

# Little Data, Big Data, No Data? Data Management in the Era of Research Infrastructures

Workshop Report  
26-27 April 2018  
Hyytiälä Forestry Field Station, Finland

Helena Karasti, Andrea Botero, Karen S. Baker, Elena Parmiggiani  
MULTICS Project  
University of Oulu



# **Little Data, Big Data, No Data? Data Management in the Era of Research Infrastructures Workshop Report**

Conducted 26-27 April 2018, at Hyytiälä Forestry Field Station, Finland

Helena Karasti, INTERACT Research Unit, University of Oulu, Finland  
Andrea Botero, INTERACT Research Unit, University of Oulu, Finland  
Karen S. Baker, INTERACT Research Unit, University of Oulu, Finland and School of  
Information Sciences, University of Illinois at Urbana-Champaign, Illinois, USA  
Elena Parmiggiani, INTERACT Research Unit, University of Oulu, Finland and  
Department of Computer Science, Norwegian University of Science and Technology,  
Trondheim, Norway

University of Oulu  
Department of Information Processing Science  
Series B 70, Working papers  
ISSN 0786-8421  
ISBN 978-952-62-2006-2 (electronic version)

Multi-scoped infrastructuring project (MULTICS, <http://interact.oulu.fi/multics>) acknowledges funding by Academy of Finland (grant # 285903).

## **Suggested Citation:**

Karasti, H., A. Botero, K.S. Baker, and E. Parmiggiani (2018) Little Data, Big Data, No Data? Data Management in the Era of Research Infrastructures. Workshop Report, April 2018, Hyytiälä Forestry Field Station, Finland. University of Oulu, Finland. ISSN 0786-8421. ISBN 978-952-62-2006-2. Available at: <http://hdl.handle.net/2142/100870>.

## Table of contents

Introduction .....	3
Workshop preparations.....	5
Pre-workshop survey - brief summary of the responses .....	7
Day I (Sessions 1-3) .....	8
Session 1: Introduction .....	8
Session 2: What is data?.....	10
Walking tour to SMEAR II station .....	11
Session 3: What is data management?.....	12
Day II (Sessions 4-5).....	13
Session 4: What is data infrastructure? .....	13
Session 5: Workshop Wrap Up and Next Practical Steps Forward .....	14
Post workshop survey results.....	16
Post workshop reflections and recommendations .....	18
References .....	19
Acronyms .....	21
Appendices .....	21
Appendix 1. Agenda .....	22
Appendix 2. Pre survey .....	24
Appendix 3. Post survey.....	32
Appendix 4. Handouts.....	34

## Introduction

This workshop was organized to provide members and other stakeholders of INAR RI Ecosystems with an introduction to data management in the ecological and related sciences. The notion of local data management was used as a starting point to discuss data management activities taking place at or close to the origins of data, and to envision how data was coordinated within and across boundaries of a variety of related contexts.

**INAR RI Ecosystems**<sup>1</sup> is a project funded by the Academy of Finland Research Infrastructure (FIRI) program 2017-2021. The aim of the project is to propose and consolidate an umbrella for environmental and ecosystem research infrastructures (RIs) in Finland (Bäck et al. 2017, ENVRIplus 2017). The consortium is led by University of Helsinki and composed of key ecosystem research components in Finland including Universities of Helsinki, Eastern Finland, Turku, Oulu, and Jyväskylä, as well as three national research institutes including Natural Resources Institute Finland (LUKE), Finnish Environment Institute (SYKE), and Finnish Meteorological Institute (FMI). Figure 1 shows the INAR RI Ecosystems components and depicts which of the locations are ecosystem observation stations, experimental field stations, biological as well as ecophysiological laboratories, or co-locations of these.

The aim of INAR RI Ecosystems is to 1) upgrade existing platforms and construct new platforms and data structures for analysing the functional relationships between ecosystems and the environment, 2) strengthen national ecosystem research and its linkages to atmospheric and environmental sciences, and 3) build a national scale, coordinated RI which enables the development and participation of Finnish partners in international RI initiatives such as ICOS, AnaEE and eLTER as well as data RIs such as EUDAT CDI and Lifewatch. Thus, INAR RI Ecosystems contributes as a national focal point for European Strategic Forum on Research Infrastructures (ESFRI) RIs.

---

<sup>1</sup> For general overview of the project see: <https://www.helsinki.fi/en/inar-institute-for-atmospheric-and-earth-system-research/inar-ri-ecosystems-0>

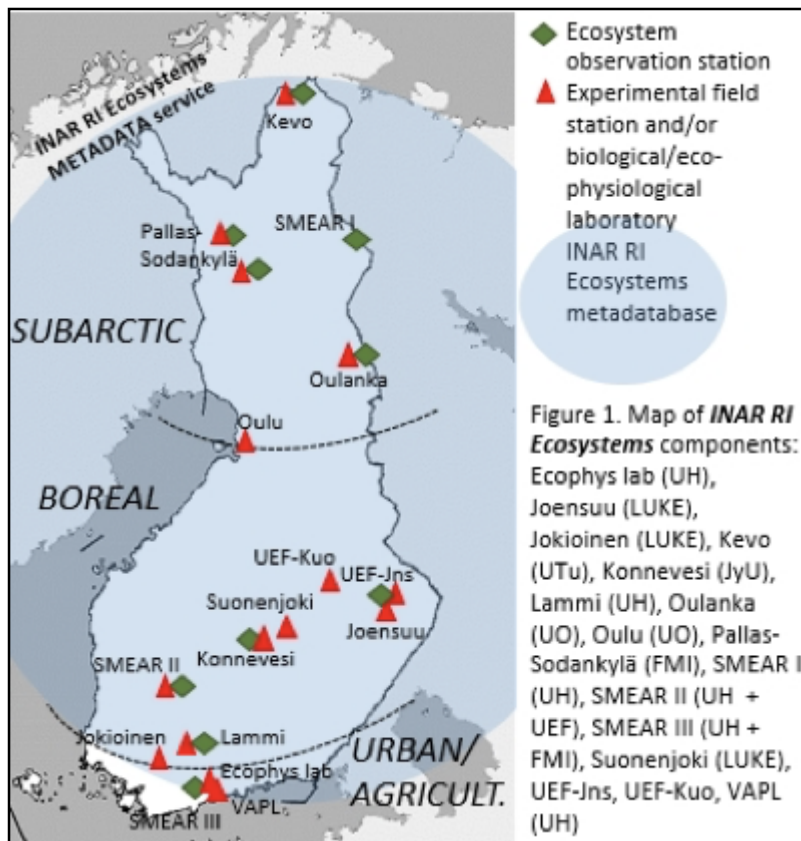


Figure 1. Map of *INAR RI Ecosystems* components. From *INAR RI* (nd).

The **Multi-scoped infrastructuring project** (MULTICS<sup>2</sup>) planned the topic, content, and ways of working for the workshop. MULTICS is an Academy of Finland funded research project that studies the formation of research infrastructures (RIs), a.k.a. information infrastructures, knowledge infrastructures, and cyberinfrastructures. The project continues a lineage of investigations focused on ecological and environmental research domains with longitudinal empirical engagements with Long-Term Ecological Research (LTER) networks, including the Finnish Long-Term Socio-Ecological Research (FinLTSER), and through them with *INAR RI Ecosystems* more recently. Volunteering to organize the workshop was a way for MULTICS' participants to contribute back to and give thanks to those in *INAR RI Ecosystems* who gave their time for interviews and discussions.

The MULTICS project has engaged in a research relationship with *INAR RI Ecosystems* as part of the national integration efforts in ecological and environmental research infrastructures (Bäck et al. 2017). This collaboration has included following the local preparations as well as the national consortiums' efforts so that the eLTER proposal submission would be included in the European roadmap for RIs. These engagements have also directed MULTICS group's interest in the ESFRI initiative, a strategy-led approach to policy-making on RIs in Europe, that is in charge of coordinating this roadmap.

