Design of automate SAR processing system in LAPAN to support easier and fast data services

Marendra Eko Budiono[1], Haris Suka Dyatmika[1], Novie Indriasari[1], Qonita Amriyah[1], Rahmat Arief[1], Rachmat Maulana[1], Shadiq Ali[1]

[1]Indonesia Institute of Aeronautics and Space (LAPAN)

e-mail: marendra_lapan@yahoo.com

Abstract: One of the Indonesia's National Institute of Aeronautics and Space (LAPAN) task is to supply and distribute remote sensing data based on Indonesia national needs as mandated by space law no 21 year 2013. One of the satellite image acquired and distributed by LAPAN including TerraSAR-X/TanDEM-X, a twin satellite from Germany launched in 2007. Acquisition of the TerraSAR-X/TanDEM-X satellite began in 2017 with aim for coverage of whole Indonesian area. With respect to the data size and data availability, an automate data processing method is needed to process all of the available TerraSAR-X/TanDEM-X data. This paper will discuss a system design that automate the LAPAN's TerraSAR-X/TanDEM-X data processing and method of presenting the data to the users as tiled data.

1. Introduction

Indonesia National Institute of Aeronautics and Space (abbreviated as LAPAN in bahasa) is Indonesian government office in charge of receiving and managing national remote sensing data especially satellite data in Indonesia. As mandated by Space Law no 21 year 2013 [1], LAPAN has duty on supplying and distributing remote sensing data, based on Indonesian national needs. This task is handled by deputy of remote sensing affairs in LAPAN. The acquired satellite data consists of optical and SAR data and categorized into low, medium and high resolution. One of the satellite data acquired is TerraSAR-X/TanDEM-X, a twin SAR satellite managed by Germany launched in 2007.
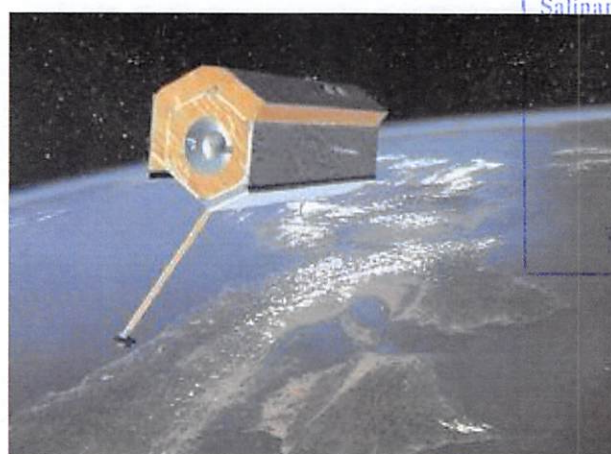
Figure 1. The view of deployed TerraSAR-X/TanDEM-X spacecraft platform [2].

The antenna of TerraSAR-X/TanDEM-X satellite has several imaging modes which make the data could be varies in resolution and specification. The available options of TerraSAR-X/TanDEM-X data is presented in table 1.

Table 1. The available imaging mode of TerraSAR-X/TanDEM-X [3].

Imaging Mode	Standard Scene Size [km]	Azimuth Resolution [m]	Polarization
Staring SpotLight (ST)	4 x 3.7	0.24	Single (VV or HH)
HighRes SpotLight 300 MHz (HS300)	10 x 5	1.1	Single (VV or HH)
HighRes SpotLight (HS)	10 x 5	1.1 (Single pol) or 2.2(Dual pol)	Single (VV or HH) Dual (HH & VV)
SpotLight (SL)	10 x 10	1.7 (Single pol) or 3.4 (Dual pol)	Single (VV or HH) Dual (HH & VV)
StripMap (SM)	30 x 50 single pol 15 x 50 dual pol 100 x 150	3.3 (Single pol) or 6.6 (Dual pol)	Single (VV or HH) Dual (HH & VV, HH & HV, or VV and VH)
ScanSAR (SC)	100 x 150	18.5	Single (VV or HH)
Wide ScanSAR (WS)	270 x 200	40	Single (VV, HH, HV or VH)

After the data is transmitted to the ground station, it is then processed into standard product software on the ground station. There are 2 standard products processed in LAPAN ground station, which are Single Look Slant Range Complex (SSC) and Enhanced Ellipsoid Corrected (EEC). The basic difference is SSC data is still on raw data state, which still has all the information like phase amplitude in the basic COSAR format, while EEC data is already multi looked, geocoded to WGS-84 reference ellipsoid [3].

