**Full Data Management Committee Meeting**

**April 25th and 26th 2012**

**CEH Edinburgh**

*Present*

**Sue Owen**, **David leaver, Bill Bealey**, **Mhairi Coyle, Ralf Kiese**, **Arnaud Frumau**, **Lisa Emberson**, **Klaus Larsen**, **Diego Guizzardi,** **Gertjan Reinds**, **Clare Howard**,

*Apologies*

**Adrian Leip**, Wilfried Winiwarter, Christophe Ammann, Albert Bleeker

Abbreviations: Data Managers=DMs

|  |  |
| --- | --- |
| **Item** | **Action** |
| 1. **Overview on the ECLAIRE website: Bill Bealey**
* Some of the pages labels and menus items need clarifying. Photos are needed on Home page at least.
* The Description of Work needs to be more visible
* Details of Meetings should be posted: please send details to eclaire@ceh.ac.uk
* Further information on the sites would be useful. Also some checking is needed as some of the details are incorrect. Whim and Castelporziano should be added to the sites list
* We need to add further information to the database pages, including database news
* Add links to other projects… ?i.e. PEGASOS
* Implement a file manager to allow users to post documents and then provide access – file structure could be by workpackage
* Later, Bill and Clare discussed the scenario pages which are on the website. Scenarios information has to be on the website because it is a deliverable. Links to the relevant scenario databases (i.e. land-use, meteorology and climate) are also included.
* Gert-Jan Reinds asked whether the scenarios information was more suited to the AFOLU database.

 Alessandro Cescatti noted that Component 4 are providing meta-data and links to external databases used for C4 in the AFOLU database. This avoids double storage of data, and avoids the need to cope with datasets which are TB large. Clare would contact Winifried Winiwarter and Adrian Leip and copy in Bill and Sue to suggest this/ask if we have understood the connections correctly. | Clare/SueClareAllBill/SueSueBill/ClareBill/DavidClare |
|  |  |
| 1. **Data Management WP21.4 Milestones: Sue Owen**
* Sue discussed the milestones.

ÉCLAIRE Data Management Plan & Data Policy Documents (month 6)ÉCLAIRE Data Portal (month 6)Database training sessions for users – online tutorials (month 8)Database documentation and guides for users (month 8)* Measurement takers (in the main, the WP leaders) need to talk with DMs to devise templates.
* Diego said that the JRC group are developing a wiki page to provide user information, handbooks etc.
* CEH are also looking into a wiki for information on data policy etc. and to encourage user questions which will form the basis of user documentation.
 | SueSue/David/Bill |
| 1. **Cross-cutting CEH-JRC data centre issues: Sue Owen**
* A. Responsibilities for information transfer between workpackages was discussed.
	+ It was noted that Component Leaders, with the support of workpackage leaders should be ultimately responsible for this, with the support of the Data Management team.
	+ It was noted that in NEU the quality ramped up at the end of the project – so we need to think about the lessons learned now. The increased interest occurred as paper writing was beginning.
	+ We propose a Data Quality Management section within the Data Management Plan.This will include a ‘lessons learned’ section to encourage earlier planning of synthesis publications.
	+ Quality control checks are a Deliverable for Month 12,

It was suggested that another QC check could be at the ‘model testing stage’ – rather than waiting until the papers are being written. This can be done through joint workshops to address this and discuss synergies.The conclusion was that data should be QC tested as soon as possible – C3 are already planning it. * + Sue outlined a data exchange support group plan (appended), to be taken to the ESG for agreement. We can investigate webforms for reporting by the C leaders to this committee.
	+ Any issues can be taken up with the ESG, passed through Clare.
 | Sue to list lessons learnedSue/allSueSue/LisaSue/Clare |
| **(3b) Cross-cutting CEH-JRC data centre issues: Lisa Emberson*** Measurement wish-lists (2 in circulation):

-(i)C1-C3 from Lisa-based on Ed Rowe’s work, -(ii)C2/C4 models from Almut for Global/European scale models.* C1 had been discussing measurements with Christof Amman, still in progress.
* Ecosystem models data needs more description.

It was noted that the input for the ecosystem models will be coming from three different places:* Data mining (Sue is currently supporting Gina with this)
1. Literature based
2. Database data
* (iii) Measurement data
 | Lisa to communicate with field workers once the list is finalisedGertjan to carry forward to modelling group as appropriateMhairi to liaise with Christof and EikoSue to support Gina |
| 1. **Components - C3: Lisa Emberson**
* Lisa outlined the work and data flows for C3 – and listed the DMs.

Her data flow diagram was good and could be produced for each ECLAIRE component.The DOSE model will be used in two components:* C1 to model the actual deposition
* C3 to model impacts of that deposition
* Lisa has a list of ecosystem models to be used in C3 – some overlap to those in C4. For C4 they will run regionally, and output compared to the earlier runs for verification etc.

Mhairi , Ralf and Lisa need to chat about C1/C3 links * Because a large number of models are performing a variety of tasks – and in some cases a model is used in more than one component, Clare suggested that it would be useful to add a modelling section to the website under Task 21.3, with information on all of the models, applications, key contacts etc.
* Lisa discussed the Data Mining of “Data” excel template.

