GeoDatabases

Managing Distributed Data
1 Replicas and GeoDatabases ........................................... 6
1.1 Replication types .......................................................... 7
1.2 Data sharing in connected or disconnected environments ........................................ 8
1.3 Getting ready for replication ............................................. 9
   1.3.1 Add GlobalID column ............................................... 10
   1.3.2 Registering data as versioned ..................................... 11
   1.3.3 Define the data to be replicated .................................. 12
   1.3.4 Replication and Metadata ......................................... 13
   1.3.5 Replication and Raster Data ...................................... 13
2 Creating Replicas ......................................................... 14
2.1 Creating replicas in a connected environment ............................................. 15
   2.1.1 Creating a one-way or two-way replica ........................ 15
2.2 Creating replicas in a disconnected environment ........................................... 23
   2.2.1 Replica creation in the source geodatabase ........................................ 24
   2.2.2 Replica transfer to the destination geodatabase ............................. 27
   2.2.3 Replica creation in the destination geodatabase ............................ 27
3 Data Synchronization .................................................... 30
3.1 Synchronizing Replicas in a connected environment ........................................ 31
3.2 Synchronizing Replicas in a disconnected environment ..................................... 35
   3.2.1 Export a data change message ..................................... 37
   3.2.2 Import a Data Change Message .................................... 39
   3.2.3 Export an Acknowledge Message .................................. 41
   3.2.4 Import an Acknowledge message ................................... 42
4 Managing changes in the layers structure .................................. 46
4.1 Applying schema changes in connected environments .................................. 47
   4.1.1 Compare replica schema ............................................ 47
   4.1.2 Import replica schema ............................................... 49
4.2 Applying schema changes in disconnected environments ............................... 50
   4.2.1 Export replica schema ............................................... 51
   4.2.2 Compare replica schema ............................................. 52
   4.2.3 Import replica schema ............................................... 53
5 Replica Management

5.1 Turning a disconnected replica in a connected one
Introduction

The GIS training “Managing distributed data” is part of the collection of trainings on geodatabase management produced by the Geospatial Support Unit (GSU), at the Emergency Preparedness and Response branch, to support the implementation of the GIS Infrastructure in the World Food Programme. This specific training module has been designed with the aim of guiding the user through the usage of a set of tools provided by ESRI to exchange information between GeoDatabases distributed in different locations. In particular we will use replica and synchronization tools described in this training module to share GIS datasets between WFP offices where GeoDatabases are in place.

Building a network of distributed geodatabase at HQ, Regional Bureau and Country Office level has been identified as a solution to improve data availability and performance by alleviating server contention and slow network access to a central server. Each GIS officer in WFP can work on data stored in a Geodatabase located in his office and get data produced by others or share his data with others by taking advantage of Geodatabase replication and synchronization techniques.

In this training module we’ll describe different types of replication (Checkout\Check in , One-way, Two-way), different options available to create replica workflows in connected and disconnected environments, how to synchronize data in different scenarios and manage existing replica.
Software needed

All the tools described in this training module allow replicating and synchronizing data across ArcSDE GeoDatabases; therefore, you must have access to a Desktop, Workgroup or Enterprise GeoDatabase to go through this training. In addition, such tools are provided by ESRI in a toolbar in ArcMap, so make sure you have ArcGIS Desktop 10.2.2 Advanced installed in your machine. In case one of these constraints is not respected, please get in touch with your regional GIS officer or with the GSU unit at the HQ, by sending an email to wfp.esriservices@wfp.org.

1 Replicas and GeoDatabases

Geodatabase replication allows you to create copies of data across two or more geodatabases such that changes to the data may be synchronized. Replica creation involves the user defining the data to replicate from a source geodatabase, and then running a process to create replicas. The process copies data from the source geodatabase to a target geodatabase and creates a replica in each geodatabase. The replica describes what data has been replicated and contains the information needed to synchronize changes. The replica in the source geodatabase is called parent replica, and the replica in the target geodatabase is the child replica.
Each ArcSDE geodatabase can host multiple parent and child replica in order to allow splitting data to replicate in more replicas and run synchronizations with different frequency and schedule according to frequency to data editing.

### 1.1 Replication types

Three different types of geodatabase replication are available: check out/check in, one-way replication and two-way replication. According to your specific needs you can choose one or the other when you create the replica.

**Checkout/Check-in** replication allows you to replicate a set of data from a parent geodatabase to a child geodatabase, edit layers in the child and synchronize edits back to the parent geodatabase when edit session is complete. The main constraint with this type of replica is that once the data has been synchronized, you can no longer synchronize additional edits. If additional edits are required, you must create a new checkout replica.

**One-way** replication allows you to replicate a set of data from a parent geodatabase to a child geodatabase and synchronize data changes multiple times in a single direction, either from parent to child or from child to parent.

If you create a one-way replica from parent to child you can edit data only in the parent, the data in the child is considered read only. In case edits are made in the child, they are overwritten if they conflict with edits applied during synchronization. One-way replication child to parent works in a similar manner, but in the opposite way.

One-way replicas persist after synchronization, allowing you to continue sending data changes.
Two-way replication allows you to establish a bidirectional link on a set of data in a geodatabase in order to send data multiple times between two geodatabases, from the parent replica to the child replica and from child to parent. If the same feature is edited in both replica geodatabases, it is detected as a conflict at the time of synchronization and specific policies can be established to decide how to resolve conflicts. As for one-way replicas, two-way replicas continue to exist after synchronization.

Specific workflows have been identified in WFP to exchange GIS data between offices, taking advantage of the different types of replication provided by ESRI. In particular for all data produced and maintained at the WFP Headquarters, one-way replication will be put in place to share it with Regional Bureaux and Country offices. The same will happen, but opposite way, for all those data produced at Country Office Level, while two-ways replication will be used to exchange data that could be edited by more GIS officers at the HQ, RBs or COs, such as logistics assets, WFP facilities or warehouses locations.