LAPAN started the acquisition of TerraSAR-X/TanDEM-X on 2017 with contract of 1.050.000 km2 coverage per year to get coverage of whole Indonesian area while also serving for other specialized acquisition tasks such as when image of special data needed for emergency and disaster monitoring or urban monitoring. For the coverage of Indonesian area, the data used is StripMap mode in HH single polarization. This is to ensure that the coverage of TerraSAR-X/TanDEM-X data will have enough detail (3.3 meter) and wide enough coverage (1500 km^2) per data. Additionally, there is some special situation that needed another imaging mode, for example when there is an emergency in a smaller area such as volcano eruption that needed higher resolution, thus the High Resolution (HS) mode is used or even a wider area need to be monitored, this time it is ScanSAR(SC) or WideScanSAR(WS) is chosen. The coverage of TerraSAR-X/TanDEM-X data from 2017 up to August 2019 is depicted in figure 2.
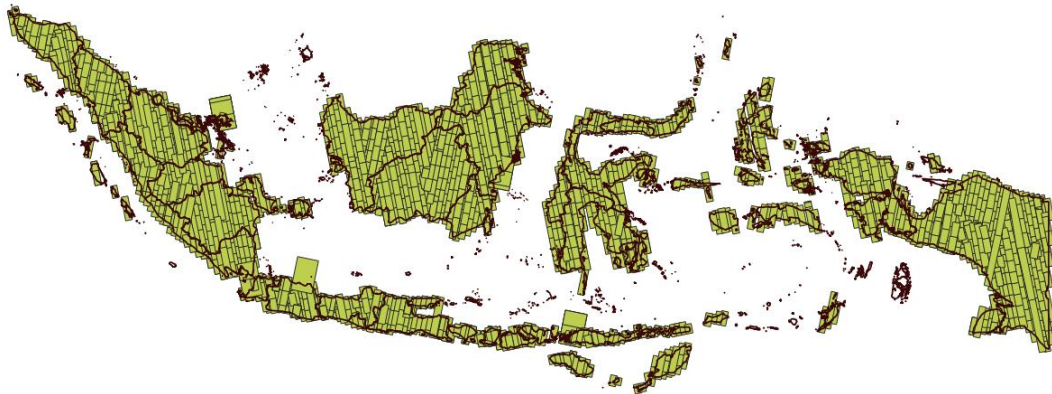
Figure 2. The coverage of TerraSAR-X/TanDEM-X acquisition from 2017 until August 2019

According mandate in PP law no 21, LAPAN also tasked with distributing the acquired remote sensing data to the user, which is in this case is other ministries and institutes in Indonesia [1]. To comply to user needs, LAPAN has to provide acquired satellite data including the TerraSAR-X/TanDEM-X images in some forms such as raw data from satellites, pre-processed for analyzed-ready data or mosaicked images from a certain period of time. TerraSAR-X/TanDEM-X itself is categorized as a high resolution satellite, which means the information held in the data is very dense that will affect the size of the data and in the end will make processing time of the data longer, thus an automation is needed in processing all the available data to make it ready for users to use the data. This paper will discuss a design of automation in processing all the available data of TerraSAR-X/TanDEM-X and a method to simplify in presenting the data to the users. Also a calculation of processing time and space of storage needed by the system to process all the available data is presented.

2. Method
In this paper, the system is designed to process the TerraSAR-X/TanDEM-X data in LAPAN. That is why the designed system will be using state that already established in LAPAN as an approach. In the manner of input and output of the system, it will be using directory structure that already structured in LAPAN and the data processing sequence applied on the system will be using the established processing sequence on LAPAN. The process that need to be automated consist of listing all the input files, do data processing including output conversion into tiles that make it ready for the standard of the data presentation system on LAPAN then describing the structure of the output directory structure and filename, lastly a database system is added in order to improve robustness of the system.

