

Outline for Final Reports – GBADs Bill & Melinda Gates Foundation and FCDO Phase IIa and Phase IIb investments

Thank you in advance for your dedication to completing the final reports. The WOAAH team is available for any questions or clarifications you may need. Please review the following information carefully:

Per the terms and conditions of the sub-grant and to meet donor requirements, **each consortium partner is required to submit separate reports for Phase IIa and Phase IIb (including the FCDO supplement). This means that WOAAH expects to receive two reports from each sub-grantee.**

The purpose of these reports is for sub-grantees to report on the delivery or progress towards delivery of activities and outputs detailed within ToRs. This includes information on results achieved; challenges, delays and any changes made; perspective on lessons learned or takeaways, and any other inputs deemed necessary. Notes are provided in grey within the outline to support completion.

To ensure timely delivery of the final reports, **the WOAAH team requests that you adhere to the following timelines** as there is no room for flexibility in the final deadlines given the closing of both investments.

Narrative report

- **Final document due on 30 April**
- By 5 April: submit relevant information to WOAAH, focusing on the deliverables outlined in the attached Terms of Reference for each investment phase, which corresponds to section B in the outline.
 - This information will be used by WOAAH and UoL to provide BMGF and FCDO with updates on progress towards deliverables at a meeting planned for 9 April.
 - For activities not yet completed but expected to be finished by the end of April, please indicate this in the achievements section.
- By 19 April: Submit the first draft of reports to WOAAH.
 - Complete Sections A, C, D, and E, and make any desired modifications to Section B.
- Between 22 to 25 April: WOAAH will review the drafts and seek any clarifications needed from sub-grantees.
- On 30 April: Sub-grantees to provide the final report.

Financial report

- **Final document due on 31 May**
- By 30 April: Submit the first draft to WOAAH for review and discussions between WOAAH and sub-grantees.
- On 31 May: Final financial report to be submitted.

Your adherence to these timelines is crucial for ensuring the successful closure of these investments. Sincere thanks in advance for completion.

A. Executive Summary

The GBADs Informatics team is based at the University of Guelph. The major objectives of the team that have been achieved to date include:

- Provision of Cloud (AWS, GCP) computing infrastructure for the storage of data and provision of Knowledge Engine software products such as dashboards, modelling software (e.g. AHLE, DMP), and APIs to relational databases, graph databases and ontologies.
- Provision of information and advice about the computation and data science aspects of the Knowledge Engine.
- Provision of software products developed by the informatics team and other GBADs team, e.g. dashboards, models, data
- Monitoring of the infrastructure to provide 24/7 access to data and software products.
- Monitoring of the number of accesses to the Knowledge Engine and where those accesses are originating.
- Provision of all relevant documentation for the Knowledge Engine and its components.
- Ongoing development of a Data Governance Handbook, available as a *living* document to all users and collaborators of GBADs.

Remaining challenges for Informatics revolve around three areas:

- Continuing improvement of the User Experience (UX) with regards to user engagement with the Knowledge Engine and the component products of the Knowledge Engine. This activity cannot have an end date since user engagement will need to be adjusted as more and different users and collaborators interact with the Knowledge Engine.
- Sustainability and maintenance of the infrastructure and products of the Knowledge Engine. This is a long-term funding issue that cannot be solved by the Informatics team alone.
- Training materials are minimal at this stage and mostly consist of documentation, presentations (slides and videos), and some instructional videos. Provision of more media was hampered by the illness and death of one of the co-PI's (Dr. Theresa Bernardo) who was heading up this effort. Dr. Bernardo did leave her thoughts on how more materials should be produced and her ideas will form the basis for more work in future phases of GBADs.

B. Deliverables as per ToRs of relevant investment

Provide a narrative with as many details as possible to reflect the activities, achievements, challenges, lessons, etc. As this report is specific to BMGF/FCDO, please refrain from including activities or outputs not covered by the relevant investment (e.g., IDRC, ACIAR, country self-funded activities, etc.