1.2 Data sharing in connected or disconnected environments

Data replication and synchronization tools provided by ESRI have been designed to work in different operating contexts; in scenarios where connectivity between offices is strong and reliable, ensuring that replica geodatabases are accessible on the network at all times, and environments where internet connectivity is weak or not available.

When planning which replication type or synchronization mechanism to use, you must assess the amount of resources available in terms of network bandwidth and latency between offices to choose of creating replicas with procedures thought for
connected environments and then synchronize changes on-line or put in place “disconnected procedures”.

In general the WFP’s network cannot be considered very reliable and in many offices Internet capacity is very weak. For this reasons the data sharing strategy must be designed carefully to ensure data availability at all times.

In the second chapter we will describe all options available to create replicas and then a specific workflow, which makes use of a combination of various techniques to create replica without a strong use of Internet connectivity and then synchronize data on-line when internet connectivity is available and use “disconnected” workflows when geodatabases in different locations are not connected.

1.3 Getting ready for replication

Before creating the replica you must determine which datasets to replicate and prepare them ensuring they meet the following requirements:

- The database user in the connection file used to create the replica must have write aces to the data.

- Each dataset must have a GlobalID column.

- All data must be registered as fully versioned; It cannot be versioned with the option to move edits to base.

- Spatial data must be stored in high-precision spatial reference. Most of the layers you have should be already stored in a high precision spatial reference. In case you get an error related to spatial reference
precision during the creation of a replica you can migrate your data to high precision follow instructions available on the ESRI portal:  

1.3.1 Add GlobalID column

You can add global IDs to a dataset in a geodatabase by right-clicking it in the Catalog tree and clicking Add Global IDs as reported in the figure below.

This command is available for stand-alone feature classes, feature datasets, tables and attributed relationship classes in a geodatabase but it cannot be executed on individual datasets in a feature dataset. If you add a feature class to a feature datasets that already has GlobalIDs for all feature classes, you have to run again the Add GlobalID command on the feature dataset. It will add the GlobalID only to the feature classes without GlobalID columns.
1.3.2 Registering data as versioned

You can only register your data as versioned when it is not in use by other people or applications, because an exclusive lock is required to ensure the dataset is not changing while the adds table is being created.

To register data as versioned follow steps below:

1. Start ArcCatalog or ArcMap and connect to the geodatabase that contains the feature dataset, feature class, or table you want to version, using a connection file created with the owner of the object you want to version.
2. In the Catalog tree, right-click the dataset, point to Manage, then click Register As Versioned.
3. The Register As Versioned dialog box opens.
4. Make sure the checkbox Register the selected objects with the option to move edits to base IS NOT selected.
5. Click OK.
1.3.3 Define the data to be replicated

You can limit the amount of data to replicate for each dataset by choosing to replicate only a subset of rows, a certain extent of your layers or even just the schema.

You can apply three types of filters on the data you replicate:

**Spatial:** You can limit the extent included in the replica by using the current view extent of the ArcMAP document you are using to create the replica. To do this you have to select a specific option in the Advanced Create Replica option that will be described in the chapter 2. All the features intersecting this filter will be included in the replica and will be affected by synchronization as per image below:

![Replica Area](image)

1. Selections: If there is a selection of certain objects in the datasets you are including in a replica, only these features will be included in the replica.

2. QueryDefs: Definition queries can be used to filter the content of feature classes and tables.
If more than one filter is used, the intersection of all filters is applied.

1.3.4 Replication and Metadata

Metadata for the data you choose to replicate is copied during the replica creation process. However, changes to the metadata are not applied during replica synchronization.

1.3.5 Replication and Raster Data

Raster catalogs, raster datasets and mosaic datasets cannot be versioned in ArcGIS for Desktop and, therefore, cannot be replicated. However, the Create Replica wizard will extract a raster dataset, raster catalog or mosaic dataset from the source geodatabase if it is in the Arc Map document. Since the data is extracted, any edits made to the raster will not be transferred to the relative replica during synchronization. The spatial extent of the replication area is used to determine the portion of the raster to extract. For raster datasets, the data is clipped to the spatial extent. For raster catalogs and mosaic datasets, the entire raster dataset is extracted for each raster dataset intersecting the spatial extent.

1.3.6 Replication and joined tables

In case layers are joined with tables in the same geodatabase, when replica are created including these layers, joined tables will be replicated as well but the join is not kept. It means that once the replica creation is finalized, the join between layers and tables must be recreated in the destination geodatabase.
2 Creating Replicas

All tasks related to replica creation and management can be executed through ArcMap using a specific toolbar called Distributed GeoDatabases. You can display it in ArcMap by selecting the tab Customize, then choose Toolbars and click on **Distribute GeoDatabases**.

At this point the toolbar reported below should appear:

![Distributed Geodatabase toolbar](image)

This toolbar contains a set of tools that allow you to create a replica, synchronize changes, apply schema changes, manage existing replica and more. In this section of the training we’ll explore the first tool for creating replica through a wizard. It will allow you to define what data should be replicated, what kind of replica you would like to create, and if you are putting in place a replica in a connected environment, which is the destination geodatabase.

We will initially describe the process to Create Replicas in Connected Environments and then additional steps needed in case your Internet connectivity is not reliable or insufficient.

Before starting please consider that:

- In case you want to add to the replica more layers stored in the same feature dataset, you can add just one of them to the ArcMap document because at a later stage the Creating Replica Wizard will present you a list of objects you want to include in the replica, taking into consideration the entire feature dataset for each layer in the ArcMap document;
2.1 Creating replicas in a connected environment

If you have already prepared your data for replication, you can use the Create Replica wizard in ArcMap to create checkout, one-way or two-way replicas. For putting in place a replica in a connected environment two connection files are needed: one connection file to your geodatabase and one connection file to the destination geodatabase. Both connection files but must connect to the geodatabase with a user with privileges to create new objects in the respective geodatabase, otherwise once the replica will be in place you will have limited privileges (for instance you won’t be able to change the schema of a layer adding or removing fields).

2.1.1 Creating a one-way or two-way replica

You find below the steps to create a one-way or two-way replica. In the example we have chosen to include in the replica a layer containing Earthquake events stored in a feature dataset called NHR.

