

Statistical Working System

Phase III

Project Initiation Document

31-Aug-2018

|  |  |  |
| --- | --- | --- |
| Approved by | Date | Signature |
| Laurent Thomas  Project Executive |  |  |
| Samuel Varas  Senior Supplier IT |  |  |
| Kostas Stamoulis  Senior Supplier Statistics |  |  |
| Jose Rosero Moncayo  Senior Supplier Statistics |  |  |
| Pietro Gennari  Senior User |  |  |

# Executive summary

Phase III of the Statistical Working System aims to improve the IT platform by implementing the requirements considered as high priority by the stakeholders, at a cost of $513,000 by the end of December 2018, delivering an improved input system and performance improvements, a new questionnaire manager, a new Admin Console for managing the system, improved visualization and analytics, a workflow management tool to allow users to better manage their processes, and improved management of reference data, master data and metadata. Optionally if funds are available, for a further $160,000 the project will deliver by the end of June 2019 additional functionality to manage questionnaire campaigns and on-line questionnaires, SDMX browsing, and migrate the system to the Amazon Cloud.

For the Statistical Products, by November 2019 at a cost of $410,000, the project will implement the 16 statistical processes identified in the business case. Optionally if funds are available, by December 2019 and by hiring additional resources, at an additional cost of $224,000, the project will implement an updated list of 26 statistical processes identified by the Inter-departmental Working Group on Statistics.

There is an additional $50,000 required for management and service costs, such as requirements gathering facilitation, licensing, etc. A detailed breakdown of the costs, timeline and deliverables are provided further down in this document. The following table summarizes the costs and delivery below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Option** | **Description** | **Mgmt & Services** | **IT Platform** | **Statistical Products** | **Total**  **(USD)** |
| A | IT platform high priority requirements + Statistical products original scope | 50,000 | 513,000  (Dec 2018) | 410,000 (Nov 2019) | **973,000** |
| B | IT platform all requirements + Statistical products new scope | 50,000 | 673,000 (Jun 2019) | 634,000  (Dec 2019) | **1,357,000** |
| C | IT platform high priority requirements + Statistical products new scope | 50,000 | 513,000  (Dec 2018) | 634,000 (Dec 2019) | **1,197,000** |
| D | IT platform all requirements + Statistical products original scope | 50,000 | 673,000 (Jun 2019) | 410,000 (Nov 2019) | **1,133,000** |

# Background

FAO has a generic statistical mandate stated in Article 1.1 of its Constitution. Statistical activities in FAO include the development and implementation of methodologies and standards for data collection, validation, processing and analysis. FAO statistical work has historically been developed under a decentralized model, with technical units collecting, processing and disseminating data using their own disparate tools and systems. This approach failed to address two key factors for quality statistics, consistency and comparability. The aim of the Statistical Working System is to achieve standardization for statistical work through the software platform that provides a common backbone for statistical work having the flexibility to adapt to specific requirements of diverse statistics workflows. The platform represents the corporate tool to implement statistical standards and create cross-domain and multi-disciplinary information in an efficient and comprehensive way, providing the possibility to use the same methodologies, methods, standards and approaches for statistical processes. In addition, the platform reduces duplication of statistical tools, software systems, and data repositories across the Organization as well as inter-connectional issues.

Phase I of the Statistical Working System project ran from 2010 to 2014 as a CapEx project under CIO’s management and focused on improving the collection and processing of national time series statistics by coordinating development of a corporate quality framework for agriculture, forestry and fisheries statistics; and implementation of a statistical working system for ESS that supported the framework. The total expenditure on the project was USD 2,334,000.

Phase II ran from 2014 to 2017 as a CapEx project under ESS’s management and focused on building upon Phase I achievements through deployment and extension of the platform to cater for FAO core statistical domains, development of common data processing conventions and common definitions and leveraging statistical coordination. The total expenditure on the project was USD 1,388,835.

Annex 5 lists the statistical processes that were fully implemented in Phases I and II of the Statistical Working System projects.

*The Statistical Working System – Phase III* project aims to enhance this fundamental Organizational capability through the further development of a corporate platform to manage and harmonize statistical work across the Organization automating a set of statistical processes and providing new functionalities. The project scope consists of (a) the development of major enhancements to the Statistical Working System (SWS) Platform to manage all FAO statistical processes, and (b) additional implementation of key statistical processes that generate data outputs into the platform. The list of statistical processes to be included in the project has been identified by the Inter-Departmental Working Group on Statistics (IDWS) with technical units. It is important to note that only the selected statistical processes will be within the scope of this project. These are listed in subsequent sections of this document. Additional processes may be included in subsequent phases based on interest of statistical units across FAO and fund availability.

Phase III was initiated in September of 2017, but due to an overlap with the end of Phase 2 and subsequent handover of responsibility for the project (from ESS to CIO), work only began in November 2017. The project is now being undertaken under the leadership of the IT Division (CIO) with the principal stakeholder being the Statistics Division (ESS). The budget requested in the Outline Business Case was for USD 900,000 and this project document estimates an increase to deliver the IT platform, and potentially also the statistical work if adhering to the recommendations of the Inter-departmental working group on Statistics (see Annex 4).

# Objectives and Scope

The objectives of Phase III address two areas of the Statistical Working System, which are the IT and Statistical Products, which are presented each in turn below.

# 2.1 IT Products

The evolution of the IT platform will address gaps in functionality lacking in the Statistical Working System related to flexibility, scalability and usability. High level deliverables were specified in the Outlook Business Case submitted to the CapEx Board in 2016. Given the high-level nature of the requirements expressed therein, a workshop was held with stakeholders on 11-12 July 2018 to further define these requirements and prioritize them using the MoSCoW method (Must have, Should have, Could have, Wont’ have). Please see Annex 7 for the documented outcome of the workshop, which found that some of the deliverables were considered fundamental to the success of this project whereas some others were considered less important. Therefore, the objective of this project is to implement the IT requirements that were considered “Must have” by the end of 2018, with “Should Have” and “Could Have” requirements being proposed for completion in the 1st half of 2019.

Figure 1 below shows the existing high-level architecture of the SWS, which can be contrasted with figure 2a showing the proposed high-level architecture for the platform after completion of Phase III expanding on the Input System to an integrated platform of focused application modules.

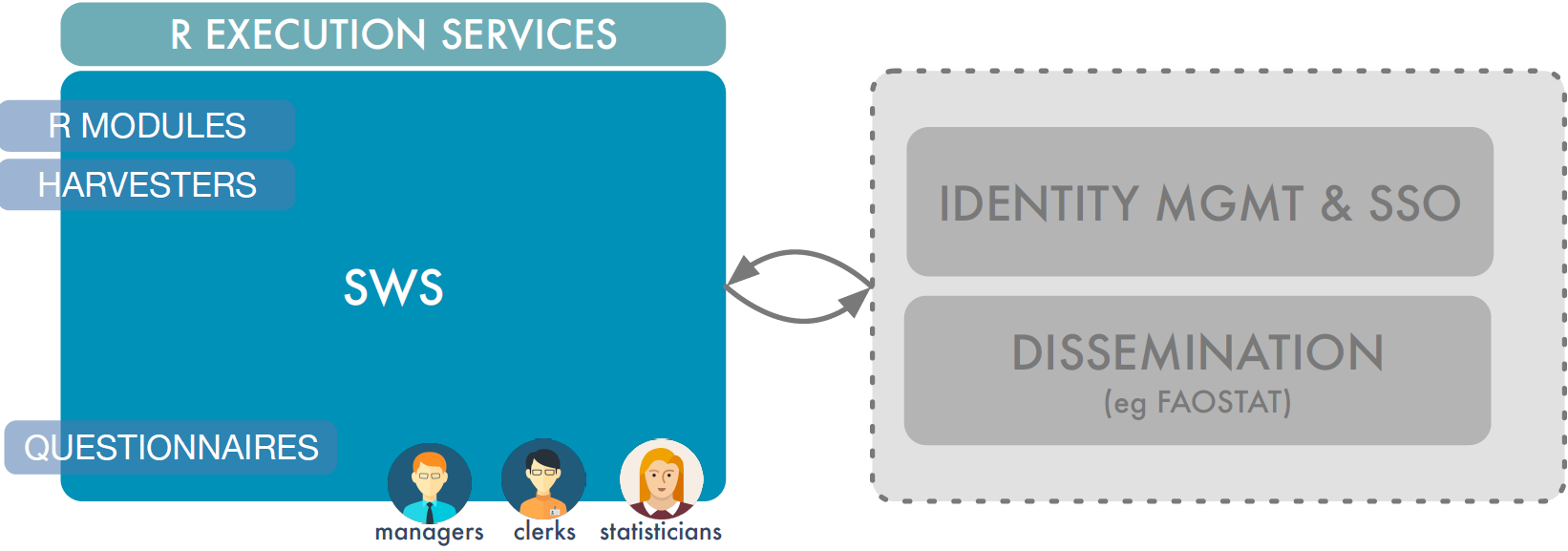


Figure 1. High level architecture of SWS after Phase II

Figure 2a represents the new architecture if all requirements of the IT platform are implemented. Figure 2b represents the architecture if the high priority “Must have” requirements are implemented.

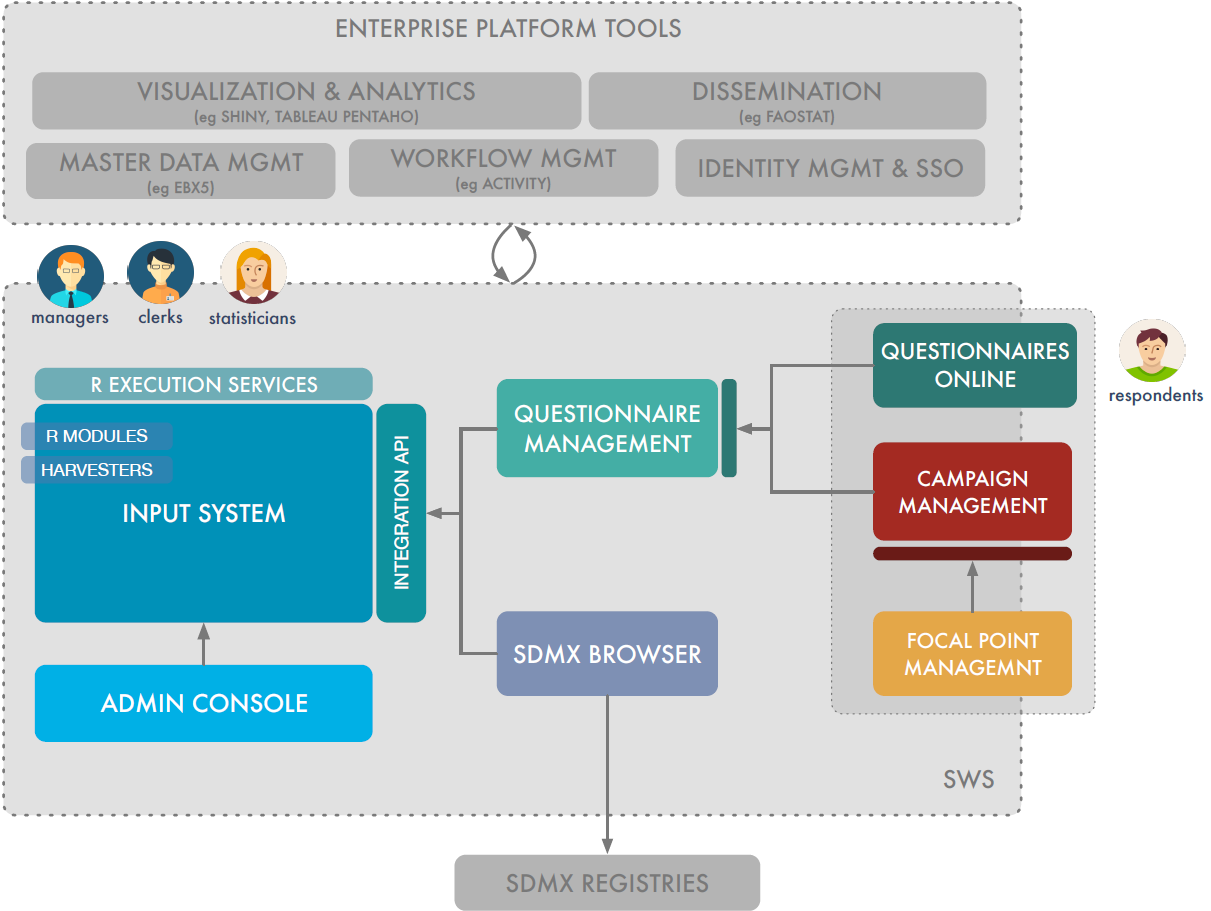


Figure 2a. Proposed High level architecture for Phase III IT work – all requirements

