IG Education and Training on handling of research data

Tuesday 1st March - 11:30 - 13:00 - Working Meeting Session 1

Notes by Heidi Laine (@heidiklaine / heidi.laine@helsinki.fi)

Agenda

1. Welcome and introductions
2. Update on recent activities and interim IG meeting on January 14-15 in Oxford, UK (Laura Molloy)
3. Update on EDISON project formal definition of Data Science profession (15 min): Yuri Demchenko
4. Practical exercise: Work in groups on reviewing data related competences for different professional groups - skipped
5. Current initiatives: please share your current projects and plans for curriculum delivery and development:
   5.1. RDA IG Archives and Records by Elise Dunham and Rebecca Grant, co-chairs
   5.2. EUDAT B2SHARE project by Ari Asmi
   5.3. Inventory of training and education resources by Amy Nurnberger, Chair for RDA US Education and Curriculum Group
6. AOB - skipped

1. Welcome and introductions

IG has four co-chairs: Yuri Demchenko, Laura Molloy, Amy Nurnberger, Christopher Jung; and looking for one more from Asia Pacific.
Next plenary: two sessions on Education & Training

Few questions & issues raised during introductions:
- what is data science (definition)?
- what are essential data science skills?
- how to reach out to researchers and engage them to change their practices, processes?
- how should we embed data management skills to existing researcher training?
- informal practices, effects of citizen science?
2. Update on recent activities and interim IG meeting on January 14-15 in Oxford, UK (Laura Molloy)

Overview of activities related to data management related skills developments:
- many different contexts, backgrounds, budgets… A lot of diversity!
- aim: spark ideas, conversations, exchanging information about existing developments, promotion of training
- for anybody who feels that handling data is part of their work, responsibilities
- scientists vs. researchers? latter is more inclusive
- what has been done so far
  - one of the tasks accomplished: defining data handling competences and skills profile: research librarian / r. administrator / r. infra manager or operator / researcher > findings are on wiki page, everyone is welcome to add material, but task is close to being done
- next steps
  - skills gathering work
  - engagement with other RDA groups
  - Possibly an WG can emerge from this IG

Discussion:
- who is interested in defining digital librarian skills?
- we have a list of competences, but how do we help people to acquire those competencies? could this IG help in moving this forward?
- Yuri: if you have ideas for WGs, raise them! WGs need to have clearly defined, achievable tasks, we might have multiple WGs
- Andrew Harrison on meeting in Amsterdam at IDCC2016 conference¹, in Europe we can envisage a joint curriculum, but that is not the case for other areas, this is the right forum to discuss how could we have a global credit / accreditation system
- Laura: Do any of the Asian participants know of Asian accreditation systems
- Donatella Castelli (Italy): infrastructures are emerging, computational but also data, what can data infrastructure provide, f. e. preservation, who decides what is preserved? We are in contact with data mining researchers, they combine the algorithm and the data. Do you have a plan to address this new way of thinking combining data, storage and platform issues? Researchers and educators need to trust infrastructure.
- Yuri: we need to approach this from the point of view of data analyst, biologists etc. They don’t know how to manage data and we need to address data literacy issues.

3. Update on EDISON project

Yuri presented the first deliverables and outcome of the EDISON project. The project goal is to formally define the profession of Data Scientist.

The project started from the definition of the Data Science Competence Framework (CF-DS), it will create a basis for defining the Data Science Body of Knowledge (DS-BoK) and taxonomy.

Based on the above we will define a Data Science Model Curriculum (MC-DS)

Univ. Amsterdam is currently implementing four programs related to data science, each of them take the basic structure proposed by EDISON and apply it thematically and based on available teaching resources.

They also consider to provide training on the basic level of competences common to all, in particular a one week course “Data management literacy” is being prepared

Academic transferability is an important issue and will allow better young people to come to research infrastructure organisations and projects.