<sup>2</sup> For more information on MULTICS see <http://interact.oulu.fi/multics>

# Workshop preparations

The aim of the workshop was to provide an introduction to data management in the ecological and related sciences as well as the place of local data management within the emerging larger data and RI context, i.e. within the data landscape. This aim was inspired by a particular observation made as we analyzed the policy-level ESFRI initiative with our interest on data management and infrastructure formation. We noted a generic invisibility issue built into the RI formation strategy, which we have tried to capture in Figure 2. The upper part of the figure depicts a simplified and animated version of the ESFRI ‘lifecycle’ model for constructing RIs. The lower part of the figure, our addition, depicts the national and/or local level of member states from which the data originate. Despite the crucial role data and data management play in RIs, the lower part of the figure has little visibility in the ‘top-heavy’ ESFRI template. Data originating from the local/national level of RIs becomes recognized in the ESFRI model when there is a need for data to flow into the central hub(s) of the RI, depicted with a highlighted arrow linking the top and bottom panels in the operation phase of the RI. The flow of data between levels represents an expectation that data management procedures are in place at national/local levels and able to produce good quality data for the ESFRI RI.

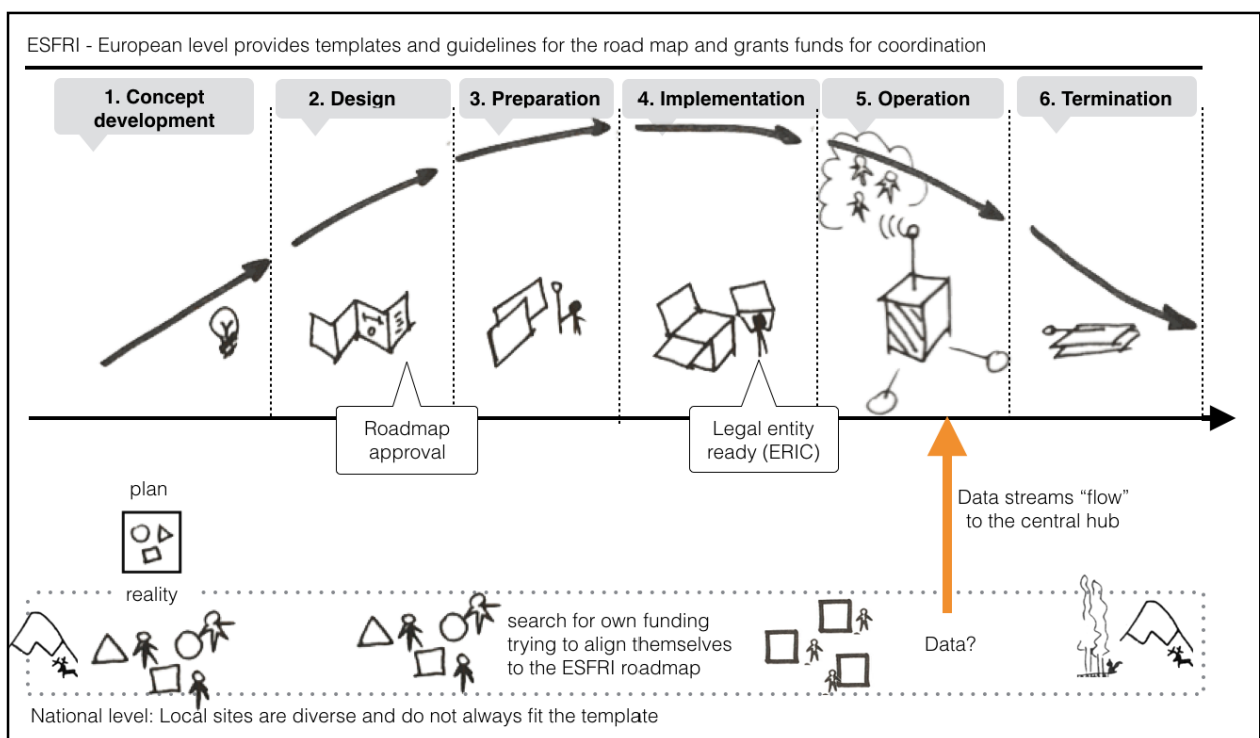


Figure 2. The upper panel shows a simplified view of the 30-year research infrastructure formation strategy adopted by ESFRI. The lower panel has been added by the MULTICS group to make visible local data management that requires not only attention but also funding if it is to design and operationalize data flow to a central hub. Source: MULTICS project - Visualization Andrea Botero.

This observation seems a particularly big assumption that is being made for the ecological/environmental RIs, given the ESFRI presentation of the upper panel only. Existing local and national infrastructure efforts are very heterogeneous and distributed and as yet often lacking in coordination both at the level of research sites and across research sites, and thus – when it comes to data management – are ‘bottom-heavy’ enterprises. Therefore, we set off to plan

a workshop that would pay attention to the national/local level as a starting point to discuss data management activities and to envision data integration across multiple boundaries and scopes.

The workshop was not planned as a traditional training session on data management, but rather as a space to collectively identify, reflect, and discuss data-related issues, resources, expectations and next steps. These topics need to be addressed and developed in order to achieve widespread, effective, and sustainable data management practices and procedures in the era of research infrastructures.

Planning the approach and content for the workshop was enhanced by the breadth of experience of our team that includes both long-term practical work with data management at LTER sites and research on data management and infrastructure work in ecological research networks and communities (see for example, Karasti and Baker 2004, 2008a, 2008b; Karasti et al. 2006, 2010). This focus is complemented by a socio-technical understanding of how infrastructures are embedded in local practices and technological arrangements; yet also connected reaching over multiple scopes, and how they emerge through longitudinal processes (see e.g.: Bowker et al. 2010; Karasti 2014). Based on this we tried to focus on the most essential data management concepts and how to present them in a palatable manner. Since data management discussions involve many ideas and technologies as well as social and institutional arrangements, workshops of one or more days may be overwhelming or incomprehensible to the uninitiated. Creating an informal atmosphere with selected content is important for generating lively discussions and thought-provoking examples that enable participants to consider how to define next steps for data management suitable to their particular circumstances. It was important for us to free individuals from the idea of seeking 'The Solution' to data management so they could grasp the diversity of data management activities, and would feel free to bounce ideas off each other, while considering what options would be appropriate for their circumstances. In planning the ways of working for the workshop, we drew on our background in the tradition of Participatory Design (see e.g. Simonsen and Robertson 2013).

Given the diversity of kinds of data work involved in INARI RI Ecosystems - spanning from nuanced hand-crafting to automated workflows as well as from novice to mature data management practices - we gave high priority to creating an appreciative atmosphere where participants could openly share their experiences and be accepted regardless of their current data arrangements. We conducted a pre-workshop survey to get an idea of the participants and their specific situations with data management. Survey responses (see next section for a brief summary) guided identification of and communication with those having specific commentaries to share. This pre-workshop communication helped some participants frame their brief presentations and begin developing an understanding of what they could contribute. Whether a participant decided they were comfortable with a five-minute commentary or a ten-minute presentation with slides, this proved to be an important first step in the engagement process. In addition, we solicited comments and feedback on our developing plans and discussed (or exchanged emails) with several workshop presenters and participants as the agenda evolved to fit our growing understanding of how to meet community needs.

The workshop was organized in five sessions over one and a half days with the first day focusing on data and data management and the second day highlighting the data landscape, that is, the larger context within which local data management takes place. Each session included short commentaries shared by the participants themselves to foster reflection around their examples

and experiences. We also arranged for short presentations by experts and invited guests, a walking tour of the station and field sampling as well as group work, interactive activities and discussions. The presentations and activities of all five sessions are briefly described below.

Following the workshop, materials including presentation slides, handouts, and photographs were posted online for access by participants so that those ready to use the vocabularies, concepts, and understandings of the data landscape would have materials at hand to refresh their memories and support their activities. A post-workshop survey was distributed, the results of which are summarized in the final section of this report.

## Pre-workshop survey - brief summary of the responses

To aid in the planning, workshop organizers designed and distributed a short survey to gather information about the state and experiences of current participants of the INAR RI Ecosystems components and other stakeholders (See Appendix 1 for survey questions). Examples and issues raised by the results of the survey served as basis to plan and structure the content and program of the workshop.

