References Report Use

References

I've come up with a report format for the references that will help get through their clean up and linking. Sample attached here. Columns:

- *Year.*
- *Note ID.* For mapping purposes.
- *Reference String.*
- *Type.* Internal or external depending on whether it points to something in MB or outside.
- *Mappable?* Indicates whether this link can be made programmatically (some assumptions made).
- *Split Check?* Indicates whether the reference string may point to multiple targets, thus need to be split.
- *Malformation?* Indicates that the format of the reference string may not comply with the normal conventions, e.g. most internalreferences start with 'MB'

The last three implement rules based on analysis of the data and what's possible with automation. I will--if we proceed with some other prerequisite work first--add a reference id column to this to make life easier mapping in links.

Using the report:

- Everything marked Y under Split Check? needs to be reviewed and corrected in the data if nec.
- Everything marked Y under Mappable? Can be left alone and the links made programmatically later (after some other processes are run, see below).
  Everything marked 'N' will need a manual link unless David comes up with some kind of index we can use to mark them programmatically. I checked the Rotunda site and found nothing I could use for this. The link target can be entered into the spreadsheet and I can map them back into the data.
- Everything marked Y under Malformed? needs to be reviewed and corrected in the data if nec.

Good news is that all internal links can be made programmatically and they constitute a big percentage of references.