21LU01 - MODERNISING THE PRODUCTION AND DISSEMINATION OF OFFICIAL STATISTICS IN THE NATIONAL STATISTICS OFFICE OF LUXEMBOURG (STATEC)

FINAL REPORT (OUTPUT 6)

REFORM/OECD FRAMEWORK DELEGATION AGREEMENT

REFORM/IM2021/006

Document revision history	Who	Date
Document initialised	OECD	03/03/2023
Document review for review by the Advisory Board – after multiple iterations	OECD, STATEC	16/05/2023
Document finalised, including additional amendments requested by the Advisory Board	OECD	TBD

TABLE OF CONTENTS

1 PURPOSE OF THE DOCUMENT

The purpose of this document is to report on the project "21LU01 - Modernising the production and dissemination of official statistics in the National Statistics Office of Luxembourg" – its outputs, activities, a summary of the lessons learnt and an outline planning for the STATEC to continue beyond the end of this project.

Overall, the "project" refers to activities and deliverables relating to the scope of the 21LU01 project. However, as this "project" correspond effectively to a broader transformation ongoing at STATEC – preceding 21LU01 and which will continue after the closure of it – it may be that in some statements, the "project" designates the broader project of which 21LU01 is only a phase, dedicated to design and capacity building activities. The 21LU01 project was preceded by the STATEC data dissemination platform migration to LUSTAT.LU (a project concurrent to the first phase of 21LU01); it should be followed by a phase of implementation that is described in this document. The context of the statement normally makes it unambiguous as to what is meant.

The *Final Report* (Output 6[1]) covers the following:

a an **executive summary** the purpose and content of the Project,
b a **description of what was accomplished** during the project, including a table of all deliverables: outputs and activities,
c a **summary of the lessons learnt** and what can be drawn from them,
d the **indicators measuring the quality of outputs** drawing on corresponding surveys, interviews, or platform analytics data,
e a **roadmap (referring to actors, tasks, and timelines) for future phases** not covered under this scope of the Project,
f the outline of a joint **communication plan**.

For readers new to the project, please refer to:

- High Level Description and Detailed Project Description documents, which correspond to the project description during the inception phase;
- Discovery report, which provides a much-refined analysis of the needs and project delivery approach, based on an extensive set of interviews and workshops. This report makes extensive reference to the discovery report.

Specifically, this report focuses on the outputs and activities carried out in 2022, following the discovery phase between December 2021 to June 2022, and leading to project closure in March 2023.

[1] Cf. Section 2 Project Outputs and Activities of the Detailed Project Description document.

2 ABBREVIATIONS AND ACRONYMS

Abbreviation / acronym	Explanation
AG	Advisory Group
COM	Communication Department
EC DG Reform	European Commission's Directorate-General for Structural Reform Support
ES	Eurostat
FMR	Stands for 'Fusion Metadata Registry' – an open-source solution developed by the BIS in collaboration with Metadata Technologies, to author and store SDMX structural metadata. FMR is widely adopted among SIS-CC members and more and more interoperable with the .Stat Suite (through the variant developed by BIS, the "FMR Workbench"). By extension, the 'FMR project' designates all activities, led by the BIS, leading to the delivery of the FMR and associated support services.
OECD	Organisation for Economic Co-operation and Development
SDMX	Statistical Data and Metadata Exchange
SDMX-RI	Suite of services constituting the SDMX reference infrastructure, developed and maintained by Eurostat as open-source solution to support primarily the reporting use case in European countries (and beyond). Amongst SDMX-RI components, the "sdmxsource library" is reused in the context of the .Stat Suite (and contributed to by OECD as well. By extension, the 'SDMX-RI project' designates all activities, led by the Eurostat, leading to the delivery of the SDMX-RI and associated support services.
SIS-CC	Statistical Information System Collaboration Community – see https://siscc.org
.Stat Suite	Suite of modules to store and disseminate date and metadata, underpinned by the SDMX standard, constituted of .Stat CORE (data storage), .Stat DE (Data Explorer) and .Stat DLM (Data Lifecycle Manager), as well as .Stat DLML excel add-in (data and referential metadata edition in excel). .Stat Suite is developed by OECD as an open-source project delivered in the context of a community (the SIS-CC). By extension, the '.Stat project' or '.Stat Suite project' designates all activities, led by the OECD, leading to the delivery of the FMR and associated support services. Sometimes referred to as 'v8', as the legacy version of this solution (OECD.Stat) was in version 7 (v7) before complete redevelopment started.
STATEC	National Statistics Office of Luxembourg
TSI	Technical Support Instrument
VIVA	STATEC internal data transformation and exchange environment
TSI	Technical Support Instrument

3 EXECUTIVE SUMMARY

As a key player in the production of official statistics in Luxembourg and in the context of the European Statistical System, the National Statistics Office of Luxembourg (STATEC) must constantly adapt to evolving demands in terms of evidence to support policy analysis and decisions, and to inform the public debate at large. Adaptation entails agility in the capacity for STATEC to evolve its data products (granularity, timeliness of data), as well as improve their accessibility, based on user feedback and through better harmonisation.

As evident during the discovery phase (see Discovery Report, section 5 – "as is state") this broader goal is hindered by the fragmentation, very usual in many statistical organisations: fragmentation by domain (different tools, process and information models across the statistical domains) as well between the stages in the process (design, collection, processing, and dissemination/reporting stages). **"Modernising the production and dissemination of official statistics in the National Statistics Office of Luxembourg (STATEC)", as stated in the project title, means, in that context, the initial design work and capacity building to enable STATEC, during later implementation phases (that are outside of the scope of this project), to provision the appropriate data infrastructure to achieve efficiencies and automation in how data is processed, as well as improve the accessibility of STATEC data for external data users**. These objectives, as framed in impact and outcome indicators (see Discovery Report, section 8) should ultimately lead to improved data user satisfaction, compliance with accessibility standards, as well as increased internal efficiency through automation and use of modern tools and techniques.

These are goals that can be achieved in the mid-long-term, and this project provides the foundations for STATEC to undertake that journey in the years to come. These foundations are the following ones:

a) **The inception of an initial governance mechanism whereby STATEC domains should be able to progressively implement the target data model and data transformations.** Initial capacity building, training, and guidelines in SDMX structural modelling have been catered for, involving power users from all business domains, as well as central IT, methodology and communication experts.
   - From the project participants survey it appears that the governance model fit for the next implementation phase should be a "centralised-consultative" model, according to which a STATeC core expert group (methodology, IT and communications), controls essentially the process for developing data models (namely, SDMX structures and code lists) for the purposes of internal sharing, reporting and dissemination – but do so in consultation with the data producers (identified as "domains").
   - Such an organisation appears however as delicate to put in place due to limited resources in STATEC and given its highly decentralised internal organisation. Arguably, an alternative model could entail more delegation and autonomy to domains, but this would require improvement on the back-office tooling – as noted by project participants –, as well as further training on tools, so that they can be used more efficiently by the domain power users.

⇨ At the completion of this project, it is not possible to conclude on this dilemma on the best governance model, and optimal organisation of the division labour for the implementation phase. More experimentation will be needed, to feed into a broader strategic analysis of data governance at STATEC, which is not limited to the question of data harmonisation addressed in the context of this project.

b) **The identification of the levers for enhanced data accessibility through greater data harmonisation, through the LUSTAT platform**. Concurrent to the project indeed, the new data dissemination platform, LUSTAT, based on the .Stat Suite open-source technology, has been successfully deployed by STATEC. The project allowed for STATEC to better understand how data accessibility on this platform could be improved

by greater data harmonisation, driven in the context of an appropriate governance mechanism. Indeed, user experience is driven by the data model (which means, concretely, search facets or filtering depending on common data dimensions and code list values). The move from a 'table oriented' dissemination to a 'search driven' exploration of a data warehouse enables STATEC to provide improved data accessibility, provided sufficient data harmonisation is achieved.

⇨ This paradigm shift – better data accessibility through greater data harmonisation based on SDMX, which can be achieved only through a STATEC-wide collaborative effort – needs to be further communicated, following this project closure, to all data producers to make efficient data modelling decisions.

c) **The high-level scoping of a platform – extended LUSTAT – that capitalises on the existing LUSTAT platform (supporting data dissemination use case), to support the data integration and internal data sharing use cases.** As it stands, the integration of various data streams in the dissemination space is essentially done manually in STATEC (files are uploaded, after having been formatted manually or as a result of an upstream data process and are essentially not mapped onto a common data model). With the definition of a target data model, and its implementation and continuous improvement in the next phase, automation will become possible, leveraging mapping mechanisms offered by the .Stat Suite, that the extended LUSTAT platform can cater for. In that context, the extended LUSTAT platform could support more use cases than only data dissemination: data integration and reporting use cases could especially be supported, allowing for greater efficiency and maintenance cost reduction. Integration with the existing STATEC file sharing platform (VIVA), and its extension to a larger portion of STATEC teams, is identified as an important success factor.

⇨ Developing the extended LUSTAT platform (including, VIVA and other compontss) to support the data integration and internal data sharing use cases at scale in STATEC, leveraging data mappings and automations, could be an objective for the next phase of the project.

