Ignore the ontological aspects of land cover at your peril: a plea for expanded metadata

Lex Comber[1], Pete Fisher[1], Richard Wadsworth[2]

[1]University of Leicester

[2]Centre for Ecology and Hydrology

Revigis Project EU\IST-1999-14189

University of Leicester

Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL

ADAS

Background

- EU REVIGIS project
  - Looking at the revision of GI
  - Many EU partners, applications - e.g. LCMGB and LCM2000
  - Applying reasoning formalisms
- ‘Revision’ defined by computer scientists
  - New information allows update
  - BUT this is not that simple for much GI including land cover
  - Integration of discordant data by modelling expert knowledge
- Identified change from LCMGB and LCM2000
  - Expert approach (stats no good!), series of papers
- Origins of data discord & inconsistency
  - Ontologies, semantics, conceptualisations, actors & networks

ADAS

What is change ?

- A subset of inconsistency between two signals
  - artefactual data differences (thematic or raw)
  - May be due to actual differences on the ground
  - So any signal of actual change must be separated from the noise of data difference (in processing, classification, etc)
- For some landscape processes e.g., geological
  - the interval between surveys is small relative to the timescale at which changes occur
- For others, such as land cover
  - Timescales are shorter & differences in methodology and thinking (O word) may obfuscate any signal of change
- This hinders development time series models monitoring & identification of locales of change

ADAS

Outline

- Sideswipe at land cover
  - Known GI-wide issues, but illustrated by land cover
- Impact of these issues on data integration
  - Particularly change detection, monitoring, forecasting
- Relevance now
  - Many users, SDIs, land cover used in many disciplines
- Bemoan current metadata standards
  - Users given no guidance
- Suggest some objectives for metadata
  - Questions about how to achieve them

ADAS

Land cover Variation

- Differences in land cover surveys
  - Technical aspects: structure, scale, resolution, input data
    - MAUP / MMU
    - Reflectance changes at different scales – leaf, branch, canopy, woodland patch
  - Our understanding changes – scientific developments
  - Political development – new mapping objectives
  - Changes on the ground – the thing we are interested in identifying

