POC BATI 2

INTEGRATION OF REGIONAL DATA

POC Bati: Proof Of Concept integration of regional data within the ITGI

Project Objectives and Scenario

Main Objective:

Test to use building data of regions to create our building objects

Sub-objectives:

Identify buildings that are present in the ITGI but are missing from the regional data.

Define methods for resolving conflicts at areas of connection between regions.

Set up a procedure for updating the building layer with successive regional data.

Approach:

To use the results and tools of the POC Bati project.

Iterative approach with continuous tests over an area of several hundred km² covering three regions.

Evaluation and adaptation phase of the model followed by iterations to integrate regional data.

Simulation of updating from recent regional data.

Final quality control based on the ITGI VRef CO_Building model.

NGI IGN

Expected Results and Deliverables

Expected results:

A layer of buildings integrated and structured according to the model of the POC-Bati project, covering a strategic area of several hundred km².

A derivative version of the layer for Quality Control in accordance with ITGI VRef Model CO_Building.

Deliverables:

Building Layer: Produced by integrating existing regional and ITGI data.

Scripts: A series of scripts for the production and maintenance of the building layer.

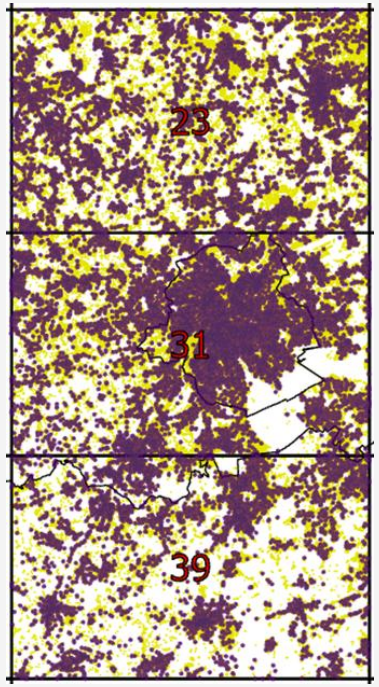
Update Procedure: Method to Update the BLD Layer with Recent Regional and IGN Data

Why use regional data?

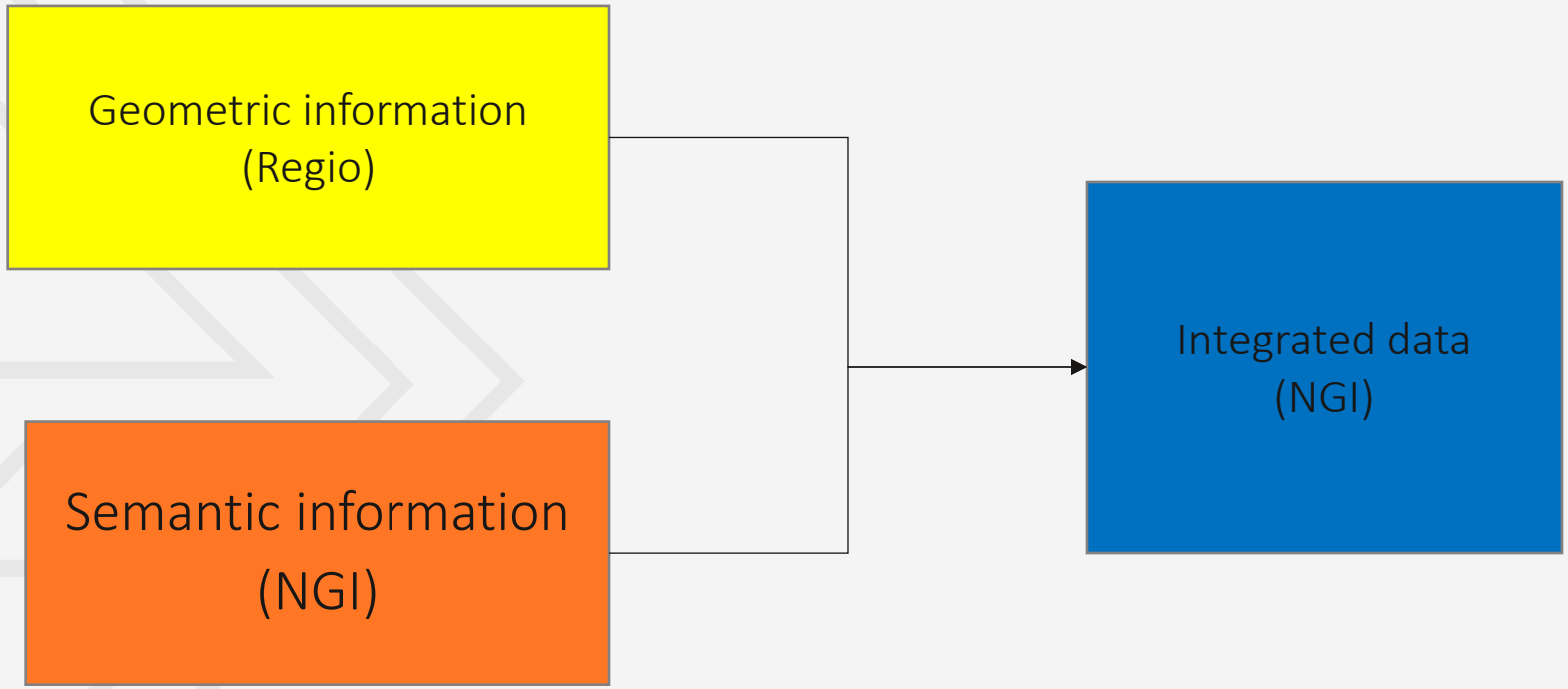
- Data exists already (no need to do twice)
- Integrator role of NGI
- Advantage of regional data: individual buildings, no building blocks

