“Full update” Guidelines

The main objective of the Land Matrix is to compile a comprehensive and consistent dataset including the most up-to-date information and it´s corresponding data sources about all large-scale land acquisitions (deals) meeting the Land Matrix deal criteria (https://landmatrix.org/faq/#what-is-a-land-deal) and targeting low- and middle-income countries according to the World Bank country group classification as of 2010.

The two primary data quality control procedures used by the Land Matrix to ensure that the information recorded is as comprehensive, up-to-date and accurate as possible and consistent with the data sources are:

1. The data review & activation workflow following a three-step sequence:

Step 1. Data creation or modification by reporters, editors or administrators

Step 2. Data review by editors or administrators

Step 3. Data activation (final confirmation & approval + publication) by administrators.

At least two different persons must be involved in this data workflow (4-eyes principle), but the data administrators are the ones who finally confirm, approve and publish the data for their region.

2. The “full update” (“fully updating”) -procedure.

What is a “full update” of a deal?
A full update is both a systematic update as well as a systematic repeated review process of a deal and consists of two main tasks to be completed:

1. Thorough data sourcing for additional or more up-to-date information.

2. Thorough and complete review of all the existing data recorded for the deal.

Thus, a recently fully updated deal should contain the most accurate, complete, and up-to-date data that the RFPs/NLOs are able to obtain and the label "Fully updated" is a data quality indicator manually assigned by the RFPs/NLOs to a deal after that complete data review & data sourcing process. Please indicate roughly in the comment field before you submit for review what you have done.

At what frequency must a deal receive a “full update”?
[As mutually agreed, all deals must be fully updated by the RFPs/NLOs at least every two years.]

This applies in particular to concluded and intended deals. Data of failed deals might for the most part be excluded from this procedure if they have already received a "full update" once in the past. (There is of course the problem of expired data source URLs in failed deals too, but for URLs, as discussed earlier, we need to find a solution other than constantly checking them.)

If you are looking for prioritization, I recommend working your way from the larger to the smaller deals, prioritizing agricultural deals and then forestry deals, later expanding to the default filtered data subset before fully updating all the rest.

What data needs to be reviewed for a “full update”?
In general, all data of the deal must be reviewed, and, of course, corrected or supplemented if incorrect or missing.

In case the deal is a subsumed deal (see “Guidelines for data entry” – version 2.5 – 1.5 Subsumed deals – page 15 / “Guidelines for data entry” - Version 2.99a - 5.2 Subsumed deals - page 23) one of the first steps should be to check if additional information can now be obtained that would allow the subsumed deal to be split into individual deals.

If it is still not possible to split the subsumed deal, a note must be created in the overall comment field: "Subsumed deal".

Special attention needs to be paid to the following variables, where data quality has partly proven to be problematic in the past, even after a "full update":

1. Data sources

1.1. Make sure that the data source file exists and contains the appropriate information.

Each of these files must be opened and the contents checked without exception. Please remember that for each individual data source such a file is mandatory (apart from a few exceptions in the case of “personal information” or “crowdsourcing”).
(To create the data source file from a webpage, use either the printer driver of your browser to “print” a pdf-file or the screenshot feature of your browser to create an image of the webpage.)

→ If the data source file does not exist or is completely corrupted and cannot be retrieved, restored or regenerated (please check the Wayback Machine (archive.org) too), also because no equivalent

new URL can be found, this data source is invalid and should be deleted in this and only in this case.

→ If the data source file was downloaded from the internet, please remember that our URL ideally leads to the file's download page and not directly to the file.

1.2. Make sure that the URL is still working.

→ If the URL has expired and cannot be replaced by an equivalent new URL, please copy the URL into the comment field of the data source together with a note that the original URL has expired and no new address could be found ("Original URL expired: http://www..."). The expired URL must then be deleted from its field.

1.3. Make sure that the date entry exists.

→ Please remember that it should primarily be the (1) creation date or (2) the publication date of the data source. Only if such a date does not exist, it should be the (3) access date - i.e. the date on which you created the PDF or screenshot copy of the web page and saved it as a data source file.

1.4. Please make sure that you have filled in the "Publication title" if it exists.

1.5. Please remember that the comment fields of the data sources ideally should roughly indicate which information has been taken from each source and where to find it in the data source file (e.g. no of page, no of table etc).

2. Location

2.1. Please check whether the geodata can be made even more accurate, especially if there is not yet a high spatial accuracy.

2.2. Please make sure that the point position and the polygons usually match. The point should be either on the main building or centrally located in the contract area. As already discussed in our Data Coordination Meetings, deviations from this may occur in certain constellations.

Please note your reasoning for deviations briefly in the comment field.

2.3. Please make sure that the accuracy level is correct.

3. General Info

3.1. Please try to make sure that there is data for the key variables under land area size, intention of investment, negotiation status, implementation status and nature of the deal. Also do a cross-check here with the data under produce info.

3.2. Please make sure that the year-based data (see “Guidelines for data entry” – version 2.5 – 1.6 Year-based data – page 16 / “Guidelines for data entry” - Version 2.99a - 5.3 Year-based data -

page 24) within the variables size under contract, size in operation, intention of investment, negotiation status, implementation status and under produce info is consistent and plausible. If possible, add further time details.

3.3. Please remember that the negotiation status "change of ownership" is only a temporary status and should not remain so.

4. Investor Info

4.1. Please check if there are unnamed or “unknown” investors involved and if so, try to find out and add the real company names. Please make sure that this does not create additional investor duplicates. As discussed, it is highly advisable not to rely on the search functions of our website for this, but to use the Excel search function as demonstrated in the Data Coordination Meeting.

4.2. Please check for missing information.

4.3. Please remember that, with a few exceptions, the operating companies are usually based in the target country of the deal.

5. VGGT scoring relevant variables.

5.1.

https://docs.google.com/document/d/1oAjvPLNmeU5Xu5wlBJNtmjO84fuLWC29txchV7Bwheg/edit#heading=h.1w6bgtgm5ydr

As always, if you have questions or need support, please bring it up in our data coordination meetings. You can also reach me on MS Teams or send an email to Christof.Althoff@giga-hamburg.de.

Created 2024-10-29 08:54:16 UTC by Christof
Updated 2024-10-29 08:55:26 UTC by Christof