ADAS

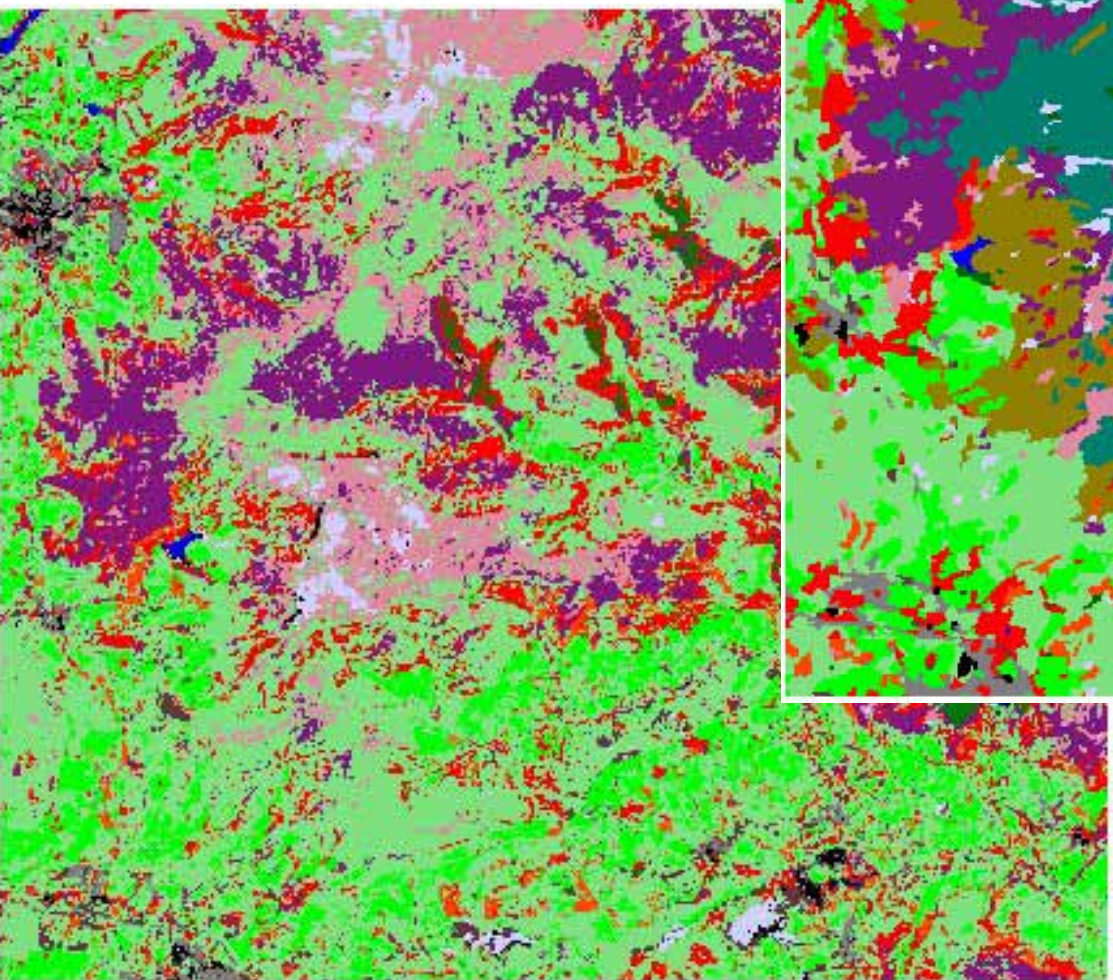
UK land cover mappings

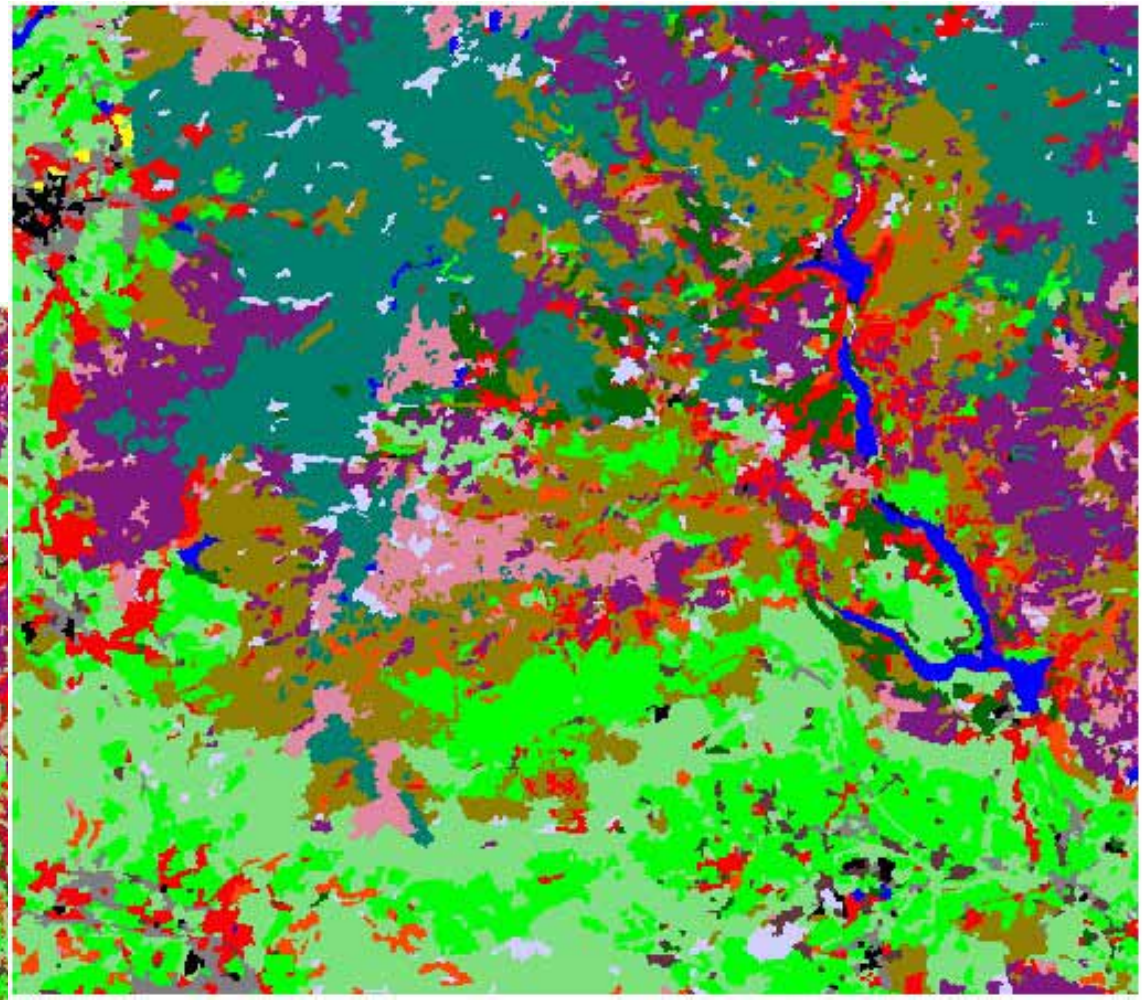
- LCMGB – raster format, 25 Target classes
  - Classification developed by scientists
  - A demo of the capability of SI
- LCM2000 – vector format, 26 BH classes
  - Tech developments (parcel, extensive metadata)
  - Policy developments
    - Habitats directives → UK BAPs
    - Stakeholder involvement → negotiation
- Different representations, concepts, objectives
- Not a criticism of LCMGB / LCM2000
  - Illustrate the issues