A high-level vision for the project implementation has been scoped, which will require further analysis and strategic decision making within STATEC, to address key risks identified at this stage:

1. **Risk 1: The risk of push back by the STATEC data practitioners impacted by the project if it would not deliver on the expected efficiencies or due to an insufficiently supported change, or insufficiently clarified roles between the business domains and the STATEC core team**. This risk was managed, in the context of this project, through an agile delivery which proved effective in adapting to each domain and individual constraints (especially time constraints) and as such, it should be pursued. As can be spotted in the project survey results (see section 4.3), while the project has delivered on its objective of developing the required skills and initiating a governance mechanism, important questions remain open as to the respective role of domains and the core team of experts – options which may lead to a different set of requirements for the needed back-office tools. In any case, sufficient dimensioning of a central support capacity will be an important mitigation factor to address this risk – that would entail a stronger STATEC core team, and/or a continued external support. Beyond the awareness and knowledge gained by STATEC through the project, more experimentation and analysis will be needed to get to a fully functioning model.
2. **Risk 2: The technical risk attached to the development of the appropriate tooling for the business domains to support all back-office tasks in a way that is effective, user friendly and as automated as possible.** Frustrations expressed in terms of shortcomings in the available back-office tooling epitomises to some extent the dilemma explained under risk one (especially, in terms of division of labour and how

expert should domains become in SDMX modelling). But an important set of platform enhancements have been identified, thanks to the project, that will have to be taken forward during the implementation phase. In the context of the SIS-CC, governance mechanisms are in place for STATEC to express priorities and monitor delivery, as a Tier 1 member of the community; this will impact the delivery of the .Stat Suite by OECD in the coming months, but also, related tools incorporated into the Suite (SDMX-RI by Eurostat; FMR by BIS). The integration of VIVA, the development of a data pipeline automation mechanism within STATEC, as well the advice given to the STATEC IT Department on how to deploy more frequently the .Stat Suite – are all important elements to take in consideration, during the implementation phase, to properly address the risk.

After the successful completion of this project, the next steps involve

- ⇨ The STATEC core team will evaluate and discuss the project findings with the STATEC Directors' Board. This exchange will have to determine the preferred course of action and the level of priority assigned to this initiative.
- ⇨ Subsequently, the STATEC core team will prepare a follow-up internal project proposal, taking into account the available resources, as well as the agreed-upon timeline. This proposal will outline the necessary steps to be taken.
- ⇨ STATEC senior management should then make a call on this proposal in accordance with the internal Project Management process.

By following these steps, STATEC will be able to establish a clearer roadmap for the future of this initiative.

4 PROJECT OBJECTIVES AND DELIVERABLES[2]

This section provides a description of what was accomplished during the project, including a table of deliverables: outputs and activities, as well as a report on the output indicators, including a short summary of the survey conducted at the close of the project with the STATEC staff who were the direct recipients of the training and support provided by the OECD experts. The project outputs were:

1. [Output 1] Kick-off meeting inception report [Delivered].
2. [Output 2] Project discovery report [Delivered].
3. [Output 3] Target data model and data process document [Delivered].
4. [Output 4] A software requirements specification [Delivered].
5. [Output 5] Training of STATEC staff [Delivered].
6. [Output 6] Final report.

This document constitutes [Output 6] of the project.

[2] Source: project high level description document (HLD); project detailed level description document (DPD).

4.1 TABLE OF DELIVERABLES

Below is the table of outputs and activities with links to (assets) documents and other materials that have been completed as part of the project deliverables.

Output(s)	*Activity(ies)*	*Asset(s)*	*Description*
1 Kick-off meeting concept note	**1.1 Kick-off meeting**	Presentation and agenda	A kick-off meeting between the OECD, DG REFORM and STATEC to discuss and refine the methodology, scope, timeline, information needs, communication arrangements, and feedback-processes and detailed plan for the discovery report and planning of the activities during the discovery phase.
	1.2 Drafting kick-off meeting inception report	Inception note	The main and agreed outcomes from the kick-off meeting between the OECD, DG REFORM and STATEC. Delivered in September 2021.
2 Project discovery report	**2.1 Managing the completion of the output**	Presentation of main discovery findings to STATEC Data Practitioners Project Advisory Group meeting summary record	Coordination with different stakeholders, resource planning and scheduling, meeting organisation, and drafting of discovery report.
	2.2 Identification of data producers and data users	Discovery Report	See Discovery Report section 5.2. Analysis of STATEC data practitioners
	2.3 Analysing the high-level requirements	Discovery Report	See Discovery Report section 5.4. Challenges and Opportunities
	2.4 Drafting the discovery report	Discovery Report	The discovery phase (Output 2) took place between December 2021 to June 2022, including the execution of one online questionnaire preparing for approximately 16 interviews taking place virtually with STATEC data practitioners, and onsite and virtual workshops with STATEC project team. Final as of July 2022 with many intermediate outputs.
3 Target data model and data process	**3.1 Managing the completion of the output**	Project Advisory Group meeting summary	Coordination with different stakeholders, resource planning and scheduling, meeting organisation, and reporting at various levels through an agile methodology. The core project team (STATEC+OECD) met on

			average on a bi-weekly basis, where they took stock of activities during the past 2 weeks and activities for the upcoming week(s).
	3.2 Analysing the target data model and process, and migration strategy	Blueprint for By Domain Activities STATEC Data Practitioners Workshop to introduce Blueprint and activities for next phase + recording	Document outlining the initial steps in establishing a data modelling practice at STATEC corresponding to activities outlined in the project discovery report, section 6.4. "By Business Domain Activities", consisting of in "laying foundations through design training and experimentation"; for each of the 4 Business Domains (MAC – Macroeconomic Statistics, SOC – Social Statistics, ENT – Business Statistics, CMP – Economic situation, Forecasting, Modelling).
	3.3 Designing the target data model and data process **3.4 Document describing the target data model and process, and migration strategy**	Target SDMX structural model and modelling process at STATEC Modelling statistical domains and exchange frameworks in SDMX using SIS-CC SDMX Matrix Generator Reference Framework for SDMX Structural Metadata Governance	The most important part of this output has been delivered in the shape of modelling guidelines – which has been drafted as generic to apply in contexts outside of STATEC. A version of the model was initiated in the discovery report (cf. section 14) which was implemented as part of output 3 and subsequently ratified for STATEC to continue to use as they move beyond the DG Reform project closure. Drawn from this, the "Reference Framework for SDMX Structural Metadata Governance" official SDMX guideline has been released In March 2023. It provides guidance to agencies for implementing the most appropriate governance architecture to maintain SDMX Structural Metadata, drawing on the experience learned with TSI 21LU01 and 21BE03 projects.
4 A software requirements specification	**4.1 Managing the completion of the output**		Coordination with different stakeholders, resource planning and scheduling, meeting organisation, and reporting at various levels through an agile methodology. The core project team (STATEC+OECD) met on average on a bi-weekly basis, where they took stock of activities during the past 2 weeks and activities for the upcoming week(s).
	4.2 Analysing data platform features **4.4 Drafting the document and stories describing data**	Features specified as STATEC requirements under output 4: Versioning of SDMX artefacts and data in DLM (nsi#213, transfer#171)	Features required by STATEC have been identified with the relevant ones documented as user stories in the .Stat Suite product backlog (see links opposite). The development of these features is outside the scope of the project.

	platform features	Snapshot of the dissemination environment (nsi#322, addin#108, nsi#300, nsi#57) Internal data sharing space for the future solution (nsi#306) Last updated date (sfs#128) and Publication date	
		Data and data process migration and integration strategy meeting recording	A technical workshop took place January 2023 to discuss the technical architecture including the infrastructure at STATEC (at least the current one) and define a deployment strategy at STATEC.
5 Training of STATEC staff	**5.1 Managing the completion of the output** **5.2 Training of STATEC staff**	Project Advisory Group meeting summary	Coordination with different stakeholders, resource planning and scheduling, meeting organisation, and reporting at various levels through an agile methodology. The core project team (STATEC+OECD) met on average on a bi-weekly basis, where they took stock of activities during the past 2 weeks and activities for the upcoming week(s).
		STATEC Data Practitioners Cross-domain meeting agenda	
		STATEC Data Practitioners Cross-domain meeting recording	Multiple on-demand meetings took place to respond to various requests for support throughout the period; for some recordings have been made which are listed opposite.
		Introduction to SDMX for Data Producers	Two .Stat Academy self-paced training courses were developed and made available to STATEC Data Producers for completion prior to the two onsite hands-on workshops to prepare and better support the training going forward. In total 13 data producers have been certified helping to develop more SDMX knowledge in the STATEC data domains.
		Introduction to SDMX Structural Modelling for Data Producers	

4.2 OUTPUT INDICATORS

Output indicators are defined in the Detailed Description Document (see [1]) and refined in the Discovery Report (see [2]).

The table below shows the Output indicators extracted from the Discovery Report having been further refined following the initial defined list published as part of the DPD document. Each one has a status of the indictor, whether it has been achieved and with a quantitative measure when relevant.