1. Open a new ArcMap document;

2. Make sure the objects you want to include in the replica have a GlobalID column and are Registered as Versioned;

3. Connect to your local geodatabase using the connection file and load into the ArcMap document the objects you want to include in the replica. In this case we have the layer: wld_nhr_EQepicenters_usgs;
4. Add any filtering criteria may be required. Spatial, Selections or Definition Queries. In this case we have zoomed in to include in the map extent only an area around Africa.
5. If the Distributed Geodatabase toolbar is not already open, click **Customize** on the main menu, point to **Toolbars**, and click **Distributed Geodatabase**.

6. If layers that can be replicated are in your ArcMap document, the **Create Replica** button on the **Distributed Geodatabase** toolbar will be active, click it to activate the Create Replica wizard.

7. If your ArcMap document includes data from more than one ArcSDE geodatabase you are prompted to choose which ArcSDE geodatabase you want to work with.

8. At this step a the Creating Replica wizard will appear, letting you choose to create either a check-out, a two-way, or a one-way replica. In this instruction we are focusing on these last two replica types, for which the process for creating are very similar. The main difference is that, in case you choose a one-way replica, you have to select if it must be parent-to-child or child-to-parent. In this example we select a two-way replica.
9. On the next panel, choose to replicate data or register existing data only. The option to register existing data only allows you to create a replica between two geodatabases containing the same data. It may be useful if your internet connection is reliable but available bandwidth is limited. In this case you can send data to the destination database by other means (FTP, physical drives etc) and then create the replica. In this example we choose to replicate also the data.

10. You can choose to replicate to a Geodatabase or an XML workspace. To create a Replica using the “Connected” workflow you must specify the connection file to connect to the destination geodatabase. While the XML option is used to create the replica using the procedure for disconnected environments, which will be described later. Browse to the connection file to which you want to replicate data.

11. Type the name of the replica you are creating.
You should select a name that then helps you identify the main information of
the replica, such as source, destination and data included. In this example we
use as replica name

**HQ_OMB_Earthquakes** because we are creating a replica between the
enterprise geodatabase at the WFP headquarters and a geodatabase at the
regional office in Bangkok including in the replica earthquake epicenters.

12. Check the **Show advanced options** box and click next to proceed.

13. The first panel of the **Advanced Create Replica Options** offers the choice
between a **Full** or **Simple** model. Keep the default **Full** model and click next.

14. The next window let you decide if you want to limit the replica to a specific
extent (the one in the ArcMap document window selected at the beginning),
use the full extent of the data or specify manually an extent.
Through this window you can also exclude individual layers or tables from the
replica, unchecking the check box associated with that layer or table.
As mentioned before, if you selected in the first step a layer stored in a
Feature Dataset, a list of all the layers in that FD will be presented.
To include in the replica only your **Earthquake** layer, uncheck all the others.

You can also change the dataset's name when it is replicated, under the **Target
Name** column. It may be useful if you are replicating only a portion of the data.

If you have previously chosen the option to register existing data only, then
each name in the **Target Name** column will offer a drop-down list of available
datasets in the target geodatabase to choose from.
Each entry in the **Check Out** column is a combo box of options. The options always include **All Features** and **Schema Only**. If a particular layer or table has a selection set or definition query defined, the options may also include **Selected Features Only**, **All Features in Def. Query**, and **Selected Features in Def. Query**.

Select All Features and uncheck the Use Spatial Extent box if you do not want to apply any data filters.

Uncheck the **Replicate related data** check box if you do not want to replicate any related data and click **Next**. This option may be useful if relationship classes are associated to the layers in the replica.

- When replicating tables the default option is to replicate the schema only; in case you want replicate the data as well, make sure to select **All Records** in the **Check Out** field, as shown in the image below;
15. For two-way replication, there is an additional panel in the Advanced Create Replica Options dialog box that lets you choose whether the parent or child will be the initial data sender. This option is only important in disconnected systems.

16. Click **Next** in the following window.
17. At this step you can choose what you want to do once the replication has been completed

**Take no further action** — This is the default option. If you choose this option, the replica will be created with the ArcMap document showing the original data.

**Change the layers and tables to point to the replicated data** — The current ArcMap document will be modified to point to the data in the replica geodatabase, preserving all symbology.

**Save a copy of this map document with the layers and tables pointing at the replicated data** — A new ArcMap document referencing the data in the replica geodatabase with symbology preserved will be created.

18. Click **Summary** to review the parameters for the current replica.

19. Click **Finish** to start replicating the data. The status of the replication operation will be monitored in a progress dialog box.
If you are interested in getting to know how to create a checkout replica please follow the guide on the ESRI portal:


2.2 Creating replicas in a disconnected environment

In this section we’ll describe the process to create a replica when Internet connectivity is not reliable or weak. After several tests at Country Office and Regional Bureau level we can say that is the process that has more chances to succeed when putting in place replicas in WFP, considering networking constraints in our organization.

The process is slightly more involved than in a connected environment simply because there are more things to remember and steps to follow.

It can be summarized in a 3 phases process:

1) Create the replica in the source geodatabase;
2) Sent it to the person managing the destination geodatabase;
3) Import the replica in the destination geodatabase.

At some point, when we described the steps to put in place a replica in a connected environment, we choose to send it to the destination geodatabase online by selecting a connection file. In this case we have to save the replica in an XML file, because we cannot ensure that the destination geodatabase is reachable remotely. Once the replica is created, it must be sent to the destination geodatabase by FTP transfer, email or physical drives and then imported in the destination geodatabase to finalize the replica creation process.
2.2.1 Replica creation in the source geodatabase

We’ll now describe the step-by-step process to create a one-way replica in a disconnected environment, simulating a scenario where the source geodatabase (Parent) is at the WFP Country Office in Myanmar, the destination geodatabase (Child) at the WFP HQ in Rome and including in the replica a layer containing WFP facilities.