ADAS

Representation

LCMGB 1990

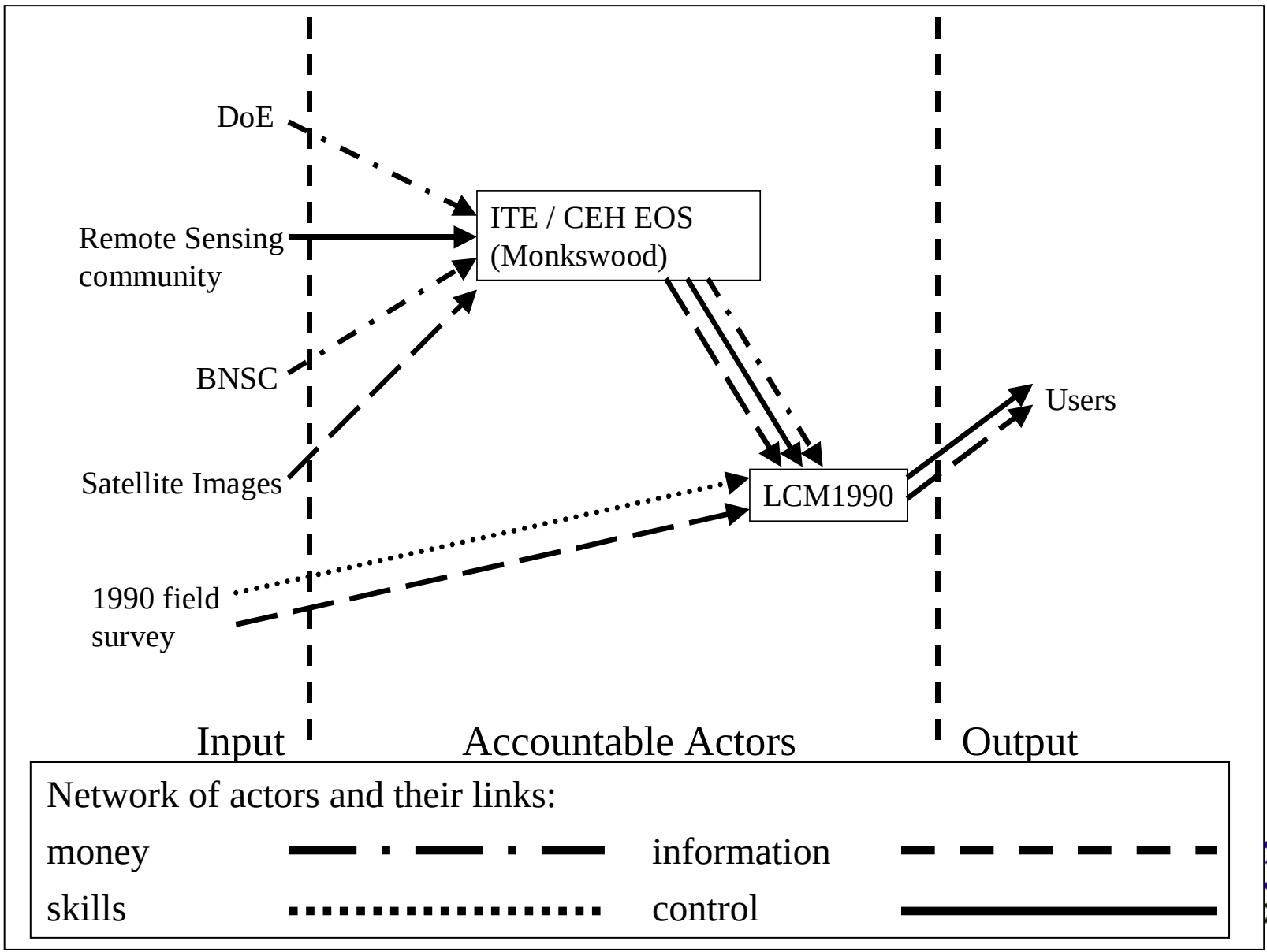




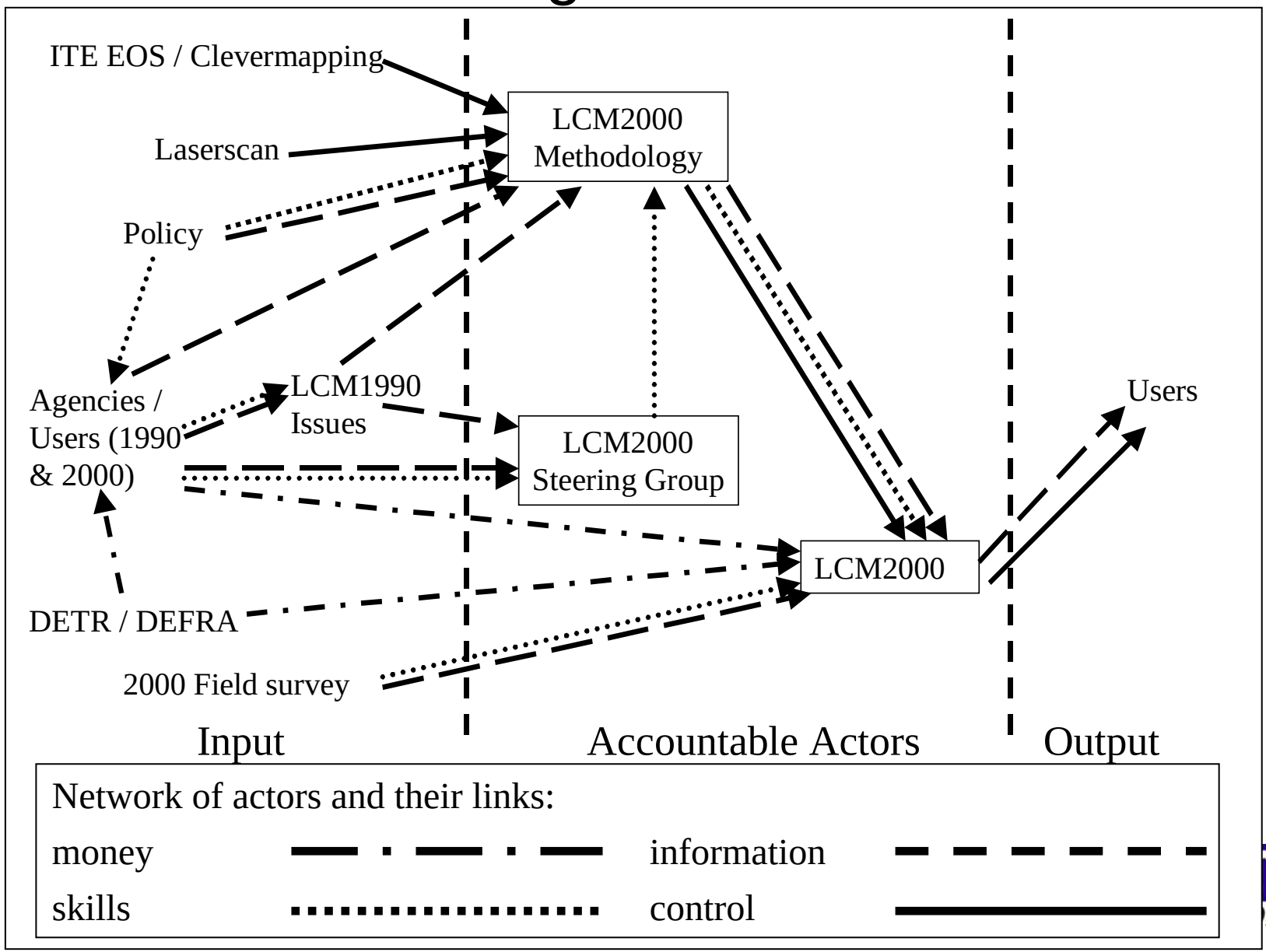
LCM 2000

ADAS

Commissioning context: LCM1990



Commissioning context: LCM2000



Politics & socio-economics

- Land cover is not neutral
  - Land cover as a political tool
  - Treated as fact by users
- Land use is even less neutral !
  - Explicitly socially constructed
  - Few uses have a single cover
- Multiple uses that overlap in time and space
  - Simultaneous - patch of forestry, used for several forms of recreation, and for grazing
  - Alternate - forest → grazing then hunting; a reservoir → flood control (spring), hydro-electric power (winter), fishing (in season) recreational boating (all year round)
- Compared to traditional classification (non-overlapping classes, Boolean assignment)