B.1. Development and refinement of methods and analytical frameworks

- Planned activities as per relevant ToRs
 - *Please indicate whether there was collaboration with other sub-grantees and specify their role in the activity as it occurred during the investment period.*

- Achievements
 - Methods established, frameworks developed, analytical process defined
 - Publications, if any
 - Please include publications specific to a method's framework along the analytical structure (e.g., classification, AHLE in the Lancet, etc.). Publications focused on a particular country case study (e.g., Ethiopia), thematic study (e.g., on farm investment into dairy cow), or global study (e.g., estimating global economic value of farmed animals) should not be included here but rather in section B.2.
 - Any additional information
- Challenges, delays, deliverables put on hold, changed or cancelled
 - *If an activity was delayed, put on hold, changed, or cancelled, please provide a reason as to why.*
- Lessons learned or any other aspects to be captured

B.2. Testing of methodologies to describe the burden of animal diseases, at production system and country level.

- National Case studies – relevant to BMGF/FCDO investments. *Please indicate the category of animals and production systems*
 - Planned activities as per relevant ToRs
 - *Please indicate whether there was collaboration with other sub-grantees and specify their role in the activity as it occurred during the investment period.*
 - Achievements
 - Publications
 - Estimates
 - Systems, process and/or developed, if any
 - Any additional information
 - Engagement with local stakeholders to structure and steer the cases study.
 - Advisory groups, etc.
 - Any additional information to be highlighted
 - Challenges, delays, or any deliverables put on hold, changed, cancelled. *If an activity was delayed, put on hold, changed or cancelled, please provide a reason as to why.*
 - Aspects needing additional work
 - Aspects not delivered / need a rethink, if any
 - Social and political factors, if any
 - Lessons learned, feedback provided to be highlighted, or any other aspects to highlight
- Global or thematic studies (if applicable). *Please indicate the category of animals, production system, and countries captured.*
 - Planned activities, outputs, and/or results as per relevant ToRs.
 - *Please indicate whether there was collaboration with other sub-grantees and specify their role in the activity as it occurred during the investment period.*
 - Achievements
 - Publications

- Estimates
- Systems, process, and/or tools developed, if any
- Any additional information to be highlighted
- Challenges, delays, or any deliverables put on hold, changed, or cancelled. *If an activity was delayed, put on hold or cancelled, please provide a reason as to why.*
 - Aspects needing additional work / not delivered
 - Aspects needing a rethink
- Any engagement with external consortium partners (e.g., private secToRs, data suppliers)
- Lessons learned, feedback provided to be highlighted, or any other aspects to capture

B.3. Informatics – Building and testing Knowledge Engine needed to deliver a sustainable GBADs

In the following narrative, the planned activities of the Informatics team will be given as per ToRs followed by the achievements.

- ***Accessible front-end user interface building on learnings from Phase II-A interface feedback***

The front end of the GBADs Informatics Knowledge Edge has been re-designed to be reactive, accessible, and easy to navigate. There are major sections for the dashboards (currently there are x dashboards), data access (via GBADs API) and data ecosystems (currently the Senegal data ecosystem is provided), and documentation (Data Governance Handbook, Guides, Dashboard documentation, media such as presentations and papers).

The interface is easy to add to as it is maintained via a GBADs Informatics GitHub repository and changes to the repository are immediately propagated to the interface.

All dashboards were refurbished to reflect comments from the participants in the two Ethiopia workshops. As well, most of the dashboards have been updated to provide similar experiences when viewed on mobile devices compared to desktop/laptop viewing, This will enable usage in the field via mobile phones and tablets.

A small number of dashboards (National Populations, Ethiopia Data Stories) also have an added section at the bottom of their interface that will allow users to provide comments about the data or its presentation. These comments are stored in a database and also sent to the GBADs Informatics Slack app so that these comments can be reviewed for appropriateness (no comment appears before this review to avoid trolling) and a response can be formulated to give back to the user. It is hoped that this will increase a sense of community amongst the users and will elicit “crowd sourced” information from our users on the ground. After a period of observation, this feature will be rolled out to all of our dashboards.

