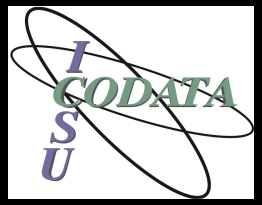
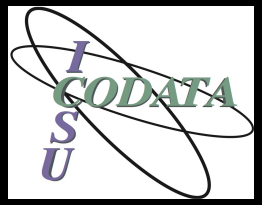


Open Research Data and Training one implementation

Open Data Skills Workshop
University of Oslo Library
Mar 14 2016

Hugh Shanahan
Hugh.Shanahan@rhul.ac.uk
@HughShanahan


I
CODATA
S
U

Motivation


RDA
RESEARCH DATA ALLIANCE


ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

Motivation

Clear need for Data Science skills amongst researchers

Motivation

Clear need for Data Science skills amongst researchers

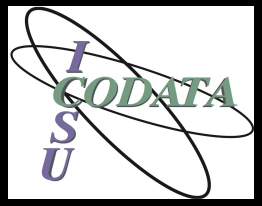
Particularly true for researchers from LMIC countries

Motivation

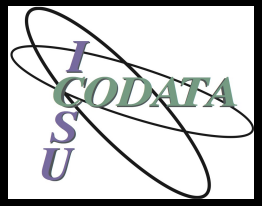
Clear need for Data Science skills amongst researchers

Particularly true for researchers from LMIC countries

Research Data Science skills≠ {Machine Learning, Big Data Analytics, MapReduce}



Research Data Science Schools



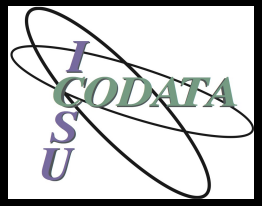
Research Data Science Schools

Curriculum focussed on researchers

Research Data Science Schools

Curriculum focussed on researchers

Based on generic skills required in the field

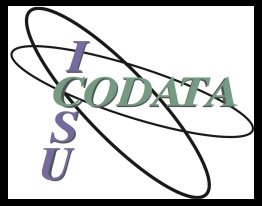


Research Data Science Schools

Curriculum focussed on researchers

Based on generic skills required in the field

Practical skills - learning by doing



Research Data Science Schools

Curriculum focussed on researchers

Based on generic skills required in the field

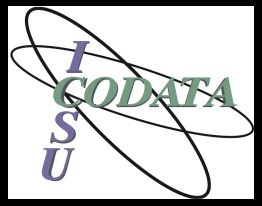
Practical skills - learning by doing

Clear set of skills dating back from original Strata O'Reilly survey

Overview of skills

Overview of skills

Annotate and publish the data



Overview of skills

Annotate and publish the data

Extract and manipulate the data

Overview of skills

Annotate and publish the data

Extract and manipulate the data

Analyse and make sense of the data

Overview of skills

Annotate and publish the data

Extract and manipulate the data

Analyse and make sense of the data

Why do Open Research?

Timetable

Week 1

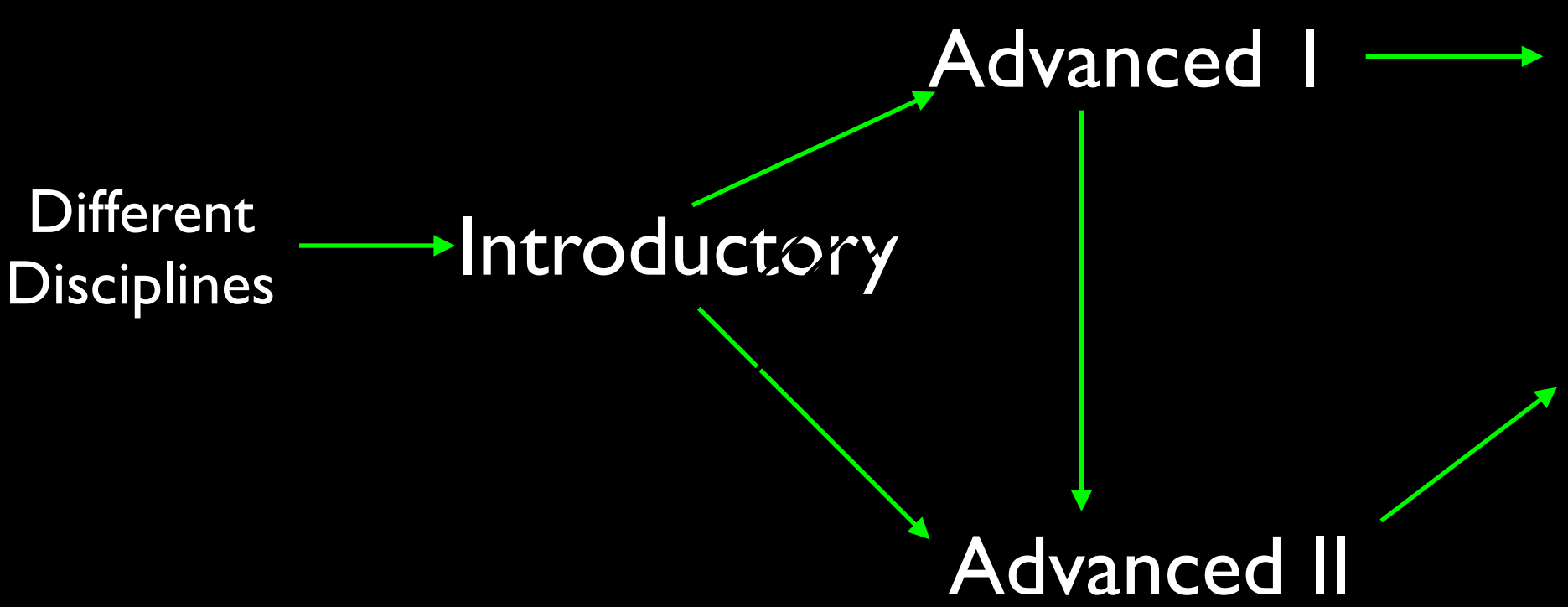
M	T	W	T	F
Open Science and Open Research Data	Data Carpentry	Data Carpentry	Software Carpentry	Software Carpentry

