

Meeting Minutes

Andy Jarvis

SESSION II

Talk by Thomas Metz on Taxonomic nomenclature

Two principle taxonomic databases (GRIN and Mansfield), and significant differences between the two (up to 20% disagreement). Taxonomies constantly being updated and changed.

Important not to throw away taxonomic determinations in histories. Sometimes resolving taxonomic identification subjective and based on expert decisions.

Discussion:

Can you also check for authority on taxonomic names?

Not in the current system, but you only need to generate the dictionary to which you compare. Could be used not only for authorities, but also gazetteers, habitats etc.

TvH: We should use this workshop to also suggest new functionalities needed for improving data quality. For example:

Author checking on authorities for species

Wish list:

Checking genebank codes and addresses

Dictionary of crop names for standardization of crops

Author checking on species names

Gazetteers (may exist in BioGeoMancer)

Translating local codes into FAO codes for genebanks

AJ: Could use systems to assign uncertainty on taxonomic identification. Fields should also be included on the taxonomic system used to identify the accession.

TM: For habitat, system helps deal with spelling mistakes, but underlying problem is the generation of a dictionary for habitats.

CL: Is there a dictionary of cultivar names?

HK: Some databases exist for breeding lines. But nothing comprehensive exists.

FA: Recommendation of workshop will outline priorities for new functionalities in tools. Might also include in technical manual as a section.

TvH: Dictionaries already exist in many databases. For example in Mansfeld. Need is for implementing in a tool like the spellchecker. Not a major undertaking.

Presentation by Fred Atieno

TM: Major attraction of R in the graphic interfaces. Might be easier to sell with graphical output.

TvH: What is the vision of using R to increase quality?

FA: Provision of packages that address specific quality issues, like outliers. For example, in spelling of locality names.

TvH: Questions whether this is the most appropriate approach to identifying outliers.

TM: Costs of using R high compared to benefits when only looking a passport data, but R does provide big opportunities for analyzing other types of data and therefore provides potentially greater benefit.

AJ and EA: DIVA-GIS offers a more user friendly interface for many of these analyses.

TH: Questions what kinds of analyses are made on passport data? Most obvious are frequency distributions.

TvH: Probably very few analyses on passport data only.

Coffee

SH: Not worthwhile to worry about translating FAO codes. Too complex.

SD: There is a process for requesting an FAO code.

TH: One output of workshop could be to suggest and streamline the FAO code requests.

TvH: Frustrated by FAO code issues, and weary of going into it.

TvH: Challenge: table has rows representing objects, and columns with properties. Something that predicts the value in each cell based on similarity matrices. This then used to identify values worthy of flagging.

TH: EURISCO could be input to such a system. But would it add value?

AJ: Possibilities include the use of self organizing maps, and Kohonen analysis

TvH: What is the technical manual going to look like?

FA: Manual would provide suggestions for minimal passport data requirements, plus other guidelines for more detailed passport information.

TvH: How will manual be written?

FA: Input from workshop will feed into process, but a team will actually write the manual. Synthesis of external references would be necessary, but should be developed jointly as opinions and knowledge is diverse.