A plan for the next version of the main KE interface will be available in May that further expands on, not just interface design, but the possibilities of UX (user experience) design for the KE. User personas and experience pathways through the KE are initially exposed for future development in Phase 3 to provide a curated user experience for different types of GBADs users: policy makers doing global or local studies, modellers and academics using the KE for their studies, developers of related software, etc.

- **26. Updated data ecosystem map for burden of disease estimation in Ethiopia.**

There was not enough time to provide a current ecosystem map for Ethiopia. The first data ecosystem map was done over 3 years ago and there needs to be an updated study. A template of data ecosystem map has recently been developed by Kassy Raymond (GBADs Informatics Technical Manager) employing the Kumu software tool. Once local information can be obtained for Ethiopia, it will be straight forward to create a data ecosystem map for the country.

- **27. Verification and quality control process, and associated living guide, for reproducible modelling and data handling; endorsing publications with version-controlled data and analytics; supporting update of the data governance handbook and process for testing and improving the user interfaces.**

All the GitHub repositories of data and code to produce dashboards and other products are attached to an AWS pipeline that will maintain the most up to date version of KE products on the KE. Developers and maintainers of the KE products, such as dashboards, can make changes, have the executable “containerized”, so that it can run on the AWS Cloud infrastructure and pushed to the AWS Elastic Container Service (lightweight mechanism for running software in the Cloud). The pipeline will ensure that only running software is presented through the KE. Versioning of the software products follows the semantic versioning protocol (reference) and is manually assigned and maintained. Most of the KE products are at version 1 but a number are already at version 2 as there has been continuous development of most products. The documentation is maintained in a similar manner to the software (in GitHub repositories or as Google documents backed up on AWS S3 Cloud storage) and are *living* documents that can be refreshed and available continuously.

The inclusion of a comments section on some of the dashboards illustrates how community involvement will be stimulated and incorporated into future versions of the KE products.

Data quality analysis work has been done by looking at the consistency of animal population data in multiple data sets and within one dataset (FAOSTAT).

Multiple data sets: Data Scientists are learning the quality of their models depends on the quality of data. Data used by the Global Burden of Animal Diseases (GBADs) is available to modellers around the world and the quality of the data provided is important as it is used in modelling disease, greenhouse gas emissions, and more. These are important topics so the data given to the modellers must be investigated and checked for internal and external inconsistencies. The goal of this analysis is to investigate data provided by GBADs to find inconsistencies in the data. Data quality was analyzed using a five-year trailing average comparison, the interquartile range for the yearly rates of change, and observing outliers on a normal distribution for the yearly rates of change for livestock populations over time. The normal distribution and interquartile range analysis is an internal data analysis that can find outliers that indicate possible data inconsistencies. The five-year trailing average helps identify external data inconsistencies between sources. Using purpose-built data analysis tools and performing analysis on the data shows there are inconsistencies in the data. These tools were developed by GBADs Informatics software developers, Ian McKechnie and Joshua Davidson, and can be found in the GBADs Informatics GitHub repositories (<https://github.com/GBADsInformatics/GBADsDataQualityInsights> and <https://github.com/GBADsInformatics/VisualizerDashboard>). The consequences of these findings show that researchers need to be cognizant of the data they are using and need to

perform their own analysis before they use it in their models as the results can show incorrect results. This work has currently been reviewed by the journal, *Data Science Journal* (<https://datascience.codata.org/>), and minor changes are being submitted in preparation for publication. The publication title is “Identifying Inconsistencies in Data Quality Between FAOSTAT, WOA, UN Agriculture Census, and National Data” by Ian McKechnie, Kassy Raymond, and Deborah Stacey.

