IVS Data Center at BKG

Anastasiia Girdiuk, Markus Goltz, Daniela Thaller

Abstract Three primary IVS Data Centers cooperate closely together to establish and maintain common data processing procedures. Our aim is to support as similar as possible data handling at each IVS Data Center. The workflow of the data acceptance is maintained to resemble one at another IVS Data Center, so that each Data Center ensures data acceptance, storage, availability of the data and compatibility with the other Data Centers. We present here the current procedures of the data acceptance at BKG and our future plans to extend the data center infrastructure. The Data Center storage capacity covers the foreseen requirements to store vgosDb files intuitively based on the data rate increase in the last year. We are interested also in the extension of the currently accepted IVS Data Center structure to additionally store the correlated data; however, the expected storage capacities need to be better defined and properly advocated. The approval process is expected to be considerable as well. While this work is in progress, the BKG Data Center uses its available capacities to serve as an exchange server for the projects of relatively small disk space demands: TPWLO and EU-VGOS. The structure of the projects repeats the IVS Data Center structure to facilitate the exchange internally. Also, it means that it would be possible to make the internal project available on the official Data Center once this is decided. The projects are maintained with restricted access, which allows us to learn the new routines to handle the data acquisition. The restricted access to our server with File Transfer Protocol (FTP) over Secure Socket Layer (SSL) is planned to be in place for all of our users according to the requirements of our internal IT Infrastructure and EU regulations on Data Security.

Federal Agency for Cartography and Geodesy (BKG)

1 Data Center Summary

The IVS Data Center at BKG works together with the representatives of CDDIS and OPAR to maintain the data center duties. The availability and accessibility of all data at any time, as well as the ability to upload the VLBI data, are provided to the users. At this moment the BKG Data Center is accessible for data downloading or listing via FTP-SSL and HTTPS as user "anonymous." The data uploading is supported by FTP-SSL connection and personal user accounts. The unprotected FTP access had to be discontinued due to security reasons. The login with the personal user account is arranged to replace the ivsincoming-user uploading procedure. We register the connections between 1 and 6 MB/sec for data uploading and 2 and 7 MB/sec for data downloading. Considering the data size the connectivity satisfies the regular data demand. The updated information about the access to the BKG Data Center can be seen on our Web page: https://ivs.bkg.bund.de/.