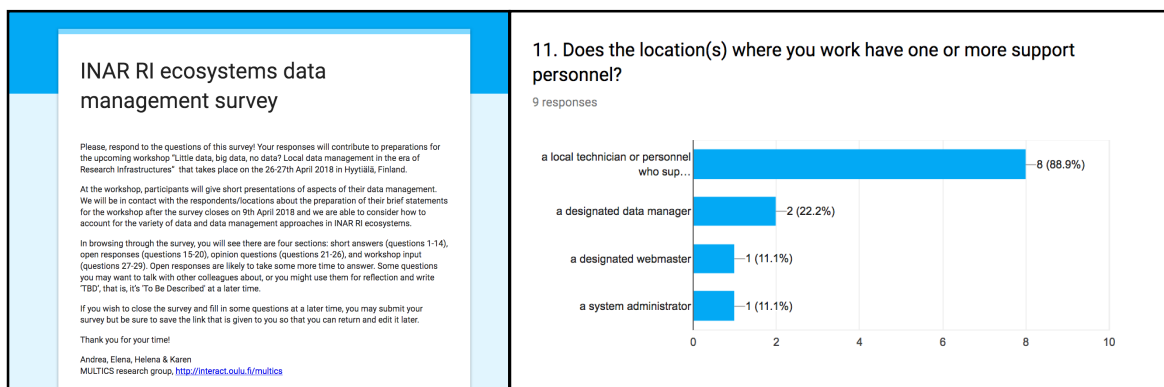


Figure 3. a. Details from the survey form; b. Responses to a pre-workshop survey question.

Insights from the pre-workshop survey in a nutshell:

- There exist a variety of 'data generators' in INAR-RI Ecosystems and their characteristics relate to the origins of data (Q3-4)
- Participants handle a wide variety of heterogeneous data, data formats, materials etc. (Q13)
- Some elements of data management are present, but none seem to be covering the whole spectrum of activities needed (metadata, controlled vocabularies & data dictionaries, data/metadata catalogs, data repositories, data policy/use statement, support personnel, file storage space, publishing data in conjunction with scientific publications) (Q5-14)
- What works well? Answers ranged from "nothing works well at the moment" to some aspects working and others not (yet) (Q19)
- Bottlenecks or problems: (Q20)
  - "No management plan, no guidelines, everyone for themselves attitude, no time for management"
  - "Some of the data is preserved by non-military service persons and they change once a year" – i.e. high turnover of temporary people



- *“We have tens of binders full of papers with old data, which are not digitized, if we only could have all the time in the world...”*
- *“There is no personnel for data management”*
- *“Relying and depending on a physical person to share the data”*
- *“Missing clear guidelines”*
- *“Documentation of many things afterwards or not at all. Ad hoc routines without considering long-term workload”.*
- *“Good tools for sharing large amounts of data directly from the station are lacking.”*

Responses bear similarities to those from a survey of FinLTSER in 2007/8 (briefly reported in Karasti 2009). An early study of the US LTER Information Management community (Karasti, Baker and Halkola 2006) also discusses data management work as dealing with and finding balances for a variety of issues even after decades of practice (see Figure 4).

Element	In practice
Data heterogeneity	“In general you are struggling with the diversity of different types of data.” (IM)
Data quality	“The QA is a big issue, in terms of like curatorship.” (IM)
Data description	“... they [scientists] realize that it will probably take 20 or 30% more time.” (S)
Open data sharing	“This was a new expectation, not just that data was managed well and documented but to be made readily available.” (IM)
Data standards	“The network has had a great influence, pushing forward a standardized approach to collecting metadata.” (IM)
Technology sustainability	“... the issue isn’t how you do it, it’s how do you maintain it and how do you make it so that it is easily maintainable.” (IM)
Information infrastructure	“And they [scientists] simply don’t appreciate the time and the energy and the effort required just to do the nuts and bolts maintenance.” (S)
Learning environment	“It’s been a massive process of sort of education and badgering...” (IM)

*Figure 4. Elements of LTER data stewardship illustrated by interview excerpts, based on an empirical study of US LTER network in 2002 (from Karasti, Baker and Halkola 2008, p. 348).*

## Day I (Sessions 1-3)

### Session 1: Introduction

The workshop opened with an overview of ongoing changes in environmental science, societal challenges, and technology developments that make issues of data management pressing to address.



Figure 5. a. Participants gathered at the old dining hall in Hyytiälä organized as the workshop venue; b. Helena Karasti introduced the background and aims of the workshop.

### Presentations

- Welcome words - Jaana Bäck, INAR RI Ecosystems
- Introduction to the workshop and pre-survey insights - Helena Karasti, MULTICS Project

### ACTIVITY BRIEF: Introductions

Those who generate, manage, and use data benefit from hearing about the emergence of new data-related roles. New data roles lead to changes in who is designated to carry out both traditional activities in support of research and new responsibilities associated with data sharing and data preservation. This part of the session was geared to build a shared and multifaceted image of the participants together with the situations and conditions of data management in the different locations of INAR RI ecosystems. To do that participants introduced themselves and their sites or organizations through a series of mobile and interactive activities where they moved themselves around the room in response to the following three key themes:

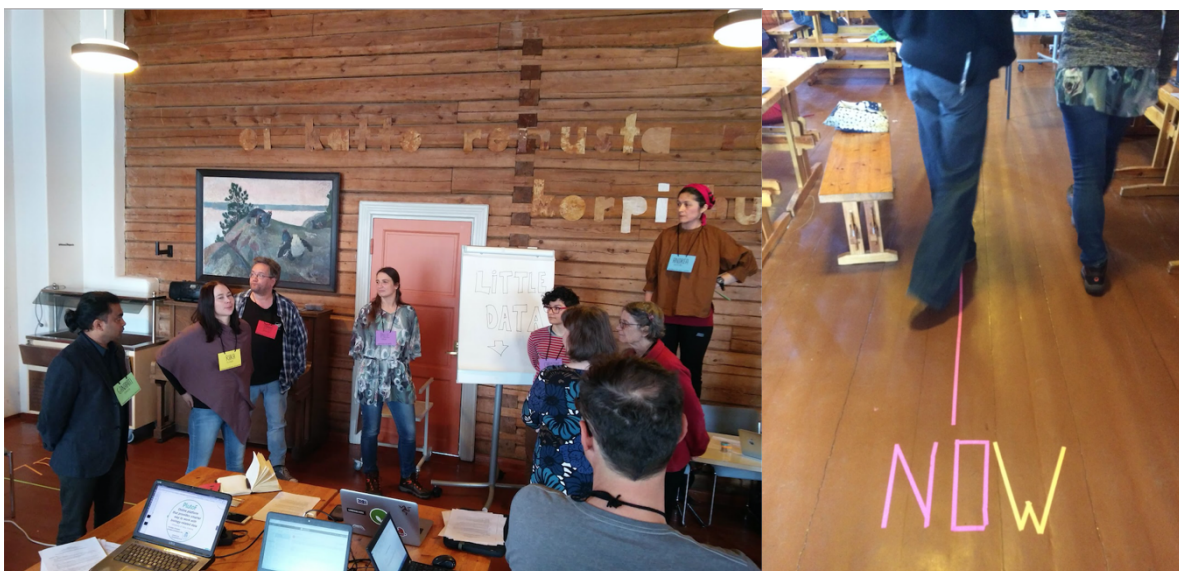


Figure 6. a. Participants introduce themselves while explaining why they located near the “little data” sign; b. Participants walk along a data management awareness timeline.

1) Signs marked with “big data”, “little data” and “no data” were located at different corners of the room. Participants were asked to position themselves according to how they would describe the

kind of data they worked with or generated (see Figure 6a). A subset of the participants introduced themselves (name, organization) and explained why they were standing in their particular location.

2) Signs marked with “No metadata”, “Local convention”, and “Metadata Standard” were located at different corners of the room. Participants were asked to position themselves near or far from them depending on whether the data they generate or use has associated metadata. Another subset of the participants introduced themselves (name, organization) and explained why they were standing in their particular location.

3) A long line made of tape was placed on the floor, traversing the room. It was marked at one end with the word ‘NOW’ and at the other end with ‘70s’ (see Figure 6b). Participants were asked to self-organize themselves along the line, standing at the time when they first encounter issues related to data management in their work. They were asked to talk with their neighbors to find the right spot in the timeline. The remaining participants introduced themselves (name, organization) and explained which year they were standing at and what was their situation with regard to data management.

