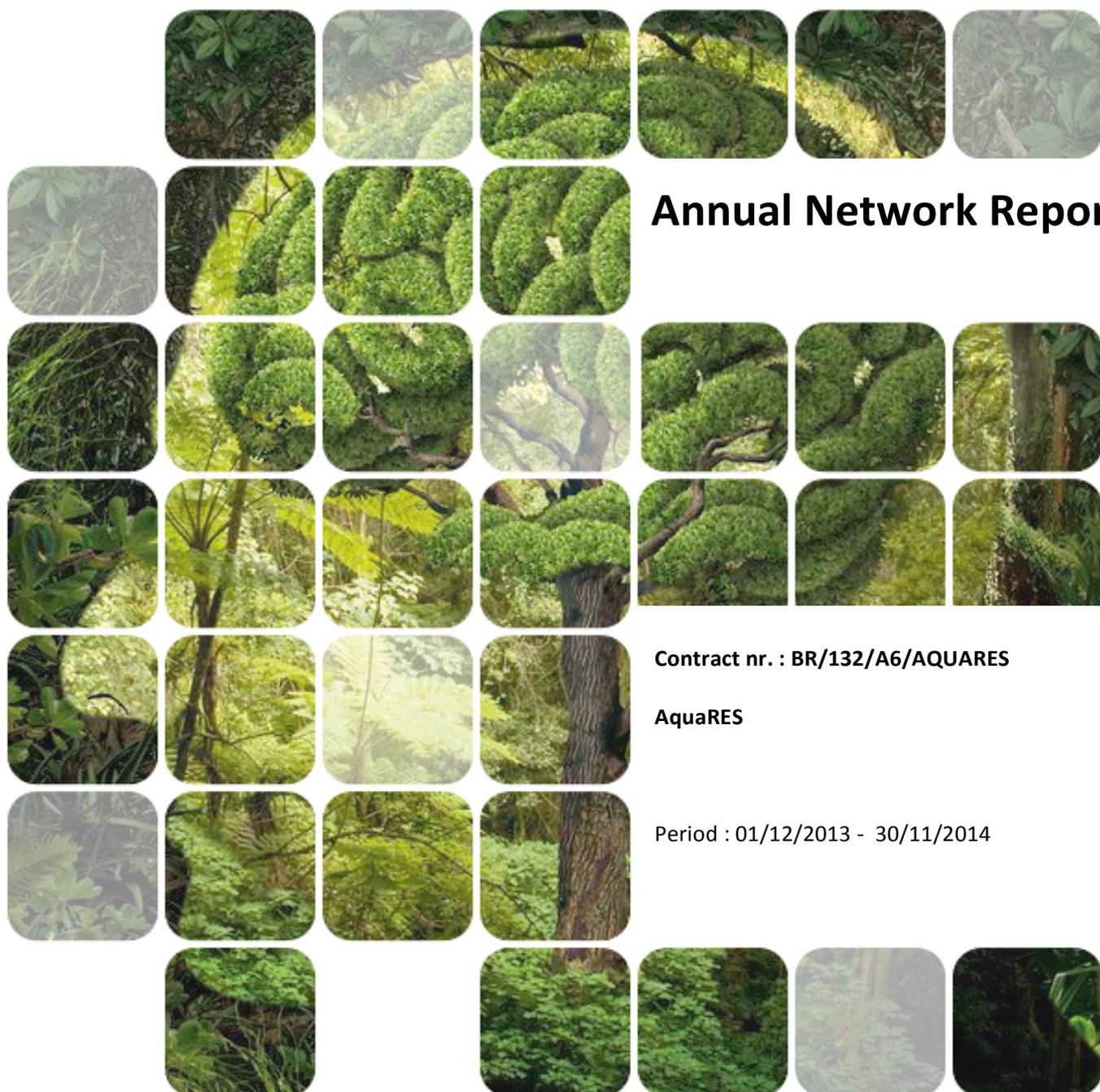# BRAIN-be

BELGIAN RESEARCH ACTION THROUGH INTERDISCIPLINARY NETWORKS

Annual Network Report

Contract nr. : BR/132/A6/AQUARES

AquaRES

Period : 01/12/2013 – 30/11/2014

NETWORK

COORDINATOR

Aaike De Wever, Royal Belgian Institute of Natural Sciences, Operational Directorate Natural Environment