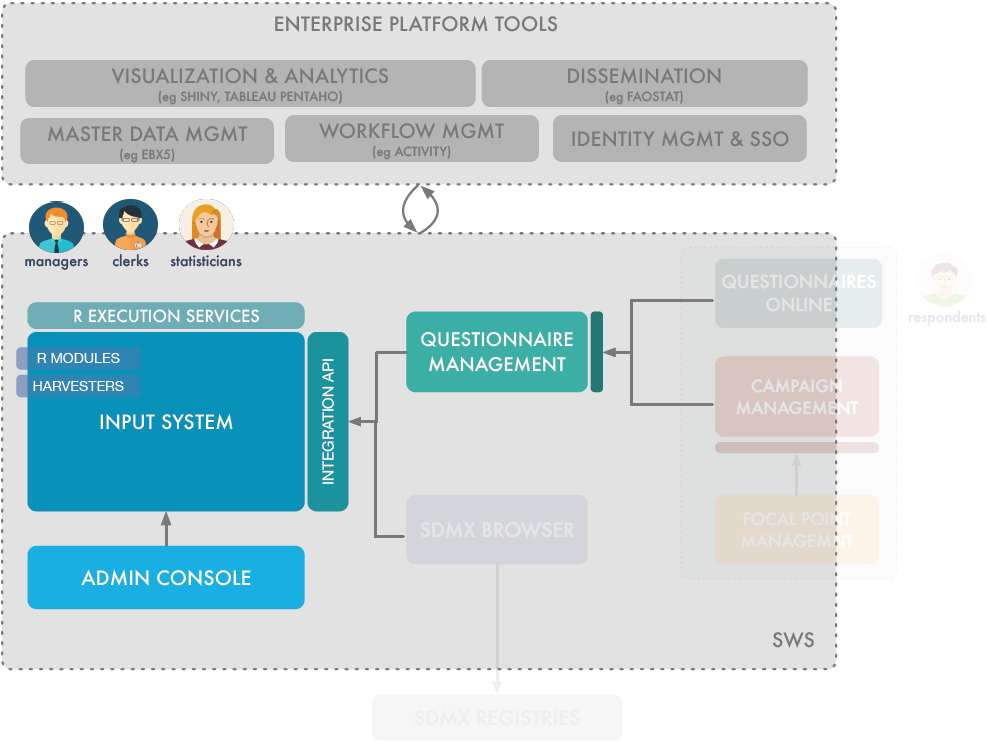


Figure 2b. Proposed High level architecture for Phase III IT work – high priority requirements

In the proposed architecture, the end product will be a combination of in-house and bespoke developed modules (Input system, Admin Console, SDMX Browser, Questionnaire Manager) integrated with existing FAO Systems (e.g. FAOSTAT for dissemination, EBX5 for Master Data Management) and COTS (Commercial off-the-shelf products) for Workflow, Visualization and Analytics, and Campaign and Focal Point Management. A series of Integration APIs (Application Programming Interfaces) will allow tight integration between the input system and the external products.

For all of the functional improvements listed below, more detailed information can be found in the workshop findings in Annex 7.

A number of enhancements and performance improvements will be made to the Input System and R Execution services (Must have). The workshop held with stakeholders highlighted the fact that the current version has several issues which are particularly felt when dealing with large datasets and large codelists, and unless these problems are resolved the benefits of introducing other functionality will not be fully leveraged. Improvements to Metadata Management: (Must have) will also be implemented in the Input and related systems.

The Admin Console (Must have) will allow end users to define and control in autonomy the diverse statistical objects that underpin their statistical workflows, thereby reducing dependency on IT staff in future.

The new Questionnaire manager (Must have) will replace the limited support for questionnaire generation which is available in the Input System with a more flexible system that can be more generically applied to all statistical domains, through interactive design and generation of offline (XLS-based) or online (HTML-based) questionnaires. Online Questionnaires (Could have) are a public extension of the Questionnaire Manager to allow correspondents from member countries to prepare and submit questionnaires online.

The SDMX (Statistical Data and Metadata eXchange) Browser (Should have) builds on domain standards allowing statisticians to browse and interactively query the data assets of remote SDMX Registries, eventually collecting relevant subsets in the Input System.

The Campaign Management and Focal Point Management (Should have) will facilitate the planning, execution and monitoring of data collection campaigns based on the questionnaires generated by the Questionnaire Manager.

Integration with a Master Data Management solution EBX5 (Must have) will harmonize the management of master and reference data for defined domain values, standardized terms, classifications and categorizations of data, such that these “sources of truth” outside the SWS can be reutilized across other systems in FAO, reducing duplication.

A platform offering Statistical Workflow Management and Reporting: (Must have) will also be integrated through the APIs, allowing sequencing and monitoring and better orchestration of statistical processes. The platform must allow users to configure the workflows, approvals and reporting, and within scope of this project is to implement workflow management for two statistical domains. Workflow for all other domains will be an operational task, not charged to this project.

Finally, the Integration APIs will permit other 3rd party tools for i) Visualization and Analytics (Must have) to work on data in the SWS in more sophisticated ways than available in a platform focused on preparing statistics for consumption downstream; and ii) perform data dissemination (Must have) to systems such as FAOSTAT.

# 2.2 Statistical Products

Phase III will also enhance the statistical content of the system by:

1. Incorporating new statistical domains: at the end of Phase II, the SWS hosts datasets from Statistics Division, and the Fisheries and Forestry departments. However, the latter uses the SWS as a data repository more than a statistical working system. The aim at the end of Phase III is that the statistical processes of several departments (Fisheries, Forestry, Climate Change and additional domains of Statistics Division) are provided by the SWS. Processes include imputation and validation routines.
2. Enabling the management and integration of multiple data sources: Questionnaire data can be incomplete because of country non-response or gaps in the questionnaires. Additional sources from national websites, on-line databases, additional data-files or manual research must be managed and merged to create a complete and integrated dataset.
3. Adding quality indicators: The SWS will compile quality indicators on relevance, accuracy, reliability, timeliness and punctuality, coherence and comparability, accessibility and clarity to meet users’ needs and the standards set in FAO’s Statistical Quality Assurance Framework.
4. Extending FAO’s implementation of Statistical Standards to the new datasets: FAO developed a set of statistical standards covering units of measure, flags, metadata, imputation methods, data revisions, terminology and definitions. Adopting the standards will be easier for datasets in the SWS, where the standards are already applied and reference files, as well as statistical routines, can be shared across domains.

As previously noted, not all statistical processes of FAO are within the scope of this project. The statistical process to be included in SWS Phase III were selected by the Inter-Divisional Working Group on Statistics (IDWS) and are shown in Table 1 below. Upon completion of the project, end users will be empowered to implement further statistical datasets and processes according to their needs.

Table 1. Statistical Processes within Project Scope

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Division/  Department | Domain of Statistical Subject | Dataset | Estimated Effort (weeks) | Estimated Cost (USD) |
| All Technical Departments/ Divisions | Data capture and processing to support the generation of the 21 indicators for which FAO is the Custodian Agency | SDG Monitoring Indicators | 36 | 45,000 |
| FIAS | Additional fisheries statistical dataset on production and trade | [[1]](#footnote-1)Aquaculture / Capture  Trade | 80 | 100,000 |
| AGL | FAO’s global water database. Covers data and information by country on water resources, water uses and agricultural water management. | Aquastat | 25 | 31,250 |
| OSP | Indicators gathered as part of the M&E process | Strategic Objective level indicators  Outcome level indicators | 40 | 50,000 |
| ESS | Economic: the domain is concerned with the collection and processing of producer prices of agricultural commodities | Producer prices in local and international currencies(annual and monthly)  Producer price indices | 31 | 38,750 |
| ESS | Environment: the domain will cover different and heterogeneous topics | Fertilizers, Pesticides, Land Use  Agri-Environment Indicators  GHG Emissions Agriculture | 54 | 67,500 |
| FOAD | Data capture and processing top support the dissemination of annual production and trade statistics for forest products | Production and Trade[[2]](#footnote-2) | 49 | 61,250 |
| ESS | Food Balance Sheets for 2018 |  | 13 | 16,250 |
|  |  | **Total:** | **328 weeks** | **$410,000** |

However, since the time that the Outline Business Case was approved for Phase III, the Inter Departmental Working Group on Statistics has revised the list that should be incorporated in Phase III. This change of scope would raise the cost and is therefore estimated in Annex 4. For reference and future planning purposes, Annex 6 lists all other statistical processes known at this time to the Inter Departmental Working Group on Statistics, that have not been listed in Table 1 above or the potential revised scope in Annex 4.

Given that statistical processes can in some cases take months to complete from start to finish, each one will be tracked according to the following list of sub-deliverables in order to ensure accurate progress can be reported:

1. Analysis
2. Requirements collection and definition
3. Methodology Proposition (requires stakeholder signoff)
4. R Script Development
5. R Script Testing
6. User validation small dataset
7. User validation full data set (requires stakeholder signoff)
8. Plugin-in
9. R Script Documentation
10. Statistical Process documentation (requires stakeholder signoff)
11. Aggregates
12. Data quality indicators

Sub-deliverables 3, 7 and 10 will have stakeholder signoff for each statistical process to ensure full accountability of the delivery.

In addition to these statistical processes, the project scope considers a cross-divisional work under: (a) Quality indicators included in each process; and (b) Implementation of Statistical Standards to the new datasets and processes.

The following list outlines items not in scope of SWS Phase III:

1. Direct data dissemination, though provides APIs and facilitates the data dissemination task of End Users
2. Workflow management, however provides APIs for integration with Workflow management system
3. Data reporting however there is a possibility of adding sub-modules developed in R.
4. Not include the following divisions, because they have no statistical processes: TCR, DPS, SP1-4, SSC, CSDM, CSF, CSAI, CSAP, DDOS, OHR, LEG, OIG, RAF, RAP, REU, RLC and RNE.
5. Support activities for both the SWS Platform and R modules after the end of the project.
6. Providing R development course for End Users.
7. Statistical Year Book (developed in R)
8. Statistical Quality Assessment and Planning Survey (QAPS)
9. Statistical Quality Assurance Framework (SQAF)

# Project deliverables

The project has the twelve deliverables documented in the table below.

Table 2. Project Deliverables

|  | **Deliverable** | **Description** |
| --- | --- | --- |
| A | Alternative methods for data collection | The SWS already offers a solution for automatic generation and import of questionnaires: however, this solution cannot be generically applied and is therefore inadequate for a number of domains. These new functionalities will not only cover the generation of XLS-based surveys, but will be expanded to include other functionalities such as: online forms, csv uploads and SDMX data collection from external agencies. |
| B | Quality Indicators | As part of the commitment of FAO to produce and disseminate statistical outputs, it is essential to ensure that statistics are disseminated to a level of quality that meets the users’ needs, and that users are informed about the quality of statistical outputs. The SWS will automatically generate a number of indicators on the quality of the data submitted by countries and on the statistical processes implemented at FAO, such as: accuracy and reliability; timeliness and punctuality; coherence and comparability; and accessibility and clarity. |
| C | Statistical Workflow Management and Reporting | The possible integration of the SWS with services offered by CIO, in particular the service bus, will allow users to design patterns of statistical activities, organize resources into statistical processes and effectively monitor their progress. Reporting functionalities will enable user to have an instantaneous glimpse of key aspects of the process according to adequately defined indicators, such as: overall process completion; number of questionnaires loaded and processed; number of modules executed and so forth. |
| D | Administration module | New functionalities must be easily configurable by end users. The overall usability of the system will gain great benefit by having a separate dedicated administration module; furthermore, this it will accrue the autonomy of the user to define objects composing their statistical workflows. This new feature will contribute also to cut the overall costs for system maintenance by reducing the numbers of IT resources dedicated to object configuration. |
| E | Focal Point/ Questionnaire Management | At present, statistical teams do not have a common way to manage addresses of responsible contacts from international organizations, government bodies and FAO representatives in the various countries. A new solution will be implemented to simplify the management. Furthermore, as requested by the IDWG Technical Sub-Group on Data Collection, this module will act as a corporate repository for storage and search of questionnaires used in different Divisions for statistical data collection. The workflow relationships/needs between focal points and questionnaires will addressed as part of package H, which is a specific new module designed to address statistical workflow issues. |
| F | Improved mechanisms for machine-to-machine data sharing | The system lacks of an efficient interface for data exchange. Some steps have already been taken in Phase II to fill this gap. However other functionalities have been requested during the latest development: data interchangeability using SDMX, the integration with other corporate systems for data collection and data dissemination, for example: FAOSTAT |
| G | Integration of reference data and comprehensive management of metadata | Reference and comprehensive metadata management was not included in the scope of Phase I and II of the SWS project. Therefore, the current system does not provide a complete mechanism for control of defined domain values, standardized terms, classifications and categorization of data. At the same time, there is a need to adopt more structured metadata to strengthen the data exchange functions as described in deliverable F. While CIO has proposed a potential solution to this problem, it will in any event require integration into the current system as well as the development of some localized features/functionality. |
| H | Comprehensive documentation targeted for different roles | Documentation in the previous phases focused mainly on online application help and methodological description, which has proven to be insufficient to meet the demands of a system that has many different facets and types of users. This will not be sufficient in the future and actually already showed evident limitations. It is therefore necessary to develop a comprehensive set of documentation targeted for the different roles and actors using the system. For example to compile a user manual for clerks; an administration manual for application administrators; a developer manual for R programmers; and a business process and workflow description for all types of users. The documentation should also be fully accessible, searchable and integrated within the application. |
| I | Ongoing improvements to the system | Activities to further enhance the features and functions of the system are required to ensure that current and future corporate needs are fulfilled and that the system remains operational in an efficient manner and is compliant with users’ requirements. |
| J | Inclusion of highest-priority statistical processes | The highest-priority statistical activities reported in the Statistical Programme of Work 2016/17 (SPW) will be implemented in the Statistical Working System during Phase III. The list of these processes is provided in Table 1. |

# Methodological Approach

The proposal methodology to deliver this project is based on the agile software development practice. This consists of iterative development, where requirements and solutions evolve through collaboration between self-organizing cross-functional teams (statistic and IT). In particular, we propose to use Scrum methodology with online tools for our development teams and customers to track, manage and monitor project progress. Therefore, entire team (statistical, IT and stakeholders) follows the same roadmap, everyone is communicating effectively in terms of where they stand each day in relation to the target date.