Another approach to data quality analysis is to look at the internal consistency of data within a single data set. We selected to look at the population data in the QCL table from FAOSTAT since this is a major resource used worldwide by modellers and analysts. A dashboard (Data Visualizer) has been developed that allows a user to select a country and a species and to see a visualization of the data contained in FAOSTAT. The dashboard shows the flags attached to each data point (official, estimated, imputed, non-FAO) and the trend of the population data over time. There is also a visualization of the percentage of data each year based on how the data point was flagged – the user can see the changes over time in how the data was obtained. As well there is an irregularity spotter that identifies points that deviate from previous points based on expected levels of change, i.e. population should not increase or decrease by more than X percent per year. This level can be set by the user. It is anticipated that this tool will provide a basis for data uncertainty analysis in the future.

- **28. Living versions of ontologies and graph databases for livestock classification, animal production and health curation – semantic interoperability, expansion, and integration into the KE.**

Version 1.0 of Trusted Animal Information portal (TAIL) has been developed and is in Alpha Launch. TAIL combines custom-built semantic technologies (ontology and graph database) and natural language processing (NLP) to allow users to search for metadata using natural language queries. The query can be enriched with semantically similar terms using the ontology. The graph database is then queried to retrieve descriptive and distribution metadata specific to the enriched query. Integration and expansion of the graph database was addressed through the creation of Extract, Transform, and Load (ETL) processes and implemented pipelines. Custom pipelines have been developed for each data source, allowing for the integration of heterogeneous data and metadata into the graph database. A Software Development Kit (SDK) and Application Programming Interface was developed to interact with the graph database, which allows the TAIL user interface to retrieve information from the graph database. The ontology and graph database are considered “living versions” as more datasets may be added in future phases of GBADs.

GitHub repositories are available for each component of TAIL including the user interface, the NLP pipeline, ETL of data and metadata into the graph database, the API for the graph database and the API for the ontology. All technologies are hosted in the Google Cloud Platform (Neo4j Aura, or through docker containers on Amazon Web Services).

Since TAIL is still in Alpha at the time of the writing of this report, a URL of this version can be reported (<https://gbadske.org/search>) but it is still under construction. A video of the portal is also available on the GBADs Informatics GitHub in the Media repository (https://github.com/GBADsInformatics/Media/blob/main/TAIL_Alpha_Video_2024-04-16.mp4) portal. The video can be downloaded and viewed by clicking on the given GitHub link and then

clicking on [View raw](#). A public Beta version of TAIL will be available in the first week of May, 2024 at <https://gbadske.org/search>.

- **29. Working paper - example of ontology as an enabler for data quality**

A paper on our semantic architecture for livestock population data and modelling has been produced in draft form and is being submitted for publication. The paper (title: *Microservice Ontology Architecture for Livestock Data*; authors: Le Nguyen, Deborah Stacey, targeted publication: *Semantic Web Journal* - <https://semantic-web-journal.net/>) describes the KE's microservice ontology architecture for livestock populations using a common meta-ontology to capture shared concepts across different data sources. Instead of relying on a single large ontology, multiple micro-ontologies with shared concepts are utilized. Various methods such as queries and mapping rules to align data with shared concepts are described. The microservice ontology structure is leveraged to extract data with similar semantics from diverse sources. This structure facilitates the use of different logical reasoners, enabling the incorporation of intricate rules for mapping and inferring domain knowledge while addressing the performance and scalability issues. Use of this semantic architecture will provide a necessary framework for modellers seeking to discover relevant data sources and a deep understanding of how livestock data can be used from multiple sources.

- **30. Guide for verification and quality control**

To aid in the development of the Animal Health Loss Envelope (AHLE) and other collaborative efforts, Informatics provides, manages, and provides user support for GitHub repositories and a deployment pipeline for all software products such as dashboards. The GitHub provided a public location for the dashboard code and allowed for Informatics to make suggestions as to how best to organize their code, data, and documentation. Informatics also provided a pipeline that could automatically take their code and produce a containerized (dockerized) version that would be deployed on the AWS Elastic Container Service (ECS). Informatics has systematically automated the management of all dashboards on the Knowledge Engine in the same way. This automation has several benefits: 1) as soon as dashboard developers upload their newest version of their code, the Informatics pipeline will create a working dashboard in a container (a container contains application code and enough operating system code to run the application in a module that can be executed in the Cloud) and then load it into the ECS. The ECS allows Informatics to run code without the expense of running dedicated VM's (virtual machines) in the Cloud to present the dashboards to users. This pipeline has allowed Informatics to reduce the cost of running multiple dashboards (currently there are 14 dashboards) by over 60% as compared to running the code on VM's (the traditional way of using computation on the Cloud).