Besides serving as an engaging approach to introductions, the exercises aimed to introduce some key ideas and basic vocabulary (e.g., ‘big data’, ‘metadata’, ‘local convention’) in an informal manner. For instance, with terms such as ‘big data’ that can refer to any number of characteristics of data such as the volume, variety, veracity, or velocity of capture, the first themed exercise exposed participants to the many kinds of data represented at the workshop. Further, the many metadata standards found in practice are typically a source of confusion yet there often exists an expectation that a first step requires identifying and using a single metadata standard. Introducing the concept of ‘local conventions’ also referred to as local ‘working standards’, is a starting point for discussion of how a local convention eventually can be mapped to different standards. Local conventions offer a first step toward coordinating description of data in a locally meaningful way while gaining experience in the use of metadata and the challenges involved in designing local processes that capture metadata.

- Commentary Session 1 - Jaana Bäck, INAR RI Ecosystems

## Session 2: What is data?

The submission of data to data facilities often assumes that data is well organized, free of errors, and accompanied by good documentation. Further, in practice, there are many kinds of data including long-term observations, campaign samplings, experimental arrangements, modeling, and data products where generation, organization, and handling are as yet ill-defined. As participants made brief statements about the data at their locations, we began building a vocabulary about the what, when, and how of data infrastructure.

### Presentations

- INAR RI Ecosystems data related insights from CSC - Jessica Parland, CSC
- What is data? A dataset? A data package? A data product? - Karen Baker, MULTICS Project



Figure 7. a. Workshop presentation; b. Workshop discussion.

### Examples from participants (selected on the basis of pre-survey responses)

We invited some participants to make short comments about the kind of data they have and the status of data management arrangements at their location. We used their own words from the survey as prompts for workshop discussion.

#### Example 1

*“no personnel for data management” and “data not described by metadata”*

Kotkanoja experimental field and Lintupaju buffer zone - LUKE, Jokioinen.

-Jaana Uusi-Kämpä (researcher)

#### Example 2

*“data described by metadata mostly”*

-SMEAR II Hyytiälä - Juho Aalto (data creator and data user)

#### Example 3

*“No metadata standard, but some sort of own that is easy to transfer to standard form” and “Data is mostly in excel-sheets located in a University server”.*

-Oulanka Research Station, Katja Sippola

#### Example 4

*“As a research technician, I do almost everything” and “We have tens of binders full of papers with old data, which are not digitized, if we only could have all the time in the world...”*

-Kevo Field station, Tommi Andersson

#### Example 5

*“What level of scale to use, or better how to define the boundaries of a data package when one has large environmental monitoring data”*

-SMEAR station - Pasi Kolari (data manager)

- Commentary Session 2 - Johannes Peterseil, Environment Agency Austria

### Walking tour to SMEAR II station

- Introduction to SMEAR stations - Jaana Bäck, INAR RI Ecosystems

Participants had a chance to visit the SMEAR II Station (Station for Measuring Ecosystem Atmosphere Relations) and to ask questions about the origin of the data, the types of data collected, the instrumentation present, and data handling at the station.



Figure 8. a. Participants exploring the forest and the SMEAR II tower; b. SMEAR station concept diagram.

## Session 3: What is data management?

At the source of the data, there are important responsibilities and opportunities today relating to data management and the growth of digital capabilities. In this session key data management activities and issues were explored by discussing examples, arrangements, and decision-making activities that accompany everyday data practices. Handouts with links to a few of the many online data management planning modules and education resources were also provided (see Appendix 4).

### Presentations

- General overview of data management at Center for Ecology - Sue Rennie, Centre for Ecology & Hydrology, UK
- What is data management? A database? A data system? - Karen Baker, MULTICS team

### Group work session

Participants divided into four groups (considering similarities in their configuration). Working around group tables fostered communication and sharing that spurred some participants in their thinking and exposed participants to a wide range of data-related concerns. The aim was to identify, first individually and then as a group, key activities of data management already happening at their locations. The participants discussed together about those they identified as important. The intention was for groups to create a list of priorities for data management (considering the time, resources, etc.), but we did not reach that level of detail. Most of the conversations centered on brainstorming around their current data management issues.

### Activity Brief: Local data management activities

Write one activity per post-it, identify if it is covered at your location or not. Use the handouts 3 and 4 as aids in your thinking. Materials: post-its, big sheets of paper, markers, handouts on vocabulary and example diagrams.



Figure 9. Working sessions by groups.

- Commentary Session 3 and Wrap Up - Sue Rennie, Centre for Ecology & Hydrology, UK

## Day II (Sessions 4-5)

### Session 4: What is data infrastructure?

#### Presentation

- Linking to the European LTER Data Network - Johannes Peterseil, Environment Agency Austria

#### Examples

- Pluto F, Kristjan Adojaan University of Tartu
- CSC data linking resources, Jessica Parland, CSC
- Linking to ICOS Pasi Kolari, Hyytiälä, SMEAR
- “Which repository to use?” Christine Ribeiro, University of Helsinki

#### Presentation

- What is data infrastructure? When and how do data infrastructures become research infrastructures? - Elena Parmiggiani and Karen Baker, MULTICS Project



Figure 10. a. Guest Speaker Johannes Peterseil's presentation; b. Working pictorial of data groupings.

### Group work session

Participants were divided into three categories: people in charge of heterogeneous data at a research station, researcher or research group working with data, and people in charge of data coming from highly instrumentalized stations. Participants were asked to review the basic set of data landscape elements identified during the workshop and draw first their own current configuration, and then their view of the near future. The exercise aimed to allow participants to imagine where they want to be in next years in terms of their own data management. In particular this exercise would make visible (for all) the relations between locations of data generation and aggregation, and what those might mean in a shared future INAR RI Ecosystem.

### Activity Brief: Data landscape mapping

Use the “template components” to make one concrete example of data infrastructure at your site(s) or based on your own situation locally. Add arrows for the connections you know and dotted arrows for the ones you "think" should be developed. Using the list of data management activities that you made yesterday, add the activities you’ve identified associated with data infrastructure (what you have, what you do not have)

Materials: Paper, pens, Handout 5: Data Management Activities & Vocabularies and Handout 6: Configurations of Data Infrastructure that references the field, local, and remote data arenas.



Figure 11. Group work discussions, activities, and reflections on the data landscape representations.

- Group presentations and reflections

## Session 5: Workshop Wrap Up and Next Practical Steps Forward

Participants agreed that the workshop was useful to start a collective conversation on the role of data management in scientific work in the environmental sciences in Finland. The discussions served to outline a few steps towards improving and developing data management plans and strategies.

Following a suggestion by Jaana Bäck, participants agreed on some practical next steps:

- 1) All attending sites work on creating or updating their profile metadata in DEIMS
- 2) Data management plans will be taken up in the next RESTAT meeting
- 3) INAR RI Ecosystems will start compiling a metadata catalogue of available data sets

4) A mailing list for issues relating to data management will be created so participants and other interested people can continue receiving and giving peer support around data management topics.



# Post workshop survey results

## What was learned? Insights into data management

During the two days, the participants considered a variety of issues relevant to data management and jointly discussed tools and strategies to link everyday work practices with data production supporting both traditional and new approaches of scientific knowledge production.

Participants had the opportunity to share their experiences and examples, and to learn from others' cases, "listening to the different 'cases' in general was rewarding" and even so that most interesting in the workshop was "other people's problems with data". Participants found it "very useful to meet people who deal with same issues and work to solve those issues" and gained insight into the existing heterogeneities within the INAR RI Ecosystems: sites/stations/institutes and their issues, data, methods, instruments, standards.

*"the different ways people are dealing with data management"*

*"types of data are different"*

*"we have "data factories" (like SMEAR ...) and more discrete and less intensive data producers"*

*"the fact that the data is scattered and heterogenous"*

*"the variety of the stations became clearer, as did the data management infrastructure and needs concerning it"*

*"the people and instrumentation at the sites"*

*"better understanding of the site issues and the importance of the metadata"*

It was successfully conveyed that there is not 'One Solution' to data arrangements and that different approaches have legitimate reasons.