Even though the adopted model is based on the agile methodology, a characterization, identification and description of the necessary requirements to complete the platform will be carried out. This process will be the first activity of the IT team, and its objective is to clarify and define the scope necessary for the correct completion of the tool. This activity must also be validated and signed by all those stakeholders currently and in the future. To carry out this activity, a first Design Thinking[[3]](#footnote-3) exercise is proposed where the stakeholders will participate to identify what is expected by the platform, then the team will work on the specification of the requirements, presenting them to the stakeholders, to end with the sign off of them. by the project team and stakeholders.

This first activity from the IT team will allow: (a) alignment of expectations between the stakeholders and the IT team; (b) clarify the scope of work that must be developed; and (c) realistically plan the development of the project.

Note that the project follows the approach defined by Generic Statistical Business Process Model[[4]](#footnote-4) (see Figure 3), and under this model the SWS only covers a part of the aspects of this process. However, this model is still generic, and its characterization to the needs of FAO must be carried out in an adequate manner. This is where the proposal of Design Thinking with all stakeholders makes sense as an exercise in alignment and necessary limitation.

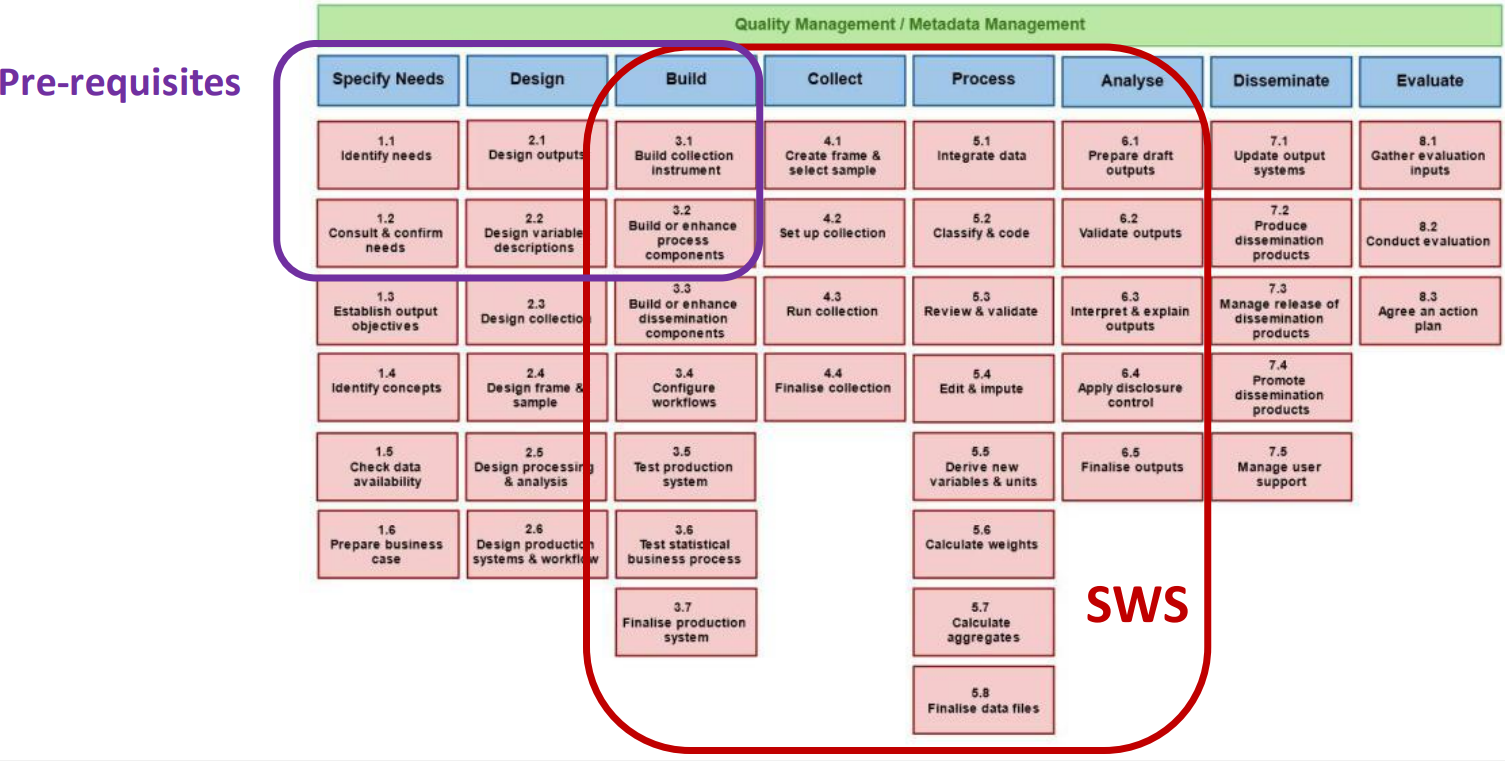


Figure 3. Generic Statistical Business Process Model

Below is detailed the methodology that each team will follow and the mechanisms of coordination and alignment necessary for the adequate achievement of the project's objectives.

# 4.1 Statistical Methodology

The following chart is the Generic Statistical Business Process Model adopted to FAO and to the SWS Project (Figure 4):

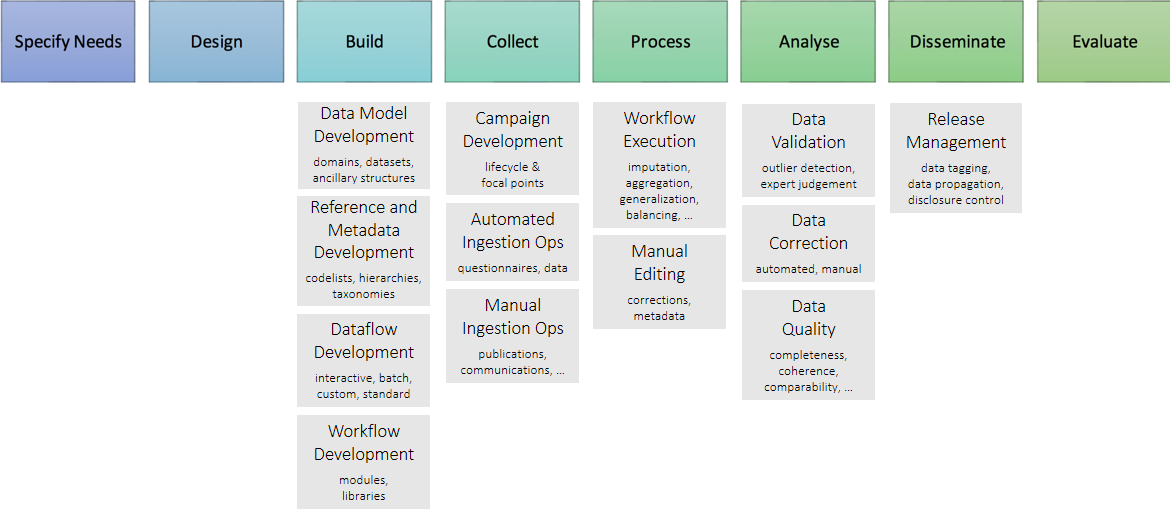


Figure 4. Generic statistical business model adopted

The following chart shows the actual status of SWS Platform in terms of covering activities associated to the Generic Statistical Business Process Model (Figure 5):

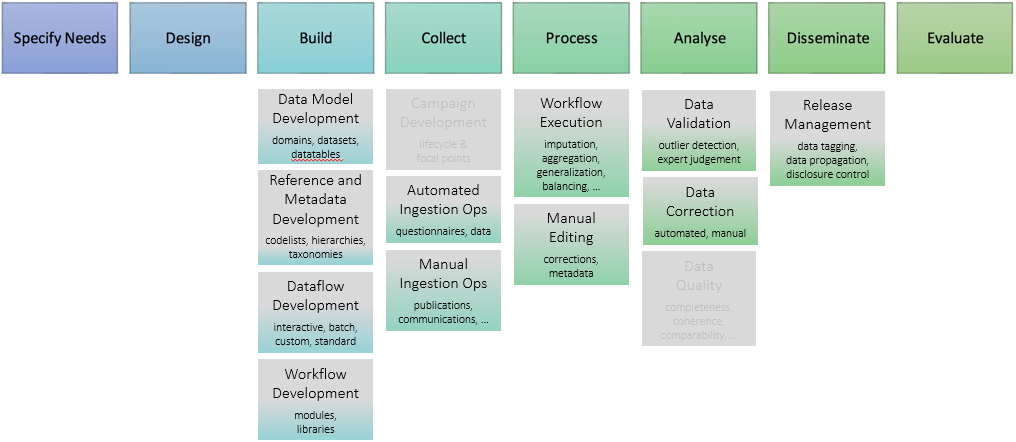


Figure 5. Generic business process model status

Note the fullness of the boxes indicates the current readiness of the area.

# IT Methodology

The proposal methodology to deliver this project is based on Agile approach (software development methodology) based on iterative development, where requirements and solutions evolve through collaboration between self-organizing cross-functional teams (statistic and IT). Agile methods or agile processes generally promote a disciplined project management process that encourages frequent inspection and adaptation, a leadership philosophy that encourages teamwork, self-organization and accountability, a set of engineering best practices intended to allow for rapid delivery of high-quality software, and a business approach that aligns development with customer needs and company goals.



Figure 6. Scrum methodology

Scrum is a subset of Agile. It is a lightweight process framework for agile development, and the most widely-used one. Scrum is most often used to manage complex software and product development, using iterative and incremental practices. Scrum significantly increases productivity and reduces time to benefits relative to classic “waterfall” processes. Scrum processes enable organizations to adjust smoothly to rapidly-changing requirements and produce a product that meets evolving business goals. The main benefits are:

* Increase the quality of the deliverables
* Cope better with change (and expect the changes)
* Provide better estimates while spending less time creating them
* Be more in control of the project schedule and state

# Coordination Methodology

Project Manager needs to ensure Scrum techniques are applied throughout the entire project. Project Manager facilitates the flawless team work, encourage to be self-organized, effective and productive; in addition to feel ownership and accountability for the final products. Project Manager constantly balances between the team stay focused while delivering and to be flexible to adapt for changes.

Project Manager should create trust in team members. Project Manager should raise collective wisdom therefore independent team members should accept diversity of opinions and the aggregation of the opinions should happen naturally.

Project Manager has to assure the team activities will take place including sprint planning, retrospective activity, team health check, team building activities, team motivation and provide constant feedback in order to create synergy.

# Timescales

See Annex 1 for the GANTT chart of the planned schedules for the Statistical Work Deliverables. In addition see Annex 4 which provides a GANTT chart for a potential revised scope of the Statistical Work proposed by the Interdepartmental working group on Statistics.

Regarding the IT Platform, Annex 2 shows the plan required to deliver the requirements considered “Must haves” in the workshop held with stakeholders in July 2018 (see Annex 7), with completion in Dec 2018. Annex 3 shows the plan required to deliver the other requirements, not considered “Must haves” in the workshop, to be delivered in the first half of 2019.