Study Area and Data Prioritization

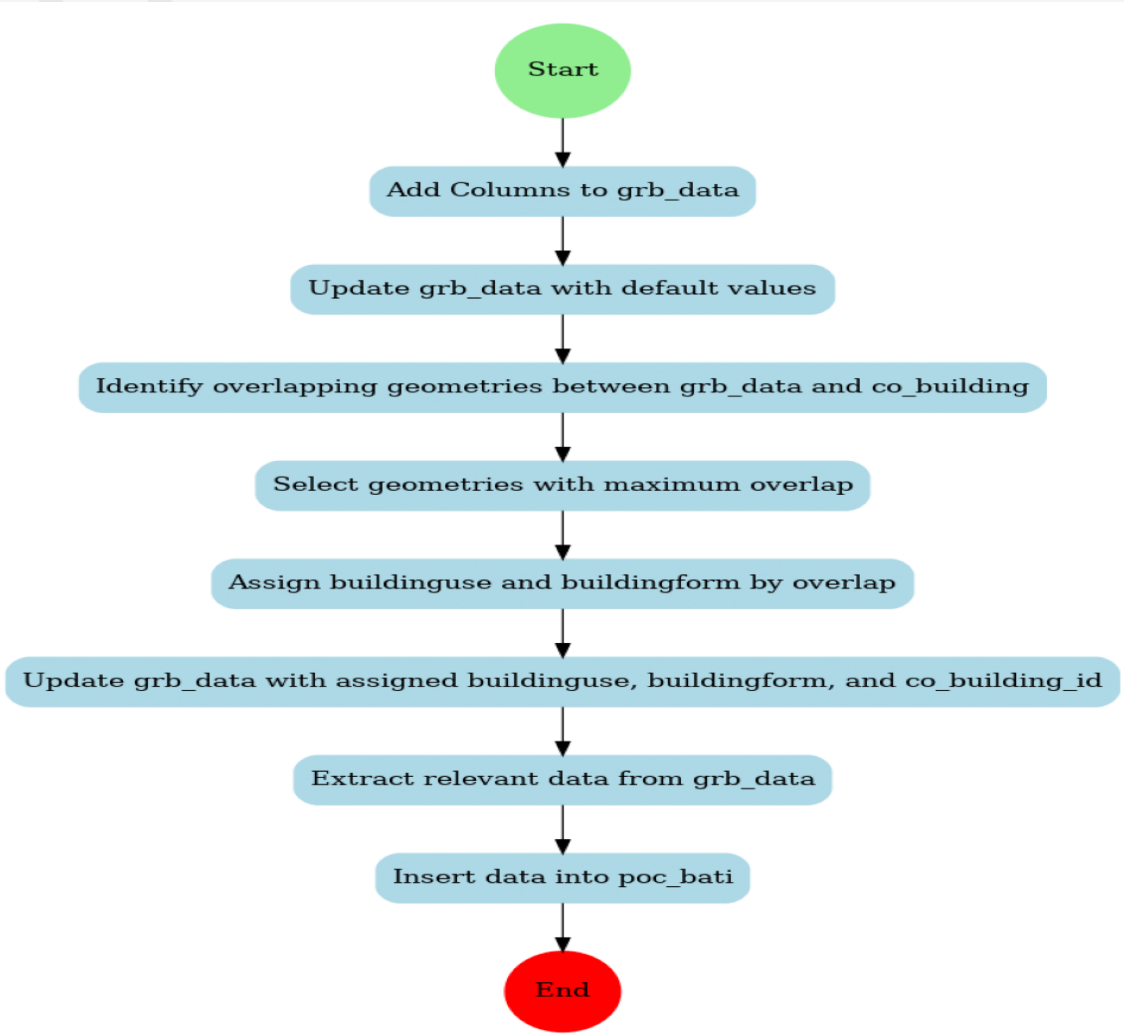
- **Study area:**
- Sheets **31**, **39**, and **23** have been selected.
- This zone covers the **three regions** : Flanders, Brussels, and Wallonia.
- The goal is to integrate data while resolving conflicts between these regions.
- **Data Prioritization:**
- **1. Flemish data (GRB):**
  - Priority because of their precision and richness.
- **2. Brussels data:**
  - Secondly, they are integrated to complement regional information.
- **3. Walloon data:**
  - Included last, because they have the fewest geometric errors.