She suggested that the template for the site data (which are to be used for the model), could be used for other information, input data (=parameterisation data) and evaluation data.The model and measurement community need to define the items and the units etc. This will be part of the Data Dictionary effort (see below)Soil layers are needed as key fields (as for NitroEurope C2)There was much discussion on types of template, but an agreement was made by the end of the meeting on the way forward, adapting the old NitroEurope C2 template. Further discussions necessary.* Further development of the data templates included storing data for ECLAIRE under recognisable Data categories eg meteorology data, flux data, site data, etc.
* There are terminology issues; eg “ parameterisation data”

 some are not directly measured, but are derived from site-specific models. We need to communicate frequently while all templates are being developed, and introduce data flags to show data derivation etc* Lisa’s template needs to be discussed with modellers and field-workers.
* Download formats need to be discussed with Sue (and later, Mhairi)
* Data gaps need to be addressed. We suggest adopting the NitroEurope protocol of flagging to indicate gap-filling method (eg interpolation, another instrument nearby …).
* Lisa had prioritised the plot scale model variable wish-list according to how essential they are in the models, and by how many models need them. This may be one way of limiting the field-worker’s task.
* Some variables are calculated from others… if so, we need common agreed methods. Units may need conversion protocols. These need to be identified
* A decision still needs to be made about which sites have the most relevant data from the NitroEurope database. Sue will try to produce a site metadata catalogue.
* NEU database issue: Data are being updated offline, so if these are to be used for ECLAIRE, it’s necessary to check with PI’s whether there are updates to that uploaded to NEU. Sue to contact NitroEurope PIs to offer support in updating their data, and will produce a report for ECLAIRE modellers.
* Be aware of the differences between NEU and ECLAIRE objectives as this may mean that NEU sites are NOT necessarily the most relevant for the work for ECLAIRE. N C GHG should be fine from NEU, but other data may have to come from other projects. Also timescale is an issue, NEU 3-5 yrs, whereas we may need more for the simulation runs.
* Lisa’s list prioritises variables needed in C3 models – but NOT what models might need later. Variables to be discarded from Lisa’s list need to be carefully checked.
* WP13 has different data needs. Gert-Jan has some site information from CarboEurope (Alexander Komarov) on C and N. CEH has ozone.
 | Lisa, Mhairi, Ralf to report on small break-out sessionClareSue to leadSue, Lisa, Klaus, Ralf, MhairiLisaSue/MhairiSue to publishChristophe Amman - Lisa/Ralf/Mhairi/Klaus SueSue/LisaSueLisa-communicate to all relevant people |
| 1. **Components - C1: WP1 & 2 Mhairi**
* Template for instruments rather than by site was suggested by Eiko
* There will be 9 heavy instrumented sites for 15 months
* Short-term 6-week intensives at 9 sites – requiring 1 template/group?
* 1-D and 2-D (time-series) data will be generated
* The 2-D includes heights, frequencies, counts, particle number distribution…etc. These will be dealt with as key-fields.
* Separate Sheets for Marga/Graegor instruments?
* PTR-MS data – very voluminous.
* Chemical species
* Data dictionaries will be important : Check the EIDC and EMEP dictionaries and build from there

[Mhairi has a list of what will be measured at the nine sites for C1.]  | Mhairi and Sue to work with Eiko and Lisa (WP13 template)Sue/David |
| 1. **Components – C1: WP 3 & 4 Ralf**
* WP3 will feed into WP6, WP3, WP12 (Sue found in WP6 description) (Ralf said WP3 would feed WP 13 and WP7?? Need to check)
* WP13 – plot scale modellers have been thinking about templates.

Should there be a common ECLAIRE/NEU plot modelling template. NEU is still waiting for plot-scale modelling, so this could minimise effort.* Templates/Provision for both model input and output are required.

[sites and models are not listed specifically in WP13, they need to be determined]. Would it be possible to upload some modelling “metadata”, rather than everything?* WP4 – will use data from WP1 & 2 to improve data processes and data from C3 to run the model. (Thus some WP13 will simulate their own ‘dry deposition’ and some will take it from other WP’s )
 | Sue/Clare check with RalfSue/RalfSue/Ralf |
| 1. **Components – C2: WP 5-8 Arnoud Frumau**
* Products will be an assessment report. Arnoud needs to contact WP leaders to make them aware of DM and Data Exchange issues
 | Arnoud/Albert |
| 1. **Components – C3 WP9 and WP 11 Sue Owen**
* WP 9 - Sue has been discussing literature data mining with Gina Mills. A template for mining literature data is under development. There may be an IPR issue regarding the upload of PDF’s to the database. Sue to discuss with Bill and Stephen Prince
* WP11 Sue Owen – will communicate again with Francesco and Elena about templates for data capture