2.1. Listing Input
The input data will be acquired from TerraSAR-X/TanDEM-X data stored on the storage server on the national remote sensing data bank section in LAPAN. The process of data acquisition from satellite to the ground station consist of several processes. Satellite capture the image, transferred to ground station and then processed into a certain level. These whole process is already automated, that is why most of the satellite data is already standardized to follow a certain set of rules in naming them. This also applied for TerraSAR-X/TanDEM-X data acquired in LAPAN which then being stored in the data bank system in LAPAN. The remote sensing data bank has its own storage system and structure on storing data, the data sorted by year and adding its own folder as depicted in figure 3, which is a necessary thing to comply with rule of its database system, causing extra folder outside the original folder structure from the ground station, depicted in figure 4. Nevertheless, all the TerraSAR-X/TanDEM-X data is stored in the uniform structure and it is a crucial aspect of the input for the automation.
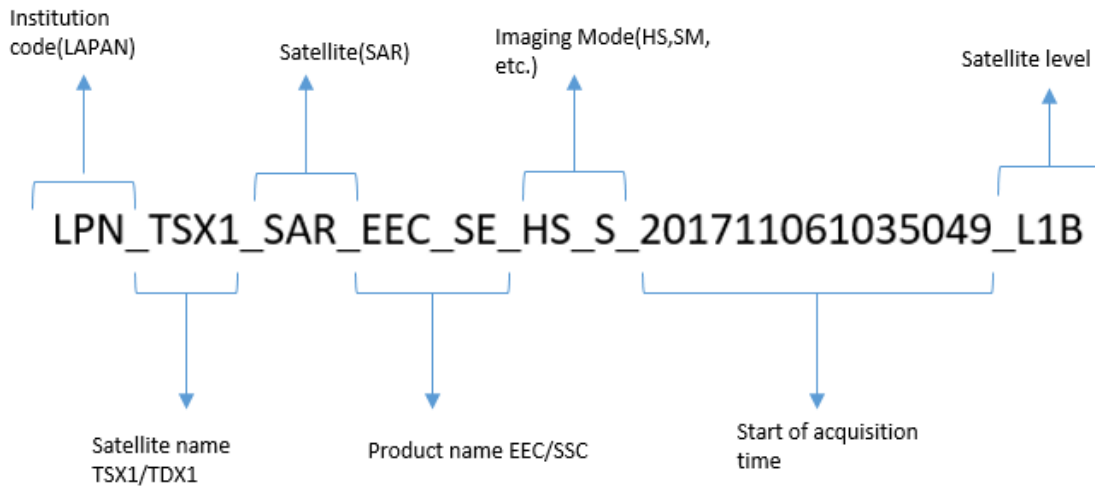
Figure 3. Naming convention of the TerraSAR-X/TanDEM-X in the storage system in LAPAN
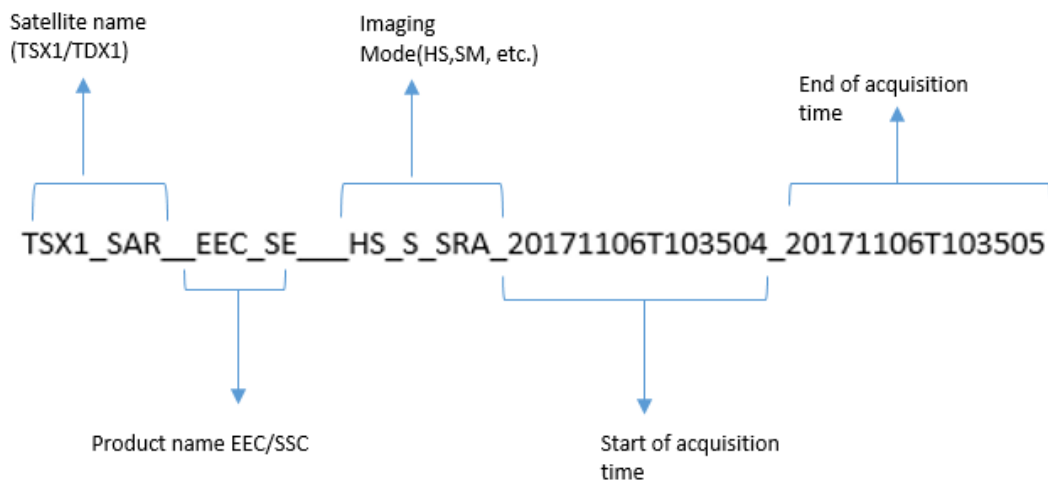

Figure 4. Naming convention of the TerraSAR-X/TanDEM-X from the ground station

2.2. TerraSAR-X/TanDEM-X processing steps

The TerraSAR-X/TanDEM-X data stored in the storage system in LAPAN consists of 2 data types, SSC and EEC, with the total amount of data up to August 2019 is depicted in table 2.

