GLOBAL AMR R&D HUB

Dynamic Dashboard – Data collection, processing, categorization and presentation

Introduction

The goal, scope and overall process for the development of the Dynamic Dashboard as well as the process for arriving at high-level categories to classify the information about funding/investment into AMR are described in the Establishing the Dynamic Dashboard –Methodology for developing the categorisation fields paper published on the Global AMR R&D Hub website.

Briefly, the goal of the Dynamic Dashboard is to present basic and applied research projects/investments from publicly and privately funded R&D throughout the research and innovation value chain on treatment, preventive measures, diagnostic products, surveillance, policy and interventions (such as stewardship) across all One Health sectors.

This paper outlines the process for collecting the information that forms the basis for the Dynamic Dashboard. Project and/or investment information is being collected and presented in a staged approach, beginning with R&D[1] relevant to human antibiotic resistant bacterial infections. Also, it has to be noted that for the launch of the Dynamic Dashboard we have not yet been able to include information on private investments. The link to the data sources is provided in the dashboard library.

The Global AMR R&D Hub gratefully acknowledges the support it has received from funding organisations around the world who contribute data to this exercise. Without the collaboration and support the Dynamic Dashboard could not have been developed.

The approaches taken for data processing and quality control are also explained. Based on the categories and definitions, which can be accessed via the dashboard library, each investment/project has been categorised. Here the specific approach to this exercise is described.

An Annex to this document highlights the data limitations and caveats. We are working to continuously improve the data collection and data quality. At launch users need to be aware of significant limitations to what the data can and cannot show.

The Dynamic Dashboard builds on a cloud-based data base implemented in the Microsoft Azure environment. Microsoft Power BI is used for the visualisations.

A separate user guide has been prepared to help with accessing and using the Dynamic Dashboard.

[1] All research and/or development will be captured by the Dynamic Dashboard and is not limited to the traditional definition for R&D for product development. This will enable collection of important research looking to expand our understanding of AMR or investigating how surveillance, stewardship, access, social changes and/or policy implementation can be enhanced.

Data sources and collection – focus human bacterial infections

The guiding principle for collecting data to be included in the Dynamic Dashboard was not to duplicate efforts that have already been taken and to make provision of data as straightforward as possible for investors/funders.

Based on information publicly available in the field, relevant data sources/websites, typically per country were identified. Many funders/countries offer such publicly available data portals. The initial list of data portals was shared with the Board and Stakeholder Group with a request for additional input.

To carry out searches in the data sources, in consultation with the Board, the Stakeholder Group and additional experts a list of standard search terms was developed[2]. Relevant terms were added over the period of about six months during which the initial data collection was carried out. The list of terms used for every source is shown in the data source table. As we are taking a staged approach, the search terms will be further updated accordingly. A standard set of data points to collect was defined[3]. For countries/funders where we could not identify publicly accessible data portals we reached out directly to the relevant ministries/organisations.

The data collection also benefited from the mapping that has been carried out earlier by the European Joint Programming Initiative on AMR (JPIAMR). The information needed to complete the core data set was collected from the data sources for the respective funders.

The data set collected was shared with the respective funder/investor to ask for permission that we (re-) publish the information, and to what extent, and to verify that we had not made a systematic mistake in data collection.

In agreement with the Board it was decided that the Dynamic Dashboard will show projects/investments ongoing on 1 January 2017 or later. This time frame was used for the searches/requests for data.

Data processing and quality control

Depending on the data source/data portal, it was possible to download Excel files with the desired information. In other instances, the information for the individual data fields had to be manually extracted.

For titles and abstracts that are not available in English, translation tools were used. The IT-solution for the Dynamic Dashboard offers an integrated translation solution. This will be implemented at a later point in time and then texts not available in English will be published on the Dynamic Dashboard.