1. Open a new ArcMap document;

2. Make sure the layer you want to include in the replica has a GlobalID column and is Registered as Versioned;

3. Connect to your local geodatabase using the connection file and load into the ArcMap document the layer you want to include in the replica. In this case we have the layer: **mmr_poi_facilities_wfp**
4. Add any filtering criteria may be required. Spatial, Selections or Definition Queries. In this case we do not add any filtering criteria.

5. Click the **Create Replica** button on the **Distributed Geodatabase** toolbar to activate the Create Replica wizard.

6. Choose to create a one-way replica Parent to child.

7. On the next panel
   - Choose to replicate also the data by checking the box **Data**;
   - Choose to create the replica into an XML document by browsing the folder where you want to save it and choose a name for your replica. You should select a name that then helps you identify the main information of the replica, such as source, destination and data included. In this example we use: 
     **Replica_Myanmar_to_HQ_WFPFacilities.xml**;
   - Type the name of the replica you are creating. In this example we use as replica name: **Myanmar_to_HQ_WFPFacilities**;
   - Check the **Show advanced options** box and click next to proceed.
8. In the first panel of the *Advanced Create Replica Options* keep the default **Full** model and click next.

9. In the next window choose to replicate the full extent of the data. In the list of items to checkout you will notice that there is not only the WFP facilities layer but also a layer containing warehouses locations. This happens because both layers are stored in the same feature dataset (WFP). In this example we uncheck the warehouses layer to include in the replica only WFP facilities.

10. Click **Next**
11. Choose to **Take no further action** and click **Finish** to finalize the replica creation process in the source geodatabase.

Once the process is completed the replica is registered in the source geodatabase and all the data needed to create the replica in the destination geodatabase is stored in the xml file.

**2.2.2 Replica transfer to the destination geodatabase**

At this step you must send the xml file produced in the previous phase to the person managing the destination geodatabase to complete the replica creation process from your side.

Please note that in most of the cases the generated XML will be too big to be sent by email. In this case the preferred option to send the file is the FTP transfer. FTP servers have been made available at HQ and at each Regional Bureau to support the implementation of the GIS infrastructure in WFP. Therefore you should get the necessary information to access both FTP servers (HQ and at your reference RB), before creating this kind of replica.

**2.2.3 Replica creation in the destination geodatabase**

Once the person managing the destination geodatabase receives the XML document containing the replica, the final phase of the replica creation process can be performed by importing all the data and registering the replica in the destination geodatabase.
Both results can be achieved through one single operation, by importing the XML document into the destination geodatabase as an XML workspace.

We report here the steps needed:

1. In the Catalog tree, right-click the geodatabase into which you want to import the replica and point to **Import > XML Workspace Document**. The window shown in the image below will appear.

2. Choose whether to import the geodatabase and its data by clicking the **Data** checkbox.
3. Click the open folder button and navigate to the XML document containing the replica, select it and click **Open**.
4. Click Next.
5. The window shown in the image below will appear. It presents a list of objects that will be imported in the destination geodatabase, any related data (Domain, Relationship Classes, Topologies etc.) will also appear.
Any naming conflicts are displayed in red. To change a suggested name in the Target Name column, overwrite the name.

6. Click **Finish** to complete the process.

Additional options are available to create replicas in both connected and disconnected environments, when for instance the same data is already available in the source and destination geodatabases or you want to replicate only the schema of your data.

If you are interested in such procedures please refer to the documentation on the ESRI platform, as they are out of the scope of this training document.

http://resources.arcgis.com/en/help/main/10.2/index.html#/Replicas_created_with_the_option_to_register_with_existing_data/003n000000vq00000000/

http://resources.arcgis.com/en/help/main/10.2/index.html#/Reusing_a_schema_to_create_a_replica/003n000000w30000000/
3 Data Synchronization

Once a replica between two geodatabases is in place data changes can be exchanged between them using synchronization procedures. Data changes include inserts, updates and delete of features in existing layers or tables but not changes in the schema structure (add or remove fields or layers etc.). These must be managed with a specific workflow called Apply Schema changes that will be described later.

Data synchronization is performed through a specific tool in the Distributed Geodatabases toolbar, after connecting to the database originator or the changes with the same database user as that used to create the replica, or the geodatabase administrator.

For two-way and one-way replication, the same filters and relationship class rules used in replica creation are applied during synchronization, with the exception of filters based on a selection set. When determining the changes to send, all edits in each replica dataset that have been applied since the last synchronization are evaluated. If an edit satisfies the replica's filters, it will be synchronized.

Additionally the geodatabase synchronization system detects errors occurring during synchronizations and rollbacks in case needed. Any changes that have been applied are removed, and the system is put back as it was before the synchronization.

Synchronization is supported in both a connected and a disconnected environment. In a connected environment, message exchange is handled by the system, while in a disconnected environment; you need to manage the message exchange. Both processes are fully described in this chapter.
If you are using advanced geodatabase features, such as geometric networks, topologies or versioning, it’s strongly suggested to have a look at the specific documentation on the ESRI portal.


### 3.1 Synchronizing Replicas in a connected environment

Synchronize connected replicas is an easy procedure that can be carried out using the Synchronization Changes wizard in ArcMap.

You can connect to your geodatabase and choose the replica to synchronize. All the required message exchanges needed to complete synchronization are executed by the system. You never have to be concerned with message exchange or which replica is the sender or receiver.

Moreover connected synchronization allows you to choose the direction in which the changes will be sent in a two-way replica. For example, you can send changes from the parent replica to the child replica or from the child replica to the parent replica, or in both directions. If you choose both directions, changes are first sent in one direction and then sent in the opposite direction, all in one operation.

Here the step-by-step process:

1. Edit the layer included in your replica;
2. Open the Synchronize Wizard by clicking the Synchronize button on the Distributed Geodatabase toolbar;
3. On the first panel of the Synchronize wizard, choose the replica that you want to synchronize. In this example we want to synchronize the two-way replica **HQ_to_OMB_Earthquakes**

Once selected, the connection information for geodatabase 2 is automatically filled in.

