

B13: Checklist of Things to Consider in Any Operational Plan for Data Integration

- Who is this tool for? Health and other social protection leaders who need to design an operational plan to implement a data linkage or integration between their population targeting datasets. It provides a checklist of things to consider, so as to avoid missing out crucial elements of data linkage implementation, based on the experience of other systems who have undergone this process.

There are many checklists by different organizations of things to consider and cover in operational plans for data integration. Some are higher level and some focus more on the technical aspects of preparing for data interoperability. Each takes a somewhat distinct angle on the questions to consider in developing the operational plan. The list below and attached Technical Note from Talend are just some examples, but provide a flavour of the issues that are commonly included and the variation between higher order checklists and more technical ones.

SMS. (2018) “Data Integration Checklist: 6 Steps to Success” The SMS Group for Data Collection Integration. <https://www.thesmsgroup.com/data-integration-checklist-6-steps-to-success/>

Step 1: Identify key stakeholders	<i>Who needs to be involved in this project?</i>
	<i>Have you involved a representative from each department?</i>
	<i>Does everyone involved have defined roles?</i>
	<i>Do you have someone who will champion this project?</i>
Step 2: Establish your key objectives	<i>Why are you undertaking this initiative?</i>
	<i>What are you hoping to achieve with this project?</i>
	<i>What are your expectations?</i>
	<i>What is your timeline for this project?</i>
	<i>What processes (both manual and automated) are currently in use?</i>

Step 3: Assess your current processes	<i>What legacy systems are in place?</i>
	<i>What (if any) cloud-based data stores are in place?</i>
Step 4: Identify your data	<i>What are you getting from your data?</i>
	<i>What do you want to be getting from your data that you aren't currently?</i>
	<i>Do you know the source of the data?</i>
Step 5: Evaluate data integration solutions	<i>Is it affordable?</i>
	<i>Will you require additional hardware?</i>
	<i>Will the maintenance costs outweigh the value it brings?</i>
	<i>Is your in-house team equipped to use it?</i>
	<i>Will it scale to your business's changing needs?</i>
Step 6: Monitor and manage	<i>Do you have metrics in place to track success?</i>
	<i>Do you see trends that will help future data integration projects?</i>
	<i>Is there a need to scale up or scale down?</i>

Pythian. (2018) "Your Data Integration Checklist" Pythian.

<https://resources.pythian.com/hubfs/Tools/Data-Integration-Checklist.pdf>

Step 1: Identify Key Players and Stakeholders	<i>Which departments will be most involved early on?</i>
	<i>Who will form your data governance committee?</i>
	<i>Which department will be your data "guinea pig"?</i>
Step 2: Take Stock of the Landscape	<i>What data solutions are currently deployed in the organization?</i>
	<i>Which still have value, and which don't?</i>
	<i>What's being done manually that can be automated?</i>

Step 3: Audit your Data	What are your most important data sources? What new data sources are you using, and do you anticipate bringing in other sources in the future?
	Is your data structured or unstructured?
	How much time is the organization spending cleaning and de-duplicating data?
Step 4: Define your Master Data Management Strategy	What terminology will be used, and when?
	Which rules will govern data use and access throughout the organization?
	Does your data management platform have the flexibility to handle varying MDM requirements?
	Who in the organization will set and enforce these rules?
Step 5: Identify the appropriate platform and tools	Will you need to scale your platform as your organization grows?
	Will you require flexibility to add new data sources and types?
	Do users require real-time data?

Hexaware. (2009) “ Data Integration Checklist – Environment Setup & Process Design” Hexaware: BI & Analytics. <https://hexaware.com/blogs/data-integration-checklist-environment-setup-process-design/>

The Hexaware checklist is a more detailed and technical one which focuses on the data itself and which would be more suitable for the IT teams involved in the data linkage project.

Data Integration Environment Setup
<ul style="list-style-type: none"> Repository setup and folder structures to hold the development objects (code) like transformations/mappings/jobs.
<ul style="list-style-type: none"> Coding standards and development process.
<ul style="list-style-type: none"> Document templates for low level design specifications and for capturing test case & results.
<ul style="list-style-type: none"> Version management process of the objects.

<ul style="list-style-type: none"> • Backup and restore process of the repository.
<ul style="list-style-type: none"> • Code migration process to move the object from one environment to the other like from development to the production environment.
<ul style="list-style-type: none"> • Recommended configuration variables like commit interval, buffer size, log file path etc
<ul style="list-style-type: none"> • User group and security definition
<ul style="list-style-type: none"> • Integration of the metadata of the database with the DI metadata and that of the DI metadata with the reporting environment
<ul style="list-style-type: none"> • Process for Impact Analysis for change request
<ul style="list-style-type: none"> • Data Security needs for accessing the production data and the process of data sampling for testing
<ul style="list-style-type: none"> • Roles and Responsibilities of the environment users like Administrator, Designer etc

<i>Data Integration Process Design</i>
<ul style="list-style-type: none"> • What are the different data sources and how are they to be accessed?
<ul style="list-style-type: none"> • How the data are provided by the source systems, is it incremental or full feed, how to determine the incremental records
<ul style="list-style-type: none"> • What are the different target systems and how would the data be loaded
<ul style="list-style-type: none"> • Validation and reconciliation process for the incoming source data
<ul style="list-style-type: none"> • Handling late arriving dimension records
<ul style="list-style-type: none"> • Handling late arriving fact records
<ul style="list-style-type: none"> • Dynamism in the validation and transformation process
<ul style="list-style-type: none"> • Error handling process definition
<ul style="list-style-type: none"> • Table structures for holding the error data and the error messages

<ul style="list-style-type: none"> • Process control or audit information gathering process definition
<ul style="list-style-type: none"> • Table structures for holding the process control data
<ul style="list-style-type: none"> • Determining reusable objects and its usage
<ul style="list-style-type: none"> • Template creation for commonly used logics like error handling, SCD handling etc.
<ul style="list-style-type: none"> • Data correction and re-entry process
<ul style="list-style-type: none"> • Metadata capture during the development process
<ul style="list-style-type: none"> • Means of scheduling
<ul style="list-style-type: none"> • Initial data load plan
<ul style="list-style-type: none"> • Job failure and restart methods.

In addition to such checklists, there are several **ISSA Guidelines on Information and Communication Technology, 2019** which outline the various dimensions of a data interoperability initiative which need to be included in any operational plan. In particular, Guideline 29 (Workplan for the implementation of interoperability-based social security programmes), Guideline 30 (Institutional interoperability application model) and Guideline 32 (Institutional semantic interoperability) and Guideline 33 (Interoperable shared data services) are helpful to look at as guides to operational elements. These are accessible for ISAA members with password protection, but most social insurance agencies should have membership and access.

Other guidance, offering more detail on each of these stages are available from the following links:

- Talend. (2014) "Data Integration Checklist: Technical Note" Talend. http://info.talend.com/rs/talend/images/TN_EN_DI_Talend_DataIntegration_Checklist.pdf
- ETL Solutions. (2021) "Preparing a Data Integration Plan: A practical guide to data integration strategy and planning" ETL Solutions. <https://www.etlsolutions.com/new/wp-content/uploads/2021/08/Data-Integration-eGuide.pdf>
- Ashraf, S. (2020) "ETL Testing Checklist: Avoid Data Integration Disasters" DataIntegrationInfo. <https://dataintegrationinfo.com/etl-testing-process/>