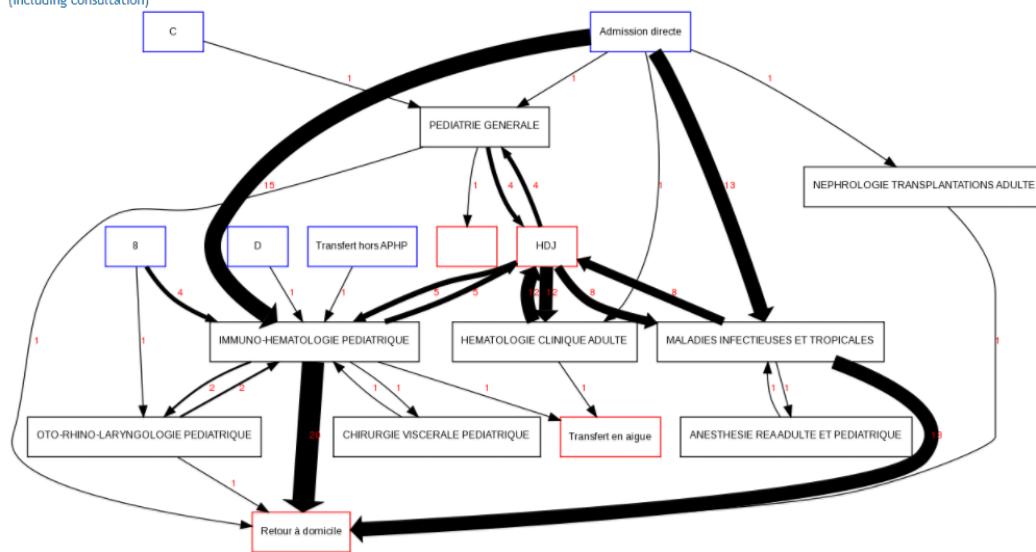


Figure 32: a patient's care pathway, at hospital department level

5.5. Cohort

The cohort tab (Figure 33) lists the cohorts in which the patient is included, excluded, queried or only imported. The system specifies which cohorts include the patient and on what date.

You can also include, exclude or query the patient in one of your cohorts by using the drop-down menu listing your cohorts.

Patient M,
20 years

Documents Biology TimeLine Journey Cohort Concepts Similarity

Cohorts of this patient

- Patient Excluded the 27/10/2015 16:11 by Nicolas Garcelon In the cohort **MICI et Hemato (>= 18 ans)** created by Nicolas Garcelon
- Patient included the 20/01/2016 18:55 by Nicolas Garcelon In the cohort **MICI et Hemato pédiatrie <=18 ans** created by Nicolas Garcelon
- Patient included the 08/02/2017 17:37 by Antoine Neuraz In the cohort **choc septique vivant** created by Antoine Neuraz
- Patient included the 03/06/2017 12:05 by Nicolas Garcelon In the cohort **eczema et thrombopenie** created by Nicolas Garcelon

Include / Exclude this patient from a cohort

Select a cohort :   

Figure 33: patient record - list of cohorts in which the patient appears

5.6. Extracted concepts

Medical concepts (medical term in conjunction with an official thesaurus) are automatically extracted from patients' records (Figure 34). This tab only displays concepts regarded as non-negative (the concept diabetes will not be recorded if it is used in a clause such as "absence of diabetes") and for which the context is the patient (if it is used for a relative, it will not be recorded either).

The table listing the concepts displays their frequency for the patient, and their TF-IDF score.

Concepts

Simplified Display - Display for validation

Concepts	Nb documents	TF-IDF doc	Patient TF-IDF
deficit immunitaire	157	22.74	0.3
dilatation des bronches	132	18.8	0.17
Aspergillose	103	11.31	0.23
Pneumopathie	98	20.05	0.12
infection	84	6	0.05
Alternance	79	6.12	0.11
Lymphopenie	78	7.48	0.12
Tolerance	75	7.22	0.06
Hypogammaglobulinemie	67	4.61	0.08
Lymphome de Hodgkin	65	4.96	0.13
Hepatosplenomegalie	56	3.73	0.07
azoospermie	55	3.38	0.11
ulcere duodenal	53	3.24	0.09
CHOP	52	5.38	0.11
Surinfection	51	4.16	0.05
CD20	51	4.07	0.06
lymphome	51	3.89	0.07
Splenomegalie	51	7.58	0.06

En attente de nick-imagine.nck.aphp.fr...