---

[2] Acinetobacter, aeroguinosa, ampicillin, AMR, antibacterial, antibiotic, antibiotic resistance, antibiotic susceptibility, antifungal, anti-fungal, antimicrobial, anti-microbial, aureus, baumannii, C. difficile, campylobacter, carbapenem, cephalosporin, clarithromycin, clindamycin, clostridia, clostridium, cotrimoxazole, drug-resistant bacteria, Enterobacteriaceae, Enterococcus, erythromycin, ESBL, ESKAPE, faecium, fungal pathogens, gonorrhea, gonorrhoea, gonorrhoeae, Gram-negative bacteria, H. influenzae, Haemophilus influenzae, Helicobacter, hospital acquired infection, hospital-acquired infection, Klebsiella, Listeria, Lyme disease, MDR-TB, methicillin, MRSA, multi drug resistance, multi drug resistant, multidrug resistance, multi-drug resistance, multi-drug resistant, multidrug-resistant, Mycobacterium, Neisseria, One Health, penicillin, pneumococcal, proteus, Pseudomonas, rifampicin, Salmonella, Serratia, Shigella lactamase, Staphylococcus, stewardship, Streptococcus, superbug, tuberculosis, vancomycin,

[3] Country of the funder/Funder/Project ID/Project title (original)/Project title (English)/Project acronym/Recipient (organisation)/Principal investigator/Country of the recipient organisation/Start date/End date/Total amount awarded/In kind contribution/Total project cost/Currency/Abstract-summary (original)/Abstract-summary (English)

The Dynamic Dashboard presents funding amounts in US Dollar (US $) and in Euro (€). The integrated currency converter of the IT-solution is used to convert the funding amounts/budgets into US $ and €. The conversion rate at the starting date of the project has been chosen.

For projects with several participating institutions, the entire budget is allocated to the institution of the principle investigator/coordinator[4].

To arrive at the investments per year, the total budgets of all investments/projects are distributed *pro rata* over the years of duration (a project starting on 1 May 2019 and ending on 30 April 2020 would see 2/3 of its budget allocated to 2019 and 1/3 to 2020). One-time investments are recorded in the year they were made.

To check the quality of the data, the Excel files created per funder were carefully checked for duplicates, missing data, and consistency of data (such as reasonable figures for the budget, texts for abstract summary, plausible dates for start and end-date etc.). A sample of projects was re-searched separately and it was checked that the data extracted are identical.

The Excel files were imported into a data base that was used for categorizing the data (see next paragraph). During import further checks were carried out (for example plausible dates and budgets). The consistency of title with the abstract/summary was checked during categorising and the content of the abstracts was checked.

Categorizing of data

To be able to present the information about the projects/investments related to different areas of AMR research, a set of categories and definitions was prepared, as described in the Establishing the Dynamic Dashboard –Methodology for developing the categorisation fields. To decide whether a project addresses research and development related to AMR, a list of bacteria with a relevant AMR issue was prepared. The list is accessible through the dashboard library. The categories will be further refined when additional areas of AMR research (such as animal-, plant and environmental health) will be added.

To be able to assign budgets to projects/investments in a given category and to avoid counting of budgets multiple times, if a project is relevant to more than one research area, R&D stage or bacteria, the project budget is split accordingly. For example, if a project is relevant to three different bacteria, the budget will be assigned to 33% to each of the bacteria. If a project falls into two different research areas, the budget will be assigned to 50% to each of the research areas. By this we make sure that the budget of a specific project will not be counted multiple times. In contrast, we did not split the number of projects. A project which is relevant to more than one research area, R&D stage or bacteria will be counted multiple times in the specific report, but this does not affect the overall total number of projects. When the Dynamic Dashboard evolves into the other One Health sectors, the budget split will also affect multiple sectors (human, animal, plant, environment) and multiple infectious agents (bacteria, fungi, viruses, parasites).