Output(s)	Intervention logic	Indicator	Values of reference *(Per indicator)* **Target** *(incl. reference date)*	Sources and means of information	Assumptions
Output 2 Project discovery report	Ensure quality of the document and appropriate involvement of stakeholders and producers.	Submission of the discovery report to the Advisory Group. Volume of interviews carried out; number of participants to workshops.	The document is finalised, and corresponding activities are executed in line with the reference timeline.	Manual	
STATUS: Achieved ***Comments:*** The discovery phase took place between December 2021 to June 2022, including the execution of one online questionnaire preparing for approximately 16 interviews taking place virtually with STATEC data practitioners. Multiple online meetings with the project core team and an onsite workshop took place (5 STATEC participants) to iteratively improve the discovery phase findings. The Discovery report was finalised as of July 2022 with many intermediate outputs. A final presentation of the main outcomes to STATEC data practitioners took place on 5 July 2022 followed by a survey to capture their feedback with the results showing an overall high-level satisfaction with the main output.					
Outputs 3 and 4 Target data model and data process document Data platform features document	Ensure quality of the document and appropriate involvement of stakeholders and data producers.	Submission of the documents to the Advisory Group. Volume of interviews carried out; number of participants to workshops.	The document is finalised, and corresponding activities are executed in line with the reference timeline.	Manual	The scope of stakeholders, data producers, to be involved will be defined in the discovery phase.
STATUS: Achieved ***Comments:*** Following a piloting with MAC business domain a blueprint was conceived to better frame the data modelling activities to be carried out with each of the business domains. An initial review of the current data migrated to the new LUSTAT data portal was completed and enabled the OECD and STATEC project teams to identify possible target models to be the subject of review and experimentation with the STATEC data practitioners. In addition, functional workshops and meetings took place to identify and capture the highest priority data platform features. In total there were: - Functional workshops and meetings: 3 - Business domain activities (incl. kick-off): 10 A satisfaction survey of the STATEC data practitioners has been carried out, the details of which are exposed below.					
Output 5	Ensure appropriate	Data producers,	Measure data producers'	Simple online satisfaction	During the discovery phase,

Training of the Data Producers in the implementation of the new data model.	training and support is provided to STATEC data producers to achieve their tasks on the LUSTAT.	recipients of training, are satisfied.	satisfaction on training provided. Number of data producers trained.	survey.	better definition of the relevant target segments of data producers within the STATEC.
STATUS: Achieved ***Comments:*** With the blueprint for the business domain activities adopted, during this phase of the project teams engaged with STATEC data practitioners to build capacity both through access to self-paced online training courses (See table of deliverables), and in-person and virtual workshops. In addition, workshops took place to support the STATEC IT infrastructure team to define a technical architecture and an enhanced deployment strategy. In total there were: - Number of data producers certified in .Stat Academy courses: 13 - IT infrastructure working group workshops and meetings: 3 sessions took place. A satisfaction survey of the STATEC data practitioners has been carried out, the details of which are exposed in 4.3.					
Output 6 Final report	Ensure the Project objectives have been achieved and process in place to measure outcome and impact in 2023-24	The .Stat Suite (LUSTAT) is installed and fully functional, in accordance with directions set in outputs 2 and 3, and a plan to progressively remodel and migrate STATEC data over 2023-24 is outlined.	Status of the .Stat Suite platform at STATEC; availability of the implementation plan in 2023-24.	Manual – statement in the final report.	
STATUS: Achieved ***Comments:*** Comments to be added after review by the Advisory Board.					

Note that Impact and Outcome indicators are also defined in the project definition, which are to be taken forward progressively by STATEC during the 2023-2024 period and once the reporting practice has been established.[3]

[3] A bid was put forward for the DG Reform TSI 2023-2024 period that would have supported a continuation of the project covering the implementation phase (*Modernising the production and dissemination of official statistics in the National Statistical Office of Luxembourg (STATEC)*).

4.3 SUMMARY OF FINAL SURVEY

Below is a summary from the final survey carried out with STATEC Practitioners at the close of the project, and who were the direct recipients of the training and support (Outputs). The group consists in:

- 3 staff from central team covering various profiles (IT, statistics, methodology).
- 7 Domain WG experts covering all areas (SOC, ENT, MAC).

They were asked several questions related to their participation in the defined activities, in particular the "by business domain activities" (workshops, meetings, trainings etc.). Two questions required a rating from 1 to 10. When a respondent scored 8 or less, they were asked to provide a short explanation. There was also a question to gauge the appetite for getting more involved or continuing in the same configuration, led by the central team with consultation. Finally, a free text box captured any additional feedback for the project team and for STATEC to note as they move beyond the DG Reform project closure.

4.3.1 AVERAGE RATINGS

What is your overall satisfaction with the activities that you took part in? (1: Very Dissatisfied, 10: Very Satisfied)

Average Rating: 7.3 (8 ratings >= 7, 2 ratings < 7)

The rating conveys an overall satisfaction with the project activities, especially SDMX related training and capacity building activities. The 2 lower ratings seem to be related to a wrong expectation as to what the project is supposed to deliver:

- One low rating (3) related to dissatisfaction with the tools provided. It should be noted that tool implementation or improvement is not within the project scope.
- One average rating (5) related to expectation of the data remodelling being effectively implemented. It should be noted that the implementation phase is not within this project scope.

Do you now have the tools and techniques to continue SDMX structural modelling beyond the initial experimentation phase? (1: Far from, 10: Yes, definitely)

Average Rating: 6.5 (7 ratings >= 7, 3 ratings < 7)

The rating conveys an overall satisfaction with the tools and techniques to continue SDMX structural modelling beyond the experimentation phase. The rating also conveys an expectation by users of an improved platform and tooling – an expectation which can considered as normal, as one of the project objectives was precisely to identify user experience gaps to guide tool implementations in the next phases. Several average ratings (7) underline that while the basics are there, improvements are needed in the tooling and training of users on LUSTAT platform usage. Many comments raise the concern of what kind of support will be provided in the next phases of the project beyond this initial phase, where OECD support has been warranted thanks to the TSI funding.

How do you see your role going forward regarding SDMX structural modelling at STATEC?

The large majority (6) would support the statement *"Happy with the work being done by the central team in consultation with me"*, with the remaining 2 respondents answering 'other', indicating no preference. This is indicative of an organisation that would rely rather on a stronger central modelling team effectively implementing the new models, and related data transformations, in consultation with data producers (rather than: data producers implementing themselves). This may be incompatible as an approach with the observed decentralised organisation of data production at STATEC; this raises especially the question of sufficient capacity in the core team (which is composed of staff from three areas: methodology, communications, IT). Several survey participants convey the notion that they were happy with the support provided by OECD but wonder what will happen next, during the implementation phase.

4.3.2 AN EXTRACTION OF THE COMMENTS COLLECTED

The comments made are provided verbatim below but reorganised by broad issues. Green colour corresponds to positive comments, orange to negative comments.

Project fundamentals: Why SDMX at STATEC?

- "I really appreciate the support provided by OCDE colleagues, as well the workshops organised by the OECD. The presentations of developments and the exchanges with other organisations allow us to learn from each other."
- "A key finding of the project which stood out, and surpassed even my first impression, was the role of modelling which is crucial in the context of .Stat and how the data is found and filtered. I'm not sure if all institutions moving to .Stat and SDMX are aware of this and even at STATEC level, this message needs to be communicated further."
- "This project was a good contribution to showing how STATEC can make better use of the LUSTAT platform."
- "I'm not sure whether the new model will ease the access to information for non-national-accountants."
- "SDMX is a standard for exchange of data, not for visualising data."

Organisation and governance: roles in STATEC and skills to grow

- "Having a relay in each unit is important to help the central SDMX team to conceptualise the DSD for each domain/thematic and to work on the disseminated data itself to fit the DSD conceptualisation."
- "Regarding the project itself, the start was very difficult as it was not clear what was expected from our side. The development of training courses and the organisations of the cross-domain meetings / workshops were very useful in this sense."
- "The governance of SDMX artifacts will be a first challenge as well as defining a plan to move forward from here with the resources available. The advantage is that we have now a methodologist among us, namely, Claude Lamboray, which has a specific previous price index domain background and knows the challenges of modelling data, metadata as well as reference metadata. He should be able to pilot initiatives to improve the data modelling in general at STATEC level. I appreciated very much the flexibility of OECD staff and the leadership of Eric and Jonathan to find the right way to help us to involve domain experts and to move forward."

Tooling and platform: which requirements?

- "The tools (DLM, Registry) should guide the user by focusing on workflow rather than keep them by SDMX lingo. A wizard or workflow-guided approach should be the goal, with a backdoor for IT specialists. Having to train users in SDMX only because the tools are not designed with the user in mind is inefficient."
- "The LUSTAT platform, more particularly the DLM, is not user-friendly."
- "The .Stat DLM, 'form-design' should only be used for training or experimentation purposes and not for actual production. Power-users do not have the required access to upload any structures into 'prod-design', which is inefficient."
- "I do not see how the metadata management could be improved by the use of the matrix generator, even if the use of SDMX standards can helps."
- "About the effective implementation: the entry cost to be able to properly manage the matrix generator is big and may not allow all the power user to be trained."