The GitHub repository contains many guides and documentation sites for supporting verification and quality control over the Knowledge Engine and its software products and the Data Portal within the Knowledge Engine and its data products

(<https://github.com/GBADsInformatics/DevopsWorkflows>,
<https://github.com/GBADsInformatics/Docs>).

- **31. Process for testing and improving the user interfaces**

To enhance the user interface and improve the design of the wider GBADs knowledge engine, the site of the Knowledge Engine (<https://gbadske.org>) was reorganized over Phases IIa and IIb to make navigation simpler and to compartmentalize the software products (mostly dashboards)

from documentation and educational media. The code to generate the Knowledge Engine is also driven from a GitHub repository making it very easy to update. To evaluate if the Knowledge Engine was, in fact, making an impact it was decided that we should monitor the requests to the ECS for our dashboards. One of our dashboards in the Knowledge Engine now shows the results of this analysis of community engagement with the Knowledge Engine.

The newest addition to the user interfaces is the *Comments* section on several dashboards (National Populations, Ethiopia Sub-National Populations, Ethiopia Data Stories). This section allows community members to submit comments about the dashboard and data and these commits are sent to members of the GBADs Informatics team (via *Slack*) who moderate the comments for appropriateness (before the comment is visible to the public) and act if warranted. This mechanism saves all comments to a database that can then be analysed to gauge community involvement and to improve the dashboards and data.

- **32. Training exercises on GBADs interfaces, including exercises and examples of updating parameters, accessing data and interpretation of results.**

This activity was limited by the illness and passing of Informatics co-PI, Dr. Theresa Bernardo, who had taken the lead on this but the Knowledge provides a full documentation site (<https://gbadske.org/docs/Welcome/>) including the Data Governance Handbook, a section on presentations and publications (<https://gbadske.org/publications/intro>), and our YouTube channel (<https://www.youtube.com/channel/UCdm3CD5v8YZdHtXbsq5WdWw>).

- **34. Refactoring of the GBADs code for the Dynamic Population Model (DPM) that is a basic component of the AHLE.**

The design and structure of an existing computer code base for the Dynamic Population Model (DPM) has been improved, without changing its external behaviour (i.e. output information is the same). The original code was produced as a collaboration between GBADs modellers in Liverpool and Ethiopia and is used as part of the system to estimate animal health loss estimate (AHLE) in Ethiopia. This code refactoring as allowed the Informatics team to detect some inconsistencies and errors in the original code (in collaboration with the original authors) and enact corrections. This has provided another layer of validation to the AHLE modelling work. The code base has also been reduced in size and complexity to allow for increased maintainability and extensibility. This new code base will form the basis for a standardised implementation for the DPM for other countries and production systems.

Other Observations

- Engagement with external consortium partners
 - There have been many interactions with developers at First Analytics. Informatics provided a repository in the GBADs GitHub so that First Analytics and the AHLE modelling team at Liverpool could store and present their code and data for their dashboards. There were fruitful discussions conducted on Python coding issues and how to containerize their work for presentation on the AWS Cloud infrastructure. First Analytics were responsible for their work and how it was presented but Informatics always provided support and requested information in a timely manner. This level of timely support and interaction was also evidenced by the speed in which the Senegal data eco-system and the Senegal AHLE dashboard were stored, deployed, and maintained by Informatics. Even though the Senegal work is not directly part of GBADs

(funded by Canada's IDRC) it does demonstrate the agility of the Informatics team for adding and maintaining new software products to the Knowledge Engine.

- A meeting with the aLive project (also BMGF sponsored) has laid the groundwork for future collaboration on Ethiopian data.

