

Data Science Projects

Structure and Handover of delivered Solutions

Enterprise Architecture *Vienna – June, 2020*

Agenda



- - A structured outcome from Data Science Projects
- Documentation

Reports and Manuals to operate the Solution with continuity

13 Template Outlook

A set of templates to standardize the structure of Data Science Solutions

1 A Handover Process

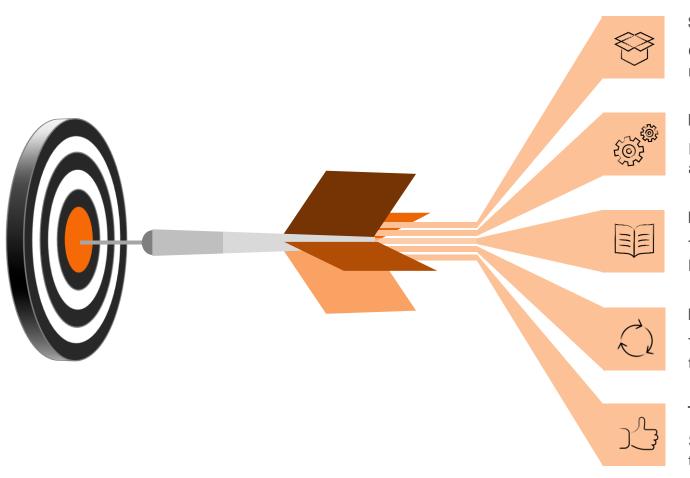
A streamline process to achieve long-term maintainability of Solutions



Data Science Solution

Data Science Projects

A structured outcome from Data Science Projects



Structured Git Repository

Organize the Solution to contain code, dependency management, documents and structured results locations.

Encapsulated Execution Logic

Provide standard scripts to execute the setup, training and scoring, despite the internal structure of the code.

Documentation

Technical reports and manuals are located together and linked to the software to guarantee up to date information.

Deployable Solution

The entire repository should be deployable as-is in the target environment to minimize manual operations.

Transparency during Handover

Streamline transfer of the Solution from the Project Team that favors long-term sustainable operations.



Documentation



Reports and Manuals to operate the Solution with continuity

Technical Report

- Business Problem definition
- Data Assets used in the delivered Solution
- Advanced Analytics and Machine Learning approaches used in the Solution
- Metrics in use and models' performance explained
- Interpretation and understanding of scored results

Operational Manual

- Description of Target Environment to deploy to
- Functional and non-functional requirements from the entire Project Team
- Deployment procedure to adopt for the Solution
- Description of format and location of expected inputs and outputs
- Training life-cycle and ordinary maintenance to be performed after handover

User Guidelines (ReadMe files in the repository)

- How to run the setup of the Software Solution
- How to run the software to (re)train the models
- How to run the software to score newly available data using trained models