Simplified version of the data model



Building Data Integration Process: From Preparation to Finalized Integration



- **Adding Columns to Table grb_data:** Preparing the table to incorporate additional information such as the date of check-in, building usage and shape, and an embedded building identifier (co_building_id).
- **Initial Record Update:** Initialize new columns with default values for all existing records, allowing for a common base before more detailed processing.
- **Overlap Identification and Resolution:** Analysis of geometric overlaps between grb_data and co_building to assign the best buildinguse and buildingform attributes to buildings, based on the percentage of overlap.
- **Data Extraction and Insertion:** Transfer of data prepared from grb_data to the poc_bati table for later use, ensuring that the data is integrated with correct and up-to-date attributes.

QC 1

- Goal = integration of buildings regions in itgi → same tests as currently used for layer CO_Building
- Geometric testing:

type of errors	-	number of errors in CO_Building	number of objects with this error in CO_Building	percentage of objects with this error in CO_Building
consecutive vertices < 0.15 m	-	133210	63389	9.27%
dimension of the object (2d) does not match dimension of the feature class (3d)	-	684083	684083	100%
hole < 0.20 m < polygon edges	-	11	11	0%
hole < 15 m^2	-	451	398	0.06%
hole < 15 m^2 - to verify (contains a geometry)	-	12	12	0%
multipart	-	37	37	0.01%
non-consecutive vertices < 0.20 m	-	275581	6873	1%
non-consecutive vertices < 0.20 m - caused by a hole too close to the outer edge	-	9	7	0%
shape_area < 1 m^2	-	174	174	0.03%
spike < 1°	-	60	56	0.01%

QC 1

- Goal = integration of buildings regions in itgi → same tests as currently used for layer CO_Building
- Topological tests:

Rules that have been tested	Nb_Object	Count_error	Count_distinct_object	Error rate	Comment
Building_Beyond_PPC_and_Building_v2	684.084	19.758	17.794	2.60%	Ces erreurs sont liées au fait que les coins des bâtiments sont à moins de 20cm l'un de l'autre. Ou les bords à moins de 20cm l'un de l'autre
Building_must_not_be_completely_within_PPC_v2	684.084	11	11	0.00%	
Building_must_not_overlap_Building_v2	684.084	74.935	49.223	7.20%	Problème de 'snapping' : arrondir les coordonnées à 4 chiffres après la virgule devrait permettre de diminuer le nombre d'erreurs
Building_must_not_overlap_PPC_v2	684.084	235	217	0.03%	
Building_must_not_to_surimpose_Building_v2	684.084	481	368	0.05%	Problème de 'snapping' : arrondir les coordonnées à 4 chiffres après la virgule devrait permettre de diminuer le nombre d'erreurs
Building_must_not_to_surimpose_PPC_v2	684.084	42	10	0.00%	
Building_have_more_15squaremetre_v2	684.084	63.687	63.685	9.31%	Certains building ont une surface de moins de 0,5m^2
Building_must_not_touch_PPC_by_point_NEW_v2	684.084	1	1	0.00%	
Building_must_not_touch_by_point_NEW_v2	684.084	19.564	19.056	2.79%	Problème de 'snapping' : arrondir les coordonnées à 4 chiffres après la virgule devrait permettre de diminuer le nombre d'erreurs
Greenhouse_have_more_60sqm_NEW_v2	684.084	98	77	0.01%	
No_Spike_ALL_v2	684.084	568	551	0.08%	
No_Z_equal_0_ALL	684.084	0	0	0.00%	La règle ne renvoit aucun objet car aucun n'a de Z....La règle vérifie que le Z n'est pas à 0 mais ne vérifie pas la présence ou non de ce Z!
ALL_must_not_duplicated_ALL_v2	684.084	20	20	0.00%	20 objets dupliqués = 10 doublons !