\* for all activities in WP11\* after checking their deliverable dates - will contact Francesco/Elena to make sure the data exchange links and communications are working, and that no unfounded assumptions are being made about Lisa (WP12 leader) interacting with WP16 and transferring WP11’s outputs | Sue/BillSue |
| 1. **Components – C3: WP 10 Klaus Steenberg-Larsen**
* 5 field sites in this WP, and several experiments at each site.
* DOSE model used to synthesise the WP10 data
* As above, it would be better to break down the NEU template structure, eg meteorological data, with 10 variables for example.
* CSV would be a better way to organise the templates, due to the large amount of data. The database can upload CSV now, but batch upload would also be a useful improvement – and there are validation issues in terms of delivery and use.
 | Klaus/Sue/LisaDavid/Sue |
|  **(10) Components – C4: WP 14-17 (Sue/Gert Jan)*** Adrian Leip has spoken with the WP leaders. No response from WP17
* WP15 – Gert-Jan Reinds is workpackage leader . He will deliver data for C5. Mostly data will go through AFOLU and relies on the results of WP13 to take it to European scale, in terms of model parameterisation etc.
* WP16 – GJR will take care of the data management. Nothing to report as yet.
* WP14 – C-stocks and C-fluxes, depending on chemical transport models (CTM), with first results this winter. Frank Dentener will generate boundary conditions for European simulations with the CTMs.

This relates to the meeting in IIASA. C4 needs land use, not only current, but scenarios. Required datasets are large and stored elsewhere, so only meta-data will be uploaded to the AFOLU datasets.  | Adrian/Gertjan to re-contact |
| 1. **Components - C5: Gert-Jan Reinds**
* C5 receives data from C2

-air pollution (O3 concentrations, O3 fluxes ...); EMEP (met.no, NO)and from C4-air pollution impacts on carbon stocks and greenhouse gas emissions (JRC, IT)-air pollution impacts on soils and biodiversity (Alterra, NL)-European maps of novel thresholds and exceedances (RIVM, NL)* Data Management Needs:

In C5 the GAINS model will be used for scenario analysis and cost optimizations. GAINS has its own data base and web-based data retrieval system. No need for extra data management. Data management is restricted to inputs for C5: large sale data sets: AFOLU DB (Note: Although Gert-Jan works with C4 not C5 he is happy to keep the role of DM for C5, but will discuss with Wilfried) | Gertjan |
| 1. **C13 / Field work Templates**
* Lisa’s modelling wish-list template was based on Ed Rowe’s wish-list.
* This was compared with NitroEurope Component 2 data template
* Identify what is extra on Lisa’s template, and then adapt the C2 template, because that is already formatted for database upload.

Modular approach to templates was discussed –When final, the template should be sent to experimentalists to approve | Sue/Lisa/Klaus |
| 1. **AFOLU Database**
* Diego demonstrated the AFOLU database
* There was a discussion regarding permissions. Outcome was the request to create a ‘group’ for ECLAIRE – Diego will look into it.
* It is possible to search across the databases in AFOLU.
* There will be URL links to relevant pre-defined download datasets in the CEH database

  | Diego/SueDiego/David/Sue |
| 1. **CEH Database and metadata:**
* The CEH database downloading facility needs to be more intuitive. This might be achieved by defining EXCEL templates, and therefore database forms based on single sheets of similar data (discussed above)
* In the CEH EIDC Information Gateway it is possible to search across databases, in a similar way to AFOLU.

An option is to store meta-data for the CEH ECLAIRE database there to provide an easy way to search for information and to request the actual data. The requests would be via URLs for pre-defined reports. Links to the database log-in would still give the user the facility to define their own download reports. * It is important to discuss the meta-data issue, even before the first data are uploaded, but that the final decision about how this should be deployed should be made by the ESG, informed by the DMC and that further investigation of the options (see 1-3 below) would be required.
1. Place the meta-data on the CEH information gateway
2. Place the meta-data on the AFOLU database
3. Provide met-data information in the CEH ECLAIRE database – although it was noted that an improvement in the database software might be needed.
* David Leaver will discuss this with John Watkins (CEH EIDC). The DMC can then take some suggestions to the ESG.
 | Sue/David/BillDavid |
| 1. The CEH ECLAIRE Database software was demonstrated (Sue and David)
 |  |
|  |  |
| 1. **Other business**
 |  |
| * Klaus pointed out that uploaded data need updating. This will be relevant for ECLAIRE and also for NitroEurope, because this is when problems with time series and missing site data become apparent.
* For ECLAIRE: WP leaders should set goals for data providers to provide a synthesis paper as early as possible. This should be built into the QC strategy.
* Lisa Emberson suggested early model runs to test raw quality.
* We agreed that early data quality workshops are needed.
* IPR disclaimer – Sue and Bill to discuss; particularly with respect to literature searches
* Remind Sue to send out the data policy docs.
* Sue will also send info re validation too.
* Sue to contact the NEU users re suggested improvements to the database
* DMC suggested that Sue should be a regular member of the ESG – or at least be asked prior to each meeting if it would be beneficial for her to attend, to discuss important issues.
 | SueSue to communicateSue/LisaSue/LisaSue/BillSueSueSueSue/Clare |
| NEXT FULL DMC MEETING: At the ECLAIRE GA, then in Spring 2013 | Sue to request a non-over-lapping time |