Figure 34: patient record - list of concepts extracted from records

5.7. Similarity calculation

The similarity calculation allows the most similar patients to be found from extracted medical concepts. The similarity distance is calculated using the Vector Space Model (VSM) method. Patients are projected in a vector space, the dimensions of which are represented by the extracted concepts. The distance between two patients is the cosine of the angle formed by the two patients' vectors.

It is possible for the user to change some settings in the options:

- Define a sub-group of patients in which to search for similar patients
- Exclude patients who have certain diagnoses
- Exclude patients contained in a cohort

The result is given in three forms:

- a graph representing the index patient and the 20 most similar patients (Figure 35). The links between patients define the similarity distance between them. Clicking on a patient opens the record. Clicking on a link displays the concepts in common between the two related patients.
- A table containing the 20 most similar patients (Figure 35), where the middle column lists the concepts in common between the similar patient and the index patient, and the right column

lists concepts only present in the similar patient and missing from the index patient. The size of the concepts is calculated according to their relevance (TF-IDF).

- A table containing the concepts missing from the index patient (Figure 36). These concepts are aggregated. Different scores are calculated and allow the concepts to be sorted by clicking on the column headers:
 - o Number of concepts
 - o Number of patients with the concept
 - o % Number of patients with the concept / number of patients in the warehouse with this concept
 - o Sum of TF-IDF scores
 - o TF-IDF scores divided by the number of patients with the concept