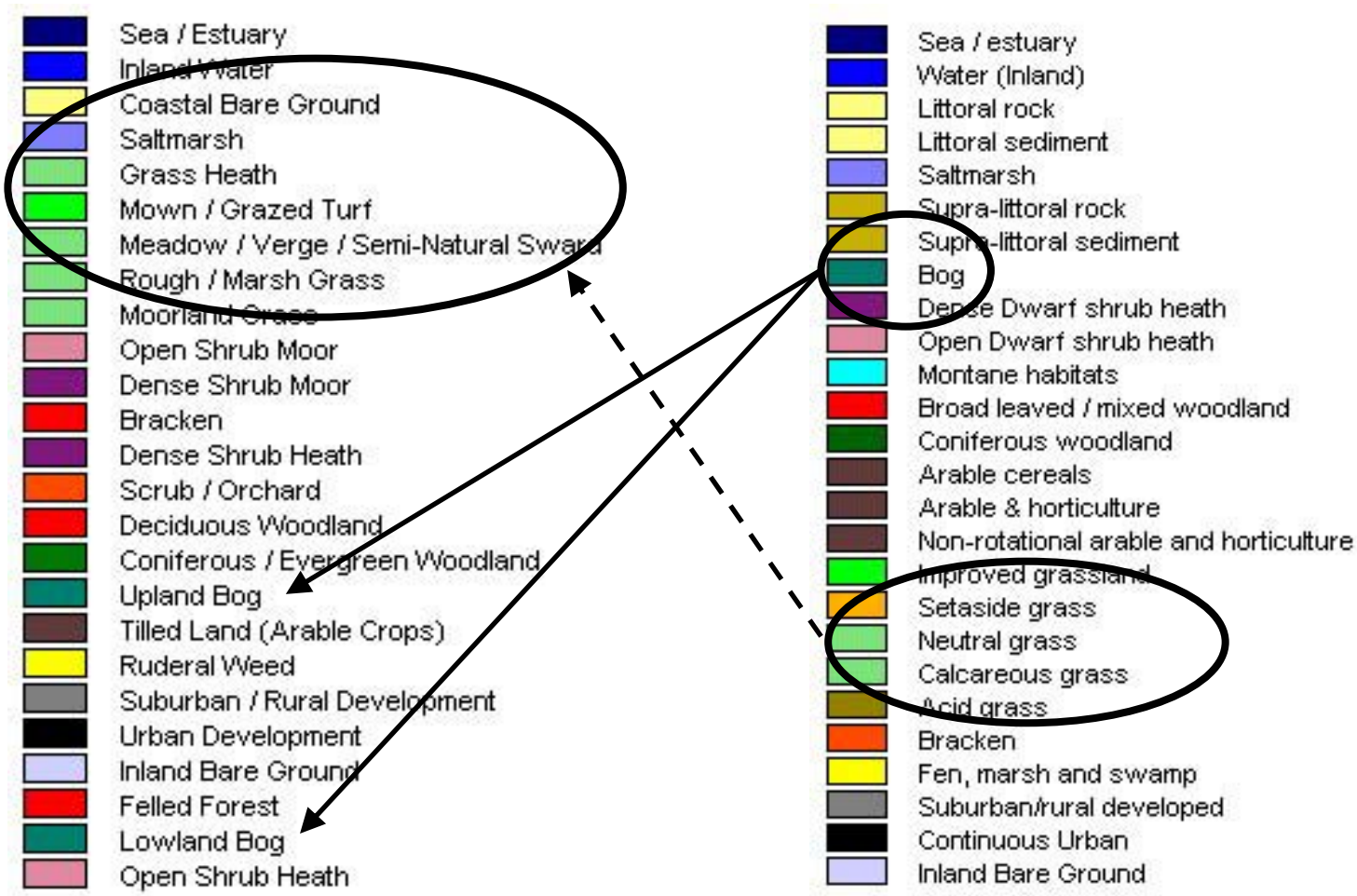
ADAS

LC / LU confusion is an issue

- Hoeschele (2000)
  - shown how confusion disadvantages subsistence farmers (grazing) vs. large land owners (forestry)
  - Differential land cover analyses (by agents of the state) as a tool for social repression over the locally defined land use
- Robbins (2001)
  - documented differences in the concept of forest held held by local farmers (grazing) and state foresters (performance)
- Land cover class definition a political exercise:
  - It determines what is recorded as existing
  - Influences future management decisions
- Makes data integration difficult (e.g. change analyses) – many : many (LC : LU)