Data acceptance is managed uniquely among the three IVS Data Centers by means of the validation procedures. As follows, all kinds of VLBI data have to adhere to the naming convention [3] as of August 2, 2021. The validation procedures are employed at BKG and OPAR by the ingest script developed by the CDDIS group [2]. In the data acceptance workflow each received file is checked with an antivirus program; this step takes the longest time in the data handling scheme as a rule. The virus-marked files are quarantined, while the passed files are forwarded to the ingest script as it is seen in Figure 1. The received data which fail the
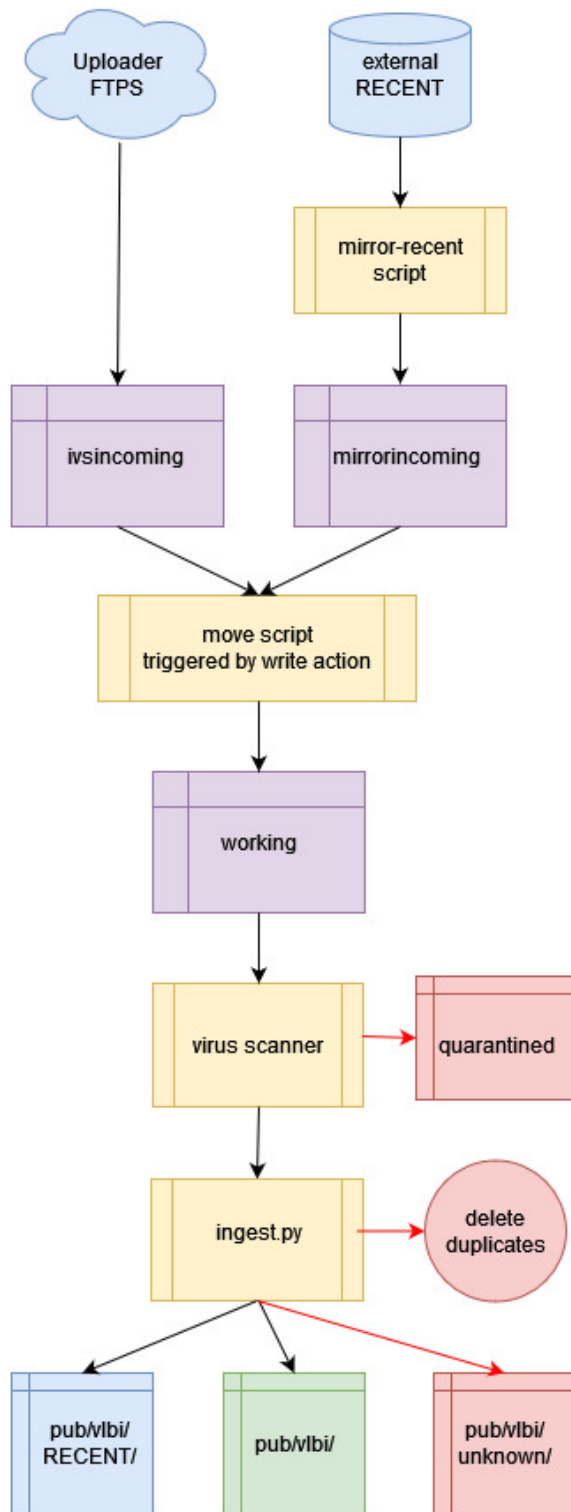
Fig. 1 The workflow of the Data Center.

validation after an attempt to ingest them at our data center can be found under the directory called "UN-KNOWN." The possibility of seeing the failed data in "UNKNOWN" is kept at the BKG Data Center for historical reasons. Meanwhile it is considered useful in the transitional period to accommodate the naming convention. The legitimate VLBI data, which has passed the validation by ingest script successfully, is placed in the appropriate directory (green box in Figure 1) according to the data type as specified in the so-called data definition file (DDF, [2]) also empowered by the ingest script. Additionally, the ingest script checks if the same data are already present in the data center. As follows, the duplicated files are dropped. The most recent data are considered the data which were received in the last 14 days. These data are copied to the "RE-CENT" directory. The data availability is secured by mirroring the data provided under this directory "RE-CENT" at the other data centers. Further, an entire data center is to be mirrored on a regular basis. The BKG Data Center structure is listed in Table 1, where additionally the vgosDb files from correlators at Bonn, GSI, Shanghai, UTAS, and Wien are collected. The USNO group was delivering vgosDb files, which were stored under vgosdb_usno. Since 2022 the corresponding vgosDb files are uploaded to the Data Center directly. The files are placed according to their type under ivsdata/vgosdb. Besides, vgosdb_bkg contains the vgosDb data processed by the BKG Analysis group.

The data acceptance workflow, along with the current BKG Data Center, was established on the new machine. It allowed for internal data ingest verification. All data from the old server was uploaded to the new one, which resulted in a significant amount of unrecognized data from the validation procedures and DDFs. The largest part of the unrecognized data is the historical tropospheric products. The Data Center Working Group has decided that these tropospheric products are recognized to be not an IVS product any longer, because the data are available at the GNSS data centers and the IVS Components have no interest in them. These products are not defined as IVS data (see [3]); thus, these data are not available at the BKG Data Center. The other large portion of the products was recognized to have been created for non-IVS sessions. The smaller parts of the unrecognized historical data have some minor differences in the name pattern. There are probably a few data types for which DDFs need to be created. The status of the different cases of the unrec-

ognized data is the subject for discussions of the Data Center Working Group.

Table 1 The structure tree of the IVS Data Center at BKG.