# Major Risks

Table 3. Main risks identified

|  |  |
| --- | --- |
| **Risk** | **Mitigating actions** |
| The lack of a technological (IT) vision | After completion of Phase II the result is a home-made IT platform with all the issues (bugs, integration problems, and high cost of supportability) that have been experienced to date as well as an out-dated user interface/user experience. The workshop held with stakeholders in July 2018 revealed that there are many improvements required. Analysis of the current platform by the IT team suggests that although the current technology is not the most modern, it can be updated to meet the new requirements. A Solution Architect will be hired to make recommendations on the platform and the new requirements (e.g. workflow, visualization). The findings may proposed a change of approach, by using Commercial off-the-shelf solutions or a different development stack, which would require a significant re-scoping of the work. |
| Requirements of IT platform not fully articulated | Some of the requirements for the IT platform are expressed at a high-level only, and therefore the estimates for effort could increase significantly once they are fully understood. An example of this is deliverable C. Statistical Workflow Management and Reporting and it is also known that there is a requirement of the user to support the product SHINY used for data analysis and visualization not currently possible in the Statistical Working System. The 2-day workshop with statistical process owners to scope the requirements of the IT platform has clarified many of the requirements, as well as prioritizing them. However, there will need to be an additional in-depth requirements analysis of 1 month to further scope the requirements of the lesser known products (e.g. statisical workflow), which will feed into the IT platform development work, but can potentially re-scope the work required. |
| Resource constraints | New constraints for hiring consultants has already impacted the project, increasing turnover and reducing the historical knowledge of the personnel now working on the project, especially on the IT platform. A technical writer will be hired to improve the standard of documentation of the platform which is not comprehensive. Outsourced personnel will augment the current team in order to deliver the platform in a reasonable time. |
| Knowledge gap in new project team members | New personnel to be hired on the project, in particular outsourced personnel, will have a steep learning curve before becoming productive, risking the timeline. In addition, outsourced personnel will be off-site to reduce cost, which causes coordinate and communications issues. To manage this, a combination of on-site and off-site personnel from outsourced providers will be arranged, which has affected the cost. |
| Change in scope of the statistical processes to be included | With respect to statistical processes listed in the Business Case to be implemented in the SWS, the Inter Departmental Working Group on Statistics has significantly revised this list, which changes the scope of work and budget. The impact of this change is documented in Annex 4 for consideration. |
| Insufficient level of commitment and prioritization by users | Management must adequately plan and prioritize the use of the SWS within staff work plans around the organization. IDWG on Statistics can play in important advocacy role here. In the workshop held in July 2018 it was clear that in the recent past there has not been enough stakeholder involvement in the testing and piloting of new features. In addition, a significant amount of time is required of the owners and users of the statistical domains to be incorporated into the platform, averaging weeks to a few months for each process. If stakeholders are not available for the statistical methodology work and testing of the IT platform, the quality and time of the project will suffer. |
| Statistical data Dissemination | The Statistical Working System is not a dissemination system, and only provides interfaces to systems such as FAOSTAT for this purpose. There is a risk that some statistical domains in scope of this project will not be disseminated by FAOSTAT, in which case an alternative dissemination solution needs to be identified, or otherwise those domains should not be implemented in this project. |
| Sustainability | A clear handover from ESS to CIO during and after the project into Regular Programme funding. During the project to identify what should be move to production and the associated cost to operate, and at the finalization the complete handover to CIO. |
| Evolving roles and responsibilities at the corporate level | Work to clearly define roles and responsibilities between ESS and CIO is ongoing. Also, it is fundamental to have a strong and robust project structure, with both the IT and statistical methodology supplier roles formally represented on the Project Board. |

# Proposed Project Structure

The proposed structure and roles will be implemented to maintain the control and responsibility of the project delivery.

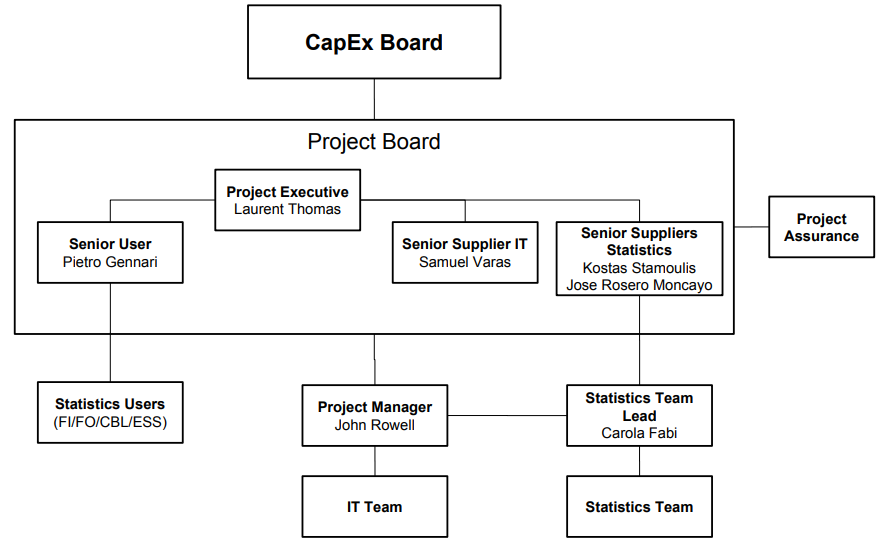


Figure 7. Project Structure & Governance

The proposed accountability and definition of the roles is presented in Table 4.

Table 4. Accountability and roles

|  |  |
| --- | --- |
| **Role** | **Role Description** |
| CapEx Board | Outside of the project, this role is responsible for the overall Project Mandate, authorizing the funds, instructing the project board to carry out the project on its behalf, and ensuring accountability of the project. |
| Project Board | The Project Board, consisting of the Project Executive, Senior User, Senior Supplier IT and Senior Suppliers Statistics, is responsible for the overall direction of the project, authorizes changes to scope, funding and timelines, and performs or delegates project assurance activities. |
| Project Executive | Final decision making on the project board decisions. Overall accountability for the success of the project. |
| Senior User | Represents all potential end users of the project outputs to the project management. The Senior User will have to liaise closely with active users, across many departments, who themselves will form a User Committee as part of the IDWG on Statistics. |
| IT Senior Supplier | Represents the work supplier. In this case limited to IT deliverables and outcomes. |
| Senior Supplier Statistics | Represents the work supplier. In this case limited to statistical deliverables and outcomes. |
| Project Manager | The project manager plans and schedules tasks and activities, oversees day-to-day execution, and monitors and report progress, raises issues and risks to the project board, brings the project to a close, and captures the lessons learned. |
| Statistics Team Leader | Responsible for ensuring the quality of the statistical outputs (statistical methods and software) and for coordinating the day-to-day activities of the team |
| Statistics Users | Responsible for collaborating with the project to adopt the Statistical Working System working closely with the Statistics Team. |

# Cost EstimationS

# 8.1 Current Budget Execution as of June 2018

In terms of budget execution and considering the USD 900,000 requested in the business case during August 2016, the current execution corresponds to USD 288,000, corresponding to 32% of the project budget. This expense corresponds to services, consultants and PSA during the years 2017 and 2018 January to June. Table 5 presents the current budget status.

Table 5. Budget Status as of end June 2018

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Budget items** | **Management & Services (USD)** | **IT Team (USD)** | **Statistical Team (USD)** | **Total (USD)** |
| **Approved** | **$50,000** | **$440,000** | **$410,000** | **$900.000** |
| Spent in 2017 | $14,000 | $20,000 | $6,000 | $40,000 |
| Spent in 2018 | $3,000 | $135,000 | $110,000 | $248,000 |
| **Total 2017+2018** | **$17,000** | **$155,000** | **$116,000** | **$288,000** |
| Available funds | $33,000 | $285,000 | $294,000 | $612,000 |
| **% budget spent** | **34%** | **35%** | **28%** | **32%** |

On taking over responsibility for the Statistical Working system platform, CIO has re-evaluated the effort to deliver the updated IT platform. In addition, a requirements workshop held in July 2018 with stakeholders (see Annex 7) revealed that there were significantly more improvements required in the current system than planned. Therefore an increase in the original estimate to complete the IT platform is anticipated, which is estimated below.

# 8.2 IT Platform high priority deliverables cost estimates

Table 6 below shows the estimated costs to complete the Statistical Working System deliverables identified as high priority (or “Must have”) requirements in the workshop held with stakeholders in July 2018.

Table 6. IT Platform – Delivery of “Must have” requirements

|  |  |  |  |
| --- | --- | --- | --- |
| **Costs for “Must have” requirements** | **Months** | **Rate** | **Total**  **(USD)** |
| Costs to date (as of June 2018) |  |  | 155,000 |
| Lead Back End Developer | 6 | 8,140 | 48,840 |
| Lead Front End Developer | 6 | 4,224 | 25,344 |
| Business Analyst/Data Manager | 6 | 8,140 | 48,840 |
| Outsourced Back End Developers | 8 | 8,800 | 70,400 |
| Outsourced Front End Developers | 8 | 8,800 | 70,400 |
| R + Shiny Expert | 2 | 13,200 | 26,400 |
| Business Analyst | 1 | 14,300 | 14,300 |
| Technical Writer | 2 | 4,400 | 8,800 |
| Solution Architect | 0.75 | 15,400 | 11,550 |
| Database Architect | 1 | 13,200 | 13,200 |
| EBX5 Support | 2 | 10,000 | 20,000 |
| **Total** |  |  | **$513,074** |

# 8.3 IT Platform optional lower priority deliverables cost estimates

Table 7 below shows the estimated costs to complete the Statistical Working System deliverables identified as lower priority requirements in the workshop held with stakeholders in July 2018. This work can be considered as optional and the additional cost would be USD 160,000, bringing the total cost of the IT platform to USD 673,000.

Table 7. IT Platform – Delivery of lower priority requirements - Optional

|  |  |  |  |
| --- | --- | --- | --- |
| **Costs for lower priority requirements** | **Months** | **Rate** | **Total**  **(USD)** |
| Lead Back End Developer | 5 | 8,140 | 40,700 |
| Lead Front End Developer | 5 | 4,224 | 21,120 |
| Business Analyst/Data Manager | 5 | 8,140 | 40,700 |
| Outsourced Back End Developers | 3 | 8,800 | 26,400 |
| Outsourced Front End Developers | 3 | 8,800 | 26,400 |
| Technical Writer | 1 | 4,400 | 4,400 |
| **Total** |  |  | **159,720** |

# 8.4 Statistical work cost estimates

The estimate to complete the statistical work of USD 410,000 is considered on target for the scope outlined in the business case. However, the Inter-Departmental Group on Statistics has requested an updated scope and this is reflected in Annex 4 in terms of deliverables and additional budget requirement, increasing the estimated cost by USD 224,000 to a total of USD 634,000.

# 8.5 Total estimated costs for 4 options

Table 8 below shows the 4 possible options for phase III of the Statistical Working System project to deliver the IT platform and Statistical products.

Table 8. Estimated costs for 4 options

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Option** | **Description** | **Management & Services** | **IT Platform** | **Statistical Products** | **Total**  **(USD)** |
| A | IT platform high priority requirements + Statistical products original scope | 50,000 | 513,000 | 410,000 | **973,000** |
| B | IT platform all requirements + Statistical products new scope | 50,000 | 673,000 | 634,000 | **1,357,000** |
| C | IT platform high priority requirements + Statistical products new scope | 50,000 | 513,000 | 634,000 | **1,197,000** |
| D | IT platform all requirements + Statistical products original scope | 50,000 | 673,000 | 410,000 | **1,133,000** |

# Quality Management Strategy

The Senior User (statistical) will ensure the quality of statistical workflows and methodologies by using peer review of methods and rigorous testing of implementations.

Key sub-deliverables of the statistical process work for each statistical process will be signed off by the owner of the data set / domain to ensure completion.

All IT deliverables will be signed off by the Senior user and Senior Supplier of the Statistics Team Lead to ensure their quality control.

# Communication Management

The work team led by the Project Manager will report periodically (weekly) to the respective Senior User, IT Senior Supplier and Statistical Senior Supplier the progress and risk in achieving the results. In addition, a monthly or quarterly meeting with the project board should be developed for evaluation.

# Implementation, Monitoring and Control

Report will be submitted to the board about the selected strategy and solution. An interim progress reports will be submitted to the CAPEX Board during the project. A final project closure report will be also submitted to the Board.

All reports will contain information on progress against the milestone activities/deliverables and expenditures compared to budget for the period covered by the report.