.Stat Suite and its deployment: more frequent updates needed

- "IT should deploy bug fixes (see bug pertaining to deleting/replacing artifacts in the DLM) to ensure the best possible UX. Furthermore, the update strategy announced by IT at the beginning of the project (i.e., two updates per year) should finally be put into action. Currently, the aforementioned bug hampers the deployment of new dataflows."
- "STATEC will have to be able to put in place the necessary resources so that releases can be installed more regularly to take advantage of the latest functionalities available."

General conclusions

Overall, these comments convey a relatively good understanding of SDMX data modelling and its implications in the STATEC landscape, and the importance of establishing a proper SDMX structural governance as part of the overall STATEC data governance landscape. From that perspective, the project fully delivered on its main goals. Clearly, frustrations are expressed in terms of tooling that can be interpreted in several ways:

a) **Expectations from the LUSTAT.LU project**: As this project overlapped with the migration project to LUSTAT.LU (from Beyond2020 to the .Stat Suite), which was indirectly linked but outside the scope of the project, some reactions are linked to the product itself not always being in line with expectations of data producers for several reasons. Here are two major ones being firstly the understanding of the importance of data modelling mentioned previously on the way data is experienced by users, but also how data can be found. Secondly, the complexity of SDMX requiring adaptations in the following domains: harmonisation, governance and process flow.

b) **Interrogation on the division of labour between domains and the STATEC core team**: Some of the frustrations may manifest the deeper interrogation, on the mind of the data practitioners, on the nature and scope of the transformation needed and the roles of the business domains versus the central team. Especially, dependent on the model adopted for the remodelling and migration work. Indeed, today and by default: this model is essentially decentralised and requiring a relatively high level of SDMX skills in domains; alternatively: this model could be more centralised (STATEC core team doing most of the work in consultation with domains, rather than assisting them doing the work), but that would require strengthening the capacity of the core team.

c) **Technical and functional limitations, that can be fixed**: Platform implementation is not part of the project scope; so, it is expected that limitations surface that should be addressed during the later implementation phase. Limitations observed relate either to delay in the deployment of new features, or missing features in the .Stat Suite: (i) The STATEC IT team has difficulty in releasing newer version of the .Stat Suite. This leads to features and enhancements not being made available to users at pace, leading to understandable frustration. An action plan should be designed to enable more frequent upgrades. (ii) Issues faced by users have been partially analysed under output 4 – but more analysis will be needed as a result of the survey outcome, in later phases of the project – also taking into account that the chosen division of labour between domains and the STATEC core team (see point b) will have a significant impact on the functional expectations.

5 LESSONS LEARNED

Below are the main lessons learned from the overall project organisation and delivery, including the STATEC-specific perspective.

5.1.1 PROJECT AGILE ORGANISATION PROVED TO BE FIT FOR PURPOSE

The decision (at the end of the discovery phase) to switch to an agile parallel delivery of outputs 3, 4, and 5 was a key success factor for the project. This allowed for a better management and planning of the delivery, as well as an identification of issues (especially, staff availability and acceptance of the project) giving the core project team more opportunity to adjust and mitigate as necessary.

The agile approach has proven to be beneficial in the STATEC context to make good progress across the different business domains and adapt to their specific constraints – especially the limited time availability of the data practitioners, and the need to develop a self-paced approach as availability may differ from one team to the other, due to varying planning constraints. This is captured especially in the output 3.2, “blueprint for the by domain activities.”

The e-learning platform allowed participants to follow the training in a flexible manner, at their own pace. Feedback during the elaboration process allowed OECD and STATEC to improve the course. The combination of online learning with onsite coaching was essential for business domain experts to consolidate their skills needed to continue with data modelling activities in the future.

The bi-weekly project meetings between the OECD and STATEC project teams ensured efficient coordination, resource planning, scheduling, and monitoring of the project deliverables. OECD interventions were scoped during those meetings and executed in short order for the work to be continually delivered by the project team. Risk and mitigations were reported, based on the previous, at the project Advisory Group meeting.

5.1.2 BETTER UNDERSTANDING OF THE LINKAGE BETWEEN GOOD DATA MODELLING AND DATA ACCESSIBILITY & USER EXPERIENCE

An experimentation phase was important to see the benefits of the data modelling and to ensure engagement from the business domains. Data modelling has an important impact on how data is searched, filtered, and consumed. To make sure the user experience is good, further evaluation of the accessibility of the disseminated data tables will have to be done on a regular basis.

The experimentation phase clearly showed the need for setting up a governance architecture around data modelling (for example the management of shared code lists), as well as awareness for business domains to think more carefully about the concepts and code lists that are used in their tables which also have an impact on the visualisation and filtering on the LUSTAT user experience. By analysing the current situation, the project triggered a more general discussion about how data should be disseminated (for example merging existing tables or creating new tables).

5.1.3 THE PROJECT IDENTIFIED SEVERAL OPPORTUNITIES AND CHALLENGES IN THE AREAS OF DATA INTEGRATION THROUGH AUTOMATIONS AND INTERNAL DATA SHARING

Potential use cases have been identified for internal data sharing. Such a space has been requested by a few business units.

The data publication process needs to be streamlined as much as possible since STATEC have no homogenous production processes in place, and this would be a first step to enable a semi-automated dissemination flow. VIVA, the existing STATEC platform for file sharing, could potentially centralise the most important data sets for internal exchange, and feed into the internal .Stat Data Explorer to enable better data integration. It could be complemented with a mechanism to automate scheduled tasks developed in R or Python (such as an “Algorithms Bank”, see below). This approach would amount to an SDMX based data integration step, feeding into internal sharing, reporting and dissemination use cases; it could bring important value but needs further development and promotion.

Automating .Stat Suite deployments is an important aspect to make different instances of .Stat Suite available rapidly and to facilitate version upgrades. This topic should be addressed by the STATEC IT Department in collaboration with the OECD .Stat Suite product team.

6 ROADMAP FOR THE EXTENDED LUSTAT PROJECT COMPLETION AND COMMUNICATION PLAN

It is important to note that the DG Reform TSI 21LU01 project was part of a larger transformation project at STATEC, with a migration from the legacy data portal to the new LUSTAT based on the .Stat Suite. The project focused essentially on the design and capacity building activities, that would provide the needed foundational elements to build on for a possible future implementation phase through the 2023-2024 (and beyond) time frame. Figure 6 shows the TSI 21LU01 project planning as well as the activities completed at project closure.

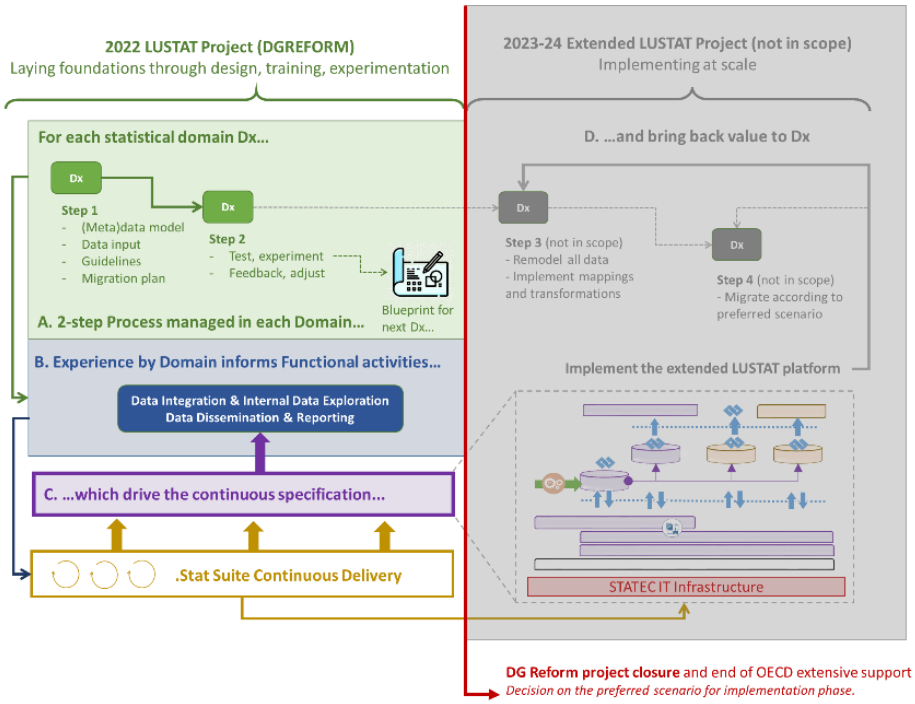

Figure 6: TSI 21LU01 project high-level plan and activities

6.1 ROADMAP FOR THE EXTENDED LUSTAT PROJECT COMPLETION

Figure 6.1 shows the draft proposal for the high-level platform architecture that was defined as part of the Discovery Report – extending potentially the existing LUSTAT platform (devoted to the dissemination and reporting use cases) to support other use cases (especially, internal data sharing and data integration). The reminder of this section describes a refined vision captured over the course of the project along three key aspects: Target Data process, Target architecture, Implementation and migration approach.