ADAS

Concepts



LCMGB 1990

LCM 2000

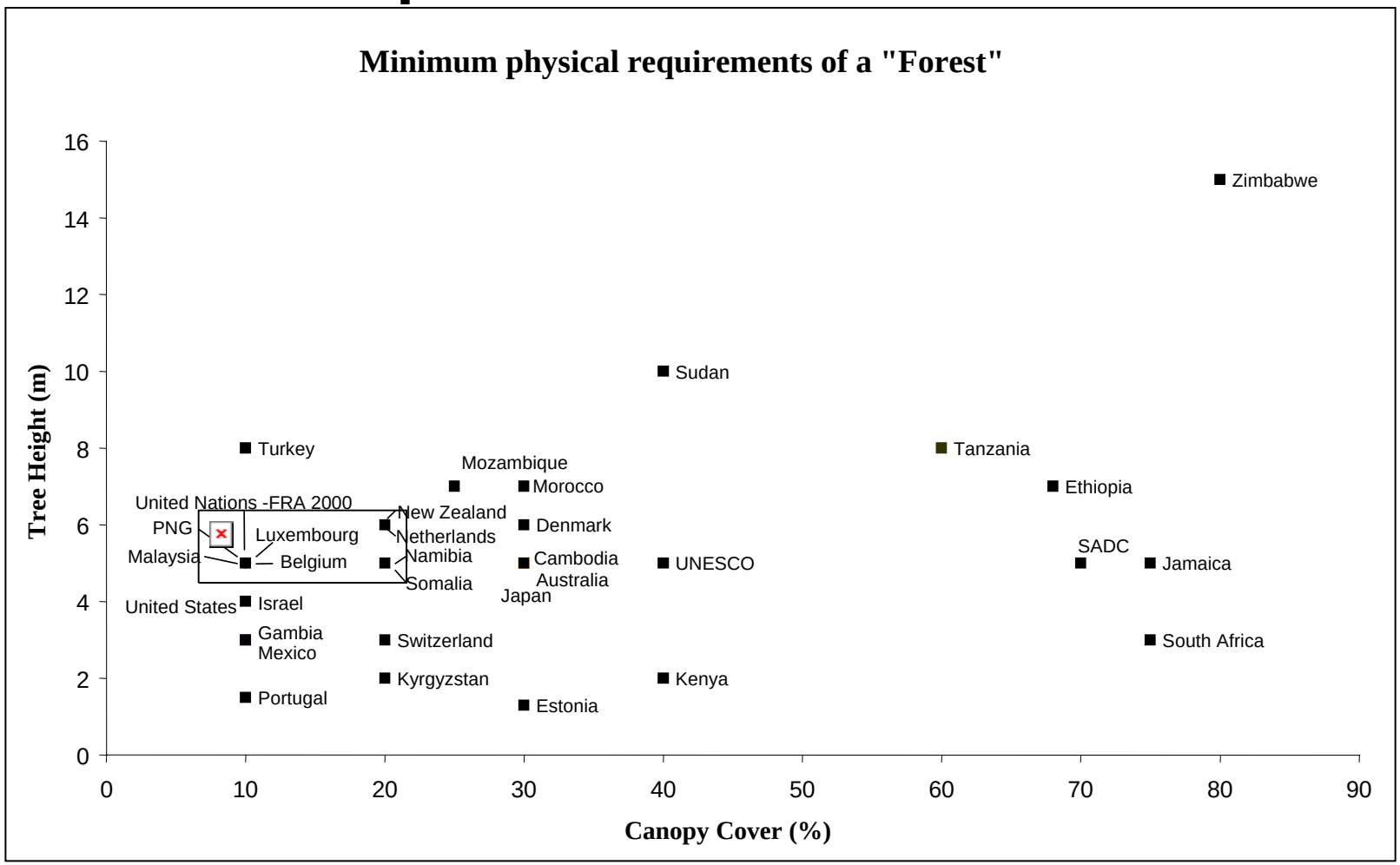
ADAS

Concepts: what is a bog?

- 1990: 12 pixels (<1 ha) in SK tile
  - permanent waterlogging, resulting in depositions of acidic peat
  - mostly herbaceous communities of wetlands with permanent or temporary standing water
  - Lowland Bogs: carry most of the species of upland bogs, but in an obviously lowland context, with Myrica gale and Eriophorum spp. being highly characteristic.
  - Upland bogs: have many of the species of grass and dwarf shrub heaths
  - characterised by water-logging, perhaps with surface water, especially in winter. species such as bog myrtle *(Myrica gale)* and cotton grass (*Eriophorum* spp.) in addition to the species of grass and dwarf shrub moorlands.
- 2000: 120728 pixels (75 km^2) in SK tile
  - Bogs include ericaceous, herbaceous and mossy vegetation in areas with peat >0.5 m deep; ericaceous bogs are distinguished at subclass level. Inclusion of Ericaceous bogs contrasts with LCMGB 1990 where bogs were herbaceous or mossy in seasonal standing water

ADAS

Concepts: what is a forest?



- Does not include species, area, strip width
- http://home.comcast.net/~gyde/DEFpaper.htm

ADAS

What's the problem…

…we are all aware of variation between data

- BUT conceptual noise obfuscates signal of change
- Increased number of GI users
- SDI's: GRID, INSPIRE, E-Science → 3rd party users
- Increased use of land cover information
  - Surrogate for a range of landscape processes
  - Different scientific disciplines (ecology, landscape ecology, land management, soil mapping)
  - They impose their own disciplinary understanding of the concepts
  - They are ignorant of them / ignore them

ADAS

Relevance

- Conceptual differences between user and data → Uncertainties
- More significant than attribute and location errors
- Metadata has a role here
  - Decreased role of metadata (Fisher 2003)
  - Users interested only in the (digital) map
  - Contrast with Coleman, Stamp

ADAS

Current Metadata Standards

- Current standards (ISO, OGC) are inadequate
  - report big 5: the easily measurable
  - shows the producer can follow a recipe
  - don’t convey the meaning (i.e. applicability) of the data to users
  - conditions users to considering only aspects related to those standards as being important

ADAS

Current Metadata Standards

- Current metadata standards do nothing useful for anyone wishing to use the data
- Most metadata and all quality statements are only in the interests of the producers
  - Who can can say they published the quality info
  - Up to the users to assess whether it is appropriate for use from that info