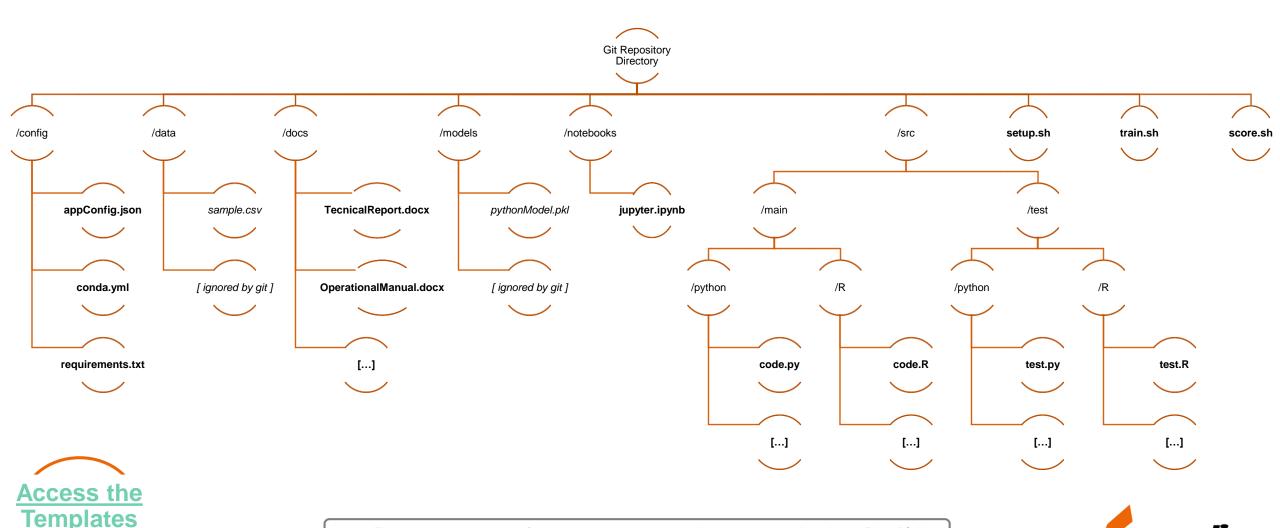




Template Outlook

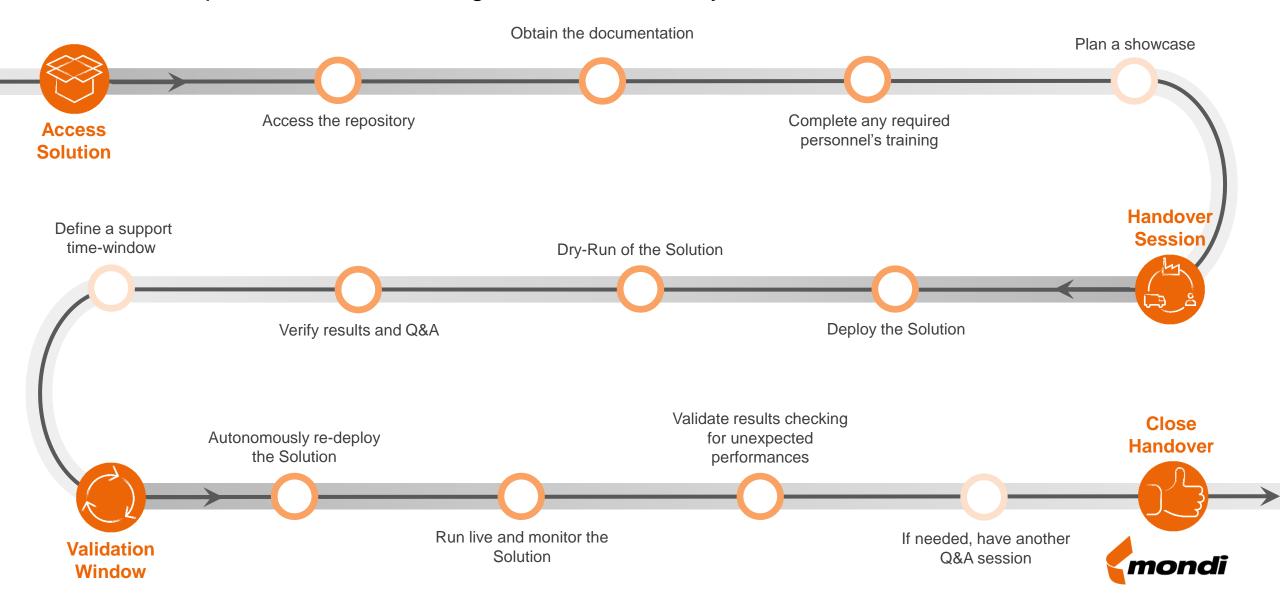


A set of templates to standardize the structure of Data Science Solutions



Note: Templates and example of documentation are also available in the repository in the "docs" folder