Publications

- Data Governance Handbook (available at <https://gbadske.org/docs/Data-Governance-Handbook-for-GBADs/intro>)
- 1 paper accepted for publication in the WOAHS Sci and Tech Review: GBADs Informatics Strategy, Data Quality, and Model Interoperability
- 1 paper in revision (*Identifying Inconsistencies in Data Quality Between FAOSTAT, WOAHS, UN Agriculture Census, and National Data*)
- 1 paper in draft form (*A Microservice Ontology Architecture for Livestock Data*)
- 2 papers accepted for oral presentation at ISVEE 2024 (*A novel approach to data quality using multiple open datasets: measuring internal and external data consistency in livestock population data; Streamlining the Discovery and Interoperability of Livestock Data and Classification Using Metadata-Driven Graph Databases*)
- 1 paper accepted as a poster at ISVEE 2024 (*A Semantic Architecture for Livestock Population Data and Modelling*)

B.4. Capacity-building efforts/activities

If your team contributed to capacity-building efforts in any aspect, please provide details here. Kindly mention the role your team played in the capacity-building effort/activities. This pertains specifically to activities funded by the Foundation and FCDO

- National/country level
 - Capacity-building activity and role in the activity
 - *Please indicate whether there was collaboration with other sub-grantees and specify their role in the activity as it occurred during the investment period.*
 - Achievements
 - Any capacity-building tools developed
 - Publications
 - Reports
 - Any key feedback to be highlighted or any comments provided by stakeholders to highlight
- Global or regional level capacity-building (e.g., SVEPM workshop)
 - Capacity-building activity and your role in the activity
 - GBADs Informatics sent the Technical Manager, Kassy Raymond to the 2023 workshop in Addis Ababa to make contacts amongst the Ethiopian community and to collect observations about the GBADs Knowledge Engine products (particularly dashboards). This workshop allowed Kassy to interact with Knowledge Engine users and to hear what they liked and did not like about the interfaces, products, and data. This was extremely helpful in establishing connections to actual users and their comments about the interfaces and dashboards (particularly the need for simplification) allows Informatics to produce second versions of many dashboards that incorporated their comments.
 - Achievements
 - Any capacity-building tools developed

- Publications
 - Reports
 - Any key feedback to be highlighted or any comments provided by stakeholders to highlight
- Technical guides
 - Use of the guides
 - Guide worked on (e.g. EV, AHLE, attribution)
 - *Please make sure to indicate whether there was collaboration with other sub-grantees and their role in the activity as occurred during the investment.*
 - Any difficulty(ies) or challenges you would like to highlight
 - Any lessons learned

C. Operations of GBADs

This section presents an opportunity to offer insights, recommendations, and lessons learned concerning the operations of GBADs. It provides an opportunity to look at the impact of aspects of GBAD operations on both your immediate team and the broader program.

C.1. Governance structures developed and implemented

- Phase IIa
 - Executive Committee (ECM)
 - Operational Management Committee (OMC), if a participant
 - Internal Steering Committee (ISC), if participant
- Phase IIb
 - Forum (former ECM)
 - ECM (former OMC), if a participant
 - Internal Steering Committee, if participant

Informatics Input

- The governance structure worked well in Phase IIa but changes in Phase IIb substantially improved governance. The addition of the Forum structure helped greatly with creating a sense of community and team problem solving by allowing for in depth discussion of specific topics.

C.2. Mechanisms, systems, and processes designed and activated to support delivery

- Working structure
 - Thematic groups (Phase IIa)
 - Working groups (Phase IIb)
- Technical review process
 - Publications review process
- Programme moniToRsing and tracking