PARTNERS

1.   Leen Vandepitte, Vlaams Instituut voor de Zee/Flanders Marine Institute

2.   Bruno Danis, Université Libre de Bruxelles

AUTHORS

1.   Aaike De Wever, Royal Belgian Institute of Natural Sciences

2.   Leen Vandepitte, Vlaams Instituut voor de Zee/Flanders Marine Institute

3.   Bruno Danis, Université Libre de Bruxelles

PROJECT WEBSITE:

http://www.lifewatch.be/aquares

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY OF THE REPORT

The AquaRES "Aquatic species Register Exchange and Services" projects builds on the specialist curated taxonomic databases *World Register of Marine Species* (WoRMS), the *Register of Antarctic Marine Species* (RAMS) and the *Freshwater Animal Diversity Assessment* (FADA), which are hosted in Belgium. The main objective of the AquaRES project is to ensure and enhance the interoperability and public availability of these aquatic species databases through the development of a set of web services. These services envisage data exchange among the 3 partner databases and with other international initiatives including Encyclopedia of Life (EoL), Catalogue of Life (CoL), Global Biodiversity Information Facility (GBIF) and e-Science initiatives such as Biodiversity Virtual e-Laboratory (BioVeL) and LifeWatch.

The activities on the AquaRES project started in May 2014. During the first 7 months of this 2-year project, we focussed on working out the basic requirements to make this project a success. As modifications to the databases, web services and tool development are ongoing, the most tangible products will only be available towards the end of the project. Progress on the planned tasks is on-track.

We completed the mapping of the database structure (Task 1.1.1), the documentation of the web services and data cache requirements (Task 1.1.2) and mapping of the web services for exchange with international initiatives (Task 4.1.1). The follow-up committee was actively involved in the discussions during this design phase during the meeting organised in June 2014. We initiated the work on improving and setting up exchange web services (Task 1.2) and Improving input and data publication services for individual datasets (Task 2.1). Data quality control and updating of the participant databases (Task 2.2) is ongoing throughout the project.

2. ACHIEVED WORK

Detailed description of the achieved work and tasks of the past reporting year

The focus of the AquaRES activities during the period between May and November 2014 was to work out the basic requirements to make this project a success. These activities included elaborating the database structure, exchange format, web services, design of the data cache and to involve the experts of the follow-up committee in this design phase. This helps us to ensure and enhance the quality, interoperability and public availability of the aquatic species databases FADA, WoRMS and RAMS and will help us to offer tools to a wide user community to access and actively use the data. Below, we provide a description of the tasks, which were addressed during this reporting period.

WP1: Data exchange and web services to ensure the interoperability among the participating databases FADA, WoRMS and RAMS

In order to **map the database structure** (Task 1.1.1) of the World Register of Marine Species (WoRMS) and the Register of Antarctic Marine Species (RAMS), which rely on the common Aphia database system, and the Freshwater Animal Diversity Assessment (FADA) database, we studied the available documentation (see respectively http://www.marinespecies.org/structure/ and http://trac.bebif.be/wiki/BioFresh/FadaDatabaseStructure). As we envisaged working with Darwin Core Archive files for the data exchange, we relied on the field definitions from the Darwin Core standard from Biodiversity Information Standards (TDWG). Starting from our existing export formats, we made a comparison of the fields used. We compiled an overview table of the 69 different terms/fields (included in the taxon core and 5 extensions) used in the two database systems. For each of the fields, we evaluated the need to include them in the export format, the use of common control vocabulary, how to generate or include information currently missing in one of the two database systems, and which fields are considered mandatory to ensure that we can successfully exchange taxonomic data. During

this exercise, we ensured that the export format for exchange between WoRMS, RAMS and FADA, can at the same time be used for including selected data in the partner database, (re)building the "AquaRES cache" (or AquaCache) as for the exchange with external initiatives (Task 4.1: Establishing and improving data exchange with international initiatives).