ADAS

Current Metadata Standards

- ISO is a money-making organization
- Their standards have built-in flexibility: ‘Profiles’
- Allows for any organization to adapt the ISO standard to their needs
- Will result in 100s of different profiles that share core elements but can be fundamentally incompatible
- In a few years we'll likely see lots of profiles claiming to be ISO standards
- The word ‘standard’ may ignored or despised

ADAS

Expert approaches can help

- Our integration work: LCM2000 & LCMGB
  - used expert (user) and object level metadata
  - LCMGB ↔ expert relations ↔ LCM2000
- Identified changes and inconsistencies
- Expert relations & metadata as a model

- Out of the range out even of the imagination of the standards writers but are clearly quality metadata!

ADAS

Improved quality metadata

- Aim to increase understanding of any dataset:
  - What is the conceptual model of the process / phenomenon?
  - Who determined the classes?
  - What was the process of negotiation?
- Different groups in the data commissioning
  - scientists, politicians, stakeholders
  - Each group has different interpretations of *what it is that is being modelled*
- Can be explored by considering actors and their networks

ADAS

1. Specify ontologies

- Can metadata describe the ‘ontological pedigree’ of our data?
  - The particular conceptualisations?
  - The who and the why ?
  - The way that the objects are measured (the epistemology) ?
  - the way that the classes are identified (semantics)?

ADAS

2. User assessments of data

- Descriptions of whether the data is suitable for use in *their* application.
  - Does it fit their conceptual model?
  - i.e. to what extent is it consistent with their model?
- *User* metadata
  - As data is used, the metadata (experience) of different users is available to other potential future users.
  - Any user-orientated specification should facilitate the resolution of differences, or discord
  - Other users decide which metadata contributions they want to use in their evaluation process

ADAS

3. Expand logical consistency

- Expand Logical Consistency reporting to include user metadata?
  - Include descriptions of how well the data fits user, expert or machine learnt models.
    - (Remember lack of geographic truth: many different and equally valid interpretations of same raw data possible)
  - This is what we do in accuracy assessments e.g. field survey to validate a satellite derived land cover map – testing the model

ADAS

BUT…there are issues

- How to describe the user assessment in a transparent and consistent way?
  - Inventing another standard ?
  - bound to miss the mark – cannot identify every conceivable, as yet unknown future potential use of the data
  - Users create their own metadata which is different than standards.
- How do we identify what is 'common' to metadata uses and creation
- How can a 'standard' be made robust enough and flexible enough to support user-created metadata?

ADAS

Back to land cover change

- REVIGIS work: data revision, integration of LCMGB & LCM2000, land cover change
  - Inconsistent to each other
  - Represent different paradigms of looking at the world
  - Expert approaches developed
- Wider context of data to understand it

ADAS

Metadata recommendations

- Needs to be put in the hands of the users, not the producers
- New classes of metadata are needed
- New forms in which metadata can be reported
- New mechanisms for establishing metadata are needed
- Land cover: a useful subject to explore an endemic problem in GI

ADAS

Contact: lex.comber@adas.co.uk

- H. Gyde Lund definitions of forestry website http://home.comcast.net/~gyde/DEFpaper.htm
- References

Hoeschele, W., 2000. Geographic Information Engineering and Social Ground Truth in Attappadi, Kerala State, India. *Annals of the Association of American Geographers* **90** 293-321.

Robbins, P. 2001. Fixed Categories in a Portable Landscape: The Causes and Consequences of Land Cover Categorization. *Environment and Planning A*. **33** 161-179.

Fisher, P.F. 2003. Multimedia Reporting of the Results of Natural Resource Surveys, *Transactions in GIS*, **7** 309-324.

Comber, A. Fisher, P. and Wadsworth, R., 2003. Actor Network Theory: a suitable framework to understand how land cover mapping projects develop? *Land Use Policy*, 20: 299–309.

Comber, A. Fisher, P. and Wadsworth, R., 2004. Integrating land cover data with different ontologies: identifying change from inconsistency. *International Journal of Geographical Information Science*, 18: 691-708

Comber, AJ, Fisher, PF, Wadsworth, RA., in press. You know what land cover is but does anyone else?…an investigation into semantic and ontological confusion. Letter to be published in *International Journal of Remote Sensing*.

ADAS