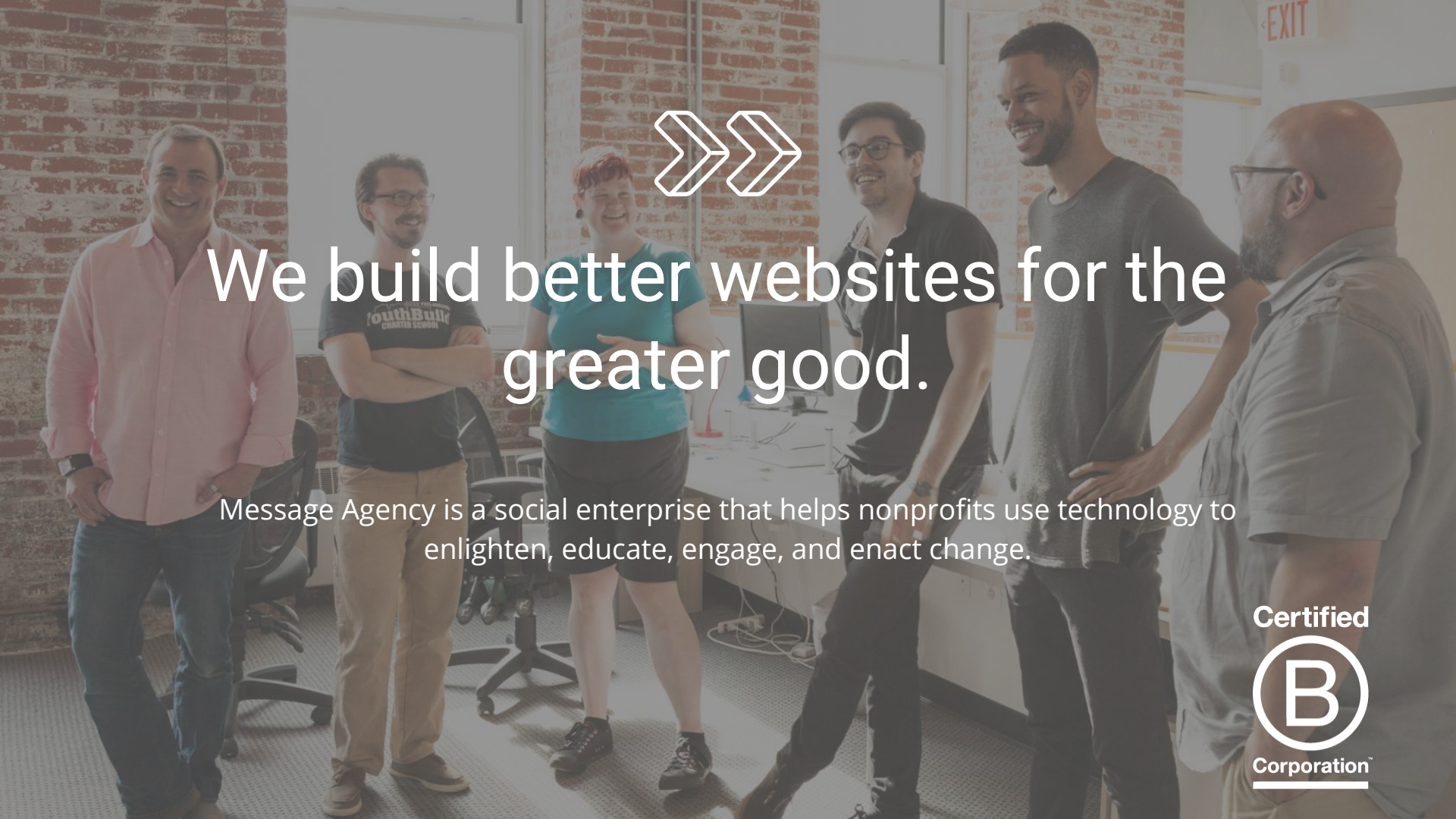
On Moving Mountains:

Lessons for Migrating Data and Content for your Next Drupal Project

MESSAGE AGENCY

Drupalcamp NJ • February 2, 2019



We build better websites for the greater good.