![Synchronize Changes Wizard](image)

4. For two-way replicas, choose the direction in which you would like to send changes. For checkout replicas, the only option available is to send changes from the child replica to parent replica. For one-way replicas, the only option available is to send changes from the parent replica to the child replica, or if it's a child-to-parent replica.
5. Click Next.

6. For checkout replicas, there is an option to reconcile and post with the parent version upon synchronization. For two-way and one-way replicas, this is always checked on.

7. Next, choose to define conflicts By Object or By Attribute.

8. Defining conflicts by object detects conflicts by row, while defining them by attribute detects conflicts by column.

9. Choose how you want conflicts resolved. (the first two are usually preferred)

   **In favor of the geodatabase1** — In this case, the edits in geodatabase1 are automatically used over the edits in geodatabase2 if there is a conflict. Since the
conflicts are resolved automatically, the replica is never in a conflict state after synchronizing with this policy.

**In favor of the geodatabase2** — In this case, the edits in geodatabase2 is automatically used over the edits in geodatabase1 if there is a conflict. Since the conflicts are resolved automatically, the replica is never in a conflict state after synchronizing with this policy.

**Manual** — With this policy, if a conflict occurs, the reconcile operation is aborted and the replica is marked as in conflict. This gives you an opportunity to perform the reconcile afterwards either manually or by running some custom reconcile code. Once the reconcile is applied and the changes posted to the replica version, the replica is no longer in conflict. While the replica is in conflict it can continue to receive changes but cannot send changes.

10. Click Finish.
3.2 Synchronizing Replicas in a disconnected environment

In case you need to synchronize a “disconnected” replica, you have to go through a process of manually exchanging messages between replicas in the source and destination geodatabases. This is needed because the geodatabase that sends changes must be aware if the synchronization on the other hand completed successfully or changes need to be synchronized again. For this reason two types of messages must be exchanged: data change messages and acknowledgement messages.

This happens through the multi-steps process described in the image below:

The data sender exports changes to be applied to the relative replica into a **Data Change Message** and send it to the receiver. The data receiver imports this message into his geodatabase and exports an **Acknowledgment Message** and send it to the data sender to confirm that data changes were received and processed. The data sender imports the Acknowledge Message into its geodatabase to complete the process.
It is important for the data receiver to export acknowledgment messages as often as possible. If no acknowledgment messages are received, the data sender resends changes by default, and maintains the information needed to resend changes until those changes are acknowledged. As a result, the data sender's geodatabase can become large, and subsequent data change messages can also become large.

If you have particular needs that are not covered by this workflow please refer to the ESRI’s documentation or get in touch with your GIS focal point in WFP.

n/003n000000v2000000/

We’ll now describe the entire process with a practical example, using the disconnected replica created in the second Chapter (Myanmar_to_HQ_WFPFacilities)

After adding a new feature in the source geodatabase we start the process to send this change to the destination geodatabase through a disconnected synchronization.
3.2.1 Export a data change message

1. Add the layer to an ArcMap document; (Highlighted in a red box the new point)

2. Open the Distributed Geodatabase toolbar and click the Export Data Changes Message button.

3. In the wizard, select the replica from which you would like to export data changes and specify if you want to save the message in a delta database or an XML file. We suggest using an XML file if there is few data to synchronize, while the geodatabase option is preferred in case you have done a lot of edits.
Specify a self-explanatory name for the Data Change message. It should contain at least the name of the Replica and the date as per the example below:

ExportaDataChanges_Myanmar_To_HQ_BaseLayers_17122014.xml

4. The last 3 check boxes are useful to manage particular needs. **In general you should leave the last 2 checked.**

For reference we report some info about these checkboxes. By combining the last two check boxes, you have three additional potential options:

1) Include all unacknowledged data changes and new data changes since the last export.
2) Include only those unacknowledged data changes that you have not received an acknowledgment message for.
3) Don't include any data change messages. This option is useful for two-way replica, for sending a message to switch roles without sending any data.

Once the Data Change Message is saved in you file system, you have to send it to the person managing the destination geodatabase as described in the section Creating replicas in a disconnected environment.

### 3.2.2 Import a Data Change Message

In order to synchronize edits made in the source geodatabase the person managing the destination geodatabase must import the Data Change Message received.

1. Add one of the layers included in the replica that needs to be synchronized to an ArcMap document and click the **Import Message** button that can be found in the Distributed Geodatabase toolbar.

2. If prompted, choose the replica geodatabase you want to import a message to. It is needed only if in the ArcMap document there are layers stored in different geodatabases.

3. Choose the delta file you would like to import. (XML or Geodatabase)

4. In case of a two-ways replica, choose between column- and row-level conflict detection and how to resolve conflicts:
• **In favor of the database** — In this case, the edits of the database importing the changes are used over the edits in the delta file if there is a conflict. Since the conflicts are resolved automatically, the replica is never in a conflict state after importing with this policy.

• **In favor of the imported changes** — Here, the edits defined in the delta file are used over the edits of the database importing the changes if there is a conflict. Since the conflicts are resolved automatically, the replica is never in a conflict state after importing with this policy.

• **Manually at a later time**—With this policy, if a conflict occurs the reconcile operation is aborted and the replica is marked as in conflict. This gives you an opportunity to perform the reconcile afterwards, either manually or by running some custom reconcile code. Once the reconcile is applied and the changes posted to the replica version, the replica is no longer in conflict. While the replica is in conflict, it can continue to receive changes but cannot send changes.

5. Click Finish to finalize the process. Now the new point is present also in the destination geodatabase.
3.2.3 Export an Acknowledge Message

Once changes are imported into the destination geodatabase, an Acknowledge message must be produced and sent to the person managing the source geodatabase to confirm that the process has been completed successfully.

1. Ensure one of the layers included in the replica that needs to be acknowledged is in the ArcMap document and click the Export Acknowledge Message button that can be found in the Distributed Geodatabase toolbar.

2. If prompted, choose the replica geodatabase you want to import a message to. It is needed only if in the ArcMap document there are layers stored in different geodatabases.