Table 2. The availability of SSC and EEC data

Data Availability	
SSC	
	2770
EEC	
	494
Total	
	3264

Both SSC and EEC need to be geometrically corrected into each surface and radiometrically corrected into backscatter value. For the SSC data, it need to be processed using GAMMA software which need the metadata file of the TerraSAR-X/TanDEM-X data which usually come as a file with .xml extension on the default data folder. This file contains all the information of the recorded data and information of the satellite positioning when recording data which are essential on processing and needed as an information on data correction. The flowchart of the data processing is shown in Figure 5.
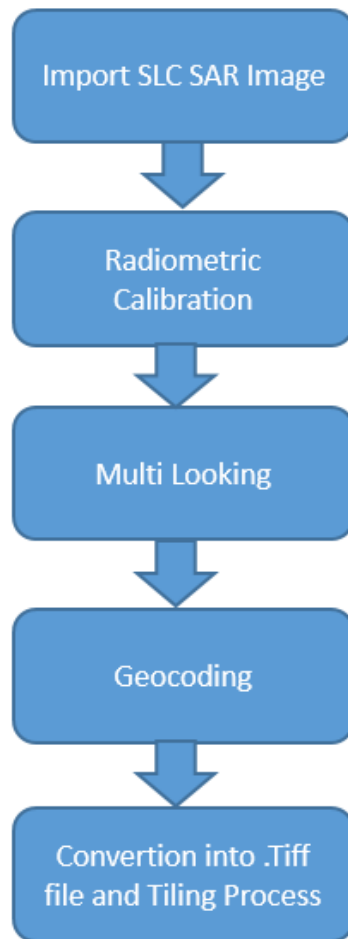

Figure 5. Flowchart of the process of SSC data [4]

This basic processing steps need to be applied on all the SSC data to make it into standard format, allowing the data to be aligned with each other. The import product will gather the information of the satellite condition using metadata file, radiometric correction is to calibrate the pixel of the data into backscatter with sigma naught (σ^0) [5] and multi looking will minimize the speckle or noise of the data. The geocoding process correcting the data geometrically into the earth surface. Lastly the create RGB step will convert the data into general Tiff image format.

Meanwhile the EEC data is already geocoded but lacking radiometric correction which will be using PCI software. On a side note, the process of the SSC is involving multi looking which change the resolution of the output data, while EEC is still on the original resolution, therefore a resampling is needed to make it into the same resolution as the SSC data after processed. The resampling process will be using GDAL library. All the aforementioned software has the ability to be running without Graphical User Interface (GUI), meaning it is possible to process the data without human intervention or with only minimal interaction.

2.3. Tiling process

A lot of space data agencies around the world are moving towards Analysis Ready Data (ARD) making, a level of data that user can do analysis to without or only minimal of additional effort [6]. Examples here including United States Geological Survey (USGS) with its Landsat satellite archives [7], Committee on Earth Observation Satellite (CEOS) with Copernicus program's data archives called Analysis Ready Data for Land (CARD4L) [8] and Commonwealth Scientific and Industrial Research Organization (CSIRO) with Data Cube program [9]. From these references, the similarity they have is that the presentation of the ARD is using data gridded in the same exact size named tiles, with each tile could contain more than 1 data.