*"there are many options how to handle the data management, not one single right solution"*

It was also recognized that the state of data management varies between participants' locations. Those participants with less advanced data management were assured/comforted that they were not the only ones feeling unsure of next steps and lacking resources for data management.

*"variety of the levels of data management in different sites and organizations, and even inside the organizations"*

*"to notice that this is not a clear thing to everyone and this really will need time and work to be done"*

The participants became more aware of how to connect with the European level.

*"aware of the resources open ... at a European level"*

*"It was interesting to get to know about DEIMS, which wasn't familiar to me before but will be useful resource/platform for me and my group in many ways."*

*"there are tools / help / standards"*

All in all, workshop participants gained a broader understanding of data management. It is not only a technical job, but a deeply socio-technical effort that necessitates a lot of communication, collaboration, planning, and eventually resources.

*“ecosystem RI community raised its readiness level to improve the data management quality and got skills to think about data management in a more comprehensive way”  
“it requires communication, thinking and planning to set up a data management system”*

Towards the end of the workshop, participants – drawing on the shared data management vocabulary, concepts, and strategies – started to think about how to shape the development of data arrangements as the final group work assignment since there are many ways to configure the multiple components of data infrastructure. They got ideas how to continue their data management work and felt more at ease with their jobs.

*“ways to improve visibility of our monitoring”  
“Improvement of issues revealed in the last group task.”  
“I feel that after the workshop I have good tools to go on with sharing our data. This does not feel as huge effort as it used to feel before, but something manageable.”*

## **What is important for continuing the process?**

One of the most important actions to work on is the promotion of awareness and recognition of the importance of data management on national and local levels. At the moment the degree of awareness is not even and there are missing resources and key responsibilities that are unrecognized and undelegated at many sites.

*“the resources to do the data management work seem to be restricted and data management tasks seem to fall from the table with least priority in the middle of all other tasks and duties. There should be some kind of change in thinking.”  
“It would be of importance also to find ways to get the ear and mandate (i.e. resources and possibilities) to take these issues forward. Help in communicating the importance.”*

Participants identified community building and more collaboration for data management on the national level as a high priority, an important step for moving forward. A data management group is relatively easy to start. For example, a group may be begun by establishing clear communication channels. However, to create a community that is sustained over time, requires development of goals, activities, and governance that contribute to a process that helps shape and consolidate the community.

*“Community building would be valuable. People will need a lot of peer support and empowerment to be able to go on in this very heterogeneous and scattered network.”  
“A regular communication channel for data managers and the opportunity to meet and share/learn from each other.”  
“Regular meeting with the network and development of common tools and solutions, sharing protocols, procedures and methods, templates, guidelines and also technical tools and platforms...”*

*"I'm looking forward the meeting in Konnevesi; we should establish a common ground for the data management work for the research stations."*

*"It might be good if somebody would follow how the things start to move in each field/laboratory/institute/project partner? Are there still barriers to start the progress? In tough questions it might be nice to get support from someone."*

*"There is still a way to go to develop the connected data management within Finland but ... the group that formed at Hyytiälä has every chance of success."*

Attainment of ambitious scientific objectives by a data generating project today, requires explicit recognition that there are different kinds of data that may require different kinds of care. As multi-group data management is embedded in different kinds of organizational and institutional settings,

*"clearly there are different groups that need different approaches."*

## Post workshop reflections and recommendations

Research infrastructure development is ongoing, at the same time our understanding of it is emergent and incomplete. The metaphor of 'doing research on a ship that is being built while in use' continues to be an appropriate image to reflect on the tasks ahead for INAR RI and others involved.

At the local level there are many interesting directions to continue the work with data. We second the proposed next steps that contribute to the community process as discussed during the workshop. This process should result in more clarity on the assignment of roles and responsibilities regarding data management. Besides the mailing list, other tools like a collective mapping of existing assets and data management practices can be carried out. This is needed to make visible what exists, give it name and recognition as well as identifying missing components with more specificity. Constructing a community focusing on data management can enable people to consolidate a shared vocabulary, relay on their collective and distributed expertise in more explicit ways and maintain the documentation and development work that needs to be done locally.

Interdisciplinary work would be needed to address the complexity and dynamics of establishing a functional, stable and sustainable local data environment and to maintaining it within the context of a larger data landscape. Therefore, it would be important that INAR RI Ecosystems stakeholders take explicit actions to raise awareness with the national funders about the central importance of data management in research infrastructure development. The message should clarify that it is not only important, but that data management requires resources/funding at many levels and that it is not supported currently by European research funding nor infrastructure sources.

There is also a need for recognizing and bridging gaps. Going back to the observation that the workshop preparations started with (see Figure 2), there is a need for bridging the gap between the European and the national/local data management levels. Finland is definitely not the only member state/country where data management procedures capable of producing good quality data for sharing and preservation are not yet in place. At this time, each member state must

struggle with figuring out how to address and support national/local data management on their own. Devising ways and processes at the European level for supporting and coordinating with data activities at the national level would be welcomed.

## References

Bäck, J., Kaukolehto, M., Rasilo, T., Pumpanen, J., Tuittila, E., Paavola, R., Laatola, K., Karjalainen, J., Suominen, O., Forsius, M., Kolström, T., Mäkipää, R., Lohila, A., Pursula, A., Juurola, A., & Heiskanen, J. (2017). *Terrestrial Ecological and Environmental Research Infrastructures in Finland - Analysis of the Current Landscape and Proposal for Future Steps (White Paper)*. Helsinki, Finland: INARI RI Ecosystems Project.

Bowker, G. C., Baker, K. S., Millerand, F., & Ribes, D. (2010). Toward information infrastructure studies: Ways of knowing in a networked environment. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International handbook of internet research* (pp. 97-117). Dordrecht, Netherlands: Springer.

ENVRiplus (2017). Further Integration of Research Infrastructures Related to Terrestrial Ecosystem Research, Including Recommendations on Co-locating Research Sites on National and International Level. Work Package 12 Report D12.3. Available at <http://www.envriplus.eu/wp-content/uploads/2015/08/D12.3-.pdf>

INAR RI (nd). Institute for Atmospheric and Earth System Research. Available at <https://www.helsinki.fi/en/inar-institute-for-atmospheric-and-earth-system-research/infrastructure/national-research-infrastructure>

INAR RI Ecosystems (nd). INAR RI Ecosystems – Integrated Atmospheric and Earth System Science Research Infrastructure. Available at [http://www.syke.fi/en-US/Research\\_Development/Research\\_and\\_development\\_projects/Projects/Finnish\\_LongTerm\\_SocioEcological\\_Research\\_network\\_FinLTSER/INAR\\_RI\\_Integrated\\_Atmospheric\\_and\\_Earth\\_System\\_Science\\_Research\\_Infrastructure](http://www.syke.fi/en-US/Research_Development/Research_and_development_projects/Projects/Finnish_LongTerm_SocioEcological_Research_network_FinLTSER/INAR_RI_Integrated_Atmospheric_and_Earth_System_Science_Research_Infrastructure).

Karasti, H., & Baker, K. S. (2004). *Infrastructuring for the Long-Term: Ecological Information Management*. Proceedings of the 37th Hawaii International Conference on System Sciences 2004.

Karasti, H., Baker, K. S., & Halkola, E. (2006). Enriching the Notion of Data Curation in E-Science: Data Managing and Information Infrastructuring in the Long Term Ecological Research (LTER) Network. *Computer Supported Cooperative Work (CSCW) – An International Journal*, 15(4), 321-358.

Karasti, H., & Baker, K. S. (2008a). Digital data practices and the long term ecological research program growing global. *International Journal of Digital Curation*, 3(2), 42-58.

Karasti, H. & Baker, K. S. (2008b). *Community Design: Growing One's Own Information Infrastructure*. In Proceedings of the 10th Participatory Design Conference (PDC'08), Oct 1-4 2008, Bloomington IN. Computer Professionals for Social Responsibility, pp. 217-220.

Karasti, H. (2009). FinLTSEr Network formation and the state of information management. *Ecological Circuits - Eco-informatics in Action, Issue 2, pp. 15-18.*

Karasti, H., Baker, K. S. and Millerand, F. (2010). Infrastructure Time: Long-Term Matters in Collaborative Development. *Computer Supported Cooperative Work (CSCW) - An International Journal*, 19(3-4), 377-415.