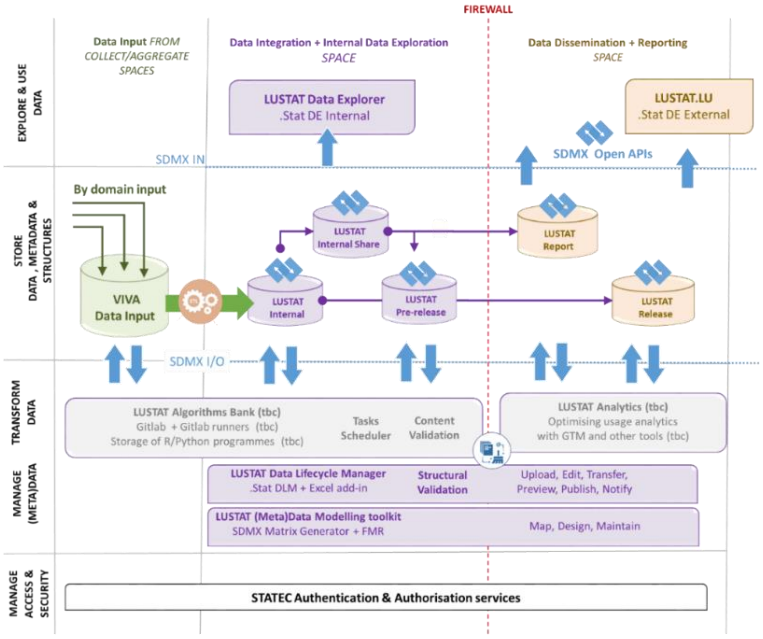

Figure 6.1: High-level view of extended LUSTAT, target technical architecture (as per Discovery report)

6.1.1 TARGET DATA PROCESS IN THE EXTENDED LUSTAT

As described in the discovery report, different statistical domains use various data preparation processes, leading to a siloed approach that has so far limited the possibilities of automation and process streamlining.

Several business domains were identified as using the DLM Excel plugin to update data, but, for example, the MAC business domain is creating CSV data files directly via a pre-defined data process. The output is mainly for sharing data with Eurostat. This is not necessarily an issue, but each siloed process would ideally, at the very least, follow a standard protocol and feed into a single data integration pipeline for serving the downstream data spaces including for dissemination.

As part of the future data process and to migrate from existing (legacy) structures and data process to the new model(s), structural mapping is to be put in place, in a way that allows for mapping automation and repurposing across teams. Figure 6.1.1 summarises the steps and tools to use to create a mapping and move to the new data structures, through transformation of the legacy structures to the new structures. Structural mapping describes the relationship between data structures that are used in the transformation. The mappings need to be created and the transformation processes and system need to be implemented. The Fusion Metadata Registry (FMR) workbench and mapping microservice, integrated as a building block in the .Stat Suite, can be used to create the structural mappings and to execute the transformations. It provides mapping sets that facilitate the automated processing of data. Further details of this have been described in the "Target SDMX Structural model and modelling process for STATEC" document, section 3.2.

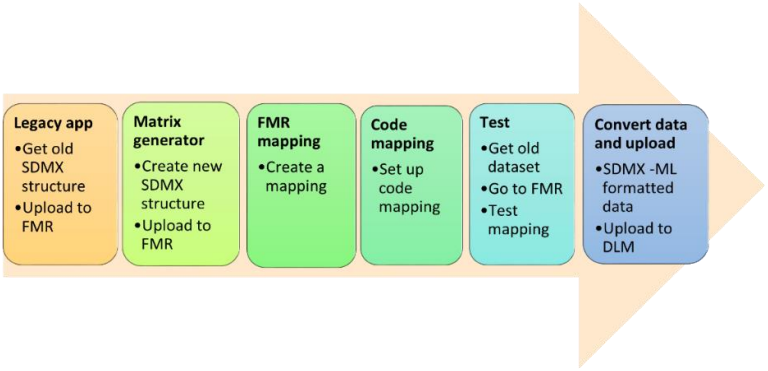

Figure 6.1.1: The target migration and mapping process

The introduction of referential metadata[4] will be an important addition to the existing process. At STATEC, there is no reference metadata structure or process defined yet for the already disseminated datasets. This is in part due to the fact that the existing LUSTAT platform was deployed at a time when referential metadata was not an available feature in the .Stat Suite – which it is now, allowing for the provision and dissemination of referential of metadata, provided the LUSTAT platform is upgraded and the process to upload referential metadata in in place. Additionally, STATEC will need to define referential metadata structures which do not exist yet for dissemination purposes. By default, the existing process to upload and edit referential metadata, and data, is based on the usage of the .Stat DLM Excel plugin – which should remain as a useful capacity for manual intervention, even though automations should be mainstreamed.

It is expected that the data process can be greatly improved by integrating VIVA as a hub to store and share data files and source data by loading data from each of the business domain production pipelines through utilising statistical tools that support the agreed input format. As analysed during the project, the .Stat Suite brings the referential metadata storage and management features needed, and VIVA brings the file storage and sharing service. Combining the two as part of the extended LUSTAT platform would allow to support the data integration step in a standard way, allowing for automations: data would be made available via a standard API in

[4] Reference Metadata is structured information that describes data. For example, information about a dataset, such as data source, compiling methodology, statistical concepts and definitions, how the data was collected, legal information, contact information and so on.

SDMX CSV format and transferred to an internal data space for further processing, analysis, and dissemination and reporting. So far, a decision has not been made as to whether the input file stored in VIVA is to be transformed to SDMX CSV by extending VIVA functionality, or through an external transformation service. For example, using Gitlab pipelines with scripting in R and/or Python. This vision for the extended LUSTAT target architecture, combining .Stat Suite and VIVA to better support data integration, will require further analysis.

6.1.2 TARGET TECHNICAL ARCHITECTURE FOR THE EXTENDED LUSTAT

Figure 6.1.2 shows the high-level target technical architecture, updated from the initial vision in the discovery report, following iterations and lessons learned during the subsequent phases of the project.

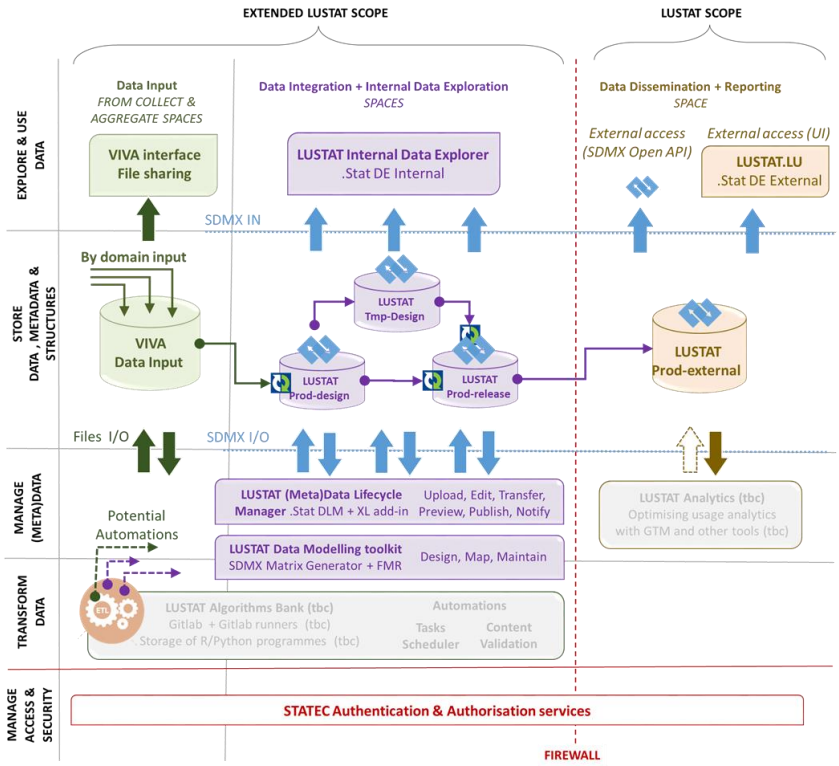

Figure 6.1.2: High-level view of extended LUSTAT, target technical architecture (per closure report)

The basic structure of the architecture remains unchanged, however building blocks have been refined, with further clarification drawing on the lessons learned during project implementation. The architecture is partially implemented at project closure (mainly, the data modelling toolkit); however, most of the implementation is outside the scope of the TSI 21LU01 project. It should hence be updated in the future based on further analysis and iterations in the actual implementations. In the following sections, key points are highlighted related to each of the layers identified in the target architecture.

Explore & use data layer (external)

- The "LUSTAT.LU Data Explorer" has been put in production in the first semester of 2022. Several enhancements have been identified and scoped as part of the project (see table of deliverables, output 4) – delivering those priority requirements will be taken forward by OECD as part of the SIS-CC governance process, in the next phases of the project. Also, delivering new features more quickly, to improve user experience on an ongoing basis, has been identified as an area of progress requiring more frequent updates of the .Stat Suite at STATEC. The Data Explorer relies on the "LUSTAT SDMX Open APIs", which can be used by external programmes as well, and have been put in production concurrently in the first semester of 2022.

