Importing Data from a p:IGI-3 File: An introduction

Table of Contents

[Summary 1](#_Toc72831339)

[Introduction 1](#_Toc72831340)

[p:IGI-3 Data Import: Maintaining the p:IGI-3 data structure in p:IGI+ 2](#_Toc72831341)

[p:IGI-3 Data Import: Data formatted to a p:IGI+ structure 3](#_Toc72831342)

# Summary

Importing data into a new or existing project is a common activity. During the transition from working with p:IGI-3 to using p:IGI+ there will likely be a need to import data to p:IGI+ from p:IGI-3 files.

With the evolution of the property model in p:IGI+, import of data from a p:IGI-3 file into p:IGI+ now offers two main import routes, and either one should not be undertaken without careful consideration**.** The decision whether to maintain the data structure seen in p:IGI-3 in p:IGI+ or format the data structure from p:IGI-3 to that used in p:IGI+ will depend on the end goal of the data import exercise and time available (short term use vs long term use/data storage).

Both approaches will be covered in this article.

# Introduction

With the expansion of the IGI property model and the use of a new data structure in p:IGI+ import of data into p:IGI+ from p:IGI-3 file should not take place without careful consideration**.**

Before importing data from p:IGI-3 users will need to decide how they wish to use and interact with their data once in p:IGI+. The decision is whether to:

A) Maintain the data structure seen in p:IGI-3 in p:IGI+.

B) Format the data exported from p:IGI-3 to fit the data structure used in p:IGI+.

The decision of which approach to use will depend on the end goal of the data import exercise and time and resources available to the user (short term use vs long term use/full data view/data storage). Both methods have their pros and cons.

A summary of the options is presented in **Fig: 1.** Within this figure **A)** in both scenarios depicts the option of maintaining the data structure seen in p:IGI-3 in p:IGI+, while **B)** shows the route of formatting the data taken from p:IGI-3 to fit the structure used in p:IGI+.

. 

# p:IGI-3 Data Import: Maintaining the p:IGI-3 data structure in p:IGI+

Users might decide to import data directly into p:IGI+ from p:IGI-3 maintaining the p:IGI-3 data structure. An example of when this might be appropriate is when a user wishes to revisit a historical project, in light of new understanding, and a short extension to the project is needed, but the user wishes to utilise the new features in p:IGI+.

The advantage of this approach is import speed; the disadvantage is the maintenance of a limited p:IGI-3 data model which will persist into p:IGI+, namely:

* A restricted/narrow standard property model with little/no context for the analytical approach used to derive the data. Standard data for the same sample will be spread over multiple rows with little option to compare data from multiple analytical methods.
* Sample molecular/gas data will be positioned over multiple rows when reported in multiple indicators, and in the case of WO-GC data multiple analysis groups if the project contains more than one GC method, e.g. Sat-GC, WO-GC, HT-GC etc.
* An inability to cross-plot properties from different analysis groups during data QC and interpretation.

The result of importing standard and molecular data directly into p:IGI+ from p:IGI-3 are shown in Fig 1a: Route A and Fig 1b: Route A.

When importing p:IGI-3 data directly, each p:IGI-3 **core property** has a proposed linking to an equivalent p:IGI+ property, indicator, unit and ratio (where appropriate). The p:IGI-3 **user defined properties**, are linked using the import systems **property name & associated alias linking scheme and then by the user if unsuccessful.**

Several assumptions were made when IGI defined the proposed mapping of p: IGI-3's core properties to the new IGI property model. These assumptions can be accessed here: [www.igiltd.com/pigi3import](http://www.igiltd.com/pigi3import) *and should be reviewed before importing any p:IGI-3 file into p:IGI+.* The listed assumptions record changes in property equations, the default analysis groups to which values will be assigned, suggestions on how to deal with mappings which result in duplicate auto-links, etc.

**Import of p:IGI-3 data to p:IGI+ using this approach:**

**Either drag and drop your p:IGI-3 file onto the p:IGI+ workspace to auto-start the import process or from the Project drop-down menu select Import data 🡪 Import p:IGI-3 data…**

**Import Stage 1 -** **p:IGI-3 file overview:** a display of the first 100 rows of data with an option for the user to select whether to import **supported artefacts. Here IGI** advises to import just data and start new with graphs & palettes in p:IGI+. ***Only properties with data are imported from a p:IGI-3 file.***

**Import Stage 2 -** **Data linking:** a display of the first 100 rows of data, accompanied with the **property linking table** with the ability to apply/save [linking templates](https://www.igiltd.com/software/igi-software-user-guides/shared-user-guide-pages/importing-data/import-data-from-excel-file-2/linking-templates/) and apply [bulk edits](https://www.igiltd.com/software/igi-software-user-guides/shared-user-guide-pages/bulk-changing-properties/) to property analysis groups, indicators and units.

It is during this stage, that users have the opportunity to correct any of the initial core property links based on the metadata present for the sample, resolve any duplicate matches and fix the mapped linking chosen for the user defined properties.

# p:IGI-3 Data Import: Data formatted to a p:IGI+ structure

With the new p:IGI+ property model, there is the capacity to store sample information on a single sample row (where appropriate to do so). With this ability, the p:IGI+ property model now provides a full multivariate analysis picture for a given sample.

Exceptions to this are where repeats have been run on the same sample in the same lab; the same sample is analysed by different labs and where samples are child samples of a bulk parent sample, i.e. picked samples. In these cases, the associated data would be seen as different samples and placed on separate sample rows.

For users who wish to create either long term projects likely to contain a mixture of historical and newly generated data or projects involved with data storage, a decision to format p:IGI-3 data before importing it into p:IGI+, would be advised.

This approach's advantage is related to data storage accuracy and ease of use moving forwards, especially when adding new data to the project and conducting comparisons across data sources. The disadvantage to this approach results from the time needed to manually reformat project data and the potential errors related to such a procedure.

Examples of importing standard and molecular data that were formatted before p:IGI+ import are shown in Fig 1a: Route B and Fig 1b: Route B.

When deciding to reformat data prior to importing it into p:IGI+, the user has several options:

1) Export from p:IGI-3, manually reorganise & split data across multiple Excel worksheets and bring in each sheet individually. One or more project merge actions would then follow this approach.

2) Return to the source of the data, organise the data across multiple Excel worksheets and bring in each sheet individually followed by one or more project merges actions. The advantage here is there may be data in the original report/source, which wasn't captured into p:IGI-3 due to its limited property model.

3) Talk to IGI – we have an in-house tool to help build rule-based mappings using key p:IGI-3 sample metadata properties to enable data to be transformed and imported correctly into p:IGI+ in a semi-automated process.

None of the options listed above is quick, but with the appropriate time commitment and an eye for detail, the result is a more informative project database, that will provide more insight and worth to the end-user over the long term. The procedure involved in migration of data using this approach would involve:

1. exporting data from p:IGI-3
2. formatting the data ready for import
3. importing the data into p:IGI+
4. merging up the data to combine data across different indicators

The result is samples stored across a single p:IGI+ sample row.

As you can see the process to rigorously bring p:IGI-3 data into p:IGI+ can be complex. However, this should only be necessary once, and then moving forward, the data will be captured in greater analytical detail, providing the means for improved interpretation. Also the data in p:IGI+ will be laid in such a way that it will be compatible and comparable with additional new data loaded to the p:IGI+ project or imported from a centralised database.