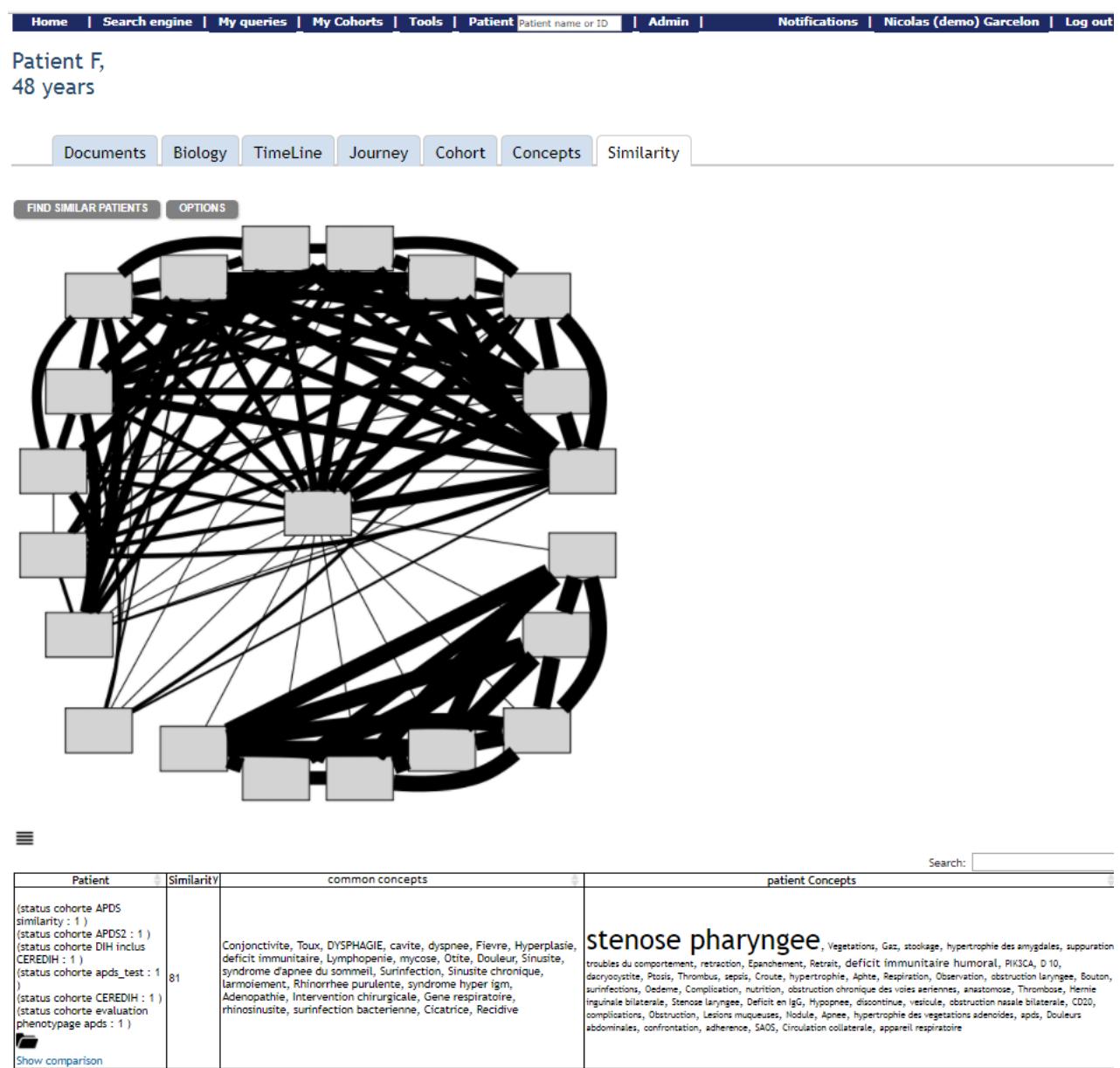


Figure 35 - patient record - similarity calculation - the most similar patients

Concepts	nb concepts	nb pat	% nb pat / nb patient entrepot	Somme TF-IDF	100 * TF-IDF / NB pat	100 * TF-IDF * NB pat / NB_PAT_ENTREPOT
apds	144	6	20	2.572	43	51.44
stenoze pharyngée	20	1	13	0.736	74	9.2
PIK3CD	49	3	12	0.619	21	7.428
pi3k	31	5	5	0.705	14	3.71053
SRSF2	12	1	6	0.546	55	3.03333
obstruction chronique des voies aériennes	3	1	20	0.115	12	2.3
oesophagite de grade ii	20	1	4	0.604	60	2.23704
Conjonctivite chronique	45	2	3	0.679	34	2.22623
telangiectasie conjonctivale	2	1	33	0.066	7	2.2
PIK3R1	4	2	18	0.108	5	1.96364
ecthyma gangrenosum	3	1	13	0.133	13	1.6625
Agammaglobulinémie	9	3	2	0.437	15	1.01628
lymphoproliferation	74	5	1	0.974	19	0.98185
Trisomie 12	25	1	1	0.85	85	0.87629
deficit immunitaire humoral	14	3	2	0.44	15	0.75429

Figure 36: patient record - similarity calculation - table of possible diagnoses

6. My cohorts

By clicking on "My cohorts" in the menu bar, you can access all the cohorts that you manage and for which you are a collaborator.

You can create a new cohort.

You can display the contents of a cohort by clicking on the cohort in the list on the left. You can search a cohort with the filter.

6.1. Viewing a cohort:

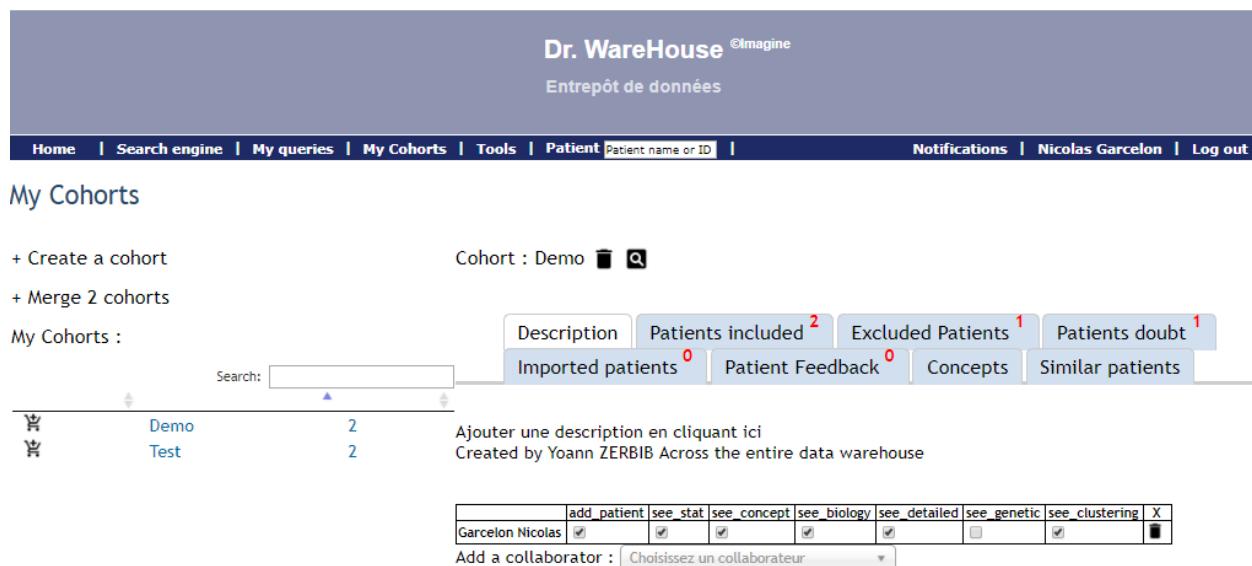
When you view you a specific cohort (Figure 37), you can:

- Change the title by clicking on it
- Delete the cohort by clicking on the icon 
- Search in all included patients by clicking on this icon  . You will get exactly the same interface as to refine a previous search. By default, all patients will be displayed, the stat tab will be calculated for all patients included in the cohort, and the same goes for the concepts, biology etc.

6.2. Description tab

In the description tab (Figure 37), you can change the description of a cohort by clicking on the description.

In the description tab, you can add collaborators with specific access rights to the cohort.



Dr. WareHouse ©Imagine
Entrepôt de données

Home | Search engine | My queries | My Cohorts | Tools | Patient | Patient name or ID | Notifications | Nicolas Garcelon | Log out

My Cohorts

+ Create a cohort Cohort : Demo  

+ Merge 2 cohorts

My Cohorts :

	Demo	2	Ajouter une description en cliquant ici Created by Yoann ZERBIB Across the entire data warehouse
	Test	2	

Search:

Description Patients included 2 Excluded Patients 1 Patients doubt 1
Imported patients 0 Patient Feedback 0 Concepts Similar patients

Garcelon Nicolas	<input checked="" type="checkbox"/>	<input type="checkbox"/>	X					
------------------	-------------------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	---

Add a collaborator : Choisissez un collaborateur

Figure 37: Viewing a cohort - description tab

6.3. "Patients included", "Patients excluded", "Patients queried" tabs

In the "patients included" (Figure 38), "patients excluded" and "patients queried" tabs you can view the list of patients included, excluded or queried, with the date and who added each patient.

For each list of patients, you can export them to Excel by clicking on this icon or copy and paste them by clicking on this icon . You can also change the status of a patient by clicking on the corresponding icons:

Include

Exclude

Query

Add a comment. The icon is red if a comment already exists.

The screenshot shows the Dr. WareHouse interface with the following details:

- Header:** Dr. WareHouse ©Imagine Entrepôt de données
- Navigation:** Home | Search engine | My queries | My Cohorts | Tools | Patient | Notifications | Nicolas Garcelon | Log out
- Section:** My Cohorts
- Actions:** + Create a cohort, + Merge 2 cohorts
- Cohort Selection:** Cohort : Demo
- My Cohorts:** Demo (2), Test (2)
- Search:** Search:
- Tab Bar:** Description (selected), Patients included (2), Excluded Patients (1), Patients doubt (1). Other tabs include Imported patients (0), Patient Feedback (0), Concepts, and Similar patients.
- Patient List:**
 - Patient F., 37 years (le 27/07/2017 15:03 par Nicolas Garcelon)
 - Patient F., 42 years (le 27/07/2017 15:03 par Nicolas Garcelon)

Figure 38: Cohort - patients included tab

6.4. "Patients imported" tab or Import patients

The "Imported patients" tab (Figure 39) allows you to import patients into the cohort on an *ad hoc* basis. To do this, just copy and paste the list of your patients, respecting the order of information to be provided:

Patient hospital ID	Surname	First name	Date of birth
---------------------	---------	------------	---------------

The separator between each variable can be a semi-colon (;) or one tabulation. Copying and pasting from Excel works without problem. You can give only the Patient hospital ID without the name and date of birth. Conversely, you can leave out the Patient hospital ID, and the system will automatically try to find the Patient hospital ID using the name and date of birth. The data loading log will inform you of any rejected patients.

You can decide to import the patients without deciding their final status (included, excluded); they will be displayed below. You can directly include patients or exclude them by clicking on the appropriate button at the time of import.