- It is important to note that LUSTAT.LU is part of a broader communication and engagement experience developed by STATEC as part of their corporate website, statistiques.public.lu, based on AEM (Adobe Experience Manager). While a first phase of the web integration has taken place, the project has revealed more opportunities for improvement – especially, in facilitating search of LUSTAT.LU data flows from within the overall website search – and the reverse, possibility to return certain pages and resources (pdf files especially) as part of the LUSTAT.LU search experience. As OECD is renewing its web environment based on the same technology (AEM), there might be interesting opportunities for better integration to explore between STATEC and OECD going forward.
- LUSTAT Analytics is identified as the module to support quantitative analysis of user experience on LUSTAT, and progressively establish a feedback loop where data users' insights (on the container-platform features, as well as the content-data products). LUSTAT Analytics typically features market standard tools such as, Google Tag Manager (GTM) as well as flash survey tools (e.g., crazyeggs) to monitor the profile of data users and the nature of their usage of STATEC data. This work could greatly contribute to the future implementation project impact measurement objectives. However, this configuration (shipped out of the box from the .Stat Suite open-source project) is different from the toolset currently used for the LUSTAT data portal KPIs. More analysis would be needed on the STATEC side to define the best holistic approach to analyse the data user experience.

Explore & use data layer (internal)

- The "LUSTAT Internal Data Explorer" allows the user to access data from all spaces in the STATEC internal network zone, prior to their dissemination. It may allow for sharing data pre-embargo or data that is never released publicly. It allows to preview the data in its final form before it is released. The internal data Explorer works also very much as a companion to the data management tools, for data practitioners to search and view data and metadata more easily at earlier stages than dissemination.
- VIVA being positioned as the corporate file storage and sharing environment across all STATEC domains in combination with the internal Data Explorer, is a relevant way of covering non-SDMX data sharing (accessible through VIVA interface as files) and SDMX data sharing (through the search interface of the Data Explorer).
- Dependent on the usefulness, the retrieval of SDMX data from any of the spaces (listed below) would possibly be able to use the .Stat DLM Excel add-in (see below).

Store data, metadata & structures layer

- The "LUSTAT.LU Prod External" has been put in production in the first semester of 2022, feeding the externally facing Data Explorer. It is a publicly exposed space outside of the firewall where "*released*" data is made available through the SDMX API and the external Data Explorer (LUSTAT.LU). It is important to note that the updating of the data held within this externally facing environment (Prod-External) so far been managed through a manual back-up and restore procedure carried out directly by the STATEC IT team. For the sake of a more efficient and secure procedure, is proposed to use the out-of-box feature of the .Stat DLM Transfer service, allowing a central data/release manager to perform the task directly with incremental updates applied to the target data space.
- To support the data integration use case, two principal spaces are to store the data, metadata and structures as "native SDMX" that can be retrieved and updated only through standard SDMX web services, input and output. These are: ""LUSTAT Prod-Release" mirroring LUSTAT.LU Prod External, for data practitioners to validate data before effective dissemination and availability to external users, and "LUSTAT Prod-Design" for the preparation of the data – including mapping from sources structure to internal or external dissemination and reporting structures, as well as potential editing of the data and metadata.

- For as long as the migration takes place from the current data models and process to the target data model and process, a temporary space is to be commissioned, "LUSTAT Tmp-design" to support the ongoing remodelling exercise, and to continue to protect the existing production data[5]. This new data space will be used for remodelling, validation, and preparation of the data ahead of any future migration. The migration is to happen only when deemed appropriate, and with the relevant communication to minimise impacts on data consumers. Once the complete remodelling has been completed and migration has taken place, this space will be removed.

Manage data and metadata

- The .Stat Data Lifecycle Manager (DLM), plus its Excel add-in for editing through an Excel interface data and metadata, has been deployed initially as a part of the LUSTAT.LU back-office, and potentially for managing data, metadata and structures for all spaces. It is used in combination with SDMX structure authoring tools (SDMX Matrix Generator for the initial design work and SDMX training of data practitioners; FMR for the maintenance of meta(data) structures and structure mappings).
- SDMX Matrix Generator is expected to remain a key tool in the continued work to remodel existing, and new, data structures by each of the STATEC business domains, and further acquiring SDMX structural modelling skills. FMR, so far not well utilised, is expected to become a key component in the mapping and future ongoing maintenance of data structures, likely to only be used by the STATEC central team.
- While this toolkit does serve well the usage by a central team of experts, it does not necessarily cater for a larger network of data practitioners – somewhat aware of SDMX structural modelling issues and contributing to it. This is exactly the situation in STATEC – as revealed by the interactions and formalised in survey results. The usage of this toolkit by a broader network of data practitioners involved in data modelling activities will require a functional simplification and better integration of the various tools. Some steps have been taken (materialised in the much better integration of .Stat Suite and FMR, for example[6]) but more needs to be done, and this project hints at the direction STATEC and the broader Community should take.

Transform data

- Data transformations (for integration purpose) take place fully manually in the current setting (e.g., data practitioners manipulating files – download and upload, transforming them by executing appropriate algorithms, and then uploading the corresponding result in the appropriate space, and checking the results for validation). A large degree of automation could be achieved by introducing standard protocols for orchestrating those tasks – including tasks scheduler and content validation.
- During the project, OECD demonstrated how this could be done using open-source tools (Algorithms Bank based on Gitlab repository, Gitlab runners, to manage and schedule the execution of algorithms coded typically in R/Python, manipulating SDMX objects stored in the various spaces, and executing transfers and mappings through web services).

[5] Note that, during the project, a "Training space" has been in use to support the data modelling training activities (output 5), as a temporary, experimental environment, isolated from the published data in the new LUSTAT data portal. This environment is considered non-stable and is to be decommissioned.

[6] As part of the sister TSI 21BE03 project, the FMR has been enhanced in close collaboration between the OECD (SIS-CC) and BIS to implement a new structural editor, FMR Workbench, that can sit over any SDMX 2.1 data store, including the .Stat Suite meta(data) storage. Therefore, removing the need to deploy an additional registry storage and reducing the need for duplication of structural data.

- Decision on the potential implementation of an Algorithms Bank at STATEC and entailed work on promoting source coding and data pipeline automation among STATEC data practitioners is something to include under the next phases of the project. This component could capitalise on the experience acquired in the STATEC IT team using Gitlab as source coding environment for IT projects, as well as the experience obtained by the OECD through their own implementation of the OECD Algorithms Bank and the replication of it under the sister TSI 21BE03 project in the National Bank of Belgium.
- VIVA being positioned as the corporate file storage and sharing environment across all STATEC domains, its combination with a STATEC Algorithms Bank could be a powerful lever (effectively, an ETL or Extract, Transform, Load) in order to automate data integration between upstream data pipelines and the LUSTAT environment, allowing for file storage and feeding LUSTAT (Prod-Design).

Manage access & security

- "FIREWALL" delimits the external network zone (like a DMZ) from the STATEC internal network zone.
- "STATEC Authorisation & Authentication services" designate services as managed in the STATEC IT infrastructure, to support access control on data and structures internally (segregation between the various domains and the core team with higher privileges), as well as access control on data disseminated under embargo to specific audiences. Notably in the past months has been the convergent efforts by the OECD, Eurostat, and BIS to adopt a common framework for all SDMX tools based on standard services OpenID – which should progressively allow for a much simplified and interoperable landscape.

6.1.3 EXTENDED LUSTAT DEPLOYMENT APPROACH

The current LUSTAT deployment approach adopted by STATEC requires a heavy manual operation that has proven to be less than optimal with only one upgrade of LUSTAT per year. This has led to an already outdated version of the underlying .Stat Suite platform being deployed into LUSTAT production environment early in the project, and subsequent feedback received during the surveying of STATEC data practitioners at the close of the project. Figure 6.1.3a shows how this is currently done.