EDISON framework is built on the European standards: e-Competence Framework version 0.3 (e-CFv3.0), ICT profiles CWA 16458: 42 profiles, recently introduced standard ESCO that defines structures, digital librarian / data scientist or similar not yet listed, currently data science under database management skills, this isn’t right and it will need to be changed

In the EDISON study, almost all vacancies require data management skills, and also research methods and business methods; data scientists need to have an overview of all processes in the relevant organization. This is depicted as inner circle & outer circle of the diagram, business also adopted constant improvement cycle that is implemented in the new type of agile data driven enterprises (ADDE). Important issue is linking between data science representing abstract data concepts and subject domain that is characterized by its specific (1) data types and structures, models, metadata, vocabulary, classification, and also organizational relations. “Death valley” between science domain and data science needs to be crossed.

Discussion:

- Data scientists need to have data literacy on different disciplines, even when they themselves don’t produce data
- Edison report on CF-DS in public domain, DS-BoK draft will be posted mid March 2016 http://edison-project.eu/

4. Update on existing initiatives

RDA IG Archives and Records by Elise Dunham and Rebecca Grant, co-chairs

- How to make value judgements concerning data preservation? > archivists have valuable expertise > IG started from a BoF in RDA6 Paris
- Discussion populated by librarians, archivists can bring other useful things to the table
- Potential outputs: how can we engage archivists more? Something concrete, a WG?
- Filtering ideas of appraisal to RDA
- Question: general guidelines or discipline specific? > the ultimate questions, general guidelines is a good place to start, some disciplines need raw data, others use only processed data

EUDAT B2SHARE project by Ari Asmi
EUDAT pilot on data storage within laboratory course  
Idea to also include an element of data management in lab work training > implement DM and metadata in the courses, so that they don’t think it’s something strange and new  
We need specific courses PLUS implementing DM everywhere!  
Question: How open are teachers for working with this type of approach? Do you have a program for introducing this approach to other areas?  
Answer: Old professors know very little about DM, so layered teaching needed

Inventory of training and education resources (http://bit.ly/1UvnVEg) by Amy Nurnberger, Chair for RDA US Education and Curriculum Group

- Training teachers and professors, how to get digital literacies to their normal teaching  
- Root problem: how do make this a natural learning area, natural skill like reading  
- Research data handling training, how many times are we inventing the wheel? Asking the community in building the spreadsheet, looking to people who are creating these learning opportunities: basic information, all participation welcome!  
- Some vocabulary work involved  
- Question: is it suitable for one-off things?  
- Answer: It’s for resources, or projects looking for resources

5. Work in groups and AOB skipped

Next session is a Joint meeting of IG Education and Training on handling of research data, WG RDA/CODATA Summer Schools in Data Science and Cloud Computing in the Developing World: International Coordination of Research Data Education and Training

- Including discussion on the first draft program and using software carpentry and data carpentry for sharing educational materials and building network of educators.

Meeting conclusion:

1) IG co-chairs will meet during RDA7 to discuss meeting outcome and proposed a set of actions, including possible Working Groups that may come out of IG and will be focused on specific topics and deliverables.
2) Initiate discussion on the mailing list to exchange initiatives and information.

Results of the co-chairs meeting

IG-ETRD will initiate establishing 3 Working Groups:

1) WG on Research Data Management (RDM) literacy that will propose model curriculum and a number of template programmes for training on Research Data Management for different group of professions and researchers.
This group will benefit from using current RDA output related to Research Data Management (RDA) (counting also on support by RDA experts to draft and teach training courses) but will extend them to academically suitable courses.

2) WG on Research Data Handling Competence and Skills that will formalize and finalise the current IG-ETRD initiative and activity on defining competences and skills for at least 4 groups of occupations: research librarians, research administrators, Research Infrastructure managers/operators (technical), researchers.

3) WG on Data Science related accreditation schemes comparison that will study and analyse accreditation and certification schemes and practices for Data Science and data management related professions in different countries.