Dr. WareHouse ©Imagine
Entrepôt de données

Home | Search engine | My queries | My Cohorts | Tools | Patient Patient name or ID | Notifications | Nicolas Garcelon | Log out

My Cohorts

+ Create a cohort Cohort : Demo

+ Merge 2 cohorts

My Cohorts :

	Description	Patients included 2	Excluded Patients 1	Patients doubt 1	Imported patients 0
	Patient Feedback 0	Concepts	Similar patients		

Search:

Demo 2
Test 2

To Import Patients
Copy paste a list of patients
Respect the order of the columns :
Patient ID;NAME;FIRST NAME;DATE OF BIRTH
Use as a separator the comma, the semicolon or tab (a copy paste from Excel will work)
You may only have the Patient ID or the name + first name + date of birth without the Patient ID :
jgarcelon;nicolas;13/05/2007
9999999

Figure 39: Cohort - Import patients in the cohort from an external list

6.5. "Patient comments" tab

This tab allows you to view all comments on patients, with the date created and the author of the comment.

6.6. "Concepts" tab

This tab allows you to view all extracted concepts from the records of patients included in the cohort. The table displays the concepts, number of patients and the number of occurrences of the concept.

6.7. "Similar patients" tab

For each patient included in the cohort, the system will calculate the 20 most similar patients (using the same method as for the similarity with a patient). It will then calculate the number of occurrences of similar patients and display them in descending order. It also calculates a mean similarity distance. The user can then sort by number of occurrences of a similar patient, or by mean similarity score.

The similarity may be run for all patients in the data warehouse or just on patients imported into the cohort (included, not excluded).

The user can select cohorts for which included patients will be excluded from similar patients.

7. My queries

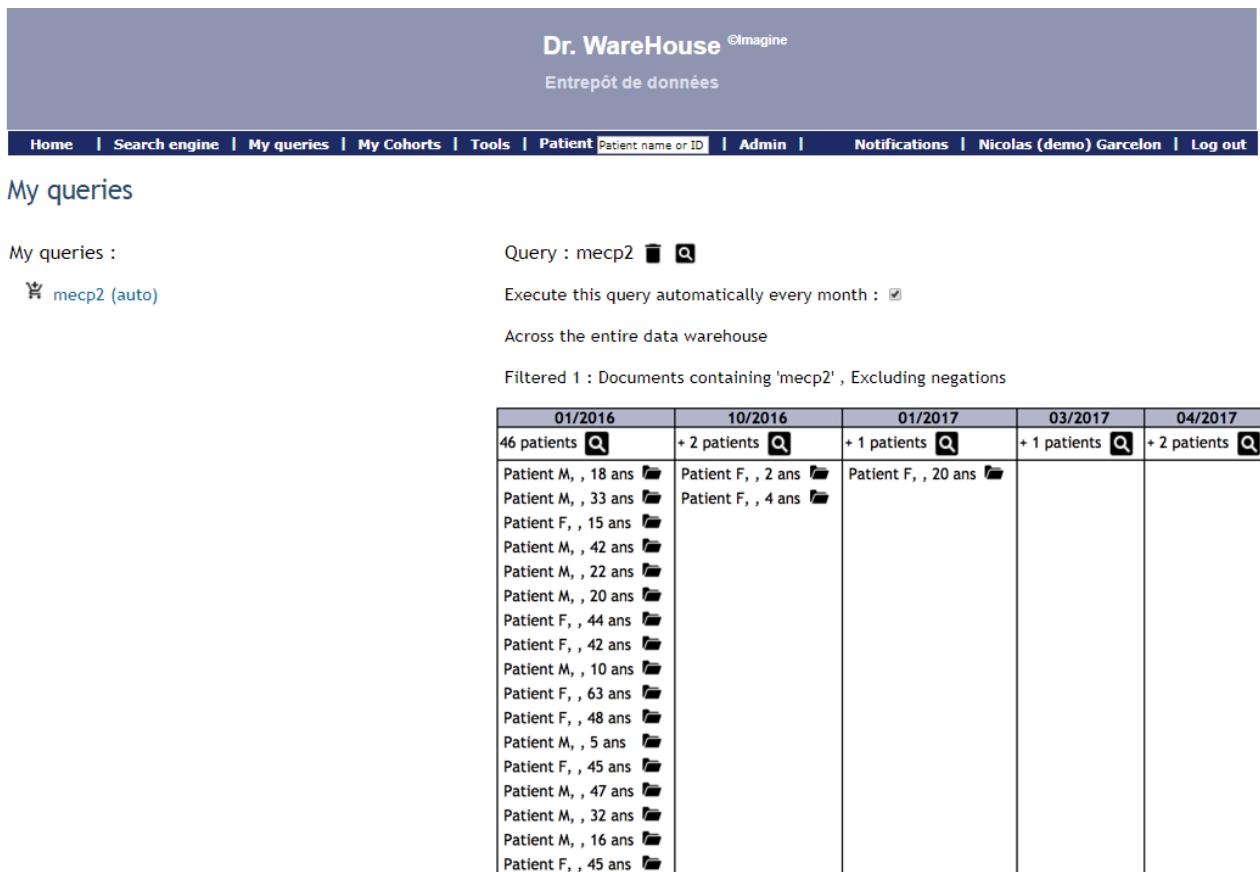
The "My queries" tool, accessible from the menu bar, allows you to view your saved queries and the patients found each month (Figure 40).

