

GO FAIR Initiative and GO-TRAIN

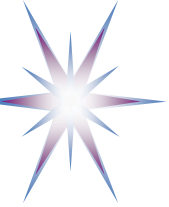


EDISON
building the data
science profession

RDA 9 IG-ETHRD meeting
5 April 2017, Barcelona

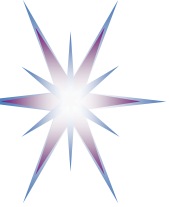
EDISON – **E**ducation for **D**ata Intensive
Science to **O**pen **N**ew science frontiers

Grant 675419 (INFRA-SUPP-4-2015: CSA)



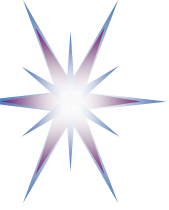
GO FAIR and IFDS

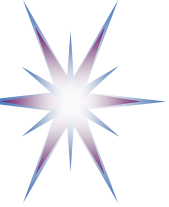
- Global Open FAIR
 - Findable – Accessible – Interoperable - Reusable
- IFDS – Internet of FAIR Data and Services = EOSC
- GO FAIR implementation approach
 - GO-BUILD
 - GO-CHANGE
 - GO-TRAIN: Training of data stewards capable of providing FAIR data services
- A critical success factor is availability of expertise in data stewardship
 - Training of a new generation of FAIR data experts is urgently needed to provide the necessary capacity.



Meeting to Discuss a GO-TRAIN Node

- Meeting to Discuss a GO-TRAIN Node
 - Hosted ICSU-CODATA, Paris, 3 February 2017
- Attended by:
 - CODATA
 - LIBER
 - EOSCpilot
 - FOSTER Project
 - EDISON Project
 - DTLS, Netherlands
 - DANS, Netherlands
 - DCC, UK
 - UCL, UK
 - Univ Manchester
- Way forward: Collaboration between projects and National initiatives

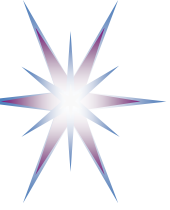




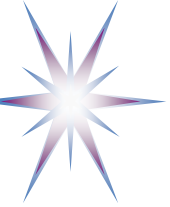
HLEG report on European Open Science Cloud (October 2016) – Demand for

Realising the European Open Science Cloud. First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud, October 2016

- **Recommendation: Allocate 5% grant funding for Data management and preservation**
- **Estimation: More than 80,000 data stewards to serve 1.7 mln scientists in Europe (1 per every 20 scientists)**
- **Industry: IDC Report on European Data Market (2015)**
 - Number of data workers 6.1 mln (2014) - increase 5.7% from 2013
 - Average number of data workers per company 9.5 - increase 4.4%
 - Gap between demand and supply 509,000 (2014) or 7.5%
- Core data experts need to be trained and their career perspective improved

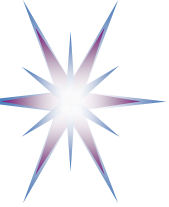


- **Clash of cultures**
 - between domain specialists and e-Infrastructure specialists (i.e. IT/Computer Science)
- New data experts come from **scientific** and **engineering** cultures
 - with very different reward systems and incentives,
 - different jargons and very different skill sets.
- **Consequences**
 - Evident a divide between researchers and those that support research with data processing and software
 - Two communities that are both essential to Open Science have not closely co-evolved and do not converge
 - Lack of data scientists that venture out from classical computer or data science departments into other scientific fields.



HLEG EOSC Report Essentials – **Core Data Experts**

- **Core Data Experts** is a new class of colleagues with core scientific professional competencies and the communication skills to fill the gap between the two cultures.
 - **Core data experts** are neither computer savvy research scientists nor are they hard-core data or computer scientists or software engineers.
 - They should be technical data experts, though proficient enough in the content domain where they work routinely from the very beginning (experimental design, proposal writing) until the very end of the data discovery cycle
 - Converge two communities:
 - Scientists need to be educated to the point where they hire, support and respect Core Data Experts
 - Data Scientists (Core Data Experts) need to bring the value to scientific research and organisations
- Implementation of the EOSC needs to include instruments to help train, retain and recognise this expertise,
 - In order to support the 1.7 million scientists and over 70 million people working in innovation.



EOSC Report Recommendations – Implementation on training and skills

- **I2.1: Set initial guiding principles to kick-start the initiative as quickly as possible. -> Bridge two cultures/communities**
 - A first cohort of core data experts should be trained to translate the needs for data driven science into technical specifications to be discussed with **hard-core data scientists and engineers**.
 - This new class of core data experts will also help translate back to the **hard- core scientists** the technical opportunities and limitations
- **I3: Fund a concerted effort to develop core data expertise in Europe.**
 - Substantial training initiative in Europe to locate, create, maintain and sustain the required core data expertise.
 - **By 2022, to train (hundreds of thousands of) certified core data experts** with a demonstrable effect on ESFRI/e-INFRA activities and prospects for long-term sustainability of this critical human resource
 - Consolidate and further develop assisting material and tools for Data Management Plans and Data Stewardship plans (including long-term preservation in FAIR status)
- **I7: Provide a clear operational timeline to deal with the early preparatory phase of the EOSC.**
 - **Define training needs for the necessary data expertise and draw models for the necessary training infrastructure**