In parallel to the review of the database structure and exchange format, the **requirements web services and data cache** (Task 1.1.2) were discussed and worked out in detail (see Fig. 1 under "intermediary results"). Overall, Task 1.1 has been successfully completed. Some of the choices reflected during this design phase may require changes to the database(s) (e.g. including record timestamp and marine flag in FADA). This will be addressed as part of Task 1.2 (Improve and set up exchange web services)

The work on **improving and setting up exchange web services** (Task 1.2) is ongoing. A test export has been produced for WoRMS and FADA, but several bugs still need to be addressed. The current focus is on producing file-based web services. The need for record-level web services, as initially envisaged, will be evaluated while testing the database synchronisation (Task 4.1).
As the work on Task 1.2 is progressing, and the **design of the data cache** (Task 1.3.1) has been worked out, we are currently planning the **implementation of a central cache for common web services** (Task 1.3.2). The **synchronisation of updates for taxonomic groups managed in one of the participating databases** (Task 1.4) is planned for months 6 to 18. At this stage, preparatory steps have been undertaken by the individual databases.

WP2 Tools/services for improving taxonomic checklist data

Except for the **data management** task (Task 2.2) most of the tasks were planned to start after 6 months or later in the project. Hence, the fact that this description focuses on the preparatory steps taken so far does not reflect any delay in the execution of the project.

As activities under Task 2.1 (Improving input and data publication services for individual datasets), RBINS is currently reviewing its import procedures and working out a specifications document for the FADA import tool and plan to discuss this with the AquaRES partners to get their input and exchange ideas on how to tackle specific data quality control steps during the import phase. VLIZ on the other hand is investigating the checklist publication tool (and its code) available through FADA and planning its implementation.

Under Task 2.2 (Perform data quality control and updating procedures), RBINS processed new checklists for several insect groups including beetles (Coleoptera) and flies (Diptera). While these data have not been fully integrated in the on-line database, we took the opportunity to review and improve our import scripts, which will subsequently be incorporated in the FADA data import tool. For WoRMS, VLIZ has been actively working on the Mollusc data through the establishment of a MolluscaBase editor community (http://www.marinespecies.org/news.php?p=show&id=3685). In addition, the WoRMS Mollusc data were further extended through the establishment of the FreshGEN "the Freshwater Gastropods of the European Neogene" database (see news item: http://www.marinespecies.org/news.php?p=show&id=3971) and the incorporation and quality checking of data on the freshwater and terrestrial snails from New Zealand. WoRMS data for the Digenea (parasitic flatworms), including marine and freshwater representatives, were completed and fully quality checked, making these data available for any exchange in the framework of AquaRES. The inclusion of FishBase data in WoRMS was improved through the import of the higher classification taxa up to genus level (both marine and freshwater). Later, this will be used as a basis for incorporating species data via exchange web services. Similar exchange mechanisms will be considered for including freshwater fish data from Fishbase into FADA and the AquaRES cache.   For RAMS, data management activities included small routine updates and verification of literature sources.

Task 2.3 (Taxamatch tools and implementation of fuzzy matching algorithms) and 2.4 (Develop interface for entering and validating distribution information for species) have not been explicitly addressed during this period, but selected topics were covered during work meetings and the follow-up committee meeting. E.g. the approach for implementing the TAXAMATCH tool (Task 2.3.1) and the selection of geographic regions

(Task 2.4.1) have already been discussed at length. This preparatory discussion will definitely enable us to move faster once we start working on these tasks.

The **Workshop to test tools and procedures for checking taxonomic lists, tools to complete distribution data and address identified gaps and overlaps between datasets** (Task 2.5) is planned for the 2nd year of the project. Nevertheless, since September, all partners have started to identify and contact several potential workshop participants to check on their interest to contribute to AquaRES and participate in the workshop. From December onwards, ULB is hiring Quentin Jossart who will coordinate the organisation of this workshop. Currently, we are investigating to hold the meeting back-to-back to the "Aquatic Biodiversity & Ecosystems" conference (http://www.aquaticbiodiversityandecosystems.org/), which we see as an added value for the contributing editors.