Nb Objects :	684.084
Rules with 0 error :	1
Rules with at leat one error :	12
Features with at least one error :	136.887

QC 1

- Goal = integration of buildings regions in itgi → same tests as currently used for layer CO_Building
- Visual tests:

TEST Reference (unique ID)	DATA QUALITY ELEMENT	DATA QUALITY SUB-ELEMENT	Test source	Sample size	Result of the test	Extra comments	Reference to the results
Source_GQC-Test_01	Age of the data	-	NGI/G/QC	Full dataset	Modifdate tussen 2006 en 2024	Grote verschillen per gewest!	
Source_GQC-Test_02	Completeness	Commission	table D.3 in [ISO19157]	Full dataset or sampling via ISO 2859-1	0.80%		Results_QA_POC_Bati.gdb\Error_VC
Source_GQC-Test_04	Completeness	Omission	table D.7 in [ISO19157]	Full dataset or sampling via ISO 2859-1	Wallonië: 1.00% Brussel: 0.20% Vlaanderen: 192%	Goed voor Wallonië en Brussel, zeer slecht voor Vlaanderen!	Results_QA_POC_Bati.gdb\Error_VC
Source_GQC-Test_11	Positional accuracy	Absolute or external accuracy	NGI/G/QC	Full dataset or sampling via ISO 2859-1	13.04%	2-10m: 121 10-25m: 24 > 25m: 18 Testobjects: 1250	Results_QA_POC_Bati.gdb\Error_VC
Source_GQC-Test_19	Usability estimation	For production	NGI/G/QC	-	Er zijn veel te veel problemen gevonden om de data rechtstreeks te integreren in onze itgi. De gegevens kunnen wel gebruikt worden om bijvoorbeeld ontbrekende gebouwen terug te vinden, maar dan nog moet er een correctie gebeuren om de geometrie en topologie in orde te krijgen volgens onze specificaties. De Vlaamse gegevens daarentegen zijn niet bruikbaar zoals we ze nu gekregen hebben.		

QC 1

- Goal = integration of buildings regions in itgi → same tests as currently used for layer CO_Building
- Missing buildings in Flanders!

Problematic: Topological Issues Detected After Quality Control (QC)

Problem:

After a Quality Control (QC) process of the geospatial data, several topological problems were detected.

These issues include polygon overlaps, unconnected lines, duplicate points, and incorrect intersections.

Impact:

These topological errors compromise the use of as is

Objective:

Find an effective tool to correct these topological errors while maintaining data integrity.

Comparison of Topological Data Cleaning Solutions

- Option 1: Standard GIS Tools (ArcGIS/QGIS)
- **Pros:** Intuitive user interface, large support community.
- **Disadvantages:** Limited topology management, risk of residual errors after correction, dependence on specific scripts or plugins (licensed).
- Option 2: Manual Correction Tools
- **Advantages:** Full control over each correction.
- **Cons:** Very time-consuming process, prone to human error, difficult to apply on large databases.
- Option 3: GRASS GIS with v.clean
- **Pros:** Strict topology management, advanced tools for automatic cleaning, highly customizable with multiple processing options.
- **Cons:** Initial learning curve, less intuitive interface for new users

Why GRASS GIS was the Best Option

- Topological Force:
- GRASS GIS maintains rigorous topology management, ensuring that all spatial relationships are properly managed and corrected.
- This avoids residual problems that are common in other, less specialized solutions.
- Advanced Tools:
- v.clean is a powerful module that offers a full range of options for automatic cleaning of geometric data.
- Enables complex and large databases to be processed accurately.
- Versatile controls:
- Multiple commands (break, snap, bpol, etc.) allow you to customize the cleanup based on the types of errors detected.
- The -c command ensures strict topological verification, minimizing the risk of new errors.