Informatics Input

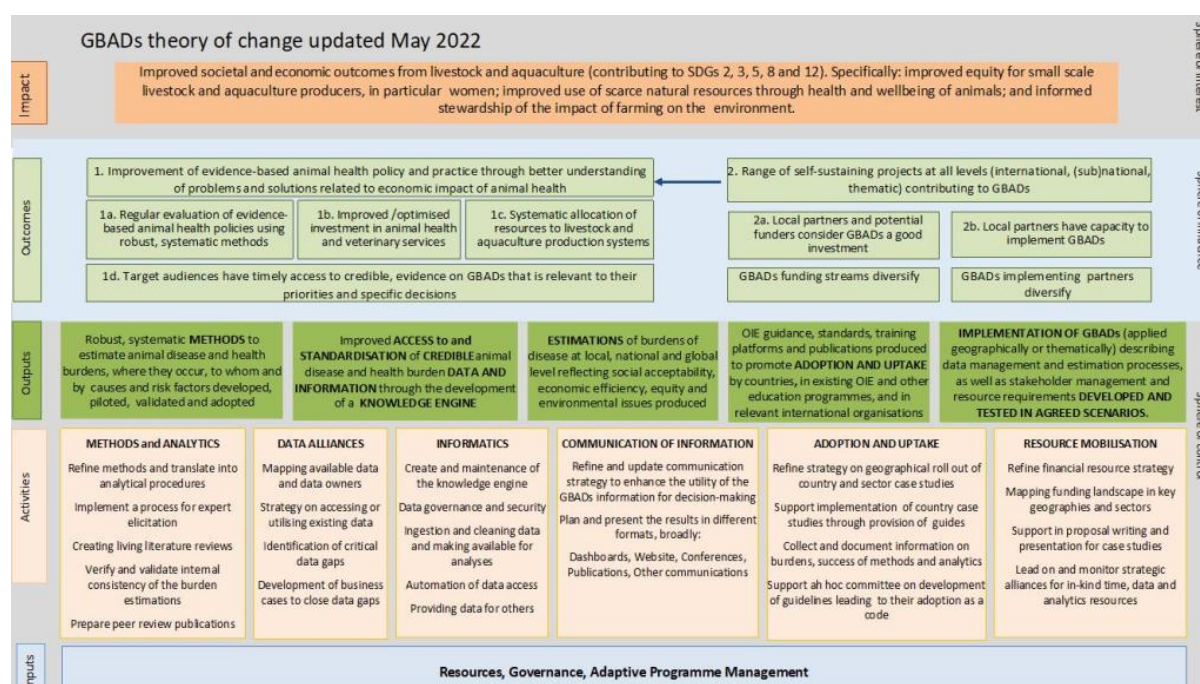
- The working structure was satisfactory in both phases although it would have been better if there had been a dedicated theme or working group that helped organize promotion,

training, and media for GBADs. This activity is large and was handled by the Liverpool team of Hannah Davies and Mike Cooper. Hannah and Mike needed much more support, particularly with regards to the needs of scientific and in country teams. It would have been better if a theme/working group called Education and lead by an expert in Tech Transfer had of been developed to make use of the talents and hard work of Hannah and Mike.

C.3. Internal and external communication

D. Reflections on alignment with overall programme goals and objectives

This section provides an opportunity for each sub-grantee to reflect on their contributions to the outcomes outlined in the program's Theory of Change and to the programme mission based on deliverables within each investment. As the close of an investment, it is important to recognize the significance of aligning our efforts with the program's overarching goals and objectives. This section provides an opportunity for all sub-grantees to provide their thoughts and perspectives on this.



our **MISSION**

The Global Burden of Animal Diseases (GBADs) programme will “measure to improve” animal health at a local, national and global level.

Our livestock keepers, consumers and the environment need:

- **Investment plans** for animal health systems
- **Allocation of resources** to manage animal health problems
- **Evaluation** of animal health investments

Contributions of Informatics to the GBADs Mission

- Data is the basis for many of the GBADs activities and so interaction with all members of GBADs is essential to understand the data needs for all teams. There have been wonderful interactions between Informatics and other teams, particularly with those in Australia, that have provided a template for all future interactions. These interactions across teams became better and better as the project progressed.
- Communications are essential for getting the GBADs message out to partners and potential users. The monitoring of the Knowledge Engine shows that over the past year, the GBADs dashboards have been visited 9775 times from 94 unique countries.

E. Conclusion