# Change Management

Change can arise from a very wide range of sources. Anything that could have a detrimental or beneficial effect on the Initiative or the Organization will be considered an Initiative Issue. These can include:

* Change in scope or requirements
* Change in the Organization’s environment
* Unanticipated problems
* Anticipated but unavoidable risks
* Suggestions to improve the Project’s Deliverables
* Problems occurring in work completed or currently under way
* New risks

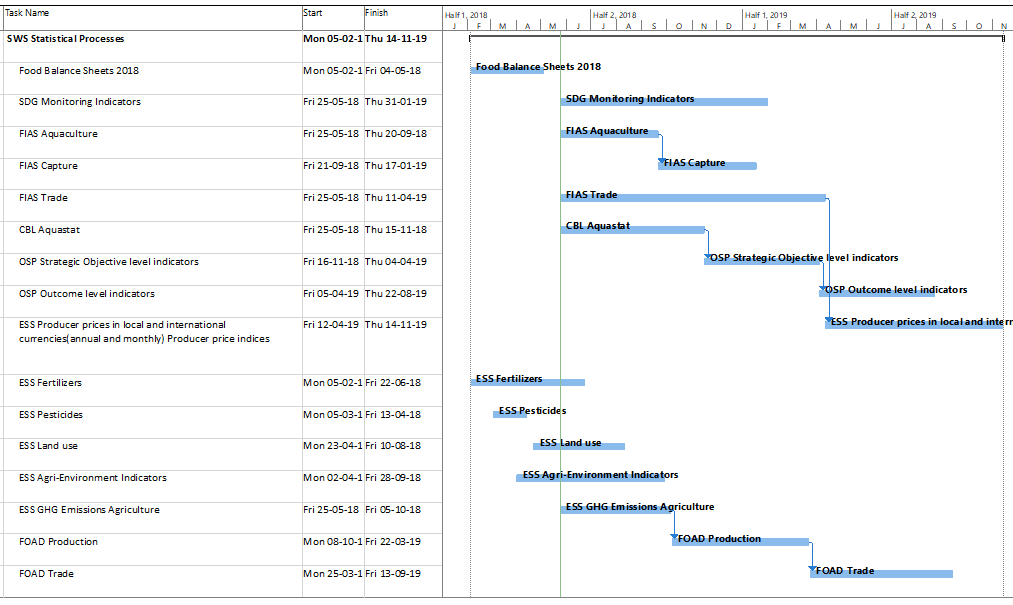
All Issues will be captured, logged and categorized in an Issue Log and an initial assessment will be made as to the nature of each issue by the Initiative Leader. The assessment will indicate a priority as follows:

* Must have – the Final Deliverable will not be accomplished without this change
* Should have – its absence would be very inconvenient, although a work-around is possible for a period of time
* Could have – a nice-to-have but not vital
* Won’t have – does not involve a change

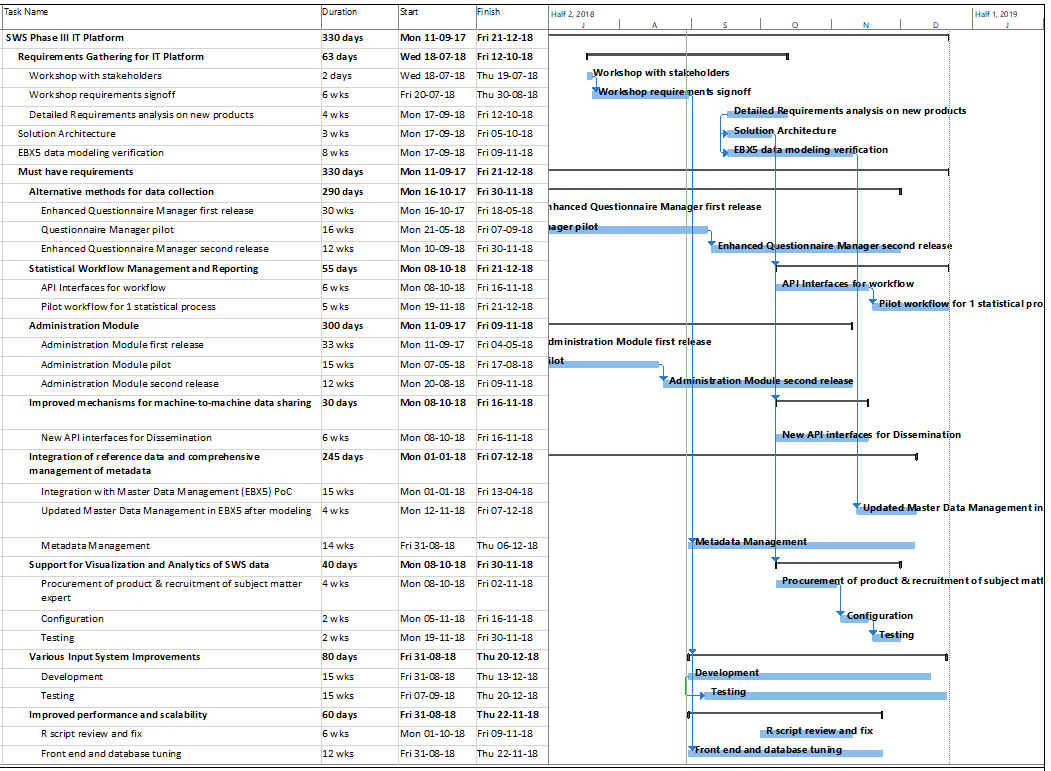
An Issue that results in a change that exceeds the cost or time tolerance of the Project will be escalated to the CAPEX Board.

**ANNEX 1. Gantt Chart of Statistical Process Work**

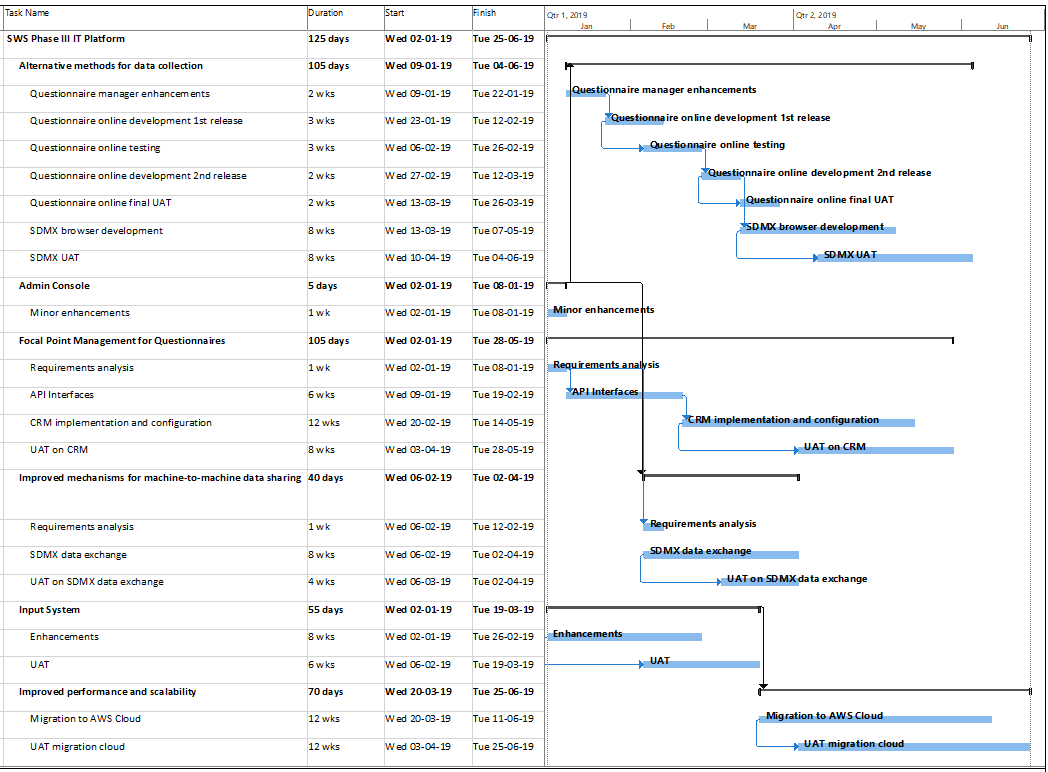
(Based on Outline Business Case)



**ANNEX 2. Gantt Chart for Delivery of IT Platform – Must have requirements only**



**ANNEX 3. Gantt Chart for Delivery of IT Platform – lower priority requirements (Optional)**



**ANNEX 4. Statistical Processes Recommended by the IDWG on Statistics in May 2018**

Since the time that the Outline Business Case was approved for Phase III, the Inter Departmental Working Group on Statistics has revised the list that should be incorporated in Phase III. This change of scope would raise the cost and time to complete the work so is estimated here in Annex 4.

The list of revised statistical processes is below, with the final column indicating those which are new or removed compared to the Approved Outline Business Case. Note that statistical processes 27-29 would not be in the revised scope.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **High priority, resource intensive statistical processes** | | | |  |
| **Div/Dep** | **Domain of Statistical Subject** | **N.** | **Datasets** | **Status** |
| **FIAS** | Data capture and processing of annual Production, Trade and Food Balance Sheets statistics of Fish and Fish products | 1 | Aquaculture |  |
| 2 | Capture |  |
| 3 | Trade |  |
| 4 | Regional Bodies | New |
| 5 | Commodities database | New |
| 6 | Incorporation of FI Commodities into FBS | New |
| **CBL** | FAO’s water database: data cature and processing | 7 | Aquastat |  |
| **FOAD** | Data capture and processing of annual Production and Trade statistics for forest products | 8 | FO – Production |  |
| 9 | FO – Trade |  |
| **ESS** | Economic Statistics. Data capture and annual processing of producer prices and government expenditure to agriculture. | 10 | ESS – Agriculture Producer Prices (annual, monthly) and Price Indices |  |
| 11 | ESS – Government Expenditure to Agriculture | New |
| **ESS** | Environment: Data capture and processing of annual agricultural input data. | 12 | ESS Environment – Fertilizers Balance |  |
| 13 | ESS Environment – Fertilizers Trade |  |
| 14 | ESS Environment – Pesticides Trade |  |
| 15 | ESS Environment – Pesticides Use |  |
| 16 | ESS Environment – Land |  |
| **Medium priority, low resource statistical processes** | | | |  |
| **Div/Dep** | **Domain of Statistical Subject** | **N.** | **Datasets** |  |
| **All** | All domains | 17 | Compilation of aggregates | New |
| **ESS** | Secondary data needed to compile many derived indicators | 18 | ESS - Macro-economic indicators | New |
| 19 | ESS – Population | New |
| **ESS** | Compilation of indicators using datasets in the SWS | 20 | ESS – Agriculture Production value and index numbers (PIN) | New |
| 21 | ESS – Agriculture Trade index numbers (TIN) | New |
| **FIAS** | Data capture and processing of annual datasets | 22 | FI – Disposition | New |
| **ESS** | Modelled output using datasets in the SWS. | 23 | GHG Emissions from Agriculture |  |
| **ESS** | Data processing to support the generation of the SDG indicators | 24 | SDG 2.8.1 - Agriculture Orientation Index |  |
| **CBL** | 25 | SDG 6.4.1 - Water useefficiency |  |
| 26 | SDG 6.4.2 - Water stress |  |
| **Low priority processes, excluded from the project scope** | | | |  |
| **Div/Dep** | **Domain of Statistical Subject** | **N.** | **Datasets** |  |
| **OSP** | Indicators gathered as part of the M&E process | 27 | Strategic Objective level indicators | Removed |
| 28 | Outcome level indicators | Removed |
| **ESS** | Environment | 29 | Agri-Environment Indicators | Removed |

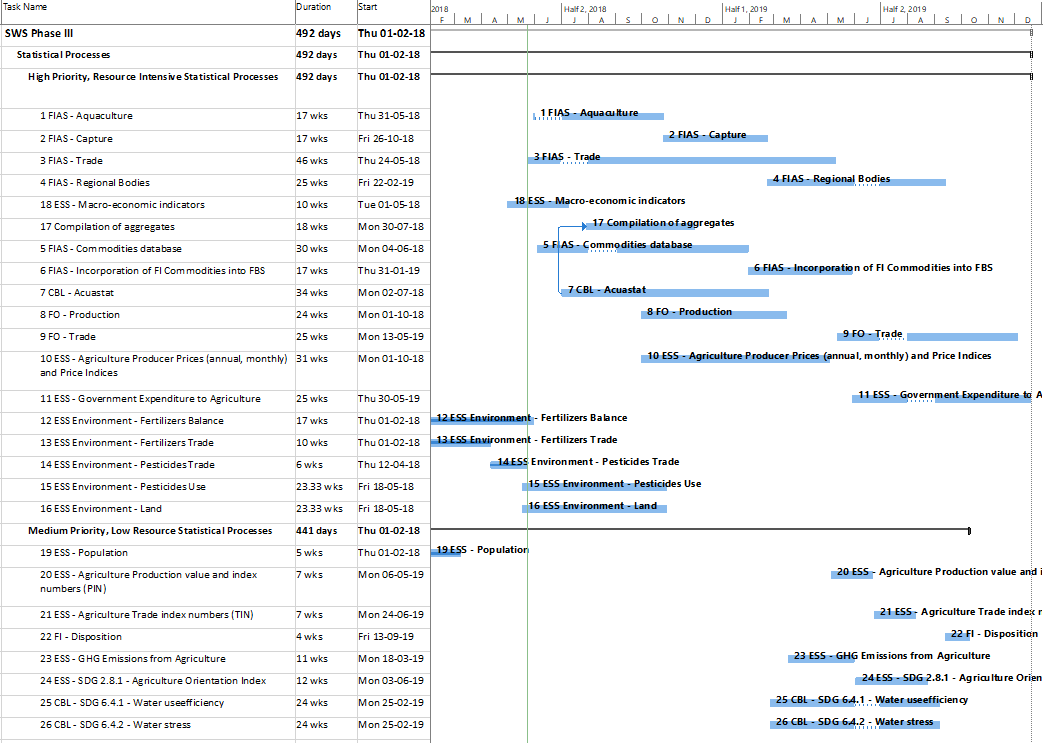
In addition to the above, ESS and OCS has identified the following statistical process work to be included in the scope:

|  |  |
| --- | --- |
| **ESS** | Food Balance Sheet (continued from Phase II) |
| **ESS** | Derived Food Commodities |
| **OCS** | SDG indicator 12.3.1 |
| **ESS** | Trade Module Enhancement |
| **ESS** | Production support, bug fixes of Fertilizers and Food Balance Sheet |

If the scope were to be revised for the statistical processes, the table below indicates the budget requirements for this change.

|  |  |
| --- | --- |
| **Budget Items** | **USD** |
| Presented to CapEx Board | $410,000 |
| Total budget required to complete revised scope | $634,000 |
| **Additional Budget required for revised scope** | **$224,000** |

The GANTT below presents the schedule necessary for the revised scope, with the project completing in December 2019 based on a team of statisticians consisting of up to 6 staff.