Categories and definitions accessible through the dashboard library were prepared to ensure consistency in categorising. The staff of the Secretariat of the Global AMR R&D Hub have categorised the data. At the start of categorisations regular meetings in the team were held to raise questions and align on the approach. Questions related to categorising projects were discussed in the team.

Projects that are relevant to AMR R&D but not related to human bacterial infections have been parked for now. Projects not falling in the expected time frame (project active on 1 January 2017 or later) or not

[4] For projects supported by the Innovative Medicines Initiative (IMI), the entire budget of the project has been allocated to the institution that is the managing entity for the contribution from the European Union.

addressing AMR have been marked as “not relevant to AMR”. They are nevertheless kept in the data base for later checking. When categorizing it was possible to assign the category “unsure”. This means that categorising the investment/project will be looked at again by several colleagues.

We have started to record additional details, which will be used for future presentations on the Dynamic Dashboard. For example, if a project is funded in a given country and carries out research in a different country, this has been manually recorded. Also, products under development listed in the abstract/summary have been manually extracted.

Creation of visualisations

We identified key audiences and their needs. We carried out a public consultation and interacted directly with key stakeholders to understand their needs and the expected utility of the Dynamic Dashboard. Based on this input and the available data and the categorisations we created a first list of desired visualisations. This was discussed with the IT company and they created the first visualisations. In an iterative process the visualisations were developed by the IT company. The draft (static) visualisations were shared with the Board and the Stakeholder Group and their feedback was incorporated to the extent possible within the chosen IT-solution.

Feedback of users is invited and the presentation of the Dynamic Dashboard will be regularly updated. Several further functionalities still to come (see next paragraph).

Further developments

The approach to the further developments of the Dynamic Dashboard has been described in the Establishing the Dynamic Dashboard –Methodology for developing the categorisation fields paper.

At launch the Dynamic Dashboard shows projects/investments from public and philanthropic funders as regards human bacterial infections.

Translations of titles and abstracts that are not available in English into this language will be implemented using the integrated solution of the Microsoft Azure environment.

Also, a search function for users to select their own parameters will be provided and we will offer the possibility for users to register, so that preferences of how they want to interact with the Dynamic Dashboard are saved.

A systematic plan for updating the data will be developed.

Annex

Caveats/limitations

Interpretation: Care should be taken in drawing conclusions about the AMR R&D landscape from the information currently presented in the Dynamic Dashboard especially in regards to identifying gaps in research needs and making statements on global activity or product-specific R&D. This is because of data gaps including geographical representativeness and the current lack of information about private sector or institutional investment.

Comparison: We are presenting a first version of the Dynamic Dashboard. It is currently not possible to compare investments or the number of projects between different countries or over time. This is because the search terms have evolved during the period of initial data collection, the data extraction methods differ between sources, due to the nature of the data source, and the timing of data collection and updates from funders.

Updates: Due to the nature of the Dynamic Dashboard, both the data and definitions for categorisation will be updated regularly and is subject to retrospective revision. Therefore, both new and existing data may vary between the update dates. When referencing information from the Dynamic Dashboard it will be vital to include the last update date.

As shown in the list of data sources, accessible through the dashboard library, extraction of data was done at a given point in time. Projects/investments awarded after this date are currently not recorded. A process for regularly updating the data base will be established.

Data extraction: When interpreting information for the Dynamic Dashboard the data considerations, sources, definitions should be reviewed with each extract/viewing as, these will also be regularly reviewed and updated.

Data Scope / Maturation: The Global AMR R&D Hub takes a staged approach to presenting information. The information is currently limited to R&D regarding human bacterial infections.

Completeness (by type of funder): Currently only research and development funded by public funders and philanthropic organizations is included in the Dynamic Dashboard. Work is underway to obtain and represent private sector R&D funding, which is vital to be able to see the true AMR R&D landscape.

