Visit Report

Wim Hugo

8 April 2014

# Research Data Alliance Meeting, 26-28 March 2014, Croke Park, Dublin

## 26 March – Day 1

Morning Session: *Attended* Plenary Presentations and Discussions

Lunchtime Meeting: Program Committee meeting – RDA Plenary 4 in Amsterdam, September 2014

Afternoon Session: *Participated* in RDA-WDS **Publishing Data** Interest Group session on behalf of ICSU-WDS

Evening Session: *Participated* in RDA-WDS **Publishing Data Workflows** Working Group on behalf of ICSU-WDS

## 27 March – Day 2

Breakfast Meeting: WDS and DataCite (M Diepenbroek, M Mokrane, B Minster, W Hugo, J Brase – DataCite) on (1) collaboration between DataCite and WDS, and (2) imminent merger between re3data, DataBib and DataCite

Early Session: *Attended* Plenary Presentations and Discussions

Morning Session: *Participated* in Interest Group on **Certification of Digital Repositories** (RDA-WDS partnership) on behalf of ICSU-WDS

Afternoon Session: *Co-Chair* of RDA-WDS Working Group on **Certification of Digital Repositories** on behalf of ICSU-WDS

Teatime Meeting: CoDATA organizing committee for CoDATA/ WDS Conference, November 2014, India, and CoDATA Workshop on Preservation of Data in Developing Countries, August 2014, Kenya.

Evening Session: *Participated* in RDA Interest Group on **Data in Context** on behalf of ICSU-WDS

Late Session: *Attended* Plenary Feedback

## 28 March – Day 3

Early Session: *Attended* Plenary Presentations and Discussions

Morning Session: *Participated* in Birds of a Feather Session on **Global Registry of Trusted Digital Repositories** behalf of ICSU-WDS and *Participated* in RDA Working Group Session on **Brokering** on behalf of ICSU-WDS

Afternoon Session: *Participated* in RDA Working Group Session on **Brokering** on behalf of ICSU-WDS

Evening Session: *Attended* Plenary Feedback and Close

## Outcomes

1. WDS is confirmed as an important participant in and stakeholder in the Data Publication and Citation landscape.
2. WDS will be able to extend the re3data and DataBib repositories to suit its requirements for a global registry through its involvement in DataCite as members. WDS and DataCite will continue in parallel to the RDA process.
3. WDS will continue to work with Data Seal of Approval to develop a coherent schema for certification of trusted repositories, and contribute to the RDA process. W Hugo will join the working group and contribute/ refine earlier work on Maturity Models for Trusted Digital Repositories.
4. The WDS Knowledge Network project will form an important part of the RDA ‘Data in Context’ interest group. Peter Fox reconfirmed, and Jul
5. WDS and W Hugo in person will be active participants in the Brokering Working Group, and collaboration will continue with ESSI Lab irrespective of RDA progress.
6. W Hugo will assist with the workshop programme for the CoDATA event in Kenya in August 2014.
7. W Hugo will contribute to the sessions on Data Preservation in Developing Countries and Meta-Data Management at the CoDATA/ WDS ‘SciDataCon’ conference in New Delhi, November 2014.

# WDS Scientific Committee Meeting, 31 March – 2 April, ICSU, Paris

Agenda: Attached

## Outcomes

1. Continued participation in the Knowledge Networks effort;
2. Continued participation in the effort to create a Global Registry of Trusted Digital Repositories;
3. Renewed participation in the effort to derive a comprehensive schema for certification of Trusted Digital Repositories;
4. Assistance with a session at SciDataCon 2014 to discuss Knowledge Networks and one to discuss Trusted Digital Repositories;
5. Request to assist with the recruitment of additional WDS Members in South Africa.

# Visit to CIRAD, 3 and 4 April, Montpellier

CIRAD has developed, within their EU-funded programmes, software and tools for the following:

1. Open structure databases and tools for the storage of plant observation data of varying complexity and structure (Pl@ntNet-DataManager);
2. Image-based identification software that can be used to confirm taxonomy of a large variety of plants, based on reference collections (Pl@ntNet-Identify);
3. Management tools for creation and testing of reference image collections;

CIRAD also participates with a number of other partners in the LifeClef project, aiming to extend the Pl@ntNet software suite to allow taxonomic identification of other media formats (sounds, video, etc.) using the same approach.

## Day 1: 3 April

1. Overview of CIRAD
2. Overview of the Pl@ntNet project
3. Overview of SAEON
4. Overview of SAEON Data Infrastructure
5. Details on Pl@ntNet-DataManager Implementation

## Day 2: 4 April

1. Details on Pl@ntNet-Identify
2. Details on test and reference databases
3. Overview of LifeClef project and its aims
4. Discussion of Opportunities and Potential Collaboration

## Outcomes

1. Pl@ntNet was used for a pilot implementation in the Kruger National Park (invasive Species). Options to be explored:
   1. SAEON to continue using a service offered by CIRAD;
   2. MOU to be agreed, allowing SAEON to host parts of the software and data; and
2. AND/ OR SAEON can download and fork the open source project to continue from there, in loose collaboration within LifeClef, to allow
   1. Video, sound and image collections to be managed;
   2. Maximum value to be derived from the media-based observation data that is available.

## Recommended Actions

1. WH to write a short rationale for the use and extension of Pl@ntNet (Annexure A)
2. SAEON to discuss feasibility and options internally and with SANPARKS/ SANBI – follow up with CIRAD in May 2014.

# Annexures

## A: Pl@ntNet

CIRAD has developed a suite of software and tools for the management of botanical observation, strongly based on image collections and automated image identification. They have used the suite, amongst other things, to develop a demonstration application for invasive species identification in the Kruger National Park (where they made contact with Ndlovu Node), and Wim Hugo visited CIRAD in early April 2014 when an opportunity arose to do so.

Discussions with CIRAD about the technical aspects of their software indicated some useful elements that are not immediately apparent:

1. CIRAD built the software for image-driven data management because they were frustrated with the narrow focus and schema constraints of existing software such as Specify and Brahms. In doing so, they have looked at the good and bad characteristics of available software, and they are confident that they have made an improvement on these offerings.
2. While CIRAD focused Pl@ntNet on botany collections, there are two significant extensions:
   1. Accommodating different media, such as video and audio, through LifeClef, and
   2. Accommodating non-botanical data– again, through LifeClef.

SAEON has a significant need for software to address a variety of functions that are partly provided by Pl@ntNet. In short, these are:

1. Defining and managing complex environmental observation data sets, especially if this involves biological observation. These range from simple observation (occurrence, abundance), to complex (ecosystem structure, traits, phenology, etc.). (Possible in Pl@ntNet-DataManager)
2. Managing non-structured material in support of the observations, such as images, video, audio, and similar. (Partly Possible in Pl@ntNet-DataManager)
3. Automation of some of the classification and tagging of non-structured data, specifically automated identification of taxa from images, video, and audio. There is also a need to identify and tag organs, specify life stages, etc. (Possible for some flora, in South Africa for Kruger Park invasive species)
4. Production of formatted outputs and publication of subsets of the data, linked to properly standardized meta-data. (Partly possible)
5. Integrated tools are required for field work/ data collection and could also be used by the education officers;
6. The system needs to link with global and regional infrastructures – SABIF/ GBIF, GEO, MetaCAT/ DataOne/ ILTER, etc.;
7. It should assist with limiting the divergence of tools that are in use at present, but be capable of interfacing with them – including software such as Brahms, Specify, and I-Spot.