
Data cleaning checklist for CGIAR Ongoing Research

1. Remove duplicate projects. Duplicates: eg, same title and two different lead centres
2. Fill up incomplete data, overview, research partners, missing tags, external links
3. Update: remove scientists who have gone
4. Check / verify / notify the Principal Investigator of any major addition or change
5. Clean the spelling of names and tags. Remove misspelled tags or reassign the most popular ones. Avoid country names and same labels as research areas, names of scientists, acronyms
6. Overview:
 - a. clean the formatting
 - b. paste clean with paste from word
 - c. source text from the proposal and change the tense, have the scientists or PI review it for completeness
7. External links
 - a. give it a title that reflects the title of the Web site/page
 - b. complete URL, please test if it works ok in preview
 - c. Collection of outputs in Mahider (ILRI Repository): project titles used to create collections of outputs, each collection has a permanent URL (but with a query). Each collection has a RSS feed.
 - d. Grants database: not an ideal source of outputs, because there may be information that is confidential or for internal communications
8. How to select partners (conventions)
 - a. CG center: full English or Spanish name (acronym in brackets), remove donors from the research partners if you have entered them here
9. Workflow:
 - a. Verify info with the PI and scientists
 - b. Invite them to contribute for preview and or edit
 - c. Introduce them to user profile and PI/scientist list pages. Ask them to provide brief bio profile and a link to a more complete profile.
10. Check available partners, otherwise request new term to be added via this form:
 - a. <http://shortlinks.cgxchange.org/request-partner-entry>
11. Titles not in CAPS

Created: 9 November 2010

Updated: 10 November 2011

For additional info and updates, please contact cqmap-support@cgiar.org