Karasti, H. (2014). Infrastructuring in Participatory Design. Participatory Design Conference. In Proceedings of the 13th Participatory Design Conference (PDC '14), 6-10 October 2014, Windhoek, Namibia. Research Papers - Volume 1. ACM, New York, NY, USA. pp. 141-150.

Simonsen, J., & Robertson, T. (Eds.). (2013). *Routledge International Handbook of Participatory Design*. Routledge.

# Acronyms

ACTRIS - European Research Infrastructure for the observation of Aerosol, Clouds, and Trace gases

AnaEE - European Research Infrastructure for Analysis and Experimentation on Ecosystems

CSC – Finnish IT Center for Science

DEIMS - Dynamic Ecological Information Management System

DEIMS-SDR - Dynamic Ecological Information Management System - Site and Dataset Registry of the European Long Term Ecological Research Network

DEIMS-US LTER - Drupal Ecological Information System of the US Long Term Ecological Research Network

eLTER - European Research Infrastructure for LTER-Europe

ESFRI - European Strategic Forum on Research Infrastructures

EUDAT - European Data Infrastructure

FinLTSER - Finnish Long-Term Socio-Ecological Research Network

FIRI - Finnish Research Infrastructure program

FMI - Finnish Meteorological Institute

ICOS - Integrated Carbon Observation System

ILTER - International Long Term Ecological Research Network

INAR RI - Institute for Atmospheric and Earth System Research

INAR RI Ecosystems - Component of INAR RI that studies major ecosystems in Finland

LTER-Europe - Long-Term Ecosystem Research in Europe Network

LUKE - Natural Resources Institute Finland

NRI - National research infrastructure

RESTAT - Finnish Research Stations involved in INAR RI Ecosystems

RI - Research Infrastructure

SMEAR - Stations Measuring Atmosphere Ecosystem Relationships

SYKE - Finnish Environment Institute

US LTER - United States Long Term Ecological Research Network

# Appendices

- Appendix 1. Agenda
- Appendix 2. Pre survey
- Appendix 3. Post survey
- Appendix 4. Handouts

## Appendix 1. Agenda

### **AGENDA: Little data, big data, no data?** **Data management in the era of research infrastructures**

26-27<sup>th</sup> April 2018; Hyytiälä, Finland  
Organisers: INAR RI ecosystems and MULTICS project

#### Thursday 26 April

7.00-8.00 Breakfast

#### 8.30-09.50 **Session 1: Introduction**

- Welcome words - *Jaana Bäck, INAR RI ecosystems* (10 min)
- Introduction to the workshop and pre-survey insights - *Helena Karasti, MULTICS Project* (15 min)
- INAR RI ecosystems! Participants will introduce themselves and their sites/organizations to set the stage for discussions of data management (60 min)
- Commentary - *Jaana Bäck, INAR RI ecosystems* (5 min)

10.00-10.10 Health break

#### 10.10-11.00 **Session 2: What is data? - Part 1**

- INAR RI ecosystems data related insights from CSC - *Jessica Parland, CSC* (5 min)
- What is data? A dataset? A data package? A data product? - *Karen Baker, MULTICS Project* (15 min)
- Examples from INAR RI ecosystems participants: *Jaana Uusi-Kämppe, Jokioinen, LUKE; Juho Aalto, Hyytiälä - Helsinki University; Riika Ylitalo, Finnish Meteorological Institute* (5+5+5 + 5 min questions)
- Metadata story - *Karen Baker* (10 min)

11.00-12.00 **Lunch**

#### 12.00-13.30 **Session 2: What is data? - Part 2**

- Collective Clinic - *One example from Oulanka research station by Katja Sippola will be given some concrete tips for metadata by the group* (45 min)
- Examples from INAR-RI ecosystems - *Tommi Andersson, Kevo research station, Pasi Kolari, Hyytiälä, SMEAR* (10 min + 10 min)
- Commentary - *Johannes Peterseil, Environment Agency Austria* (5 min + 5 min group reflections)
- Introduction to SMEAR stations - *Jaana Bäck, INAR RI ecosystems* (15 min)

13.30-13.50 **Coffee break**

13.50-15:30 **Walking tour to SMEAR station (“talking data”) + Group Photo**

#### 15.30-17:30 **Session 3: What is data management?**

- General overview of data management at Center for Ecology: *Sue Rennie, Centre for Ecology & Hydrology, UK* (20 min + 10 min)
- What is data management? A database? A data system? - *Karen Baker, MULTICS Project* (10 min)

- *Handouts 1 & 2*
- Group formation (10 min)
- Group work session (Mapping session - *Facilitator, Andrea Botero, MULTICS Project* (20 work + 10 presentation + 20 work + 10 presentation)
  - *Handouts 3 & 4*
- Commentary and wrap up - *Sue Rennie and Multics* (10 min)

18.30 Dinner (Old Dining Hall)

20.00 Sauna and kota

### Friday 27 April

7.00-8.00 Breakfast

#### 8.30-9:30 **Session 4: What is data infrastructure? - Part 1**

- Linking to the European LTER Data Network - *Johannes Peterseil, Environment Agency Austria* (20 min + 10 min discussion)
- Pluto F, *Kristjan Adojaan University of Tartu* (5 min) and CSC data linking resources, *Jessica Parland, CSC* (5 min)
- Linking to ICOS *Pasi Kolari, Hyytiälä, SMEAR* (10 min)
- *Example*, *Christine Ribeiro, University of Helsinki* (5 +5 min)

9.30-9.40 Health break

#### 9.40-10:40 **Session 4: What is data infrastructure? - Part 2**

- What is data infrastructure? When and how do data infrastructures become research infrastructures? - *Elena Parmiggiani and Karen Baker, MULTICS Project* (15 min)
- Group work session - *Facilitators Jessica Parland, Johannes Peterseil, Sue Rennie, MULTICS* (30 min)
  - *Handouts 5 & 6*
- Group presentations and reflections (15 min)

10:40-11:00 **Session 5: Wrap-up and next steps- Collective reflection. What have we accomplished? Where do we want to go?-** *Jaana Bäck,; Johannes Peterseil, Sue Rennie, MULTICS project, and all participants* (20 min)

11.00-12.00 Lunch

12.00 **Departure**



## Appendix 2. Pre survey

### INAR RI ecosystems data management survey

---

Please, respond to the questions of this survey! Your responses will contribute to preparations for the upcoming workshop "Little data, big data, no data? Local data management in the era of Research Infrastructures" that takes place on the 26-27th April 2018 in Hyytiälä, Finland.

At the workshop, participants will give short presentations of aspects of their data management. We will be in contact with the respondents/locations about the preparation of their brief statements for the workshop after the survey closes on 9th April 2018 and we are able to consider how to account for the variety of data and data management approaches in INAR RI ecosystems.

In browsing through the survey, you will see there are four sections: short answers (questions 1-14), open responses (questions 15-20), opinion questions (questions 21-26), and workshop input (questions 27-29). Open responses are likely to take some more time to answer. Some questions you may want to talk with other colleagues about, or you might use them for reflection and write 'TBD', that is, it's 'To Be Described' at a later time.

If you wish to close the survey and fill in some questions at a later time, you may submit your survey but be sure to save the link that is given to you so that you can return and edit it later.

Thank you for your time!

Andrea, Elena, Helena & Karen  
MULTICS research group, <http://interact.oulu.fi/multics>

---

#### Short answers section:

---

#### 1. Name(s) and email(s) of the respondent(s) \*

You can fill this survey together with a colleague if it is easier to answer jointly.

Short answer text  
.....

#### 2. With which INAR RI ecosystems partner are you associated? \*

Short answer text  
.....

#### 3. With what INAR RI ecosystems data (generating) location(s) are you associated? By location we mean, for example, a research site, a research station, an experimental site, a university, a repository or an archive?

Long answer text  
.....

**4. How would you describe your primary role(s) with the INAR RI ecosystems data?**

- Supporting field work of researcher participants at a particular field station
- Generating/analyzing in situ data at a field station on a continuing basis
- Generating/analyzing in situ data at a field station on a short-term basis
- Generating/analyzing data from some type of experimental location
- Generating/managing data from an instrumented research tower/station
- Aggregating/preserving/disseminating data as a repository platform
- Other...