**ANNEX 5. Statistical Processes fully implemented in Phases I and II**

The table below lists the statistical processes already fully implemented in phases I and II of the Statistical Working System project.

| **Statistical processes implemented in Phases I and II** | | |
| --- | --- | --- |
| **Div/Dep** | **Datasets** | |
| **ESS** | Agricultural Production | Crop Production |
| Livestock Production |
| Live Animal Products |
| Trade | |
| **ESS Supply Utilization Accounts and Food Balance Sheets datasets** | Feed | |
| Seed | |
| Stocks | |
| Losses | |
| Industrial Utilization | |
| Tourism consumption | |
| Food | |
| Standardization and Balancing | |
| Production of derived commodities | |

**ANNEX 6. Statistical Processes fully implemented in Phases I and II**

The majority of the statistical processes listed in the table below are considered unsuitable for implementation in the Statistical Working System. However, some may be reconsidered in future.

| **Statistical processes not included in Phase III** | | |
| --- | --- | --- |
| **Div/Dep** | **Datasets** | **Suitable for SWS?** |
| **AGA** | Domestic Animal Diversity Information System (DAD-IS) | No |
| Global Livestock Impact Mapping System (GLIMS) | No |
| **AGP** | World Information Sharing Mechanism for the conservation and sustainable use of plant genetic resources for food and agriculture (PGRFA) | No |
| **CBD** | Survey on Access and Benefit-Sharing for Genetic Resources for Food and Agriculture (ABS) | No |
| **ESA** | Market price monitoring system | No |
| Integrated Food Security Phase Classification (IPC) | No |
| Resilience Index Measurement and Analysis (RIMA) | No |
| Nutrition surveys | No |
| Urban and Internally Displaced Persons (IDPs); urban food security surveys; urban food security rapid assessments | No |
| Rural food security rapid assessments covering the agriculture and Livestock sectors | No |
| Climate resilience in Sri Lanka | No |
| Resilience Surveys in Somalia | No |
| Resilience Surveys in Niger | No |
| Resilience Surveys in Uganda | No |
| **ESN** | Collection and compilation of compositional data of foods | No |
| Global Individual Food consumption data Tool (GIFT) | No |
| **ESS** | Development Flows to Agriculture (formerly called Official Development Assistance (ODA) to Agriculture | No |
| Development Flows to Agriculture | No |
| Consumer price indices and food price indices of agricultural commodities | No |
| Credit to Agriculture | No |
| Land cover datasets for SEEA land cover, land cover change accounts and quality assessment | No |
| Agriculture Capital Stock Database | No |
| Climate Change Relevant Statistics--Temperature Change | No |
| Foreign Direct Investment in Agriculture | No |
| Country Investment Profiles | No |
| FIES data collection | No |
| Pilot data collection on new crops production | No |
| Rural Livelihoods Information System | No |
| **EST** | Cereal country balance sheets (CCBS) | No |
| Banana and citrus country balance sheets (BCBS); Tropical fruit country balance sheets (FCBS); Tea country balance sheets (TCBS); Hides and skins country balance sheets (HCBS); Jute and hard fibres country balance sheets (JCBS); Dairy country balance sheets (DCBS); Meat country balance sheets (MCBS); Oilseed complex supply and utilization balance sheets (OCBS); Sugar country balance sheets (SCBS) | No |
| Food aid transactions | No |
| FAO Food Price Index and FAO Commodity Price Indices | No |
| Commercial trade data for all LIFDCs and other African and Asian countries | No |
| Food Price Monitoring and Analysis (FPMA) Database | No |
| Agrometeorology/Remote Sensing Database (Seasonal vegetation Indicators, vegetation Indices, Precipitation) | No |
| **ESP** | A mixed-methods study to identify the migration dynamics in the Kiambu county (Kenya) in relation to social protection and value chains. The study will utilise a survey of about 1,000 households. | No |
| The endline survey data collection for the Impact Evaluation of UN Joint Programme for Accelerating Progress towards the Economic Empowerment of Rural Women in Ethiopia (UNJP-RWEE). | No |
| **FIA** | Ad hoc-collection of data on the value of capture fisheries | No |
| Global Fisheries and Aquaculture Employment Statistics | Yes |
| Global fleet statistics | Yes |
| Statistical data in GLOBEFISH | No |
| Fisheries and Resources Monitoring System (FIRMS), and support to SDG14.4.1 dissemination and monitoring | No |
| Atlas of tuna and billfish catches | No |
| Global Tuna Catches by Stock | No |
| **FOA** | Survey on installed pulp and paper production capacities | Not now |
| Recovered paper data survey | Not now |
| Survey on NWFP consumption | Not now |
| National Forest Monitoring | No |
| Second Report on the State of the World's Forest Genetic Resources | No |
| Global Forest Resources Assessment | No |
| Monitoring the implementation of the Global Plan of Action for the Conservation, Sustainable Use and Development of Forest Genetic Resources | No |
| **OCC** | FAO Country Profiles | No |
| Quality Assessment and Planning Survey (QAPS) | No |
| Develop and carry out a survey on satisfaction of users' of FAO statistics | No |

**ANNEX 7. SWS Requirements Requirements Gathering Workshop Findings**

Statistical Working System

Requirements Gathering Workshop Findings

11-12 July 2018

|  |  |
| --- | --- |
| Version | v1.0 |
| Date | 28/8/2018 |

**Document Signature Table**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Name | Function | Date |
| First draft | Daniele Olivotti | Busines Analyst | 16-Jul-18 |
| Second draft v0.1 | John Rowell | IT Officer | 27-Jul-18 |
| Changes proposed | Carola Fabi Stefania Vannuccini Javier Montero Serrano | Statistician Senior Fishery Officer Statistician | 31-Jul-18 01-Aug-18 01-Aug-18 |
| Final draft v.02 | John Rowell | IT officer | 24-Aug-18 |
| Reviewed by v1.0 | See Participants list below |  | 28-Aug-18 |
| Approved by |  |  |  |

|  |  |
| --- | --- |
| **Participants (\*indicates attended first day only)** | |
| Jose RoseroMoncayo, ESS\* | Jean Munyeshyaka, ESS |
| Carola Fabi, ESS | Stefania Vannuccini, FIAS |
| Francesca Rosa, ESS | Thomas Berger, FIAS |
| Francy Lisboa, ESS | Sara Montanaro, FIAS |
| Salar Tayyib, ESS | Arvydas Lebedys, FOA |
| Natalia Golini, ESS | Roberto Bargigia, FOA |
| Irina Kovrova, ESS | Virginie Gillet, CBL |
| Nathan Wanner, ESS\* | MariaMagdalena Heinrich, SP4\* |
| Javier MonteroSerrano, ESS\* | John Rowell, CIO |
| Giulia Gonnella, ESS | Daniele Olivotti, CIO |
| Francesca Schiavello, ESS | Matteo Terrinoni, CIO |
| Philip Chan, ESS\* | Rahul Dogra, Facilitator |
| EunJeong Lee, ESS |  |

# Purpose and scope

This document summarizes the requirements for phase III of the SWS project, gathered during the workshop held 11-12 July 2018, prioritizing them according to the MoSCoW method (Must have, Should have, Could have, and Won't have). While the main purpose of the workshop was to define and prioritize enhancements to the SWS proposed in the Business case submitted to the CapEx board, the outcome demonstrated that there is a significant amount of work required to fix bugs and pain points in the current version. Indeed, several stakeholders emphasized that fixing the issues in the current system should take priority over the new proposed enhancements.

# Description of the Statistical Working System

As background to the description of requirements in this document, this section contains a quick overview of the Statistical Working System (SWS) and its modules.

## Input system

The users can query, view and edit the data through the Input System. It is the operative interface of the system.

The Input system has two main views: the query tool which allows users to define slices of the datasets; and the session view where the actual data is presented as a grid (Excel-like) and can be edited either manually or through execution of R scripts.

The data editing for a dataset is always done using “sessions”; sessions are private working areas where users update values before committing them to the main database. Sessions are created by selecting a subset of the whole dataset using the Query tool.

The main features of the Input system are:

* Create “working sessions” on datasets;
* Input/Edit data in datasets (values and flags) through an “Excel-like” interface. A log (or history) of all changes is maintained;
* Input/Edit metadata in datasets (e.g. sources, comments). A log (or history) of all changes is maintained;
* Input/Edit data in datatables using an “Excel-like” interface. History is not currently stored for changes to datatables. Datatables are used for various auxiliary purposes such as reference data or temporary data;
* Data and metadata import and export for datasets;
* Data import and export for datatables;
* Execution of R plugins;
* Execution of Harvesters (to read and transform data coming from external sources);
* Create data tags (snapshots of the data);
* Live helper (basic chart generation);
* Perform data validation using R scripts;
* Perform aggregations and calculations using R scripts;
* Grant/revoke users’ access to datasets;
* Conflict resolutions (when two or more users update the same data in different sessions);
* Approval cycle for data (updates are added to a pending approval queue; the changes must be approved before being committed to the main database).

This is the oldest module of the SWS; many of the “pain points” reported by users are focused in this module.

## Admin console

The Admin Console, developed in 2017/18, allows users to manage system objects. Before the existence of this module, users were dependent on the IT team for the creation of any system objects.

The main features of the Admin Console are to allow creation, updates and deletions of the following objects: domains, datasets, datatables, R Plugins, harvesters and metadata models, and initiate reference data updates through an EBX5 connector.

Through the Admin Console administrators can also manage users’ rights on the objects, for example provision of create, edit, delete rights for domains, datasets, datatables, RScripts, and execution rights for R scripts.

## Questionnaire Manager

The Questionnaire Manager allows users to define questionnaires based on their datasets in preparation for dissemination to reporters around the world.

The main features are:

* Template definition
* Questionnaire generation and prefilling
* Data import from returned questionnaires

## SDMX browser

The SDMX browser allows users to browse data made available by external organizations that support the SDMX standard and harvest that data.

# Glossary/additional information

**Codelist:** a list of allowed values for data elements (e.g. the list of countries in the form of code + country names, a list of agricultural products, etc). In a dataset each dimension has a linked codelist ensuring that only known items can be used.

**Metadata**: are used to provide information about a value, a set of values or an object; examples of metadata are: comments, source, contact, coverage, etc.

In the SWS metadata can be attached to a single value (observation metadata), a set of values (block metadata) or to a system’s object (object metadata).

**Observation metadata** refers to a particular version of the value, if the value is updated the old value is moved to the history with the linked metadata.

**Block metadata** refers to a group of data. When a value in the group is updated the block metadata remains attached to the value. In other words, block metadata refers to the keys (e.g. 2010, wheat, all countries) not to the values themselves.

**Object metadata** can be attached to a dataset, a datatable, an R Script, etc.

**R**: R statistical software is integrated in the system; R scripts can extract, process and write data in the SWS (<https://www.r-project.org/>). R scripts are executed on an RServer.

**Shiny**: is an R package (see <https://shiny.rstudio.com/>) that allows users to query raw or processed data from the SWS platform, perform table-pivoting, perform analysis, evaluate impacts of changing values, and so on. The integration of the Shiny package with the SWS is limited: the free version is currently used which does not permit integration with FAO’s Authentication framework, and so data is read using R and fed into the Shiny tool. The tool, because of the lack of authentication features, cannot write data back in the SWS.