Message Agency is a social enterprise that helps nonprofits use technology to enlighten, educate, engage, and enact change.

Certified B Corporation™

“Hindsight is 20/20.”

“Hindsight is 20/20.”

Considering migration too late?

- Cost Overruns
- Blown Schedules
- Redesign and Rework
- Poor Change Management
- Outdated Documentation

Mistakes We Make

- Waiting until the site is built to start planning the migration.
- Ignoring key source data or content when designing architecture.
- Assuming consistent quality of source data.
- Reproducing legacy architecture in your new solution.
- Underestimate effort and resources required for data transformation and cleanup.

Begin at the Beginning

- Plan for migration on Day 1. It cannot be an after-thought.
- Arrange for access to source data at your kickoff.
- Consider migration explicitly in your project plans.
- Make sure you schedule adequate time to import, validate, refine scripts, and re-import if needed well before launch.

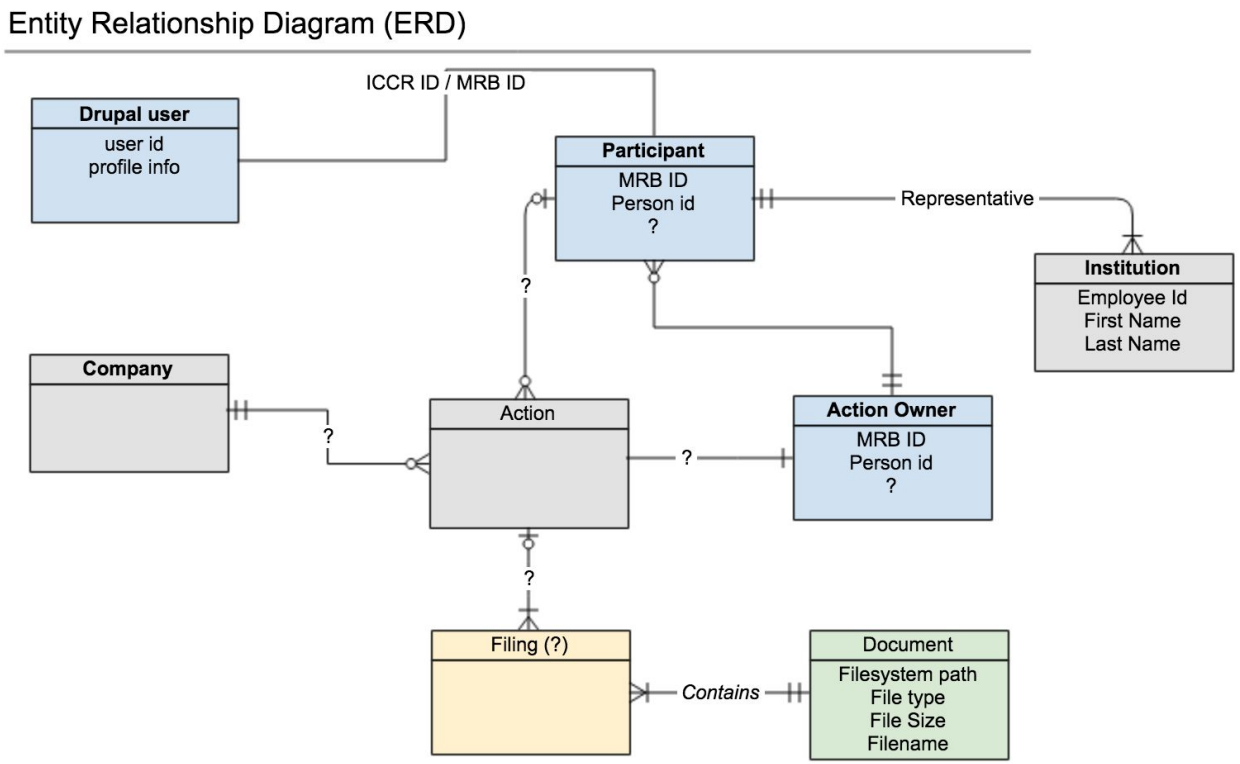
Consider the Source

- Have content strategists conduct a content inventory and developers do data audits.
- Go to the source: find the person responsible for the schema and unearth documentation.
- Identify the outliers first.
- Prioritize what needs to be migrated, truncated, or deprecated.

