

Introduction and Module 4 Transcript

Below you find the transcript for the presentation "Presentation_Module4.pptx". This document is part of <https://doi.org/10.4121/21399975>, created by Cindy Quik and Luc Steinbuch, and licensed under CC-BY-NC 4.0

Slide	Transcript
1	We will continue with Module 4 where we will discuss FAIR in relation to domain specific issues. Note that Maarten Storm, main data steward for ESG, will do a part of the presentation.
2	<overview>
3	First some figures. According to the portal research.wur.nl , which should keep track of all scientific output of WUR, we produced in ESG until now ca 600 datasets and 50 software depositions. Which is not that much compared to the primary research output of reports and peer reviewed articles. You can also see that the concept of formally sharing data in a scientific context is something which came up in the last decade. Of course, WEnR predecessors like STIBOKA did already create Dutch soilmaps for a long time.
4	In 2020, a master student investigated qualitatively the use of data repositories by ESG researchers. Almost everybody used existing data, for example from the big spatial databases. Many ESG-researchers liked to share their data out of idealistic motives, however that also takes extra effort, it is not always clear if the datasets are actually used, and there are concerns that the dataset in itself is not properly "reusable" – which makes me wonder if that is a shortcoming in the readme file?
5	We tried to guess which kinds of research output comes from ESG. First of all, our laboratories produce measurements. For example, the NCL lab – luminescence dating – basically measures foton counts over time, in different contexts. This are huge datasets in a specialized format, which are processed into a few numbers' summary per observation. We assume this will hold for many analysis: there is a bunch of raw data, and the outcomes which probably fit into a .csv file. For work on the methodology, one likes to have access to the raw data; otherwise, the final outcomes, with an estimation of their uncertainty, are fine. Think also about spectral data for soil analysis. There will also be photos, for example from tree-ring dating. In grey what we expect, but are not sure about: Autocad

	<p>We assume that there is also some Social sciences like data, or as they call it sometimes: arts and humanities. This can be qualitative types of data.</p> <p>I see sometimes an exhibition of beautiful posters made by students, I assume that is somehow related to spatial planning etc and I am sure these are not made with MS-Paint.</p> <p>The groups with *Ecology* in their name do sometimes work with genetic data – see the next slide.</p> <p>Many of us will work with geographic data, and also many of us will do coding.</p> <p>In the following, we will very shortly say something about genetic data, but mainly focus on geographic data and in the last part of this module also on code.</p> <p>Did I miss something, do you have anything to add?</p>
6	<p>About Genetic data we can be short and honest: we don't know so much about this. As far as we know, it's about huge amounts of data, which make sense to be coupled in often calculation-intensive processes. There are specialized repositories for genetic data, although some might be focussed rather on human medical applications.</p> <p>Perhaps, it is worth to organise a workshop "how to FAIRize genetic data" together with for example plant, animal and nutrition & health scientists?</p>
7-17	< no transcript >
18	Now about code
19	On research.wur.nl, "software" is part of "research output" and actually a very small part of the total.
20-35	< no transcript >
36	<p>As a nice end, a slightly different topic: I want to point that ISRIC, while relative independent, is somehow also part of ESG. Sharing soil data is part of their mission. Apart from digital maps with point- or pixel data, they also have physical objects such as monoliths and soil samples. A nice idea, related to sharing data, is this prototype of a thin section (soil slice, <i>bodemslijpplaatje</i> in Dutch) microscope. ISRIC has collected many thousands of thin sections. Because one can zoom in quite far, use different light polarizations and directions etc etc, is just too much work to digitize all the raw information that is in those. They are working on a prototype, so that you can contact ISRIC, and they will put the thin section of your choice under this special microscope which can be controlled by webinterface. I really like this intention and approach to open a physical dataset to the world.</p>