# High Level Requirements Prioritization

The deliverables proposed for phase III of the Statistical Working System in the Business Case submitted to the CapEx board in 2016 were reviewed and prioritized by the participants and are summarized in the table below. Nine IT deliverables were discussed, with the majority of the discussion revolving around deliverable 8, Various Input System Improvements.

|  |  |  |  |
| --- | --- | --- | --- |
| Id | Deliverable | | Priority |
| A | Alternative Methods for Data Collection | a) Questionnaire Manager | Must have |
| b) Questionnaires online | Could have |
| c) SDMX browser | Should have |
| B | Statistical Workflow Management and Reporting | | Must have |
| C | Administration Module | | Must have |
| D | Focal Point Management for Questionnaires | a) Campaign Manager | Should have |
| b) Focal Point Manager | Should have |
| E | Improved Mechanisms for Machine-to-Machine Data Sharing | a) SDMX data exchange | Could have |
| b) Dissemination (e.g. FAOSTAT) | Must have |
| F | Integration of Reference Data and Comprehensive Management of Metadata | a) Integrations with a Master Data Management (MDM) solution (EBX5) | Must have |
| b) Improved Metadata Management | Must have |
| G | Support for Visualization and Analytics of SWS data | | Must have |
| H | Various Input System Improvements | | Must have |
| I | Improved performance and scalability | | Must have |

The following pages document each deliverable as described in the business case, followed by a table providing more detail on the requirements and their priorities, as articulated in the workshop. The deliverables did not all receive the same amount of attention during the workshop and so some will need additional requirements gathering before being implemented.

# Alternative Methods for Data Collection Requirements

The proposed requirements for “Alternative Methods for Data Collection” were described thus:

*“The SWS already offers a solution for automatic generation and import of questionnaires: however, this solution is not adequate for all domains. New functionalities would be:*

*a) Questionnaire Manager: End-to-end management of questionnaires for data collection covering flexible data and layout models; point-and-click interactive design; template lifecycle; questionnaire lifecycle management; online and offline publication; SWS integration; Integrates with a Campaign Manager for questionnaire lifecycle management.*

*b) Questionnaires online: development by end users of questionnaires directly in the system rendering them in HTML for focal points in the field to submit responses, thus minimising compatibility issues. In addition, must support input validation and authenticated submissions.*

*c) SDMX Browser: facilitates interactive inspection and data collection by the end users from SDMX registries, resources outside of FAO, such as UNSD and EUROSTAT, and incorporate them into SWS database. In addition, includes dynamic provider registration; asset browsing; interactive query definition, management, and execution.”*

The Questionnaire Manager has already been developed and tested by FIAS users, and should be reviewed and tested also by users of other divisions to ensure it meets the requirements of all of the stakeholders. The table below lists additional requirements identified to date for the Questionnaire Manager and other methods for data collection.

| Alternative Methods for Data Collection Requirements | | |
| --- | --- | --- |
| Requirement | Description | Priority |
| Import additional metadata | The Questionnaire Manager can import some of the metadata contained in the questionnaire. The Questionnaire Manager should be able to import additional metadata fields, such as the metadata form. In addition, the Feedback form should also be stored for future reference, preferably in the campaign management. | Must have |
| Copy & Paste of multiple values in questionnaires | The Questionnaire Manager automatically generates questionnaires and prefills them with data contained in datasets if available. The Questionnaire Manager can reimport data from the submitted questionnaires directly into SWS’ sessions; in order to allow the reimport of data, there are hidden Excel cells embedded in the Questionnaires generated. It has been observed that the hidden Excel cells prevent the reporters from easily copying and pasting multiple values. A solution needs to be found in the Questionnaire generation such that copy and paste is possible on multiple values. | Must have |
| Formulas in questionnaires | Questionnaires generated by the questionnaire manager are static, in that they contain values but not formulas. Users should be able to define and embed formulas when designing the questionnaires. | Must have |
| Language and Logo support | The questionnaire manager should be able to manage multiple languages and present the FAO logo for the language being used. | Should have |
| Handle text data in questionnaire | The Questionnaire Manager can import data in datasets and some sets of metadata. Questionnaires could contain textual data, and descriptions. The questionnaire manager should be able to import this kind of data. | Should have |
| SDMX Browsing | The SWS has an SDMX browser available. The SDMX data harvesting capabilities should be expanded. | Should have |
| Change decimals’ option presentation | When designing a questionnaire, disabling the Decimals options still shows the number of decimals in the UI making it ambiguous. | Could have |
| Online questionnaires | The questionnaire manager should have an online version.  Note that some participants expressed the view that on-line questionnaires would not be feasible in some countries with poor connectivity. | Could have |

# Statistical Workflow Management and Reporting Requirements

The proposed requirements for “Statistical Workflow Management and Reporting” were described thus:

*“Workflow to allow users to design patterns of statistical activities, organize resources into statistical processes and effectively monitor their progress. Reporting functionalities will enable users to have an instantaneous glimpse of key aspects of the process according to adequately defined indicators, such as: overall process completion; number of questionnaires loaded and processed; number of modules executed and so forth.”*

No additional requirements were proposed during the workshop.

# Admin Console Requirements

The Administration Module, also referred to as the “Admin Console” has already been developed based on the requirements that were described thus:

*“New objects must be easily configurable by end users. The overall usability of the system will gain great benefit by having a dedicated administration module; the end user will have autonomy in defining objects composing their statistical workflows. In other words, interactive and domain-driven administration of domains, datasets, data tables, plugins, harvesters and metadata models/instances. In addition, includes object lifecycle management, input validation, access controls.*

*These new functionalities will contribute in reducing the overall costs for IT support in managing object configuration.”*

Additional requirements were presented by participants which are documented in the following table.

|  |  |  |
| --- | --- | --- |
| Admin Console Enhancements | | |
| Issue | **Description** | **Priority** |
| View object owner | All objects in the SWS have an owner. It should be possible for all users to see who the object owner is. | Must have |
| Administrators management of objects | Sometimes the account of the owner of an object is not available (e.g. locked out, expired, etc). SWS Domain Administrators should be able to manage other users’ objects within their domain. | Must have |
| Centralize and review permissions management | After the release of the Admin Console users are able to manage the SWS objects without the support of the IT team. However, the ability of users to grant access to datasets is still managed in the Input system which requires IT support. All permissions management should be managed in the Admin Console for consistency and reduced IT support requirements. Permissions should be organized in layers with each domain having an administrator who can give access to other users. It should be possible to generate an organigram showing who has permissions to what in the various domains. | Must have |
| Attach files to metadata | Metadata are used to provide additional information about a value or an object. The SWS allows users to attach metadata to values, groups of values and some of the objects. However, users should be able to also attach or link files to metadata and should be able to search those files in several ways: by name, by tag, by category, for example. (Also affects Admin Console) | Must have |
| View all the available datasets | Users are not allowed to see datasets in domains they don’t have access to. They should be able to browse all the available datasets in read only mode. This feature will allow users to see what is available and eventually ask for access to the domain’s owner. | Must have |
| View all the available plugins | Users cannot see plugins in domains they don’t have access to. Users should be able to browse and see the all the available plugins in read only mode. This feature will allow users to see what is available and eventually contact the plugin’s owner. | Must have |
| Object metadata templates | Users can define the model for the object metadata independently from the IT team; this gives freedom on one side but allows users to define non-standard models. Standard metadata formats should be defined and validated in a managed way (for example by the Office of Chief Statistician), and made available as templates. | Must have |
| Object metadata UI | Users can define the model for object metadata independently from the IT team. The object metadata’s model is described using the JSON schema standard. The SWS should provide a User Interface to allow users to define the object metadata model themselves. | Must have |
| Integration with EBX5 for metadata | Some of the object metadata could be defined in EBX5. The SWS should be integrated with EBX5 and pull the information from it. | Must have |
| Admin manual | Documentation on how to use the various features of the Admin Console should be provided, preferably available from within the Admin Console itself. | Must have |
| Codelist export | The Admin console allows users to see codelists’ contents. Users should be able to export the codelists contents. | Should have |
| Edit rights of R scripts | Domain administrators should be able to delete unused R scripts created by any user of their domain. Domain administrators should also be able to change ownership of R scripts. | Should have |
| Remote access | There are occasions when users need to access the SWS from outside FAO. Currently the Admin Console cannot be accessed through FAO’s remote access gateway. | Should have |

# Focal Point management for Questionnaires Requirements

The proposed requirements for “Focal Point Management for Questionnaires” were described thus:

*“At present, statistical teams do not have a common way to manage addresses of responsible contacts from international organizations, national government bodies and FAO representatives in the various countries. A new solution will be implemented to harmonise the management. Furthermore, as requested by the IDWG Technical Sub-Group on Data Collection, this module will act as a corporate repository for storage and search of questionnaires used in different Divisions for statistical data collection. The workflow relationships/needs between focal points and questionnaires will be addressed as part of the workflow implementation. Includes:*

*a) Campaign Manager: facilitates the possibility to send out many questionnaires, identified and selected from the Questionnaire module, at the same time to many focal points, identified and selected from the Focal Point Manager module. Campaign Manager should facilitate the follow up on each and every questionnaire sent out.*

*b) Focal Point Manager: Facilitates the possibility to manage focal points in the system, manage all related relevant information about them and relate them to questionnaires and campaigns.”*

The table below lists additional requirements for the Focal Point Management for Questionnaires.

|  |  |  |
| --- | --- | --- |
| Campaign / Focal Point Manager Requirements | | |
| Requirement | **Description** | **Priority** |
| Campaign manager | The campaign manager facilitates the possibility to send out many questionnaires, identified and selected from the Questionnaire module, at the same time to many focal points, identified and selected from the Focal Point Manager module. Campaign Manager should facilitate the follow up on each and every questionnaire sent out and monitor the progress of the campaign throughout, for example, informing on percentage of respondents, and facilitating follow-ups to those who have not responded. There can be more than one focal point for each questionnaire per country, to be sent as To: and Cc: in the emails distributed. | Could have |
| Template and language support | The campaign manager should allow management of email templates, reminder templates, languages and logos displayed according to the language. | Could have |
| Focal point management | At present, statistical teams do not have a common way to manage addresses of responsible contacts from international organizations, national government bodies and FAO representatives in the various countries. A new solution should be implemented to harmonize the management of focal points. | Could have |
| Focal point management | The Campaign manager should include focal point management. The campaign manager should be able to extract the contact details of the focal points and use the information to generate bulk email messages. | Could have |

# Improved Mechanisms for Machine-to-Machine Data Sharing Requirements

The proposed requirements for “Improved Mechanisms for Machine-to-Machine Data Sharing” were described thus:

*“The system currently lacks an efficient interface for data exchange. Functionalities required include: data interchangeability using SDMX (explained in deliverable 1), and integration with other corporate systems for data collection and data dissemination, for example: FAOSTAT.”*

The table below lists additional information.

|  |  |  |
| --- | --- | --- |
| Improved Mechanisms for Machine-to-Machine Data Sharing Requirements | | |
| Requirement | **Description** | **Priority** |
| Standard dissemination tool | The data dissemination is not in the scope of the SWS project; data contained in the SWS is usually disseminated using FAOSTAT, and the SWS provides the APIs to permit FAOSTAT to query and disseminate information. However, not all the datasets can be disseminated using FAOSTAT. Users need a corporate dissemination platform to disseminate data, but this is not within the scope of the project. | Should have |

# Integration of Reference Data and Comprehensive Management of Metadata Requirements

The proposed requirements for “Integration of reference data and metadata” were described thus:

*“The provision of a complete mechanism for control of defined domain values, standardized terms, classifications and categorization of data. In addition there is a need to adopt more structured metadata to strengthen the data exchange functions described in “Improved Mechanisms for Machine-to-Machine Data Sharing”. This deliverable will focus on:*

*a) Integrations with a Master Data Management (MDM) solution (EBX5). MDM solution is a harmonization tool. It facilitates to use the same master data in a broad variety of information systems in FAO.*

*b) Metadata Management: The module can store and facilitate transactions on metadata.”*

The table below lists additional information. Note that many of the requirements relating to metadata management listed below can also contribute to “Admin Console” and “Various Input System improvements”.

|  |  |  |  |
| --- | --- | --- | --- |
| Integration of Reference Data and Comprehensive Management of Metadata Requirements | | | |
| Requirement | **Description** | | **Priority** |
| Attach files to metadata  (also in input system enhancements) | | Metadata are used to provide additional information about a value or an object. The SWS allows users to attach metadata to values, groups of values and some of the objects. However, users should be able to also attach or link files to metadata and should be able to search those files in several ways: by name, by tag, by category, for example. (Also affects Admin Console) | Must have |
| Merge observation and block metadata | | In the input system, observation metadata and block metadata are located in two different tabs. All the metadata should be accessible in a single point to avoid switching back and forth between the two tabs. | Must have |
| Highlight cells with metadata | | The SWS shows the observation metadata and the block metadata only when the cell (value) containing them is selected. Users have to click on all the cells (values) to check if any metadata has been attached. The SWS should highlight the cells containing metadata, providing the option to select the element of metadata of interest (e.g. highlight cells with metadata GENERAL/COMMENT attached). It would also be useful to be able to highlight cells according to other criteria, e.g. according to the flags - highlight “official” values, or dim cells associated to “no data available” flag (which currently require value = 0). | Must have |
| Search by metadata | | Users should be able to search the metadata of the SWS and receive resulting lists of metadata matching the search criteria. | Must have |
| Metadata categories | | Metadata are organized in categories and sub-categories that cannot be changed by the users. It has been reported that new categories and sub-categories of metadata are required in the SWS. Agreement will be required on who can insert such categories and sub-categories. (Also affects Admin Console) | Must have |
| Object metadata  (also in input system enhancements) | | It is currently possible to easily attach metadata to values. However, the system should also allow users to easily attach metadata to other objects such as datasets, datatables, R scripts, variables, codelists, etc. (Also affects Admin Console) | Must have |
| Public/private comments | | Metadata attached to observations or to blocks of observations can contain comments. Users should be able to define comments as both public (can be disseminated) and private (for internal use, not for dissemination) on the same observations. | Must have |
| Metadata on empty cells | | The users should be able to attach metadata to empty cells (cells without a value). | Must have |
| Custom hierarchy codes | Codes can be organized in hierarchies. Each item at any level in the hierarchy should have a unique code. Some codes are introduced only for the purpose of creating groups (e.g. European countries). The system should be able to handle groups without introducing new codes to avoid codelists divergence from the standard ones. The system should be able to handle groups without adding “dummy” codes. | | Should have |