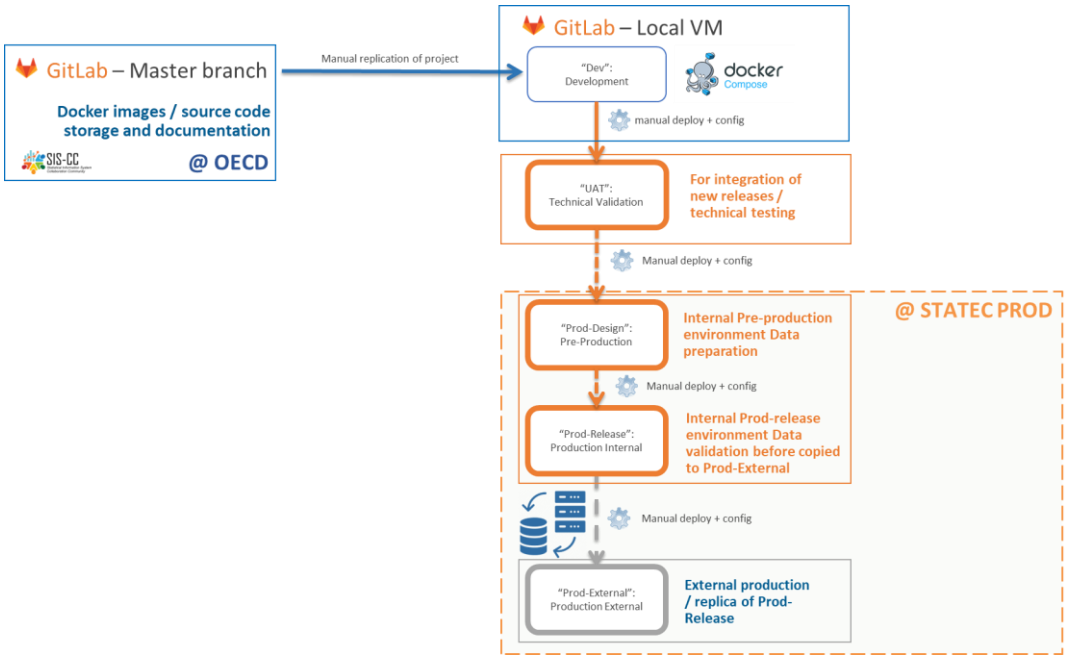

Figure 6.1.3a: Extended LUSTAT target deployment approach technical

- **Dev "local"** is used and accessed by one STATEC developer to manually retrieve and install the .Stat Suite docker images from the SIS-CC Gitlab project where checks are made against a local configuration.
- **UAT** is used for technical integration and testing of new versions of the .Stat Suite before upgrading of LUSTAT production. A manual deployment is performed by the STATEC IT team with support from the STATEC developer in charge of retrieving and installing the .Stat Suite docker images in the **Dev "local" virtual machine**.
- **Prod-Design** and **Prod-Release** instances are updated once technical validation is completed on UAT. This update is performed manually by STATEC IT team, currently one per year.
- **Prod-External** is updated following the updating of the internal instances and internal user testing is complete. This process is manual with configuration applied. It should be noted that the Prod-External instance is not accessible from the internal services therefore the moving of validated data for dissemination is moved using a technical back-up of the Prod-Release (internal) database with a restore to the Prod-External database. This task is carried out by the STATEC IT team.

The way forward

Through joint discussions and knowledge sharing between the OECD experts and STATEC IT Team, an adoption of DevSecOps practice with automated upgrades through containerisation technology is foreseen to materialise in 2023. This enhanced deployment approach will bring about increased automation and the possibility to take advantage of more frequent updates of the underlying .Stat Suite platform, moving from one per year to at least two, and hopefully more as the team builds greater capacity and knowledge in the process. This will also address the related comments received as part of the final survey of STATEC data practitioners. Figure 6.1.3b shows how the target deployment approach would look like.

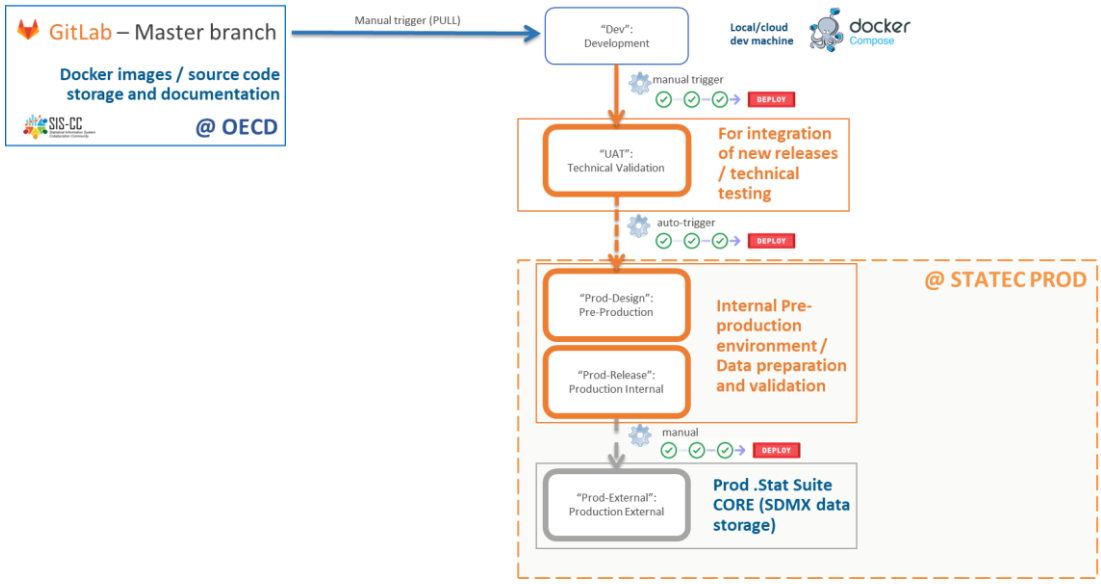

Figure 6.1.3b: Extended LUSTAT target deployment approach technical

Most of the instances remain as is except for **Dev "Development"** that is to be centralised on common infrastructure and could be extended to serve a wider number of developers if STATEC wish to build more expertise in developing internal services and/or contributing to the .Stat Suite project.

The distinct differences introduced with this adjusted approach include automation through DevSecOps adoption enabling each instance to be updated almost simultaneously and thus removing the current manual procedures carried out by STATEC IT team.

In addition, the current manual technical back-up of the **Prod-Release** (internal) database and restore to the **Prod-External** database is to be moved to a non-technical procedure utilising the out-of-box feature of the .Stat DLM Transfer Service, allowing a central data/release manager to perform the task directly with incremental updates applied to the Prod-External data space.

How far STATEC go in automating the upgrades of Extended LUSTAT (new versions of .Stat Suite) will depend heavily on business drivers, automation of integration tests, validation of new features by STATEC data practitioners, security testing, and technical capacity.

6.1.4 IMPLEMENTATION AND MIGRATION APPROACH

It is apparent that the extended LUSTAT is not just a platform or an application. It's an environment consisting in multiple components – leveraging generic .Stat Suite components, but also the existing file sharing environment VIVA, and others (FMR, potentially an Algorithms Bank, etc.). What is important when implementing the future target technical architecture is to:

- **a)** maintain a holistic view of the environment – instead of fragmented approach resulting from years of legacy layers adding up with no global approach,
- **b)** preserve a service-oriented approach, whereby each component can be later replaced by another component featuring the same services with limited impact on other component (decoupling) and, finally,
- **c)** model-driven, model-at-the-centre approach.

The **SDMX information model drives the architecture** not only because it enables component decoupling (b), but also because it provides the *lingua franca* for all STATEC business domains to share the same infrastructure and to capitalise on similar data pipeline automation practices and codes, and for statistical concepts to automatically translate into technical artefacts (such as database) and, ultimately, improved user experience through data accessibility.

The outcome of the DG Reform TSI 2021 project has helped to define the way forward for an implementation as part of a future project to be launched. It is important to note that the implementation has already been well-described as part of a request put forward for the DG Reform TSI 2023-2024 period that would have supported a continuation of the project covering the implementation phase (See: "Modernising the production and dissemination of official statistics in the National Statistical Office of Luxembourg (STATEC)" – TSI 2023 request).

A future project could consist in the **implementation of the target data model and process** on the .Stat Suite by STATEC with the appropriate continued assistance in training and support. A full migration of the 4 business (statistical) domains to the new environment, the automation of data transformations, and an upgrade of the target platform to its final production version. This is to be achieved through a number of activities building on the work done under TSI 2021 project, and the established business domain groups and data modelling governance. These are to include continued training and coaching on SDMX data modelling and good practice of STATEC data practitioners; support implementation of data transformation algorithms, tools usage (.Stat Suite); refined specification, validation, and testing of data platform features; and on demand interventions and ad-hoc workshops based on feedback received from STATEC data practitioners or consumers.

For STATEC **to implement the target data model and process**, they must first consolidate the SDMX governance. A basic governance model goes hand-in-hand with the data migration and will avoid the detrimental effects the current decentralised model brings whereby each business domain maintains their own codelists and other SDMX artefacts. This can be seen most prominently in the LUSTAT Data Explorer with duplicated, yet conceptually identical concepts and facets, making it difficult for end users to find and use the data.

The governance model in figure 6.1.4a is a tentative suggestion for STATEC, considering their specific needs and constraints[7]. These options need further analysis and adjustment in the STATEC context, to identify the most appropriate compromise between the different criteria (metadata quality, metadata reuse, efficiency) considering the main constraints (low resources and siloed production). Further details of the target structural model have been described in the ***Target SDMX Structural model and modelling process for STATEC*** document, section 3.1.