**comment:**

Long answer text

---

**5. Are the datasets you work with described by metadata?**

Long answer text

---

**6. Do you use metadata standards (local or formal)? If so, please describe.**

Long answer text

---

**7. What controlled vocabularies or data dictionaries are used for your data? If online, provide links.**

Long answer text

---

⋮

**8. If you contribute to or work with a catalog of data (metadata or datasets), please name them or provide the link(s):**

Long answer text

---

9. Is there a particular data repository where you currently or in the future plan for data to be made available?

Long answer text

---

10. If your work is framed by a data policy and/or data use statement, please describe. If online, provide the link(s):

Long answer text

---

11. Does the location(s) where you work have one or more support personnel?

(please check all that apply)

- a local technician or personnel who support those working in the field
- a designated data manager
- a designated webmaster
- a system administrator
- Other...

comment:

Long answer text

---

12. Do(es) the location(s) provide researchers file storage space?

- Yes
- No

⋮

comment:

Long answer text

---

13. In your work, please indicate materials, data, data formats, and other categories that are used.

(please check all that apply)

- Non-digital unstructured (e.g. hand-written notes, sketches, logs, figures, maps)
- Digital sources unstructured sources (word files, ppt, etc)
- Physical samples (e.g. biological, organic, inorganic)
- Digital still images
- Digital video and audio recordings
- Field observations and measurements
- Spreadsheets and other flat files (e.g. ascii, binary)
- Databases (relational or other)
- Data from in situ sensors (e.g. streaming, models)
- Data from satellites (e.g. streaming, models)
- Survey data involving human subjects
- Spatial data (e.g. GIS)
- Long-term data (e.g. time-series, monitoring)
- Data from computer simulations and/or models
- Data from drones or other Remote Operating Vehicles (ROV)
- Classified data (e.g. military)
- Other...

⋮

**comment:**

Long answer text

---

#### 14. What do you do when you publish research based on data?

(please check all that apply)

- Include all the data in the publication
- Include all the data elsewhere such as in a technical report or on a website
- Include an example, a subset, or a summary of the data
- Provide an online link using a unique identifier (e.g. a DOI)
- Not applicable
- Other...

**comment:**

Long answer text

---

#### Open Response Section

---

#### 15. How are (your) data managed in conjunction with projects, programs, and locations?

Long answer text

---

#### 16. What individual and collective guidelines (e.g. agreed practices, procedures, standards) are in use at the location(s) for data collection, quality assurance, storage and long-term archiving?

Long answer text

---

17. How are data saved, where are they located, and how are they accessed?

Long answer text

---

18. How are data shared with partners, programs or outside archives?

Long answer text

---

19. Describe one or two aspects of data management that work well at your location.

Long answer text

---

20. Describe one or two bottlenecks or problems relating to data management at your location.

Long answer text

---

### Opinion Section

---

21. How important is it to make data reusable, that is, well-documented with metadata, clear data use licensing information, information on provenance, and consistent with domain-relevant community standards?

	1	2	3	4	5	6	7	
Not very important	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very Important

22. How challenging is it to make data reusable, that is, well-documented with metadata, clear data use licensing information, information on provenance, and consistent with domain-relevant community standards?

1      2      3      4      5      6      7

Very easy                          Very difficult

23. How important is it for researchers to be able to find what data are available from data repositories?

1      2      3      4      5      6      7

Not important                          Very important

24. How challenging is it for researchers to be able to find what data are available from data repositories?

1      2      3      4      5      6      7

Very easy                          Very difficult

25. How important is it to transform the conduct of science through an increasing number of shared/open data sets/sources?

1      2      3      4      5      6      7

Not important                          Very important

26. How challenging is it to transform the conduct of science through an increasing number of shared/open data sets/sources?

1      2      3      4      5      6      7

Very easy                          Very difficult

## Workshop related input

---

**27. Who will represent your INAR RI ecosystems partner/location at the data workshop in Hyytiälä?**

Short answer text

---

**28. What topics would you like to have addressed in the workshop, given the overall theme of the workshop?**

Long answer text

---

**29. Other comments and concerns?**

Long answer text

---



## Appendix 3. Post survey

# Little data, Big data, No data? In the era of research infrastructures - Closing workshop survey

---

Thank you for participating so actively in the workshop. We ask you now to give us some feedback on the workshop and some general thoughts that are useful for our Multics project research including our planning for future data management workshops.

---

Your role:

Short answer text

---

Your name (optional)

Short answer text

---

1. What are the key thoughts or insights that occurred to you at the workshop?

Long answer text

---

2. What was most interesting to you in the workshop content/sessions?

Long answer text

---

### 3. What worked well at the workshop?

Long answer text

---

⋮

### 4. What did not work so well at the workshop?

Long answer text

---

### 5. What workshop outcomes would you like to see happen?

(e.g. a mailing list for data management issues? etc.)

Long answer text

---

## Appendix 4. Handouts

### Handout1: Some Centers and Repositories

- **US LTER**

Palmer LTER site-based data system

<http://pal.lternet.edu/data>

US LTER Network Information System

<https://portal.lternet.edu/nis/home.jsp>

DataOne Ecological Domain Aggregator

<https://search.dataone.org/#data>

- **Instrumented national networks and/or central facilities**

International FLUXNET

<https://fluxnet.fluxdata.org/about/>

Global SMEAR

<https://www.helsinki.fi/en/inar-institute-for-atmospheric-and-earth-system-research/infrastructure/global-smear>

- **National & European Research Infrastructures**

ICOS Integrated Carbon Observation System

<https://www.icos-ri.eu/>

<https://eudat.eu/communities/integrated-carbon-observation-system>

PEEX Paneuropean Experiment

<https://www.atm.helsinki.fi/peex/>

ACTRIS Aerosol, clouds, and trace gases research infrastructure

<https://www.actris.eu/>

ANAEE Analysis and Experimentation on Ecosystems

<https://www.anaee.com/>

eLTER Long-Term Ecosystem Research in Europe

<http://www.lter-europe.net/elter/data>

euDat Research data services

<https://eudat.eu/data-management>

CSC Finnish IT Center for Science, supercomputing center

<https://www.csc.fi/>

### Some Registries

DEIMS-SDR Repository for Research Sites and Datasets

<https://data.lter-europe.net/deims/>

Re3data registry for repositories

<http://www.re3data.org>

INSPIRE registry

<http://inspire.ec.europa.eu/registry>

## Handout 2: Some Online Data Management Resources

- Data Management Training (DMT) Clearinghouse, an online registry of learning resources of ESIP, USGS, DataOne, ICSU, Digital Preservation Network  
<http://dmclearinghouse.esipfed.org>
- ESIP commons: Data Management Short Courses  
<http://commons.esipfed.org/datamanagementshortcourse>
- USGS DM Training Modules  
<https://www2.usgs.gov/datamanagement/training.php>  
<https://my.usgs.gov/confluence/display/cdi/Data+Management+Training+Modules>
- DataOne: How to manage ecological data  
<https://www.dataone.org/education-modules>  
<https://dataoneorg.github.io/Education/>  
<https://www.dataone.org/esa-2011-how-manage-ecological-data-effective-use-and-re-use>
- Metadata Standards Directory
  - Alex Ball, Metadata Standards Directory. Research Data Alliance.  
[https://www.youtube.com/watch?v=Lh8w2\\_TpFP8](https://www.youtube.com/watch?v=Lh8w2_TpFP8)  
<http://rd-alliance.github.io/metadata-directory/standards/>
- Controlled Vocabularies
  - NISO. (2017). *National Information Standards Organization, Issues in Vocabulary Management. Report TR-06-2017.*  
<https://www.niso.org/standards-committees/vocab-mgmt>
  - EcoPar Parameters and Methods for Ecosystem Research Monitoring  
<http://www.ufz.de/lter-d/index.php?en=42566&contentonly=1>
  - SeaDataNet Common Vocabularies  
<https://www.seadatanet.org/Standards/Common-Vocabularies>
  - NISO. (2017). *National Information Standards Organization, Issues in Vocabulary Management. Report TR-06-2017.*  
<https://www.niso.org/standards-committees/vocab-mgmt>
  - NOAA feature type conventions
    - <https://www.nodc.noaa.gov/data/formats/netcdf/v1.1/#templatesexamples>
      - feature type templates
      - Point (CDL template-point)
      - Timeseries (CDL template – orthogonal)
        - (CDL template – incomplete)
      - Trajectory (CDL template – incomplete)
      - Profile (CDL template – orthogonal)
        - (CDL template – incomplete)
      - TimeSeries Profile
      - Trajectory Profile
      - Swath
      - Grid