3. In the wizard, select the replica from which you would like to export the acknowledge message and specify its name.
Specify a self-explanatory name. It should contain at least the name of the Replica and the date as per the example below:

**AcknowledgeMessage_Myanmar_To_HQ_Earthquakes_20012015.xml**

4. Click **Finish**

Once the Acknowledge Message is saved in your file system, you have to send it to the person managing the source geodatabase as described in the section Creating replicas in a disconnected environment.

### 3.2.4 Import an Acknowledge message

This last phase finalizes the process of synchronized data in a disconnected environment. The person managing the source geodatabase must import the received Acknowledge Message by following the steps below.

1. Add one of the layers included in the replica that needs to be synchronized to an ArcMap document and click the **Import Message** button that can be found in the Distributed Geodatabase toolbar.

2. If prompted, choose the replica geodatabase you want to import a message to.
It is needed only if in the ArcMap document there are layers stored in different geodatabases.

3. Choose the Acknowledge Message you would like to import.

4. Click Finish to finalize the process.
3.2.5 **Synchronize “connected” replicas automatically**

In a system where there are many replicas and/or frequent synchronizations, guarantee that data is efficiently synchronized can be very time and bandwidth consuming.

For support these use cases, it is possible to create replication agents to automate the synchronization process. Replication agents work by automatically connecting to replicated geodatabases and performing synchronizations. In this case, end users do not have to explicitly synchronize their databases, as synchronization happens automatically.

In a connected environment, replication agents can be built using geoprocessing tools, export them to Python scripts and then execute them using scheduling software, such as the Windows scheduler, so that they can be run on a regular basis. For example, you may want to schedule a synchronization between two enterprise geodatabases once a week at a nonpeak time.

Below we report an example of script produced to synchronize a replica containing natural hazard datasets from the geodatabase 1 at the WFP HQ and the geodatabase 2 at the regional office in Bangkok (OMB).

You need to provide to the script:

- The connection file to connect to the Geodatabase 1 (**`wfp_hq_esri_HQ.sde`**)
- The connection file to connect to the Geodatabase 2 (**`hq_hq_esri_OMB.sde`**)
- Replica name (**`HQ_to_OMB_NaturalHazards`**)
from arcpy import env
import sys
print(sys.version)
print("arcpy imported")

# Set workspace
workspace = r"C:\wfp_sdi_confs\synchro_script"

# Set local variables
replica_gdb1 = r"C:\wfp_sdi_confs\synchro_script\wfp_HQ_esri_HQ.sde"
print("defined geodb 1")
replica_gdb2 = r"C:\wfp_sdi_confs\synchro_script\hq_HQ_esri_OMB.sde"
replica_name = "wfp.HQ_to_OMB_NaturalHazards"
sync_direction = "FROM_GEODATABASE1_TO_2"
conflict_policy = "IN_FAVOR_OF_GDB1"
# Not applicable for one way replicas, there is not conflict detection.
conflict_detection = "BY_OBJECT"
# Not applicable for one way replicas, there is not conflict detection.
reconcile = ""
# Only applicable for Checkout replicas

print("Synchronization starting at " + str(datetime.datetime.now()))
print("=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-")
print("Please don't close this window, wait for next message.")
# Execute SynchronizeChanges.
arcpy.SynchronizeChanges_management(replica_gdb1, replica_name, replica_gdb2,
    sync_direction, conflict_policy,
    conflict_detection, reconcile)

print("Synchronization ended at " + str(datetime.datetime.now()))
del arcpy
print("Now safe to close. Press ENTER.")
# raw_input()
4 Managing changes in the layers structure

When a replica is created, data and schema are copied from the source geodatabase (Parent) to the destination geodatabase (Child). The data includes the rows to be replicated from the datasets in the replica. The schema consists of the fields, domains, subtypes, and other properties that describe the replicated data. Initially, the schemas are identical on both replicas, but over time, changes might be applied to the schema in each geodatabase.

Specific procedures have been designed by ESRI to support replica and synchronizations when the schema of one or more objects in the replica is changed (for instance when you add or remove a field). Even though, it is best to avoid schema changes because they can lead to inconsistencies among replicas, and the extra task of applying schema changes can increase performance costs. However, there are cases when schema changes must be applied.

In these cases before synchronizing the replica you have to run through a Schema Change process. It first compares the schema of layers in both geodatabases (source and destination) then modifies the structure of the geodatabase receiving the Schema Change by modifying the structure.

These procedures are slightly different for replicas in connected or disconnected environments.
4.1 Applying schema changes in connected environments

In case you are working with a “connected” replica you can execute a schema change using the **Compare Replica Schemas** tool to check differences in the structure of the layers in the two geodatabases and then the **Import Schema Changes** tool to apply changes on the destination geodatabase.

4.1.1 Compare replica schema

In this example we apply a Schema Change after having added a field (schematest) to the Earthquake layer in the source geodatabase.

1. Connect to the destination geodatabase and add the layers included in the replica that needs to be synchronized to an ArcMap document and click the **Compare Replica Schema** button that can be found in the Distributed Geodatabase toolbar.
2. Specify or browse the connection file to the source geodatabase.

3. Use the Replica name combo box to choose the replica.

4. Specify or browse for an output replica schema changes XML file.

   Specify a self-explanatory name. It should contain at least the type of message (SchemaChange in this case), name of the Replica and the date as per the example below:

   **SchemaChange_HQ_TO_OMB_Earthquakes_20012015.xml**

5. Click Finish.
4.1.2 Import replica schema

1. Connect to the destination geodatabase and add one of the layers included in the replica that needs to be synchronized to an ArcMap document and click the **Import Schema Changes** button that can be found in the Distributed Geodatabase toolbar.

2. Browse for the replica schema changes file. Replica name and Replica type will be automatically filled in.

3. Click **Next**
4. The second dialog box lists the differences between the two schemas. Check the check boxes under the Apply column for the changes you want to apply to the replica schema.

5. Click Finish.

At this point the Earthquake layer in the destination geodatabase should have the “schematest” field and any synchronization should work properly.