Level 1	Level 2	Description
ivscontrol/		IVS DC control files
ivsdata/		VLBI data files
	vgosdb/	VLBI data in vgosDb format
	vgosdb_bkg/	vgosDb processed by BKG AC
		vgosDb prepared by:
	vgosdb_bonn/	Bonn Correlator
	vgosdb_gsi/	GSI Correlator
	vgosdb_shao/	Shanghai Correlator
	vgosdb_utas/	UTAS Correlator
	vgosdb_wien/	Wien Correlator
	vgosdb_usno/	submitted by the USNO group
	db/	mk3db data storage
	ngs/	NGS-cards data storage
	aux/	session supplementary data
	swin/	SWIN files
ivsdocuments/		IVS DC documents
ivsformat/		master-format
ivsproducts/		analysis products
	crf/	source coordinate
	trf/	station position
	eops/	24h EOP time series
	eopi/	Intensive EOP time series
	daily_sinex/	SINEX of 24 h sessions
	int_sinex/	SINEX of Intensives
	trop/	tropospheric products
gsfc/		software-related input data
	ancilliary/	minimal set of a priori data*
	sked/	frequency catalog (SKED)
RECENT/		all data of the last two weeks

* in the file format in use by nuSolve and Calc/Solve.

2 Data Center: Main Area

The VLBI data takes about 700 GB at the moment, while the main area of the Data Center has altogether 7 TB. Besides, disk space of 20 TB is assigned to the SWIN files. The Shanghai correlator is uploading the SWIN files to us directly. The rest of the SWIN data are to be acquired from CDDIS.

Along with the main area of the Data Center, the vgosDb files from different correlators are collected as shown in Table 1. We mirror twice per hour the public data storage referred to the VLBI data at the Bonn, GSI, Wien, and UTAS Correlators. At this moment the Shanghai and USNO Correlators are deliver-

ing vgosDb files to our Data Center. Here, it is worthwhile to note that we register and allow with respect to the ingest script procedures a replacement of the vgosDb files at any Data Center. Let us follow the analysis chain. First of all, the correlator that is responsible for a session uploads its vgosDb file set, in which at least the correlator version 1 of the wrapper file is provided. Next, the Analysis Center (AC) that is responsible for this session according to the master file uploads the analyzed vgosDb file set, in which the version from the correlator is retained. This version of the vgosDb file set is recognized as official and final, unless new correlation or fringe fitting is prepared. If the correlator uploads a version of a vgosDb file set which is already present at a Data Center, the vgosDb file set gets replaced. At this point it makes no difference whether an AC or a correlator has uploaded the vgosDb file set; the newer version with respect to the stored digital fingerprints of the file (md5sum and shasum verify file integrity) will be chosen to be placed in the Data Center. This version from the correlator, however, does not include either the previous correlation version or the corresponding analysis version. Thus, the responsible AC will upload only the analysis of the last available version of the vgosDb file set to the Data Center.

The analyzed vgosDb files of the ACs other than the responsible one are not available within the vgosDb file set. That is why the BKG Analysis Center provides its own analysis version of the vgosDb files at our Data Center under ivsdata/vgosdb_bkg.

3 Data Center: Additional Activities

Considering the VLBI data increase, we have disk space left to provide our service for the projects demanding a relatively small disk space. At this moment we work with the European VGOS (EU-VGOS, [1]) and the Test of the continuous Piece-Wise Linear Offset parameterization (TPWLO, [4]) projects. The disk space and IVS-like storage environment with protected data access are provided. The chosen login authorization has been seen to complicate the procedure; thus, a simple login sharing routine was enabled for the IVS uploading accounts. The IVS compatible data structure is initialized so that the data can be accepted with the same procedures as if they were submitted to the main area of the Data Center. As an end effect, the

Table 2 SINEX files are to be received from five IVS ACs under the TPWLO project.