LAPAN as Indonesian data satellite provider, also begin to adopt this method of data presentation. Some initial research has been conducted by LAPAN's researchers resulted in initial grid size for the tiles of LAPAN's ARD which will be using geodetic projection as grid [10]. In order to lining up with this program, the TerraSAR-X/TanDEM-X data also need to be able to be presented in gridded size. To decide the grid size of the tiles, it is need to be taken into account that one TerraSAR-X/TanDEM-X data is 1500 km^2 wide, a single TerraSAR-X/TanDEM-X data will not suffice into a grid of 1x1 degree, which is roughly 111x111 km^2. Hence, it is decided by making one TerraSAR-X/TanDEM-X data into tiles with the size of sub-grid, which is smaller than 1x1 degree. At this point, a tile with size 0.1x0.1 degree is chosen. As support of the later those tiles will be merged with its neighboring tiles from other TerraSAR-X/TanDEM-X data into larger size such as 1x1 degree. Making the final state of TerraSAR-X/TanDEM-X data into tiles will also help when making mosaic as it will only need to merged all the tiles that belong on the target area that need to be mosaicked.
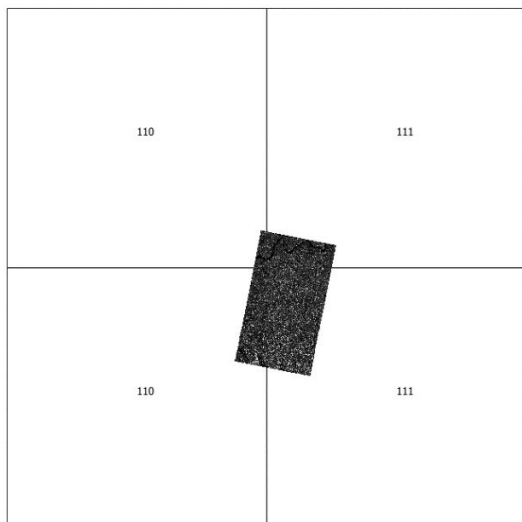

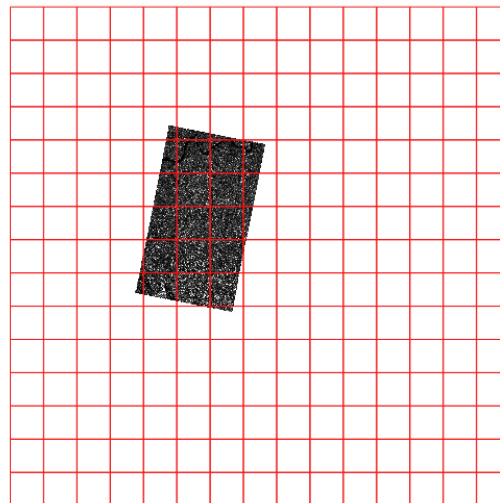**Figure 6.** Illustration of the 1x1 degree tile compared to the TerraSAR-X/TanDEM-X data

Figure 7. Illustration of the 0.1x0.1 degree tile compared to the TerraSAR-X/TanDEM-X data