4.2 Applying schema changes in disconnected environments

In case you are working with a “disconnected” replica, before comparing the schemas and import schema changes you must export your schema to a file, send it to the destination replica where it can be compared and imported. At this scope we’ll use an additional tool provided in the Distribution Geodatabase toolbar called Export Replica Schema.
For this example we’ll remove the **int_staff** field from the **WFPFacilities** layer which we included in a disconnected replica in the Chapter 2

### 4.2.1 Export replica schema

1. Connect to the source geodatabase, add the layer included in the replica that needs to be synchronized to an ArcMap document and click the **Export Replica Schema** button that can be found in the Distributed Geodatabase toolbar.

![Export Replica Schema](image)

1. Specify a name for the XML file that is to be created.

![Export Replica Schema Wizard](image)

2. Choose the replica to export.
3. Click Finish.
At this point you have to send this export to the person managing the destination database who will compare the schema and import changes.

### 4.2.2 Compare replica schema

1. Connect to the destination geodatabase and add the layers included in the replica that needs to be synchronized to an ArcMap document and click the **Compare Replica Schema** button.

2. Browse the replica schema file received.
3. Use the Replica name combo box to choose the replica.

Specify a self-explanatory name. It should contain at least the type of message (OutputSchemaChange in this case), name of the Replica and the date as per the example below:

```
OutputSchemaChange_Myanmar_to_HQ_20012015.xml
```
5. Click Finish.

4.2.3 Import replica schema

1. Connect to the destination geodatabase and add one of the layers included in the replica that needs to be synchronized to an ArcMap document and click the **Import Schema Changes** button.

2. Browse for the replica schema changes file. Replica name and Replica type will be automatically filled in.

3. Click **Next**

4. The second dialog box lists the differences between the two schemas. Check the check boxes under the Apply column for the changes you want to apply to the replica schema.
5. Click Finish.

At this point the Schema Change process in a disconnected environment is finished and the `int_staff` field should have been removed from the WFPFacilities layer in the destination geodatabase.
5 Replica Management

Going through the previous chapters you may have noticed that working with replicas involves many procedures and things to remember. Is it a checkout, one-way or two-way replica? Who is child or parent in a replica pair? Can we synchronize data using the workflow for connected environments?

Most of these questions can be answered using a tool available in the Distributed Geodatabase toolbar called **Replica Manager**.

You can use this tool to rename, refresh, and review the properties of each replica as well as remove datasets from a replica. You may also view all replicas with a role of parent or child; list only checkout/check-in, one-way, two-way, or all types of replicas; and view the replica log. This utility is available in both ArcCatalog and ArcMap.

To open the Replica Manager, click the Manage Replicas button on the Distributed Geodatabase toolbar.

In the example below replicas created in the Second Chapter are listed.
For each replica you can access information regarding:

- **Name**
- **Owner** — The user who created the replica
- **Type** — The type of replica created: Check out/Check in, One way, or Two way
- **Role** — Shows whether the replica is a parent or child
- **Status** — Displays whether the replica is currently a data sender or a data receiver
- **Conflicts** — Displays Yes if there are any outstanding conflicts to be resolved with this replica and No if there are not
- **Version** — The replica version set during replica creation
- **Date Created** — The date and time that the replica was created

By right-clicking an individual replica, you bring up a context menu with the following options:

- **Rename** — Renames the replica; type a new name and press ENTER.
- **Unregister** — Unregisters the replica; if a synchronization version exists, it will also be removed.
- **View log** — Opens the Replica Log, which maintains a record for each data message sent or received by the replica.
- **Refresh** — Refreshes the replica; the latest state of the replica properties are displayed.
- **Validate schema** — Validates the replica schema; a progress bar monitors the status of the validation. This automatically corrects a replica that contains
invalid datasets. Datasets can become invalid if they have been renamed, deleted, or unversioned.  

- Properties—Opens the Replica Properties dialog box. Replica properties are discussed in more detail below.

This last option is particularly interesting because it allows changing some replica parameters that are fundamental for the workflows identified in WFP. It contains three tabs: General, Description, and Advanced.

The **General** tab shows some additional replica parameters, such as the replica and synchronization versions.
The **Description** tab shows the datasets that are included in the replica and allows to remove them by right clicking and choosing **UnRegister from Replica**; however, they still remains in the geodatabase.

![Replica Properties dialog box](image)

You can also see if any filters were applied to the replica dataset during replica creation by choosing **View filters**.

The **Advanced** tab in the **Replica Properties** dialog box shows information about the generation numbers associated with the replica. The generation number is a number maintained by the geodatabase, which keeps track of messages being sent and received by the replica. For example, the first data change message sent from one replica to its relative would make the current generation of the replica 1. When the relative imports that message, its relative replica generation gets set to 1. When the replica receives an acknowledgment of the data change message, its last acknowledged generation gets set to 1.
The Advanced panel displays the following information about generation numbers:

**Current Generation** — The current generation number, which describes how many data messages were sent from the replica;

**Last Acknowledged Generation** — The generation number of the most recent data message for which an acknowledgment has been received;

**Generation** (based on last received message) — The current generation number of the relative replica based on the last message imported into this replica;

The Relative Replica Connection contains the connection information of the relative replica. You can set this information by browsing to the location of the relative replica's connection file.
Replicas created using the workflow for connected environments already know how to reach the relative geodatabase because the connection files to it was provided at the replica creation time. In this case for both geodatabases the Relative Replica Connection in the Replica Manager will already include this information.

Instead replicas created using the workflow for disconnected environment are not aware of this information. The connection file to reach the relative geodatabase needs to be provided manually if you want to synchronize data using the workflow thought for connected environments, otherwise you will only be able to use the disconnected one.

5.1 Turning a disconnected replica in a connected one

In contexts with limited bandwidth capacity, big latency or internet connectivity not reliable you may want to create replica using the workflow for disconnected environments but then be able to synchronize data online, when internet connectivity allows, and use the synchronization workflow thought for disconnected environments otherwise.

In order to achieve this you need to provide to your geodatabase information to reach the relative geodatabase through the Replica Manager.