WP3: Tools/services to validate species distribution/occurrence data

Similarly to WP2, all tasks except for 3.4 (Consultations with steering committee and its members) are covered at a later stage in the project.

In June, we held a very successful steering committee meeting. Nine external specialists with a different background participated in the meeting. Following the presentation of, and discussion on the component databases, we organised the meeting to address specific questions/discussion topics to ensure that we got maximal feedback from the experts. Topics included;
- Data exchange among databases and planned improvements on the databases,
- Developments to improve data entry and validation by experts and possibilities for citation and valorisation of entered data,
- Data exchange and tools targeting regional and national initiatives: potential, possibilities for wider uptake of data products? and
- Data exchange and tools targeting international initiatives: streamlining exchange, towards improved data quality, uptake of developed tools, requirements from potential users.
A detailed meeting report is provided in annex.

WP4: Data exchange with international initiatives

During this reporting period, our efforts to establish and improve data exchange with international initiatives (Task 4.1) focussed on the mapping of the web services (Task 4.1.1), which was covered in parallel to Task 1.1 (Mapping the database structures and documenting requirements in terms of export functionalities, web services and central data cache). In addition, this exchange was one of the discussion topics addressed during the steering committee meeting (Task 3.4). Further actions under this work package will be carried out towards the end of the project.

3. INTERMEDIARY RESULTS

The AquaRES project focusses on the improvement of the taxonomic databases World Register of Marine Species (WoRMS), the Register of Antarctic Marine Species (RAMS) and the Freshwater Animal Diversity Assessment (FADA) and the development of tools to facilitate data entry and usage of the data. As modifications to the databases, web services and tool development are ongoing, the most tangible products will only be available towards the end of the project. Nevertheless, we have produced the following outputs, which reflect selected tasks described above. These include:

1. Follow-up meeting report (provided in Annex)
2. Comparison table for output fields (database mapping – Task 1.1)
3. Example Darwin Core Archive export data for WoRMS and FADA