2.4. Output directory and file naming convention

Another aspect to be considered is the structure of the output directory and naming convention of the output files. This is important because the automation process need a systematic directory and file name structure so that output generation process could work. Directory structure of the output folder and the naming convention is depicted in figure 8 and the detailed description is shown in Table 3 and Table 4
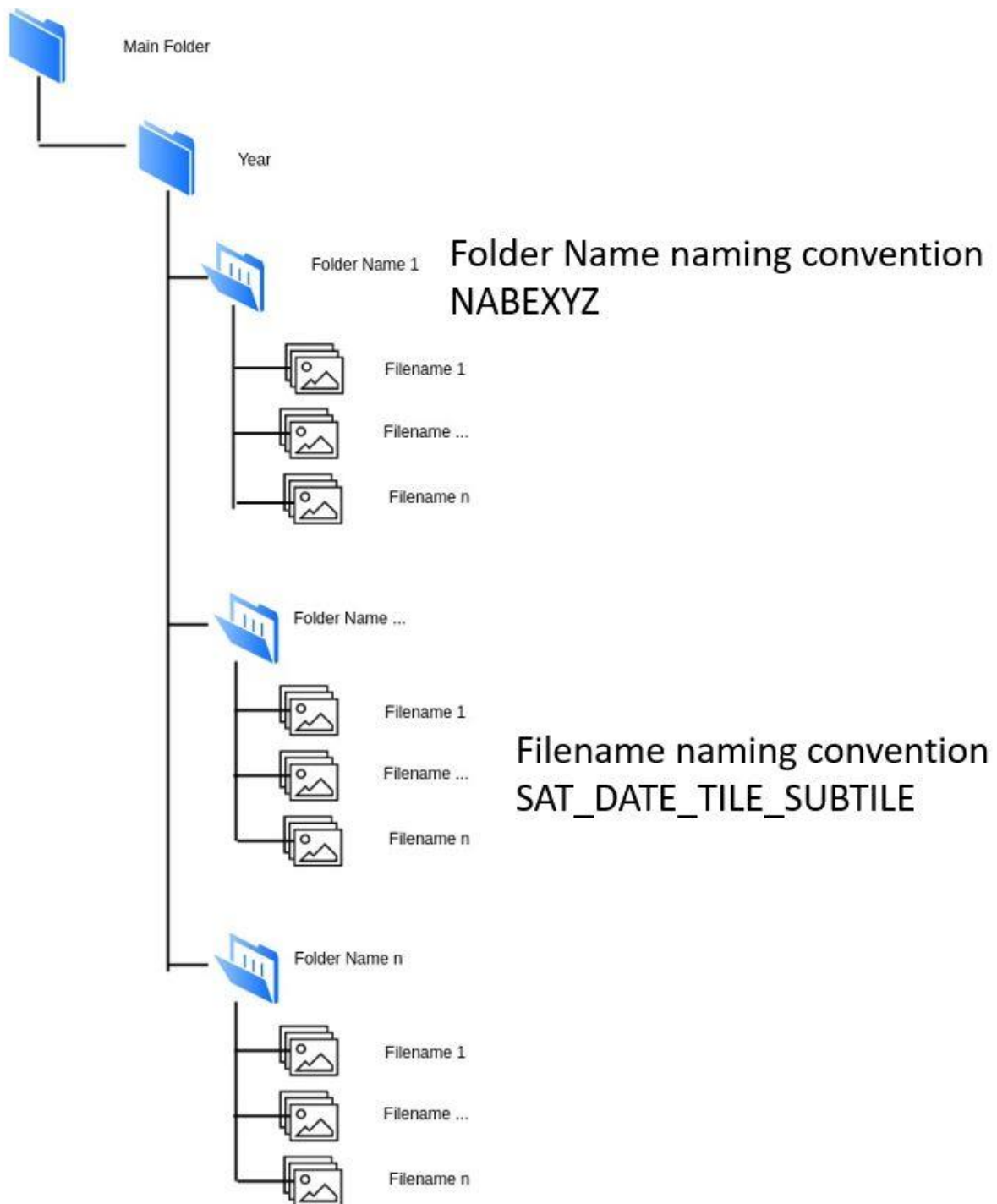
Figure 8. Folder structure and naming file output convention

Table 3. Description of the naming convention of the folder output

Symbol	Range	Description	Possible value
N	1 digit alphabet	Latitude information of 1x1 degree tile	N = Latitude information if the degree is on northern hemisphere S = Latitude information if the degree is on southern hemisphere
AB	2 digits numeric	Latitude value of 1 x 1 degree tile	A=Tens part of the latitude value (0 or 1) B=Ones part of the latitude value (0 – 9)
E	1 digit alphabet	Longitude information of 1 x 1 degree tile	E = Longitude information of the degree. It will always be "E" as Indonesia location is always on eastern hemisphere
XYZ	3 digits numeric	Latitude value of 1 x 1 degree tile	X=Hundreds part of the latitude value (0 or 1) Y=Tens part of the latitude value (0 – 9) Z=Ones part of the latitude value (0 – 9)

Table 4. Description of the naming convention of the output filename

Symbol	Range	Description	Possible value
TILE	7 digits alphanumeric	The name of the 1 x 1 tile where the file belongs to	Refers to the table x above
SUBTILE	2 digits numeric	The name of the 0.1 x 0.1 degree tile where the file belongs to	0-99
SAT	4 digit alphanumeric	The name of the satellite the data come from	TSX1 or TDX1
DATE	12 digits alphanumeric	The acquisition date of the original data the tile came from	Last 2 digit of years,2 digits month,2 digits day, time start of acquisition

2.5. Database
In order to make the system more robust, which is when a system is able to cope with error, some control mechanism need to be applied. Sometimes sequential process like batch processing could

encountering error when processing many files and because it is a sequential process, the software could crash and stopped in the middle of the process. To start it again from the beginning would be a waste of time, that is why some way to keep track of the process to avoid restarting the process each time an error occurred need to be applied. One of the mechanism that can be applied is by using a database to keep track of the process. In a database, the state of each process could be timestamped and also the metadata of each file could also be stored.