Institution	IVS AC
Federal Agency for Cartography and Geodesy BKG	BKG
Vienna University of Technology TU Wien	VIE
Onsala Space Observatory OSO	OSO
Norwegian Mapping Authority NMA	NMA
German Geodetic Research Institute DGFI	DGF

Table 3 Under the EU-VGOS project the vgosDb files are provided by the correlators at Bonn and TU Wien. And SINEX files are to be received from eight ACs and institutions.

Institution	IVS AC
Federal Agency for Cartography and Geodesy BKG	BKG
Vienna University of Technology TU Wien	VIE
German Research Centre for Geosciences GFZ	GFZ
Onsala Space Observatory OSO	OSO
Metsähovi Radio Observatory (MRO*)	–
Yebes Observatory (YBS*)	–
Norwegian Mapping Authority NMA	NMA
Finnish Geospatial Research Institute (FGI*)	–

* affiliations given for recognition at the Data Center.

project data could be forwarded to the main area of the Data Center and accepted straightforwardly. That means that the same ingest script with validation procedures and DDFs operates on the incoming data, where the database is extended to accept the project-assigned solution names and research groups which do not have an Analysis Center affiliated with IVS.

The TPWLO project, in which some of the IVS ACs participate (see Table 2), requires only the acceptance of SINEX files. The EU-VGOS project shares the SINEX files as well. Besides, the vgosDb files of the sessions observed under the EU-VGOS project are made available for the project members (see Table 3). The EU-VGOS main project goal [1], which is to test the impact of the different calibration and fringe-fitting methods for the geodetic solutions, sets an additional challenge for the Data Center procedures by requiring the possibility of sharing with the project members more than one correlation version per session.

4 Outlook

The BKG Data Center takes an active part in the joint activity of the IVS Data Center components. We have established our workflow with the ingest script and validation procedures. We support data acceptance and data redundancy by collecting vgosDb files from the correlators and providing BKG Analysis group results. Besides, the BKG Data Center's entire data set was processed with the ingest script; thus a considerable amount of data was declined by the validation procedures and DDFs. The validation workflow was adjusted in agreement with the Data Center Working Group. Further verification of the remaining unrecognized data is planned in line with a similar CDDIS activity.

The BKG Data Center is working to follow the EU regulations and provide users with convenient access. As follows, the HTTPS is expected to be crafted for up-

loading as well. The data completeness is to be checked by mirroring the entirety of CDDIS and OPAR at least once a day. On one hand, we apply our best efforts to follow the IVS recommendations about the availability of the SWIN files. On the other hand, the acquisition of new space is limited to the defined requirements, the definition of which is reviewed regularly (currently: a disk space increase of 5 TB during two years) and set in connection with the actual user demand.

Meanwhile available disk space is given to the VLBI-related projects. Compliance with the demands of these projects benefits our operations on a smaller scale than an entire IVS Data Center. In particular, experience with the personalized access has facilitated the adjustment of the setup of the IVS accounts. These projects are valuable for further development.

References

1. F. Jaron et al., Current Status of the EU-VGOS Project, IVS General Meeting Proceedings, 2022, this volume.
2. D. Behrend, M. Bérubé, J. Gipson, A. Girdiuk, M. Goltz, T. Yates, P. Michael, C. Barache, Data Ingest at the IVS Data Centers, Presented at: IVS 2022 General Meeting, 2022.
3. D. Behrend, Data Center file conventions, https://ivscc.gsfc.nasa.gov/products-data/DataCenter_File_Conventions.pdf, 2021.
4. A. Nothnagel, S. Böhm, R. Dach, A. Girdiuk, M. Glomsda, H. Hellmers, A-S. Kirkvik, T. Nilsson, D. Thaller, First results of Earth rotation parameter estimation with piece-wise linear offsets, IVS General Meeting Proceedings, 2022, this volume.
5. T. Yates, J. Woo, N. Pollack, J. Ash, J. Roark, S. Blevins, P. Michael, VLBI Data Ingest Improvements at NASA CDDIS, IVS General Meeting Proceedings, 2022, this volume.