A streamline process to achieve long-term maintainability of Solutions



Data Science Projects

Access Solution

Access Solution							
	Task	List of Requirements	Actors	Possible Status			
Ö(Access the repository	 The Solution consists of at least one code repository available and accessible in a Mondi Git service Each code repository is at least conformant with the standard introduced in the repository's template, including User Guidelines (README files) 	The team receiving the handover	All requirement are satisfiedOtherwise failed requirements			
	Obtain the documentation	 Technical Report and Operational Manual are physically available in each repository or they are explicitly linked to an internal Mondi service in a README file in the same location The documentation contains all the required information (when applicable) according to the available templates 	The team receiving the handover	All requirement are satisfiedOtherwise failed requirements			
	Complete any required personnel's training	 If the Solution involves the training of personnel which is impacted by or has to operate alongside the provided solution, the training dates and involved personnel need to be already planned The supportive training material is available and at least one round of training has been completed successfully 	The team receiving the handover	All requirement are satisfiedOtherwise failed requirements			
	Plan a showcase	 All the above Tasks are in status Green Plan a showcase session (e.g. 1-2 days) according to the complexity of the project 	The team receiving the handover	All requirement are satisfiedOtherwise failed requirements			

Data Science Projects

Handover Session

' '	Halldovel Session						
	Task	List of Requirements	Actors	Possible Status			
d (Deploy the Solution	 The Solution is deployed in the Target Environment and it is setup according to the steps in the Operational Manual Data (if needed) is available in the Target Environment for running the Solution as prescribed by the Operational Manual 	The initial Project Team	All requirement are satisfiedOtherwise failed requirements			
d (Dry-run of the Solution	 If the Solution needs to train models, it is executed for training on historical data according to the steps in the Operational Manual. The output is a set of trained model obtained without errors The Solution is executed to score data using, if required, the previously trained models in accordance with the Operational Manual. The output is a set of results. 	The initial Project Team	All requirement are satisfiedOtherwise failed requirements			
d (Verify results and Q&A	 The results are structured and interpreted as per Technical Report. If applicable, the performances stated in the Technical Report respect the documented metric and expectations During the Q&A session, questions about deployment, run and use of the expected results are answered clearly and possibly referencing the available documentation. 	All involved parties	All requirement are satisfiedOtherwise failed requirements			
	Define a support time-window	 All the above Tasks are in status Green Define a Validation time-window (e.g. 3-4 weeks) during which support can be provided by the Project Team in case of inconsistent behavior of the Solution. 	The team receiving the handover	All requirement are satisfiedOtherwise failed requirements			