3. Result and Discussion

The final design of the system is shown as figure below. It consists of 2 separate systems which share the same database to keep track of the process. The first system that process SSC data is depicted in figure 9 with the "Data Processing" block is the same as the process in figure 1. Meanwhile flowchart of the EEC processing system is depicted by figure 10.
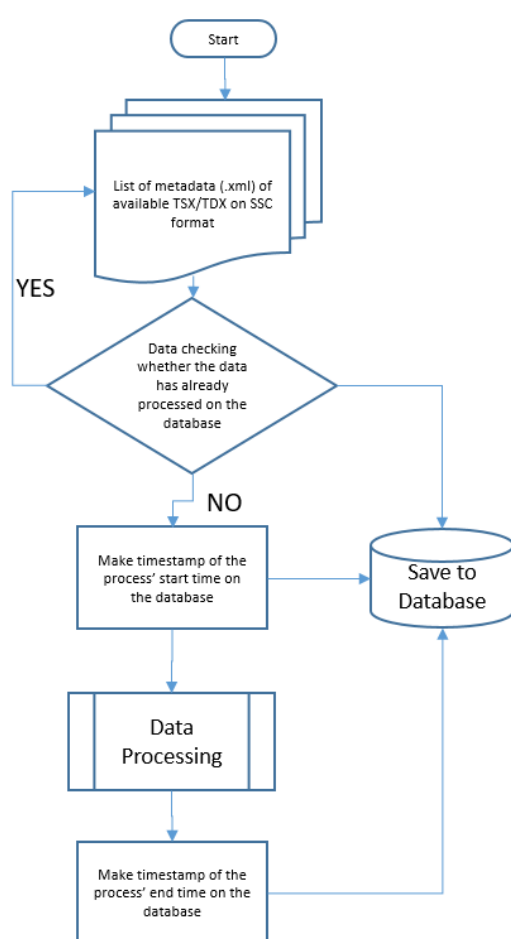

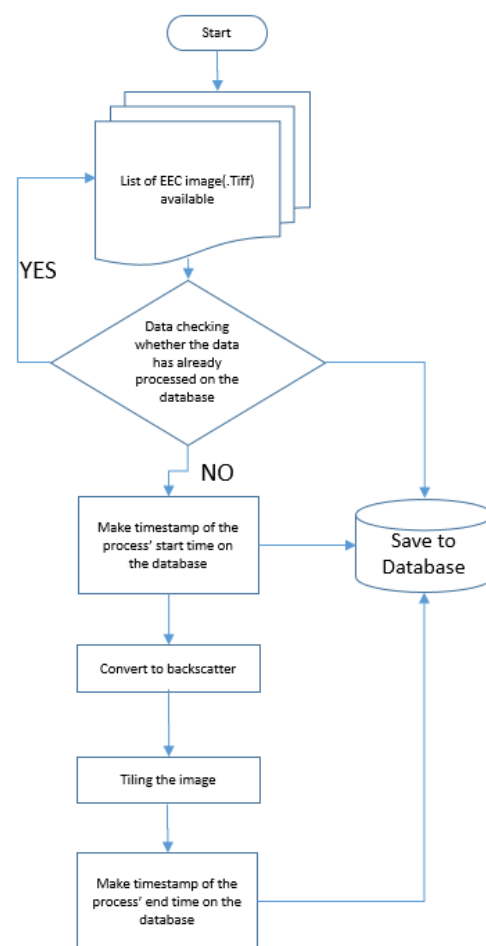**Figure 9.** Flowchart of the SSC data processing   **Figure 10.** Flowchart of the EEC data processing

The final design on the system is consist of 2 system flowcharts. This is due to different software used in processing the SSC and EEC data. The software used for processing the SSC data is GAMMA which only available under linux Operating System, while the EEC processing which convert the data into backscatter is using PCI Geomatic software, is running under Windows OS. For the SSC processing, it is then applied using Java programming language with openjdk 11.0.4 and the EEC processing is using python 3.7 programming language. For the database mysql ver 14.4 is used.