1. The person who initially created the disconnected replica must add to and ArcMap document one of the layers in the replica and Open the Replica Manager tool;

2. Right click the disconnected replica to modify and choose **Properties**;
3. Click the Advanced tab, you will notice that the **Relative Replica Creation** section is empty. Browse the connection file to the relative geodatabase;

![Replica Properties dialog box](image)

4. Check the Persist user name and password check box. If the user name and password are not persisted, you will be prompted to provide them every time you want to run a synchronization online.
6 Data sharing workflows in WFP

Specific workflows have been designed to facilitate GIS data sharing across WFP offices; they make use of distributed geodatabases and replica and synchronization tools to overcome well known networking issues in our organization.

A combination of the different techniques described in the previous chapters is suggested to allow each office to synchronize data online when Internet connection is available and reliable, while using disconnected synchronization in case the online procedure is not feasible because of poor connectivity.

In general we can consider that GIS officers at the WFP headquarters produce and maintain some datasets that may be of interest for GIS colleagues working in regional and country offices, such as layers containing information related to natural hazards (tropical storm tracks, earthquake epicenters, flood extent areas etc.). At the same time GIS colleagues working in the field collect and maintain data extremely interested for other GIS practitioners in WFP, such as population figures, food security and vulnerability data, context related information etc.

All these datasets can be shared across offices setting up one replica for each flow of information. For instance we’ll create a replica containing all data produced at the WFP head quarters in which the geodatabase at HQ is the parent and the destination geodatabase at the CO or RB is the child. One or more replica will then be created for data sitting and CO or RB level, in this case the roles are inverted, CO or RB geodatabases are parent and the one at HQ is the child.

Additionally we’ll consider a set of data that is managed at the HQ with a global extent and a standardized data structure, to which the CO can contribute by editing existing layers. These include WFP locations of interest (WFP facilities, warehouses)
and logistics infrastructures (airports, ports, unhas routes, border crossing points etc.). Such data will be included in an additional replica that from the CO will populate global layers maintained at the headquarters.

6.1 Data flow HQ → CO and RB

As mentioned before the Geospatial Support Unit at HQ maintains and daily updates a set of layers related to natural hazards. The one reported below are considered of interest for country and regional offices, therefore they will be shared through a specific replica:

- Earthquakes epicenter locations (wld_nhr_eqepicenters_usgs)
- Flood extent areas (wld_nhr_faa30days_dfo)
- Tropical storms buffers, nodes, tracks (wld_nhr_tsbuffers_gdacs, wld_nhr_tsnodes_jtwc, wld_nhr_tstracks_jtwc)

As these layers are managed at HQ with a global extent, at the replica creation time spatial filters will be applied according to the spatial extent of the Country where the destination database is hosted.

This will be a one way replica put in place between the HQ (Parent db) and the Country Office\Regional Bureau (child DB) using the process for disconnected environments described in the paragraph 2.2.

After creating the disconnected replica, we need to add the connection file to reach the destination geodatabase using the steps reported in the paragraph 5.1. This will enable the replica to be synchronized using the workflow for connected environments.
At this step the replica can be synchronized using either the workflow for connected environments or the one for disconnected environments. When this replica will be put in place for all Country Offices and Regional Buereaux, there will be to many replicas to run manual synchronization. For this reason automatic synchronization workflows will be put in place using the reference documentation in the paragraph 3.2.5.

### 6.2 Data flow CO and RB → HQ

In this section we’ll provide some additional information to facilitate the creation of replicas to send GIS related data from the Country Office\Regional Bureaux to the WFP Headquarters.

After standardizing GIS datasets and storing them in the Enterprise Geodatabase at the Country Office, one or more replicas can be put in place to share this data with other GIS practitioners at the WFP headquarters and regional offices.

The number of replica to create depends mainly to the amount of data available at the country office and how frequently it is updated. It is suggested to create group together layers with the same frequency of update, for instance layers containing administrative boundaries are not updated often while table containing secondary data maybe be updated on a daily base. In this case the best strategy would be to create one replica containing all boundaries datasets, another replica containing all tables, and additional ones according to update frequency.

All these replicas can be created using a workflow similar to the one described in the previous paragraph.
Each of them will be a one way replica where the parent geodatabase is the Enterprise Geodatabase residing at the country office and the child geodatabase in this case is the Enterprise Geodatabase at the HQ. These replicas will be created using the process for disconnected environments described in the paragraph 2.2.

After creating the disconnected replica, we need to add the connection file to reach the destination geodatabase using the steps reported in the paragraph 5.1. This will enable the replica to be synchronized using the workflow for connected environments. At this step the replica can be synchronized using either the workflow for connected environments (automatically or manually) or the one for disconnected environments.

### 6.3 Collaborative editing of global layers

Some of the layers used daily at the WFP HQ for mapping or analysis purposes are managed at a global extent. This set of layers include:

- WFP facilities \(\text{wld\_poi\_facilities\_wfp}\)
- Warehouses \(\text{wld\_poi\_warehouses\_wfp}\)
- Airports \(\text{wld\_trs\_airports\_wfp}\)
- Ports \(\text{wld\_trs\_ports\_wfp}\)
- Border Crossing Points \(\text{wld\_poi\_bcp\_wfp}\)
- UNHAS routes \(\text{wld\_trs\_unhasroutes\_wfp}\)
- Supply routes \(\text{wld\_trs\_supplyroutes\_wfp}\)
- .....


The suggested procedure to manage these layers is to adopt a standardized data structure defined by these global layers at the country office as well. To achieve these is necessary to put in place a multi-phases process:

1) Share the standardized data structure by creating a 2 ways replica between the HQ and the CO, which will include all the layers managed at a global extent. At the replica creation time filters will be applied to replicate only data of interest for that specific country. Each layer contains in the attribute table an iso3 field; for instance, when replicating data for Myanmar, definition queries will be created to replicate features containing only the iso3 = MMR.

2) At the CO, edit replicated layers by filling in information available. For instance add locations of WFP offices if not already existing or close warehouses that are not available anymore.

3) Synchronize this data back to HQ using either connected or disconnected synchronization.