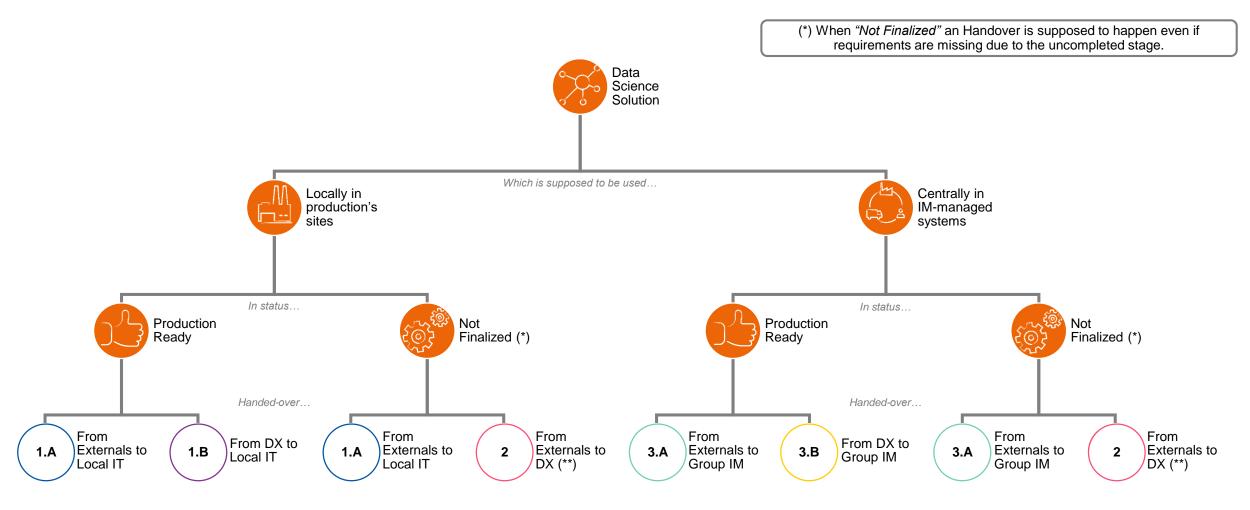
Data Science Projects

Validation Window

validation window							
	Task	List of Requirements	Actors	Possible Status			
Ö	Autonomously redeploy the Solution	 The Solution is deployed in the Target Environment and it is setup according to the steps in the Operational Manual Data (if needed) is available in the Target Environment for running the Solution as prescribed by the Operational Manual 	The team receiving the handover	All requirement are satisfiedOtherwise failed requirements			
d	Run live and monitor the Solution	 If the Solution needs to train models, it is executed for training on historical data according to the steps in the Operational Manual. The output is a set of trained model obtained without errors The Solution is executed to score data using, if required, the previously trained models in accordance with the Operational Manual. The output is a set of results. 	The team receiving the handover	All requirement are satisfiedOtherwise failed requirements			
	Validate results checking for unexpected performances	 The results are structured and interpreted as per Technical Report. If applicable, the performances stated in the Technical Report respect the documented metric and expectations for the whole validation period If prescribed by the Operational Manual, perform the necessary model re-trainings and ordinary maintenance tasks 	The team receiving the handover	All requirement are satisfiedOtherwise failed requirements			
	If needed, have another Q&A session	 If the previous tasks were not always in Status green, it might be necessary to prepare for a new Q&A session with the Project Team to clarify doubts Prepare the closing of the Handover Process and start planning for "operational mileage checks" of the Solution if necessary 	All involved parties	All requirement are satisfiedOtherwise failed requirements			

Data Science Projects

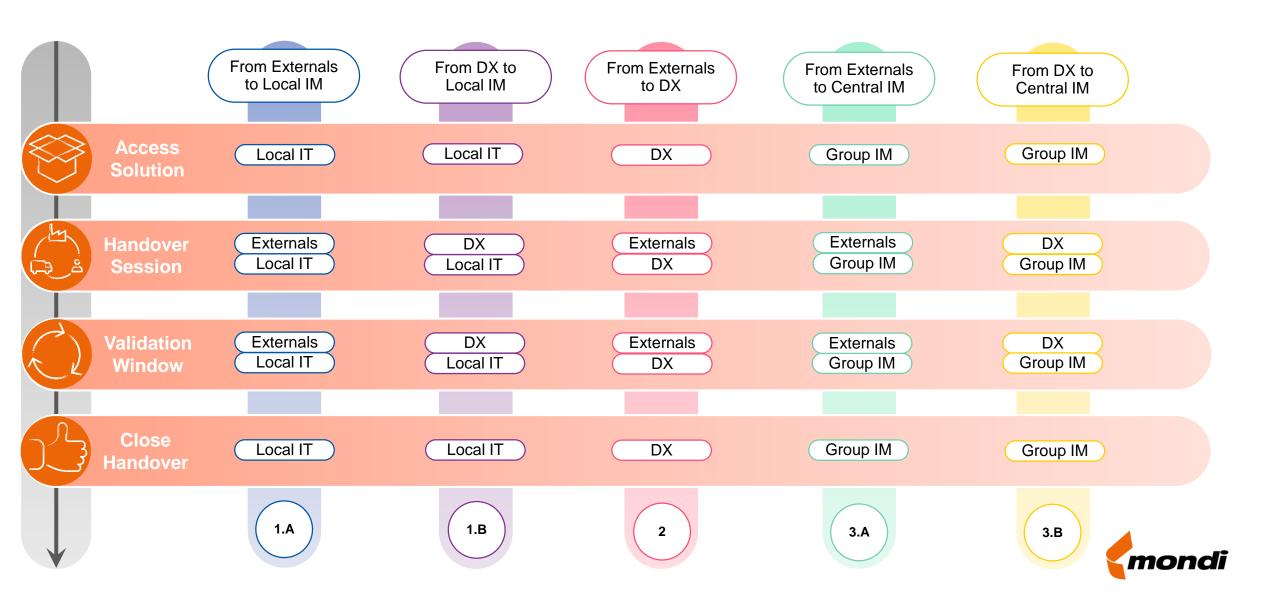
Necessary when...







Active involvement of teams





Thank you!