The title of the query can be changed by clicking on it.

The query can be deleted by clicking on this icon .

The query can be run in the data warehouse by clicking on  next to the title of the query.

The  at the top of each month enables the list of patients for the corresponding month to be displayed in the search engine (as for a cohort). So, you can carry out a search just on these patients.



The screenshot shows the Dr. WareHouse interface with the title "Dr. WareHouse ©Imagine Entrepôt de données". The navigation bar includes Home, Search engine, My queries (highlighted), My Cohorts, Tools, Patient (with a search input field), Admin, Notifications, and Log out. The user is logged in as Nicolas (demo) Garcelon.

The main area is titled "My queries" and displays a query named "mecp2" (auto). It includes options to execute the query automatically every month (checked) and across the entire data warehouse. A note indicates "Filtered 1 : Documents containing 'mecp2' , Excluding negations".

A table summarizes patient counts by month:

01/2016	10/2016	01/2017	03/2017	04/2017
46 patients 	+ 2 patients 	+ 1 patients 	+ 1 patients 	+ 2 patients 
Patient M, , 18 ans  Patient M, , 33 ans  Patient F, , 15 ans  Patient M, , 42 ans  Patient M, , 22 ans  Patient M, , 20 ans  Patient F, , 44 ans  Patient F, , 42 ans  Patient M, , 10 ans  Patient F, , 63 ans  Patient F, , 48 ans  Patient M, , 5 ans  Patient F, , 45 ans  Patient M, , 47 ans  Patient M, , 32 ans  Patient M, , 16 ans  Patient F, , 45 ans 	Patient F, , 2 ans  Patient F, , 4 ans 	Patient F, , 20 ans 		

Figure 40: My queries - viewing a saved query and patients found each month

8. Notifications

Notifications alert you when:

- someone shares a query with you. When you click the notification, the system displays the search engine with the pre-filled query.
- someone adds you as a collaborator in a cohort. When you click on the notification, the system displays the cohort.
- someone sends you a message. When you click on the notification, the system suggests responding to this person.

You can send a message via the notification tool, by clicking on "send a message", then by selecting the recipient and writing a message in the text field.

9. My profile

You can access your profile by clicking on your name next to the logout button (Figure 41).

You can see your rights, your statistics and your contact details.

You can access the list of people in your department as well as the hospital units and their modification over time.

The screenshot shows the Dr. WareHouse interface. At the top, it says "Dr. WareHouse ©Imagine" and "Entrepôt de données". Below the header, there's a navigation bar with links: Home, Search engine, My queries, My Cohorts, Tools, Patient (Patient name or ID), Admin, Notifications, Nicolas (demo) Garcelon, and Log out. A "My account" link is also present. The main content area is divided into three sections: "My profile", "My stats", and "My contact information".

My profile:

- Your profile : admin_anonyme
- This profile gives you access to :
- Fast Patient Access
- Tous les services
- Search engine access
- Anonymized data
- Statistics tab
- Biology Tab
- Result tab
- Map tab
- Clustering tab
- Genetic tab
- Administration of the directory
- Concept Tab

Authorized documents : tout

My stats:

- 13 Query
- 10 Patient files consulted
- 3 Cohorts created by you
- 0 Cohorts to which you are associated
- 353 Patients included in your cohorts
- 353 Patients not included by you

My contact information:

- E-mail :
- Phone : Add a phone number
- Change my local password : valider

At the bottom right, there are links: Contact | Logs | ETL | Credits.