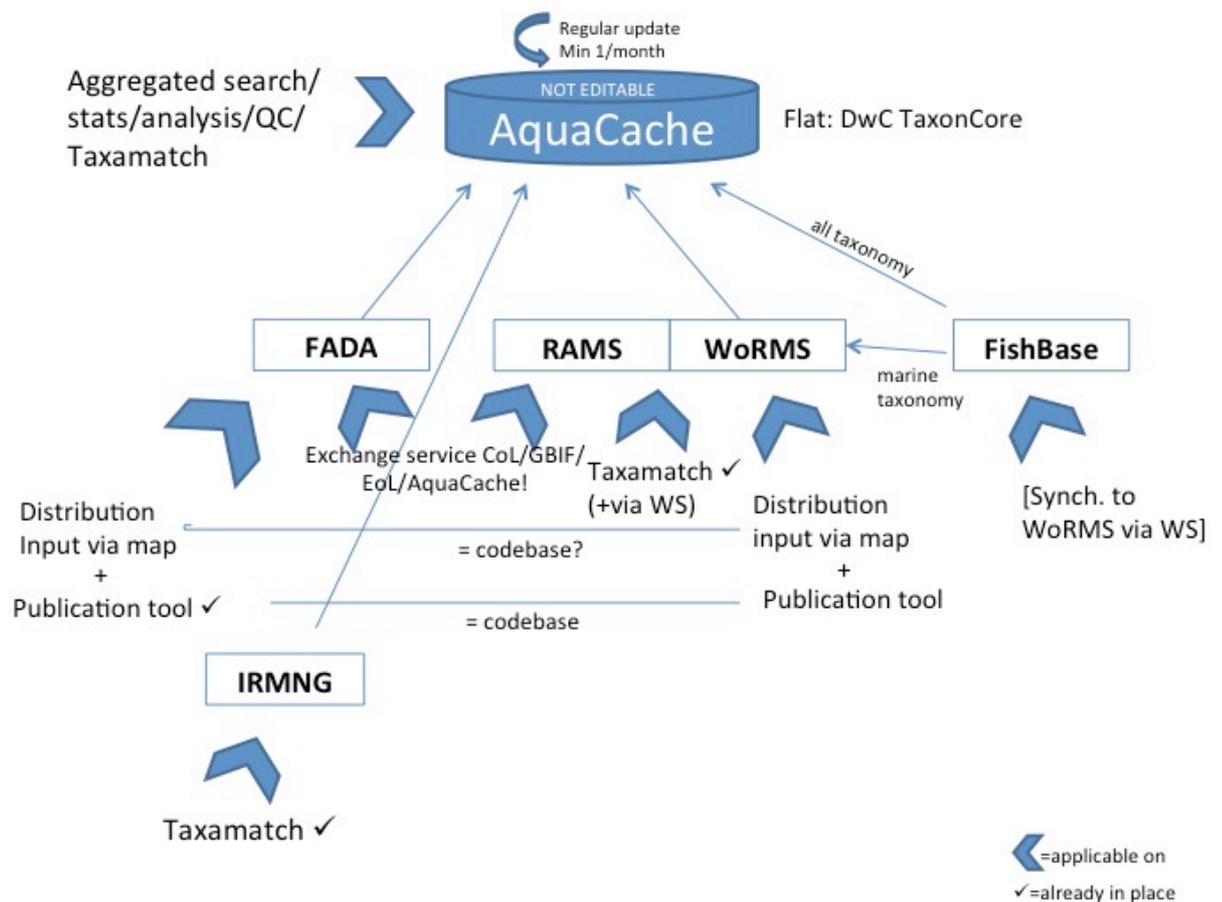4. Outline of the « AquaCache » design (see Fig. 1 below)

Fig. 1: Design of the AquaCache highlighting component databases, services and input- and output-tools to the different databases.

Currently these intermediary outputs are available for internal use, but will be made publicly available through the AquaRES website after a final review of the material over the coming months.

4. PRELIMINARY CONCLUSIONS AND RECOMMANDATIONS

During this mostly preparatory phase of the project, we were successful in comparing the database structures and designing the requirements and specifications for the exchange web services and common cache. One of the main conclusions of the discussions on this topic was that it will be much more effective to focus the web service developments on file based exchange through the Darwin Core-Archive format, rather than implementing record level web services.

5. FUTURE PROSPECTS AND PLANNING

Overview of the foreseen activities and planning for next reporting year, taking into account the actual state of the work and the intermediary results

In terms of planning, we are sticking to the originally proposed timeline, but shifted by 5 months due to the later start of the actual work on the project.

For next year the main activities include:
- Deploy new or improved data exchange web services (Task 1.2)
- Build the central data cache (Task 1.3.2)
- Test data exchange and synchronise data based on exchange web services (Task 1.4)
- Improving input and data publication services for individual datasets (Task 2.1)
- Design and develop interface for entering and validating distribution information for species (Task 2.4)
- Hold a workshop with taxonomic editors (Task 2.5; see also below)

As mentioned under "2. Achieved work" we are currently actively preparing the *Workshop to test tools and procedures for checking taxonomic lists, tools to complete distribution data and address identified gaps and overlaps between datasets*. The envisaged date for this workshop is September 2015 and we plan to announce the exact date and venue early 2015 once we have a good overview of the meeting logistics. Obviously, as we want to demonstrate and test the tools developed during this project, this workshop will be a major project milestone.

6. FOLLOW-UP COMMITTEE

Dates of the meetings and overview of the concrete contributions of the follow-up committee

In June, we held a very successful steering committee meeting. A detailed meeting report is provided in annex.