The automation of the system is successfully produce output in tiles thus satisfy in supporting LAPAN preparation in built an ARD system and ready to be mosaicked. Despite all that, there were problems

faced when developing the system that help the system into better when it has subjugated, but there were still that need to be taken as a note. Usage of the database is an example of the subjugated problem that give the system the better state. At first, there is no need of the database on the preliminary state of the system, but when doing a test run, sometimes the system facing an error when processing a set of TerraSAR-X/TanDEM-X data, that became the alarm of the need of the database to keep track of the process is surfaced. The database also has potential in helping to speed up the mosaic process. The other example is the tiling process, this is a core feature that support the both ARD and mosaic process. It is based on the fact that SSC and EEC are on a different level of processing thus the need of a system that able to uniting 2 set of spatial data is by make smaller tiles despite also make them into same final state (in backscatter value) and same resolution. Ironically, the tiling phase is also the thing that need to be reviewed, especially the size of 0.1 degree per tiles. On one note, it is sufficient because the resolution of the TerraSAR-X/TanDEM-X that up to 6.6 meter after multi looking, but on the other hand it will complicated the mosaic process as the number of tiles would be enormous and the time to merging all of them would grow.

From a test run, a process of full SSC data is taking approximately 15 minutes, while EEC data is taking approximately 10 minutes. From this test an approximation of the time needed to process all the TerraSAR-X/TanDEM-X data is depicted in table 5.

Table 5. Calculation of the processing time of the available TerraSAR-X/Tan-DEM-X data in LAPAN

	Data Availability	Time needed to process 1 scene (minutes)	Time needed to process all (hours)
SSC	2770	15	692.5
EEC	494	10	82.33
Total	3264		774.833

For the space needed, a base calculation is used, one scene is around 1500 km^2 that will be split into tiles with each tile has 0.1 degree in size. With 0.1 degree being around 11 km x11 km, or 121 km^2, a scene would be assumed to become 1500/121= 13 tiles. One tile of the final data has size of 19,304 KB, thus one data is assumed as 13x19,304 KB=250,952 KB. The final calculation is depicted in table 6

Table 6. Calculation of the space needed for processing all the available TerraSAR-X/Tan-DEM-X data in LAPAN

	Data Availability	1 scene per average (KB)	Space needed to process all (GB)
SSC	2770	250,952	695.137040
EEC	494	250,952	123.970288
Total	3264	250,952	819.107328

4. Conclusion

An automatic TerraSAR-X/TanDEM-X processing system has successfully built. The system is able to process SSC and EEC data into final product that already corrected geometrically and radiometrically, multi looked and then presented the data as tiles with size of 0.1 degree each.

Based on the test run on the system, it could be inferred that the system will take approximately 775 hours and need 819.107328 GB of storage to process all the available TerraSAR-X/TanDEM-X data up to August 2019.

5. Acknowledgements

6. References

[1]   Indonesia, P.R., *Undang-undang Republik Indonesia no. 21 tahun 2013 tentang Keantariksaan*. 2013.

[2]   DLR, T.-X., *Ground Segment Basic Product Specification Document*. 2013, TX-GS-DD-3302, Dec 2008, http://sss. terrasar-x. dlr. de/pdfs/TXGS-DD-3302. pdf.

[3]   AIRBUS, *TerraSAR-X Image Product Guide*, in *Basic and Enhanced Radar Satellite Imagery*. 2015. p. 24.

[4]   Werner, C., et al. *Gamma SAR and interferometric processing software*. in *Proceedings of the ers-envisat symposium, gothenburg, sweden*. 2000. Citeseer.

[5]   AIRBUS, *Calculation of beta naught and sigma naught for TerraSAR-X data*. 2008.

[6]   Dwyer, J., et al., *Analysis ready data: enabling analysis of the Landsat archive.* Remote Sensing, 2018. 10(9): p. 1363.

[7]   Wulder, M.A., et al., *The global Landsat archive: Status, consolidation, and direction.* Remote Sensing of Environment, 2016. 185: p. 271-283.

[8]   Lewis, A., et al. *CEOS Analysis Ready Data for Land (CARD4L) Overview*. in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. 2018. IEEE.

[9]   Ticehurst, C., et al., *Building a SAR-Enabled Data Cube Capability in Australia Using SAR Analysis Ready Data.* Data, 2019. 4(3): p. 100.

[10]   Dimyati, M., R.D. Dimyati, and P. Danoedoro, *Interpretability Evaluation of Annual Mosaic Image of MTB Model for Land Cover Changes Analysis.* Telkomnika, 2018. 16(3): p. 934-945.