QC 2

- Goal = integration of buildings regions in itgi → same tests as currently used for layer CO_Building
- Geometric testing:

type of error	-	number of errors in the database	number of objects with this error in the database	percentage of objects with this error in the database
consecutive vertices < 0.15 m	-	78648	51244	4.66%
hole < 0.20 m < polygon edges	-	65	65	0.01%
hole < 15 m²	-	548	498	0.05%
hole < 15 m² - to verify (contains a geometry)	-	142	140	0.01%
non-consecutive vertices < 0.20 m	-	10423	8313	0.76%
non-consecutive vertices < 0.20 m - caused by a hole too close to the outer edge	-	12	12	0%
shape_area < 1 m²	-	2056	2056	0.19%
spike < 1°	-	431	379	0.03%
dimension of the object (2d) does not match dimension of the feature class (3d)	-	1099031	1099031	100%

QC 2

- Goal = integration of buildings regions in itgi → same tests as currently used for layer CO_Building
- Topological tests:

Rules that have been tested	Nb_Object	Count_error	Count_distinct_object	Error rate	Comment
Building_Beyond_PPC_and_Building_v2	1.099.031	12.129		1.10%	Ces erreurs sont liées au fait que les coins des bâtiments sont à moins de 20cm l'un de l'autre. Ou les bords à moins de 20cm l'un de l'autre
Building_must_not_be_completely_within_PPC_v2	1.099.031	14		0.00%	Nos PPC silo et Covered Grandstand sont repris dans leurs buildings. Donc c'est un problème de définition de classe.
Building_must_not_overlap_Building_v2	1.099.031	0		0.00%	
Building_must_not_overlap_PPC_v2	1.099.031	414		0.04%	Problème de snaping ou problème de définition des classes
Building_must_not_to_surimpose_Building_v2	1.099.031	75.774		6.89%	Problème de 'snapping' : arrondir les coordonnées à 4 chiffres après la virgule devrait permettre de diminuer le nombre d'erreurs
Building_must_not_to_surimpose_PPC_v2	1.099.031	45		0.00%	
Building_have_more_15squaremetre_v2	1.099.031	109.402		9.95%	Certains building ont une surface de moins de 0,5m^2
Building_must_not_touch_PPC_by_point_NEW_v2	1.099.031	0		0.00%	
Building_must_not_touch_by_point_NEW_v2	1.099.031	11.521		1.05%	Problème de 'snapping' : arrondir les coordonnées à 4 chiffres après la virgule devrait permettre de diminuer le nombre d'erreurs
Greenhouse_have_more_60sqm_NEW_v2	1.099.031	0		0.00%	
No_Spike_ALL_v2	1.099.031	2.000		0.18%	Sur des polygones qui ont l'air d'être des découpes de façade de batiment ou des polygones qui sont l'intersection entre 2 batiments. Ou un batiment découpé au milieu d'un autre batiment.
No_Z_equal_0_ALL	1.099.031	0		0.00%	La règle ne renvoit aucun objet car aucun n'a de Z....La règle vérifie que le Z n'est pas à 0 mais ne vérifie pas la présence ou non de ce Z!
ALL_must_not_duplicated_ALL_v2	1.099.031	75.708		6.89%	75708 objets dupliqués = 37854 doublons !
Total		**287.007**		26%	

Nb Objects :	1.099.031
Rules with 0 error :	4
Rules with at leat one error :	9
Features with at least one error :	194.696

QC 2

- Goal = integration of buildings regions in itgi → same tests as currently used for layer CO_Building
- Visual tests:

TEST Reference (unique ID)	DATA QUALITY ELEMENT	DATA QUALITY SUB-ELEMENT	Test source	Sample size	Result of the test	Extra comments	Reference to the results
Source_GQC-Test_01	Age of the data	-	NGI/G/QC	Full dataset	Alle modifdates zijn van 2024		
Source_GQC-Test_02	Completeness	Commission	table D.3 in [ISO19157]	Full dataset or sampling via ISO 2859-1	Vlaanderen: 0.40%		Results_QA_POC_Bati_v2.gdb\Error_VC
Source_GQC-Test_04	Completeness	Omission	table D.7 in [ISO19157]	Full dataset or sampling via ISO 2859-1	Vlaanderen: 1.00%		Results_QA_POC_Bati_v2.gdb\Error_VC
Source_GQC-Test_11	Positional accuracy	Absolute or external accuracy	NGI/G/QC	Full dataset or sampling via ISO 2859-1	Vlaanderen: 12.00%	2-10m: 36 10-25m: 13 > 25m: 11 Testobjects: 500	Results_QA_POC_Bati_v2.gdb\Error_VC
Source_GQC-Test_19	Usability estimation	For production	NGI/G/QC	-	Er zijn nog steeds te veel problemen gevonden om de data rechtstreeks te kunnen integreren in onze itgi. De gegevens kunnen wel gebruikt worden om bijvoorbeeld ontbrekende gebouwen terug te vinden, maar dan nog moet er een correctie gebeuren om de geometrie en topologie in orde te krijgen volgens onze specificaties.		

QC 2

- Goal = integration of buildings regions in itgi → same tests as currently used for layer CO_Building
- Examples of problems:

	SHAPE *	OBJECTID *	fid	cat	id_pb	id_source	source_name	buildinguse	buildingform	b_type	begin_date	modif_date	integration_date
1	Polygon	1	1	13	0	WAL	31	12	1	0	16/08/2016	26/03/2024	26/03/2024
2	Polygon	2	2	14	669003	VL	31	12	2	0	15/01/2022	02/08/2024	02/08/2024
3	Polygon	3	3	18	0	WAL	31	12	2	0	19/07/2016	26/03/2024	26/03/2024
4	Polygon	4	4	21	0	WAL	31	12	2	0	19/07/2016	26/03/2024	26/03/2024
5	Polygon	5	5	22	3166539	VL	31	12	2	0	11/12/2015	02/08/2024	02/08/2024
6	Polygon	6	6	24	6442531	VL	31	12	2	0	02/10/2018	02/08/2024	02/08/2024
7	Polygon	7	7	26	0	WAL	31	12	1	0	02/08/2017	26/03/2024	26/03/2024
8	Polygon	8	8	29	0	WAL	31	12	2	0	19/07/2016	26/03/2024	26/03/2024
9	Polygon	9	9	31	4895850	VL	31	12	2	0	07/07/2021	02/08/2024	02/08/2024
10	Polygon	10	10	36	0	WAL	31	12	2	0	19/07/2016	26/03/2024	26/03/2024
11	Polygon	11	11	39	0	WAL	31	12	2	0	16/08/2016	26/03/2024	26/03/2024
12	Polygon	12	12	41	663810	VL	31	12	1	0	26/01/2009	02/08/2024	02/08/2024

Power and Limitations in Data Cleansing

- Achievements and Positive Points:
  - **Noticeable Improvement in Data Quality:** Through the use of GRASS GIS, topological errors have been significantly reduced, making the data more reliable for future analyses.
  - **Efficient Automation with v.clean:** The tool has been able to automate many complex steps, providing a robust solution for large-scale geometry cleaning.
- Challenges and Limitations Encountered:
  - **Artifact Geometries Generation:** In some cases, GRASS GIS introduced unwanted geometries (artifacts) that were not needed and could cause problems in subsequent analyses.
  - **Lack of Time for a Full Analysis:** Due to time constraints, it was not possible to perform an in-depth analysis to diagnose these artifacts or to fully assess the impact of the solutions applied.
- Conclusions :
- **GRASS GIS: A Powerful Tool with Complexities:** Although GRASS GIS has demonstrated great efficiency in data cleansing, its use can lead to undesirable outcomes that require additional expertise to be properly managed.

Recommendations and Next Steps

- **Importance of Exchanges with the QC:**
- Significant progress was made when the exchanges with the QC team intensified. Their expertise made it possible to define clear objectives and to compare the data with the realities on the ground.
- This collaboration highlighted practical challenges and helped steer the project towards more pragmatic solutions.
- **Final Recommendation:**
- **Creating a Separate Product:** It is recommended to create a separate product, named Building Integrated, separate from CO_Building.
- **Objective:** This new product would allow for the development of regional data integration while ensuring the stability of current processes and minimizing data quality risks.
- **Benefit:** This would ensure flexibility in developing and improving integration tools without disrupting existing workflows.

Thank you for your attention

Questions ?

Remarks?

contact : Jordan.ikalulu@ngi.be