- Data Carpentry Lessons  
<http://www.datacarpentry.org/lessons>
- EDI: Data Management  
Youtube: <https://www.youtube.com/channel/UCNZoWPaMG6lkEiH8xRNnrr>
  - How to clean and format data using R, OpenRefine, and Excel
  - Creating ‘clean’ data for publication
  - Drupal Ecological Information Management System (DEIMS)
  - Information management and technology at the Virginia Coast Reserve (VCR)
  - Using the PASTA + Search API and building a local data catalog
- EDI: Five phases of data publishing  
<https://environmentaldatainitiative.org/resources/five-phases-of-data-publishing/>
- EDI: EML Metadata template  
<https://github.com/EDIdorg/MetadataTemplates>
- UK DM Training on managing data  
<https://www.ukdataservice.ac.uk/manage-data>
- Repository Registry  
<https://www.re3data.org>
- Cook, R. B., R. J. Olson, P. Kanciruk, and L. A. Hook. 2001. Best practices for preparing ecological and ground-based data sets to share and archive. *Ecol. Bulletins* 82:138-141.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 160018.

## Data Management Plans

### Guides to Data Management

- Oakridge National Lab  
<https://daac.ornl.gov/PI/plan.shtml>
- MIT “Data Planning Checklist”.  
<HTTP://LIBRARIES.MIT.EDU/GUIDES/SUBJECTS/DATA-MANAGEMENT/CHECKLIST.HTML>
- JHU Data Management Questionnaire.  
<HTTP://DMP.DATA.JHU.EDU/ASSISTANCE/GUIDANCE-ON-WRITING-DATA-MANAGEMENT-PLANS/>
- Australian National Data Service: data management for Researchers.  
<HTTP://ANDS.ORG.AU/RESEARCHERS/MANAGE-DATA.HTML>
- Digital Curation Centre: Data Management Plans.  
<HTTP://WWW.DCC.AC.UK/RESOURCES/DATA-MANAGEMENT-PLAN>

### Guides to Data Curation

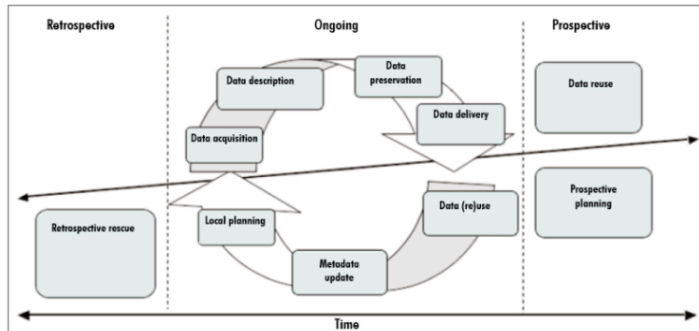
- UK Digital Curation Center  
<http://www.dcc.ac.uk/training/train-the-trainer/dc-101-training-materials>

## Sample Data Management Plans

- Natural Science Data Management plan examples:  
[HTTP://WWW.ICPSR.UMICH.EDU/ICPSRWEB/ICPSR/DMP/RESOURCES.JSP#A06](http://www.icpsr.umich.edu/icpsrweb/icpsr/dmp/resources.jsp#a06)
- U-Wisconsin:  
[HTTP://RESEARCHDATA.WISC.EDU/MAKE-A-PLAN/EXAMPLES/](http://researchdata.wisc.edu/make-a-plan/examples/)
- UC San Diego: sample data management plans spanning multiple NSF and NIH directorates  
[HTTP://IDL.UCSD.EDU/DATA-CURATION/EXAMPLES.HTML](http://idl.ucsd.edu/data-curation/examples.html)
- DataONE sample plans. Available:  
[HTTP://WWW.DATAONE.ORG/DATA-MANAGEMENT-PLANNING](http://www.dataone.org/data-management-planning)
- US LTER site profiles:  
<https://lternet.edu/site/>

# Handout 3: Session 3

## 1) Extended Temporal Horizon of Data Management/Data Curation



Karasti, H., & Baker, K. S. (2008). Digital data practices and the long term ecological research program growing global. *International Journal of Digital Curation*, 3(2), 42-58.

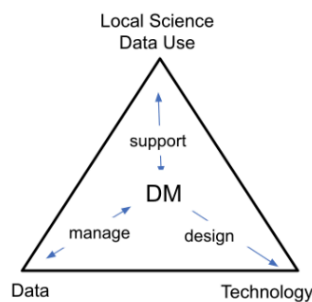
## 2) Example U.S. LTER Key Site-Based Data Activities

Activity	LTER characteristics
Data planning - local	Prospective in terms of fieldwork and local data use Site-specific ecological and social science data Largely non-reproducible observational but also experimental data Heterogeneous and complex data (sets) Largely manual data taking
Data acquisition	Attending to ongoing collection and updates over time to dynamic datasets Digital data record of experiments and process studies Retrospective recovery of legacy datasets in contemporary digital form
Data description	Local data storage and preservation Initial, intensive data & dataset description Continuing metadata description Multi-site data category building Local controlled vocabularies and dictionaries
Data use	Short-term analysis and use Site-based monitoring and innovative science Long-term network science Metadata update
Data delivery	Open, public access to data and metadata two years after collection Web interfaces for online data delivery Exchange with network partners and archives
Data reuse	Appropriate data presentation for direct use Appropriately contextualized delivery for data selection/integration Data preparation/structuring for data interoperability Unanticipated data uses
Data planning - global	Prospective in terms of data preservation Multi-community metadata standards making

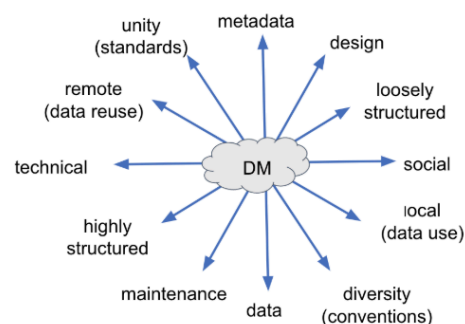
Table 2. Salient characteristics of LTER data activities.

Karasti, H., & Baker, K. S. (2008). Digital data practices and the long term ecological research program growing global. *International Journal of Digital Curation*, 3(2), 42-58.

## 3) U.S. LTER Examples DM Role



## 4) Balancing Tensions



-Karasti, H., & Baker, K. S. (2008). Digital data practices and the long term ecological research program growing global. *International Journal of Data Curation* 3(2), 42-58.

## Handout 4: Session 3

# Potential Local Data Management Activities

- Identify local data coordinator/manager
- Consider local and remote resource options
- Identify how your data and information will be made publicly available
- Project catalog
- Personnel names and contacts
- Major partners names and contacts
- Keyword list(s)
- Sampling location coordinates and maps
- Field notebook digitization
- Methods manual
- Identification of local metadata conventions
- Metadata catalog
- Dataset links in metadata catalog
- Data parameter dictionary
- Unit dictionary
- Coding scheme dictionaries
- Address data quality
- Create a Community of Practice w/annual meetings

**\*\* all items publicly available\*\***



## Handout 5: Session 4

### Potential DM Activities with Remote Connectivity

- Consider local and remote resource options
- Identify how your data and information will be made publicly available
- Identify community repositories and archives to work with including kinds of data they accept
- Consider local institutional repository for project artifacts (ie project reports, field notebook digitization)
- Identify community keyword list(s) in use
- Identify metadata standard
- Map local metadata to standard metadata
- Identify community data parameter dictionary
- Identify community unit dictionary
- Identify community coding scheme dictionaries

**\*\* all items publicly available \*\***

## Handout 6: Some Configurations of Data Infrastructure