# Support for Visualization and Analytics of SWS data Requirements

The proposed requirements for “Support for Visualization and Analytics of SWS data” were described thus:

*“Integration with an Analytic and Visualization solution facilitates end users to query (raw or processed) data from the SWS Platform and create any kind of charts, diagrams reports etc. for their needs.”*

The table below lists additional requirements.

| Support for Visualization and Analytics of SWS Data Requirements | | |
| --- | --- | --- |
| Requirement | Description | Priority |
| Authorized access and write updates to SWS | The SWS contains very basic data visualization tools for displaying simple charts. The need for a more powerful data analysis and data visualization tool led to users utilizing an alternative solution: Shiny. The Shiny application is partially working; however, the free version does not allow the full integration with FAO’s authorization framework and so authorized R plugins are used to extract data but this data cannot be directly written back to the SWS as Shiny does not currently have a licensed authentication framework. Users need to perform data analysis, change data and evaluate the impact on connected figures, create and export charts and reports, and in some cases import results back into the SWS. The users expressed the preference to continue to use Shiny as it is built on top of R; the R scripting language is known in the statistics division and it does not require learning new scripting languages. | Must have |

# Various Input System Improvements

Participants highlighted the following issues with the input system that severely affect the ability of users to carry out their work.

|  |  |  |
| --- | --- | --- |
| Input System Issues | | |
| Issue | **Description** | **Priority** |
| Scrolling problem | When scrolling through large datasets the interface skips some rows and shows empty rows; when using the “fixed first column” feature, data and items are sometimes misaligned. This renders the interface barely usable. | Must have |
| Deletion of large sessions | Sometimes large sessions cannot be deleted. Currently, users need to call IT support to delete them. | Must have |
| Session expiration | User browser sessions expire too quickly and it seems that the expiration time is random. When a session expires, users have to log back in to the SWS and extractions have to be reloaded. (also affects other modules) *Note: “session” is this case is referring to the user’s authenticated session to the SWS using the browser, and not the private working areas in the SWS where users update values before committing them to the main database.* | Must have |
| Validation tab not working | The SWS allows R scripts to be launched for data validation, the results of which are shown in the validation tab. However, the validation tab is not showing the validation results (further investigation is needed to identify situations when this occurs). | Must have |
| Displaying codes not included in the dataset | Each dimension of a dataset has a codelist attached; the only usable items for the dimension are those belonging to the codelist. If a user tries to upload data containing an unknown code the value is discarded and the user notified. Users can select subsets of a codelist when defining the dataset (e.g. use only the African countries from the list containing all the countries). In such cases, the system should discard and notify the user if a code is not in the allowed subset of codes. | Must have |
| Mark inserted values | When a user updates an existing value, the cell is marked with a small red triangle. The system should exhibit the same behavior when entering a value for the first time in an empty cell. | Must have |
| Excel import | Data import from Excel is not working. (Most likely, Excel import should be discontinued and CSV or SDMX used instead). | Could Have |

Participants requested the following enhancements be implemented in the input system to allow users to effectively carry out their work.

| Input System Enhancements | | |
| --- | --- | --- |
| Issue | Description | Priority |
| Real-time validation | The SWS contains a validation mechanism: i.e. when a user updates a value an R script can be launched on the R Server, the results of which are returned to the SWS that highlights the updates in the user interface. This allows R to be used for the input validation routines making the statisticians fully in control of the methodological rules definition. However, the time required to execute an R calculation is currently too long, and as a result the user experiences a delay on the user interface, and for this reason the feature has never been used in practice. Users should be able to define validation rules that can be performed in real-time without visible delays. | Must have |
| Real-time calculation | The same mechanism used for real-time validation can be used to perform real-time calculations in the SWS, and it has the same performance issues as described for real-time validation. As a result this feature has also never been used. Users should be able to define calculation rules that can be performed in real-time without visible delays. | Must have |
| Multi-value editing | The SWS should allow users to select multiple values and edit them together. It should be possible to select multiple cells and modify the value, the flags or the metadata in one go. | Must have |
| Copy & Paste | The SWS should allow users to copy and paste values between cells. | Must have |
| Custom order for codelists | Currently it is possible to define a custom sort order for the items in a codelist, such that the items will be sorted accordingly when presented to the users in the interfaces. However, users want the possibility to specify a “per user” sort order as different users work on data in different ways. | Must have |
| Improve query tool search | The “search by name” feature in the SWS query tool can only search from the beginning of words. The query tool should be able to find words in the middle of codelists (e.g. when searching “Samoa”, it should be able to find “American Samoa” even if the code does not start with the letters “Samoa”). | Must have |
| Add a “Select all” button for elements | In the query tool, when the dataset contains an “Element” dimension, there should be a “Select all” button to allow the user to quickly select all the items. | Must have |
| Import/export | Data, metadata and block metadata are exported in 3 separate files. Users have requested that the system should provide data, metadata and block metadata all in one export file. | Must have |
| Export empty cells | Empty cells (N/A values) should be included when exporting data. | Must have |
| Journal history import/export | The SWS maintains the history of the changes to all values. Users test features and R scripts in the Quality Assurance (QA) environment, and sometimes they have the need to migrate this data and history to the Production environment; For this reason the system should be able to export and reimport the history. Note: The system allows the export and import of journal history, but the feature has been seldom used, and should be tested to assess if it is fit for purpose. | Must have |
| Cannot delete multiple sessions | There should be the possibility to delete multiple sessions. | Must have |
| Move permissions to Admin Console  (also in Admin Console) | Permissions management should be removed from the Input system and be fully managed in the Admin Console. | Must have |
| Show empty rows behaviour | Some of the datasets can be sparse, having many empty rows for some items or countries; the default behavior is to hide empty rows which can be overridden using the “show empty rows” button. Showing all the empty rows on big datasets can block the system, be unpractical or show illogical element/item combinations (e.g. [0115] Barley associated to the element [5330] Female Reproductive Activity). Users should be able to select only the empty rows they want to see. | Must have |
| Rounding depending on the scale | The SWS allows definition of the number of decimal digits to be shown for the values of a dataset. The SWS should have an option to define the number of decimal digits based on the value input; for example: several decimal digits for small values, a few or no decimal digits for large values. | Should have |
| Data import progress | Importing data in the SWS can be a lengthy operation for large amounts of data. Therefore users should be able to see progress of the import to be sure that the operation is proceeding and estimate the completion time. | Should have |
| Multiple hierarchies | The SWS can handle only one hierarchy for each codelist; a new codelist must be created for each additional hierarchy (the list of codes is not duplicated). For example, a hierarchy for listing countries by region, may also need to be structured in a hierarchy by Fishing areas for certain datasets referencing the same countries. The system should be able to handle multiple hierarchies for a single codelist without the need of creating new codelist objects. (Also affects Admin Console, Reference Data Management). | Should have |
| Remove approval cycle for R scripts | The approval cycle feature can be optionally activated in the SWS: when values are updated and saved back to the database the changes are moved to an approval queue. Authorized users can then check and approve the change. R scripts can change large amounts of data, and so if the approval cycle is enabled the approval queue can become unmanageable. The approval step should be removed when data has been changed using R scripts. This may not be the required behaviour for all domains, therefore this should be an option on a per dataset basis. In addition it should be possible to not display unchanged figures in the pending queue. | Should have |
| Mark edited session | Sessions are created for users whether they are viewing or editing data, and these sessions are automatically saved and listed in the user interface. Users need to be able to identify those sessions containing uncommitted changes. | Should have |
| View data without creating sessions | Data in datasets can only be viewed through a session. Users should be able to view/query data without creating a session. | Should have |
| Improve search in query tool | When defining a dataset it is possible to limit the codelist to a subset (e.g. from the list of all the geographic areas allow only the African countries to be used). The query tool does not take into account the reduced codelist and returns also codes outside the scope of the query (e.g. all the countries are returned in the search results, not just the African ones). | Should have |
| Improve search in query tool | The query tool should allow the user to select multiple search results (currently the search results are cancelled after each selection). For example, if the user wants to select Austria and Australia to define the session and types “aus” the two countries are returned; However, if the user then chooses “Austria”, the search results disappear and the user has to repeat the search all over again to select the second one “Australia”. | Should have |
| Data and processes migration | Users need to move data and processes between QA and Production instances, and so they need a clear process to perform such migrations. Therefore a procedure needs to be established for data and statistical processes migration. (Also affects Admin Console) | Should have |
| Statistics on values added and edited | Users need to report the number of updated or newly inserted values. The system should keep track of the changes and generate reports on number of updates/changes applied. | Should have |
| RServe management | Updating R packages requires a restart of the RServe server in order to load the changes, and that can only be done by accessing the R server, which can only be done by the IT team. Users need to be able to reload the updated R modules without need for IT support’s intervention. | Should have |
| R packages management | The methodological team maintains the R Scripts and libraries. Updating some core libraries can cause some compatibility issues to R scripts already in production. R developers need a version control mechanism to detect conflicts and eventually rollback the changes. | Should have |
| Delete session on Save to database | The SWS allows users to update data contained in their datasets by creating a “working session”; this allows the users to commit the changes to the database only when they are satisfied with the updates made. On completion, currently the session is kept in the list of available sessions. The SWS should allow to commit the session and delete it from the list (save and delete) in one action. | Could have |
| Better storage for mapping tables | The SWS stores some datatables, used for example for mapping tables, temporary data, not validated data, R script logs. Currently datatables do not keep a history log of the changes and so if incorrect updates are made there is no record and no way to rollback. Users should be able to store and see the history of changes to datatables and have a means to roll them back. | Could have |
| Input system navigation | There should be a link to the Questionnaire manager and Admin Console from the Input system. | Could have |
| R scripts on GIT | When changes are made to R scripts they need to be installed on both production and QA environments. A more efficient approach could be to execute R scripts directly from the GIT version control system. | Could have |

# Improved Performance and Scalability

The proposed requirements for “Improved Performance and Scalability” were described thus:

*“Improved scalability and performance for:*

* *Handling large sessions*
* *data access (primarily write throughput, but also read latencies),*
* *data storage (wasteful storage model for observation metadata, with implications on I/O performance)*
* *R execution models (hardcoded on a single R server).”*

The table below lists additional information.

|  |  |  |
| --- | --- | --- |
| Performance and Scalability Requirements | | |
| Requirement | **Description** | **Priority** |
| Improve R scripts speed | R scripts performance needs to be improved. More information will need to be gathered by the IT Team and Statisticians to understand in which situations performance is a problem, and what the causes are. | Must have |
| Improve metadata reading speed | There is a significant decrease in performance when users read or write both data and metadata, compared to when just reading or writing data. Performance when working with metadata needs to be improved. | Must have |

# Other Requests of Participants

Participants also requested the following non-technical improvements.

|  |  |  |
| --- | --- | --- |
| Non-Technical Requests | | |
| Requirement | **Description** | **Priority** |
| Improve communication | Stakeholders would like to receive more communications on the SWS, in particular regarding advances in the system, presentation of new features, bug fixes, and so on. This would also allow users to be aware of the activities planned and be prepared to allocate the time needed for communicating needs, testing features or providing feedback. | Must have |
| Documentation | Users have requested more documentation | Must have |
| Jira notifications | When a ticket is updated or closed in Jira the user that opened the ticket should be automatically notified via email. | Should have |
| Training | Users would like to receive training | Should have |

1. Aquaculture and Capture have been already finalized in Phase II. They will be rolled out to production environment in Phase III [↑](#footnote-ref-1)
2. FOAD dataset has been already finalized in Phase II. It will be rolled out to production environment in Phase III [↑](#footnote-ref-2)
3. https://www.ideou.com/pages/design-thinking [↑](#footnote-ref-3)
4. https://statswiki.unece.org/display/GSBPM/Generic+Statistical+Business+Process+Model [↑](#footnote-ref-4)