Completeness (global): Data from a considerable number of sources has been collected, as shown in the list of data sources, accessible through the dashboard library. Collecting data from the Members of the Global AMR R&D Hub was prioritised. It needs to be highlighted that information from a number of funders is missing. Particularly coverage of funders from the Southern hemisphere and low- and middle-income countries is still limited. **Completeness (per country):** In addition, not all funders, investments and/or projects have been captured from countries. Both the global and coverage per country will improve as the amount information captured by the Dynamic Dashboard increases over time.

All funders/investors that have information about AMR-related R&D projects/investments are encouraged to share them with the Global AMR R&D Hub. It is also planned to develop tools for funders/investors to upload data directly.

Completeness (per type of funding): Our initial data set are likely skewed to certain types of funding steams/vehicles, such as direct grant support for projects and personnel. As our data set mature the hope is that we will be able to represent all 'push-type' funding (support for R&D inputs) regardless of the type of vehicle, this is again due the readiness of availability of initial data.

Completeness of capturing AMR-relevant R&D: The search terms used determine and may limit the range of relevant projects that could be identified.

Data accuracy: The Secretariat has relied on the completeness and accuracy of the original data from the funders. However, any obvious discrepancies, errors or gaps were investigated. The limited information contained in abstracts limits the interpretation and categorization. With the large number of projects/investments collected, it is not possible to go to more detailed information, such as detailed descriptions, reports, publications coming out of projects etc.

Multi-beneficiary challenges: For projects where several institutions are supported funding information for all individual partners is normally not available. This will lead to inflated numbers for some countries/research organisations and too low numbers for others.[5]

Limitations in data processing/presentation:

As mentioned above, if a project is relevant to more than one research area, R&D stage or bacteria, the project budget is split accordingly. However, we did not split the number of projects. A project which is relevant to more than one research area, R&D stage or bacteria will be counted multiple times in the specific report, but this does not affect the overall total number of projects.

The category 'other bacteria' includes cases where the categorized project/investment addresses bacteria for which no individual category was created and where there is a resistance issue (see the list of bacteria with a relevant AMR issue, accessible through the dashboard library), and cases where the project was relevant to bacteria more generally.

The report on subprojects with individual project IDs and individual investments were calculated as independent projects that may result in duplication of the number of projects.

It was decided to choose the exchange rates to US Dollar and Euro on the start date of the project. For currencies where the exchange rate fluctuates (sometimes strongly) over time, this may mean that investments are over- or under-estimated. This method was chosen as it can be clearly defined.

It was decided not to factor in adjustments for inflation rates, since it is considered that having reliable information about inflation rates from across the world since January 2017 is challenging. It would also have required additional processing of the budget information adding complexity to the IT solution.

Sometimes a project/investment is composed of sub-projects, each with their own entry in the data-base. It is planned to develop tools that will link such sub-projects into the overarching project.

Projects that only have a component of AMR are currently parked and will be included later.

[5] For example, for collaborative projects funded by the European Commission the entire EU contribution is allocated to the coordinator and for IMI projects to the managing entity for the EU-contribution.

As mentioned above, projects are included that are active on 1 January 2017 or later. With the time period of 2017 being far enough in the past, all projects that could be found with the search terms used will have been collected. For later years, especially as from 2019, there are still many projects that have not yet been collected because the awards have only just been announced or the search was carried out before the awards were announced. Therefore, the graphs show a decline in recorded investments from 2017. As more data is collected, the closed year two years before the current time should have completed data.

Limitations in checking data quality

The testing of reproducibility of categorisation is still to come. It will be done by re-categorising a subset of projects by a different person. Depending on the outcome of the exercise a larger effort of re-categorising may have to be done. This may impact the conclusions one can draw as to repartition of investment in AMR R&D to different sectors and categories.

The processes for checking the data quality so far are listed in the methodology paper. Additional checks will be carried out.

Moving forward, all limitations, methodological, process or otherwise will be regularly reviewed by the Secretariat to improve data collection, categorisation and presentation.