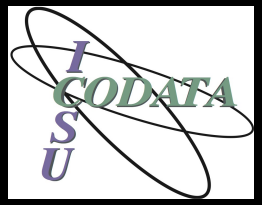
Timetable

Week 2

S	M	T	W	T	F
Visualisation and Open Science	Visualisation and Analysis	Visualisation and Analysis	Visualisation and Analysis	Analysis and Computational Infrastructures	Computational Infrastructures and Open Science

Progression (long term vision)





Open Science

Reflective sessions

Why do Open Science?

What impact does it have on your research?

Open Research Data

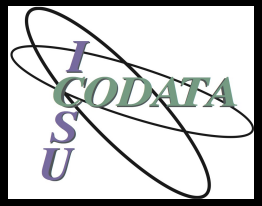
Importance of Data Sharing and Open Research

Introduction to

Data Publishing

Data Curation

Metadata

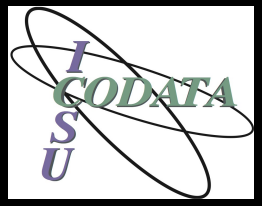


Software Carpentry

The Unix shell - data munging

Introduction to R

Using Git

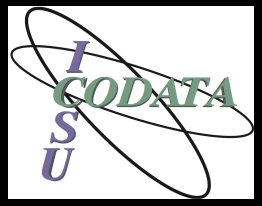


Data Carpentry

Motivation - why use databases

Introduction to SQL

How to do queries, aggregation and joins



Visualisation

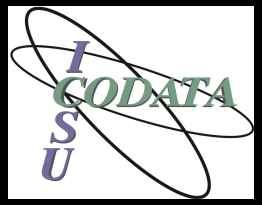
Part I Tools (using R)

Data wrangling

Plotting with tools such as ggplot2

Part II Workshop

Given a data set - plot it and then receive feedback



Analysis

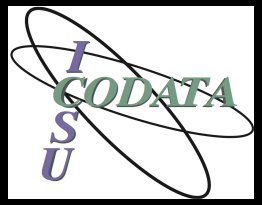
Expectation of understanding of significance tests

Significance beyond standard tests

Machine Learning

Clustering, un/supervised learning

Cross-validation, ROC, AUC



Computational Infrastructures

Introduction to concepts of cloud computing (IaaS)

How to spin up a Virtual Machine on a cloud

Use of batch scheduler or using containers (docker)

Excellent pedagogical model

Software and Data Carpentry have used this model successfully for the last 10 years.

Materials will continue to be developed by community.

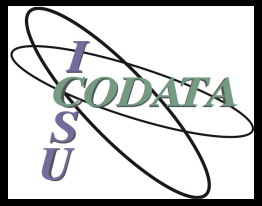
Allows the creation of a community of Research Data Scientists from different disciplines.

Train the trainers

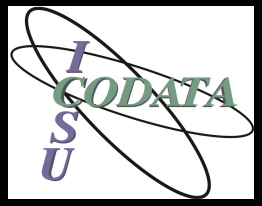
Long term success (5-10 years) -
integration by all Universities

Medium term success - scaling up of schools

Assistant instructors in one
school become instructors in next



Implementation



Implementation

First Introductory school August 1-12 2016
Trieste, Italy, ICTP

Implementation

First Introductory school August 1-12 2016
Trieste, Italy, ICTP

Initial funding from ICTP, TWAS and CODATA

Implementation

First Introductory school August 1-12 2016
Trieste, Italy, ICTP

Initial funding from ICTP, TWAS and CODATA

Up to 90 attendees - funding at present for 30-40

School web site

http://indico.ictp.it/event/7658

School web site

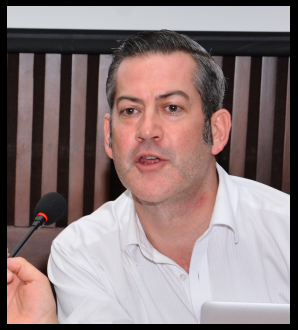
http://indico.ictp.it/event/7658

School web site

http://indico.ictp.it/event/7658

Co-chairs of RDA Working Group

Dr. Andrew Harrison, University of Essex



Dr. Simon Hodson, CODATA