**Automated Cleanup and its Challenges**

There are two ways to go about data cleanup and each applies to different situations: Manual and automated processes. There are tradeoffs either way, but in general, we try to avoid manual processes if we can. They are time-consuming and leave more consistent room for human error such as typos, losing one’s place, etc. By contrast, automated processes are faster, more efficient and generally more accurate; however, whereas each mistake in a manual process will typically affect only a small part of the database, a single mistake in an automated process can make a huge mess of one’s records. The mitigating factor is that in an automated process, there are usually fewer opportunities for making mistakes.

Large-scale cleanup through automated processes can be likened to an elaborate domino course: When you set it in motion, the action happens so fast, but it’s easy to forget that there is a great deal of setup and preparation that goes into making this possible. Also like a domino course, once you set it in motion, you can’t stop it. This is why it is so important to be painstaking and careful before running something like an Import or Global Change. If something goes wrong, we may fix one particular error in 5,000 records, but then generate errors in 1,000 of them because of a setting that was overlooked. It is very easy to ruin a data set by being hasty, neglecting to double-check everything, or failing to take certain factors into account before execution. With backups, Queries of records changed and other safeguards, we can recover from such mishaps; however, the recovery takes time and is best avoided. It’s better to spend 2 hours planning a Global Change, even if we’re just reviewing it several times, than to crank one out in 15 minutes and cause an accident that takes 3 hours to clean up.

**The Priority Paradox**

Data cleanup is important because so many other processes depend on clean data. Rarely is a data set so horribly disfigured that an organization can’t limp along from day to day without cleaning it up, but doing so is a huge waste of staff resources. In a perfect world, data cleanup would be priority #1 since clean data is the backbone of everything else we do that involves Raiser’s Edge. Despite this, data cleanup typically becomes a “back burner” job that happens when nothing else is going on. For one, it’s usually not time-sensitive and everything else typically is. In addition, for reasons stated above, it’s a task that requires a lot of focus. One may start a cleanup project, but then be interrupted by something else and lose their place, leading to a mistake that ruins the entire project and takes longer to fix than the original problem. Thus, before we set about working on a cleanup task, good RE: admins try to make sure our plates are clean so nothing gets in the way. It may make data cleanup a frustratingly slow process, but if we don’t do things this way, we risk making a bad situation worse. If some aspect of cleanup needs to be expedited, then in turn, we need to be able to set aside chunks of time to focus on cleanup to the exclusion of all else.

Nobody wants a clean system more than your database administrator, because that system is our work environment. Good admins do this sort of cleanup in their spare time without being specifically asked, for the same reason you’d want to keep your desk organized. We are sometimes very eager to get a cleanup task done, feeling it every day like an itch; experience and hard lessons have taught us that the only sensible way to proceed is carefully and patiently.