Figure 41: My profile - my rights, my stats, my department

10. Contact

The contact page gives you the contact details of the data warehouse technical hotline. Don't hesitate to contact them in case of problems.

11. Logs

This page allows you to view the number of clicks for each feature of the data warehouse (Figure 42).

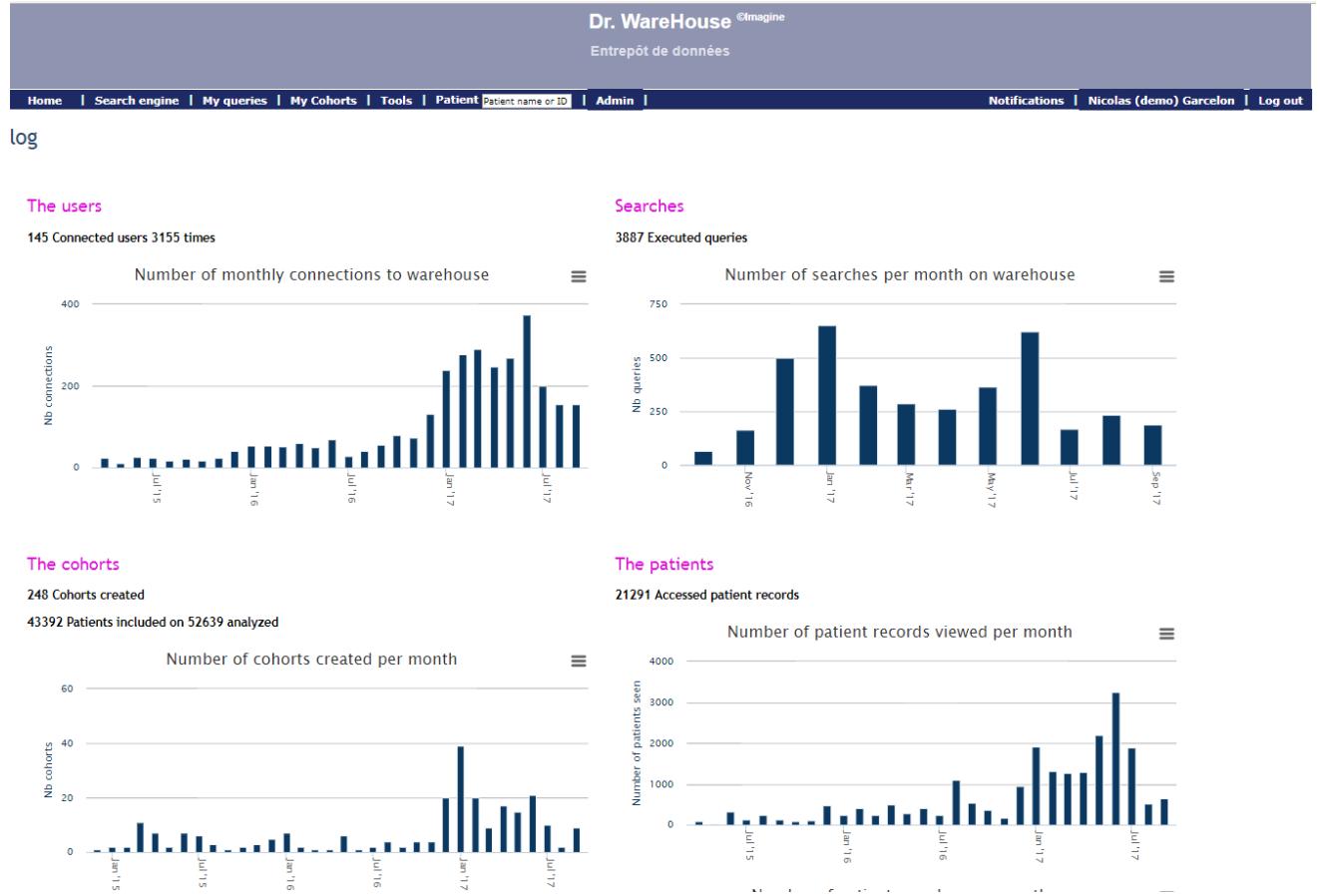


Figure 42: Data warehouse usage log

12. ETL

This page gives you access to the loading details of the data warehouse. For each built-in data source, you have the document number per year and per month.

13. Credits

The credit page refers to the different open source or free libraries used in this tool.