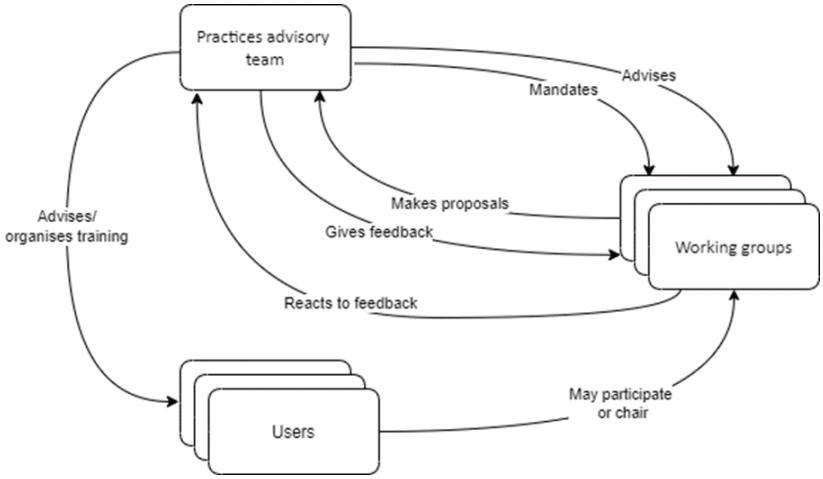

Figure 6.1.4a: Recommended SDMX structural metadata governance for STATEC

Concretely how will the migration happen?

With the current STATEC data dissemination platform already in place, and with a large amount of data, communication and careful planning is a must for the chosen migration strategy. The way in which the current disseminated data has been structured has demonstrated clear difficulties for end-users to filter, pivot, and visual data. Although most of the data structures use similar concepts or dimensionalities, the duplications, and redundancies in terms of naming and coding convention can be observed across the different business domains. The data structures do not conform to standard data modelling approaches, such as the methods described in the Statistical Data and Metadata Exchange (SDMX) guidelines. These deficiencies appeared because the existing data structures were not sufficiently well analysed and the migration to .Stat Suite had to be completed (in 2021-22) under tight deadlines with limited support and training having been conducted with the STATEC data practitioners at that time.

In addition to public data dissemination, reporting to international organisations is a key activity in the overall data process. The discovery report captured the reporting activity as being siloed with individually defined procedures, tools and formats that often rely on one individual to execute a series of manual steps. However, STATEC mainly transmits data to Eurostat via the Edamis online platform, and mostly the data transmissions are based on the SDMX standard.

Two scenarios are proposed for the migration, with each having pros and cons, but neither being fully optimal due to the constraint of having already disseminated and public access to the poorly modelled data.

[7] Challenges and opportunities encountered by STATEC data practitioners have been extensively documented in the 21LU01 project Discovery Report, section 5.

1. **Remodel in LUSTAT (Tmp-Design) and migrate in batches per business domain as they become available, replacing existing data in the public facing LUSTAT (Prod-External):**
    a. **Pros:** This allows good model data to be made available more quickly and reduce the need to maintain two parallel data update processes over a shorter period and with less duplication of data storage. STATEC could also initiate a feedback loop with end-users that can help to continually adjust the data process and target model(s). Encourage a greater level of engagement with STATEC business domain experts by quickly demonstrating the benefits of good, modelled data, thus increasing the speed at which the overall remodelling takes place. This scenario implies an agile approach that proved a key success factor of the TSI 21LU01 project delivery.
    b. **Cons:** LUSTAT (Prod-External) will host a mix of good and poor modelling that will have further detrimental effects on the user experience and the ability to cross-compare/analyse with low data reuse capabilities. Although the existing modelled data already limits the user experience and data reuse.

2. **Remodel all data for all business domains in LUSTAT (Tmp-Design) and migrate in full replacing the current LUSTAT (Prod-External) data:**
    a. **Pros:** LUSTAT (Prod-External) would remain at the current level of user experience from the existing poorly modelling with no mix of good and poor models. Communication to end-users would only be needed one-time ahead of the one-time migration.
    b. **Cons:** STATEC data practitioners would have to maintain two parallel data update processes with high duplication of data storage over a much longer period. The resource implications of this scenario would be high; this issue, however, could be potentially mitigated with the appropriate automations set-up. This scenario implies a waterfall approach without iterations and an established feedback loop with end-users that can help to continually adjust the data process and target model(s); this issue, however, could be potentially mitigated by setting up a temporary, externally facing space, to host the beta version of LUSTAT.LU with the remodelled data, during the time of the transition.

There remain some key open questions to decide and in order to develop each scenario further:

a) **How do we communicate with end-users** on a continuous basis and get their feedback especially regarding the issues they experience accessing the data? How is that feedback loop to be established?
b) **How do we govern the continuous modelling effort**, in order to make decisions progressively by domain and globally on the data model. How to facilitate for additional SDMX modelling expertise?
c) **How is remodelling and migration work organised**: by decentralised teams (domains) as the case today by default? Or by the core team, the delivery capacity of which should be strengthened? This decision has an implication on the back-office tooling, as revealed by the data practitioners' survey.
d) **How fast do we go**: What kind of assistance will be needed? How do we prioritise the business domains? How fast we can go versus the available resources?

Figure 6.1.4b summarises, at a conceptual level, the high-level view of the next phases of the project "extended LUSTAT project, implementing at scale":

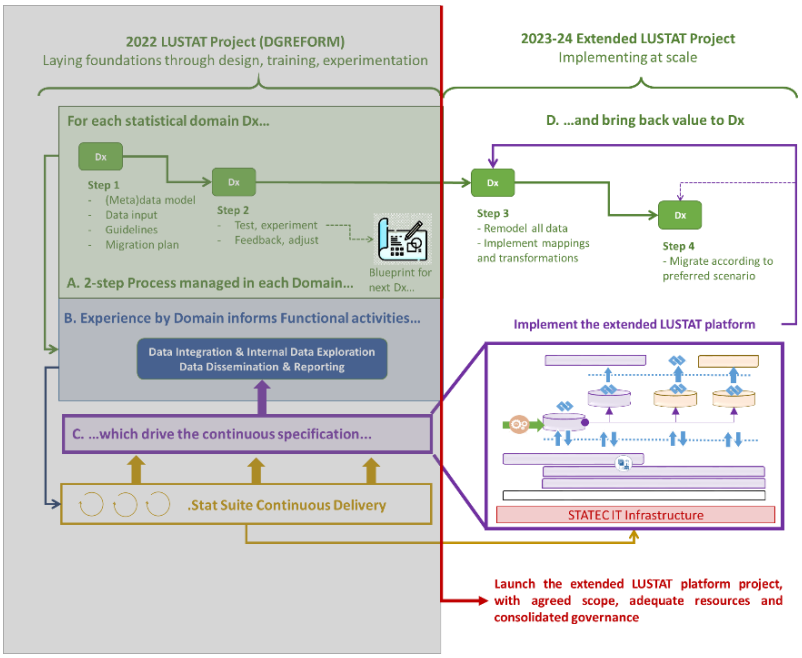

Figure 6.1.4b: High level view of the Extended LUSTAT project, implementing at scale

6.2 COMMUNICATION PLAN

The communication within this project and beyond is organised at several levels to serve different stakeholders being informed with a different level of detail and frequency according to their needs.

Internal communication: Communication activities will be launched to inform STATEC data practitioners on the outcome of the TSI 2021 project and on the next steps. As the project moves into the next phase, it is important that the STATEC data practitioners remain involved and aware of how the experimentation phase will translate into an implementation phase. As a follow-up of the design and capacity building phase, a STATEC internal project will be launched that aims at implementing the target data model and process and migrating to the LUSTAT platform. A Project Committee will steer this project. All relevant stakeholders will be involved in this project. Being part of STATEC's portfolio of projects, reporting to higher management will be done on a regular basis.

External users: While the actual implementation phase is only to start, STATEC will already communicate upfront about the project for example by publishing short reports about past and planned activities. It is important to give visibility to the modernisation efforts conducted by STATEC and to make users aware of upcoming improvements of the data dissemination. The project will also be presented within the framework of the Luxembourg national statistical system. STATEC has the lead role within this system and the other members are both users and providers of the LUSTAT platform. Once the new data tables will be released, further communication efforts must be done to ensure external data users are appropriately informed about the changes that will impact the disseminated data tables. In all these communications, appropriate reference will be made to the DG Reform support.

OECD and SIS-CC: The e-learning courses delivered through the .Stat Academy have been instrumental in building the needed capacities to work with the .Stat Suite platform. In the future, STATEC will continue to encourage its relevant staff to follow these courses. STATEC is available to organise together with SIS-CC joint communication activities to further promote the training material.

EC DG Reform: The realisation of the first phase of the project, dataset remodelling experimentation was supported by the European Commission DG Reform of via the TSI 2021. During this phase, the STATEC, OECD, and DG Reform communicated on the progress of the project via the quarterly advisory group meetings. This

meeting configuration comes to an end with the closure of the DG Reform project. STATEC, EC and OECD should liaise to set the parameters for future public communication which should take typically involve coordination of communication on the social media and the wording of it (quote), and reference to the TSI 2021 support to creation of e-learning material, namely .Stat Academy courses, to support the capacity building of STATEC data practitioners in SDMX data modelling practice.

Other Fora: The project is a good illustration of how an SDMX-based data integration environment could be developed to support data dissemination and internal data sharing use cases. STATEC is willing to share its experiences with other interested parties via international fora, such as the SIS-CC workshops, workshops and conferences organised within the European Statistical System, as well as other relevant events such as the SDMX Global Conference.

7 ANNEX: PROJECT CLOSURE MEETING SUMMARY RECORD

To be added following project closure meeting.