Main recommendations include:
- Need for regular consultations with editor groups
- Use of Catalogue of Life web services for extending databases with data available through this network
- Need to explore possibilities to build on FaunaEuropaea tools for FADA
- Explore existing cross-mapping tools for comparing names
- Further evaluate the requirements and relevance of record level web services for FADA

7. VALORISATION ACTIVITIES

7.1 PUBLICATIONS

-

7.2 PARTICIPATION/ORGANISATION OF SEMINARS (NATIONAL/INTERNATIONAL)

Oral presentation, poster… and/or organisation of workshops, symposia etc.

- 2-4/06/2014: Catalogue of Life Global Team meeting (Washington DC, US; Leen Vandepitte, VLIZ) – Presentation, with mentioning of AquaRES project in relation to WoRMS
- 08-12/09/2014: Presentation at the European Marine Biology Symposium (EMBS) (Saint-Petersburg, Russia; Leen Vandepitte, VLIZ) – Presentation, with mentioning of AquaRES project in relation to the LifeWatch project and WoRMS
- 26/09/2014: Participation in the Creativ-B Final Event (Brussels, Belgium; Aaike De Wever, RBINS) – Informal contacts with follow-up committee members and biodiversity infrastructure experts
- 06-08/10/2014: Participation in the Technical developers training workshop: 'building biodiversity workflows with Taverna' (Ostend, Belgium; Bart Vanhoorne, VLIZ; Aaike De Wever, RBINS) – Exploring potential to use Taverna Workflows in AquaRES
- 12-16/10/2014: Participation in the World Conference on Marine Biodiversity (WCMB) (Qingdao, China; Leen Vandepitte, VLIZ) – Informal contacts with marine biodiversity experts
- 26-31/10/2014: Poster presentation at the Biodiversity Informatics (TDWG) meeting (Jönköping, Sweden; Aaike De Wever, RBINS) – AquaRES was mentioned during the poster presentation "Centralising thematic (freshwater) biodiversity data using the Darwin Core standard and GBIF's Integrated Publishing Toolkit (IPT)"

7.3 SUPPORT TO DECISION MAKING (IF APPLICABLE)

-

7.4 OTHER

-

8. ENCOUNTERED PROBLEMS AND SOLUTIONS

Encountered problems/obstacles, adopted and/or envisaged solutions, unsolved problems

No significant problems were encountered so far. As the actual project start was only in May 2014 (instead of the start date of the contract in December 2013), we plan to ask for an extension of the project duration. Although we are on-track according to the original timeline, we observed that the developments of the web services have advanced relatively slowly; this will be sped up early 2014, once RBINS' developer has returned from sick leave.

9.2 COMPOSITION OF THE FOLLOW-UP COMMITTEE

The core follow-up committee with external experts consists of:
- Dimitri Brosens, Belgian Biodiversity Platform/Research Institute for Nature and Forrest (INBO)
- Wouter Addink, Naturalis/Species 2000 (represented Peter Schalk at meeting 24/06/14)
- Peter Schalk, Naturalis/Species 2000/GBIF Governing board
- Saskia Van Gaever, Federal Public Service Health, Food Chain Safety and Environment - Marine Environment
- Ward Appeltans, Intergovernmental Oceanographic Commission of UNESCO/Ocean Biogeographic Information System (OBIS)
- Wouter Los, University of Amsterdam/LifeWatch

In addition, we involved the following users and associated partners during our follow-up meeting and the project communication:
- Geoff Boxshall, The Natural History Museum/WoRMS Steering committee
- Anton Vandeputte, biodiversity.aq (through Skype)
- Nabil Youdjou, biodiversity.aq
- Koen Martens, Royal Belgian Institute of Natural Sciences (RBINS) / Freshwater Animal Diversity Assessment (FADA)

We also invited Christine Keulen (Département de l'Etude du miliey naturel et Agricole (DEMNA) – Service Public de Wallonie) for the follow-up committee meeting, but she did not join the meeting nor formally confirmed or declined interest in the follow-up committee.

10. REMARKS AND SUGGESTIONS

Concerning for example: the coordination, the use or valorisation of the results, personnel change …

-