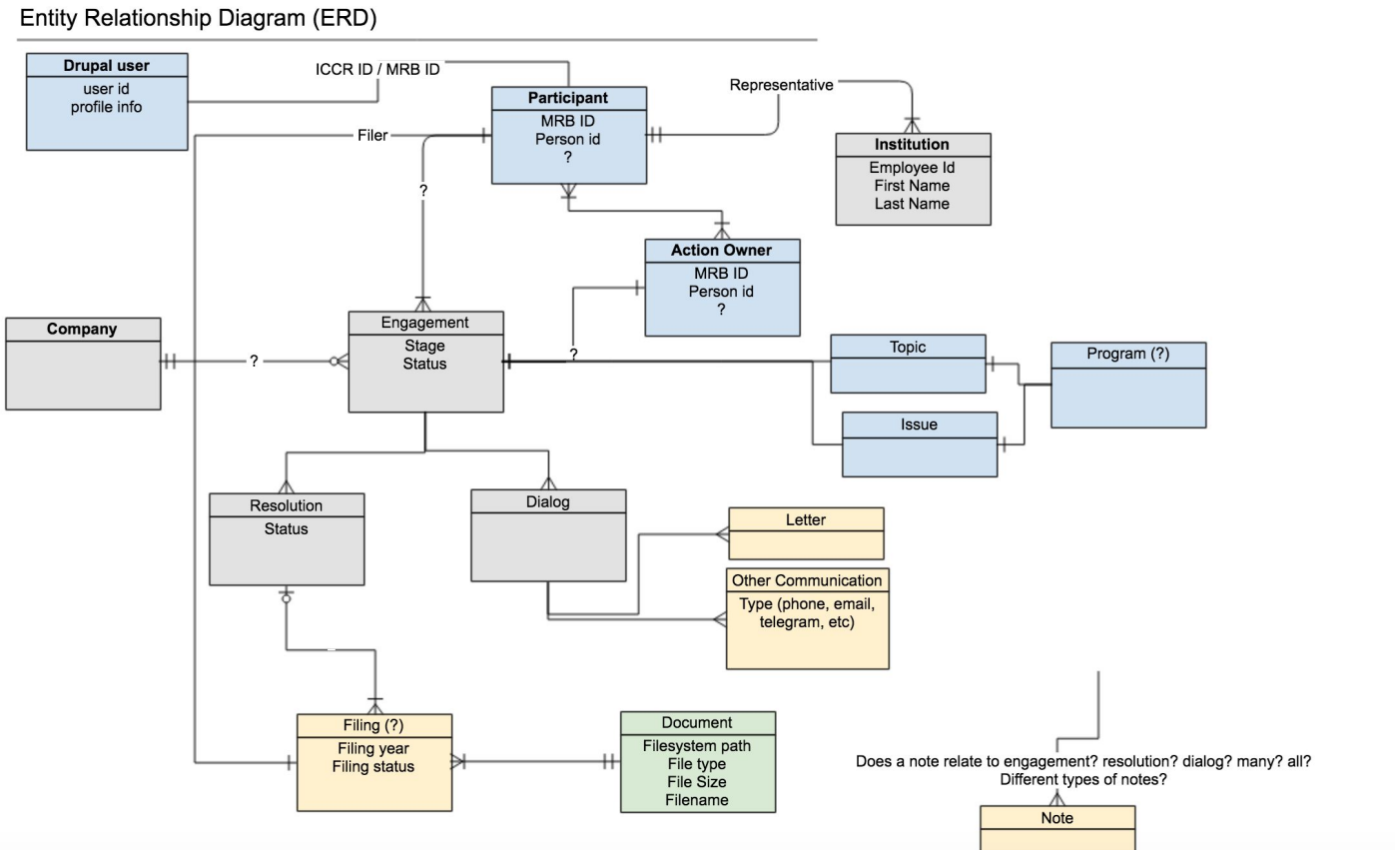
Some Things to Evaluate

Can you standardize fields across content types?	How atomic is the data? How are dates handled?
Are there attached images/media/files that need to be migrated?	Are there timestamps or string values? What's the source encoding?
Is the taxonomy well-structured? Do vocabularies cross content types?	How will data be transformed between source and target?
Are there page elements built as blocks that we need to incorporate?	Do the source and target speaking the same language in terms of data structures?
How are the body fields? Plain text? Are there embedded styles and markup in Rich Text Edtors?	Is data duplicated in places? If so, which is authoritative
Are there inline images and video?	What data on users is stored and needs to come over? What's the communication plan for deployment, e.g. resetting passwords?

ERD (current)


Entity Relationship Diagram (ERD)
ICCR ID / MRB ID
Drupal user
user id
profile info
Participant
MRB ID
Person id
?
Representative
Institution
Employee Id
First Name
Last Name
?
Company
?
Action
?
Action Owner
MRB ID
Person id
?
?
Filing (?)
Contains
Document
Filesystem path
File type
File Size
Filename

ERD (proposed)



Design for Migration

- Normalize, but let the structure of the source data help inform your design decision-making.
- Try to limit the need to transform large amounts of data.
- Map objects and fields as you go.

Migrate: Manual

- Start planning as soon as the wireframes (waterfall) or feature (agile) is approved.
- Use a tool to track what moves, what needs work, what dies—and who is responsible for the work.
- If the work is distributed, distribute the knowledge.
- Prepare to transform, not copy-paste.

Migrate: Scripted

- Start as soon as the schema is ready..
- Groom your source data, if possible, to reduce the burden on migration scripts.
- Conduct a test import once the architecture is in place and validate the results.
- If you are using Migrate API, you can continuously synch the data. Otherwise, migrate an initial set and then again immediately before launch to capture new and updated data.

Validating the Migration

- UAT is the right time to validate the migration, particularly with clients.
- Having stakeholders post content as part of UAT solves a lot of problems.
  - They test the admin interface.
  - They test the content governance workflow.
  - Their real-world content will stress test the design—during a time that changes are still possible, not immediately prior to launch.

kickoff discovery design development uat launch

Uneventful Launches

Tools You Can Use

Tool	Method	Description
Good ol' Spreadsheets	Manual	Helps structure content; lots of person hours; recommend using Google Docs for collaboration
GatherContent	Manual & Automated	Mimics Drupal content structures; module pushes content from platform to Drupal, but doesn't account for paragraphs or other features
Content Import	Automated	Lightweight import tool for turning spreadsheets into nodes, users, taxonomy, etc. Good for relatively simple, one-shot migrations.
Feeds Module	Automated	Another lightweight import tool for varied source data. Low barrier to entry, but very rigid. D8 module is alpha.
Migrate API	Automated	Very robust, relatively high barrier to entry; synchronizes sources
Custom Migrations with Drush Scripts	Automated	Requires writing PHP scripts, but provides maximum flexibility

Level of Complexity

Wrap-up: Lessons Learned

- Know the audience and design for the 80%
- Cut your losses: don’t design around emotional attachment, don't recreate the same mistakes
- Don’t force Drupal developers to become business analysts
- Do include Drupal developers in migration planning
- Don’t assume the same approach will work for every project.

Thank you!

marcus@messageagency.com
aaron@messageagency.com