1st GEDE Workshop on Digital Objects

Peter Wittenburg

GEDE-DO Co-Chairs: Dimitris Koureas, Koenraad de Smedt,
Peter Wittenburg

RDA GEDE = delegates from 47 large Research Infrastructures

ACTRIS
AGRO
AnaEE
Asterics
BBMRI
CESSDA
CLARIN
DARIAH
DiSCCo
EATRIS
ECRIN
EISCAT
ELI
ELIXIR
eLTER
EMBRC
EMFL
EMPHASIS
ENES
ENVRI Plus
EPOS
E-RIHS
ESRF
ESS
ESS - Social
EUCALL
EU-OPENSCREEN
EURO-ARGO
EURO-BIOIMAGING
FAIR
HBP
IAGOS
ICOS
INSTRUCT
KM3NeT
LIBER
LifeWatch
METROFOOD
MIRRI
MYRRHA
NIDI
PARTHENOS
RDA
SCK•CEN
SERISS
SIOS
SKA
VAMDC

CO-Chairs:

Maggie Hellström (ICOS)

Carlo Maria Zwölf (VAMDC)

Peter Wittenburg (RDA)

Started and funded in RDA Europe 3

Funded in RDA Europe 4?

GEDE Topic Groups

- GEDE includes 47 large Research Infrastructures (mostly ESFRI)
- Finalised a broadly discussed paper on PID usage (Maggie, Carlo)
- Built 4 new topic groups and just starting these
  - Digital Objects (about 130 interested experts, hope on EU funding support)
  - Citation (so many different suggestions)
  - Repositories (trustworthy repositories are in the focus, what about sensor chips with memory)
  - Blockchain (how can we make use of it in science)
- Goal is always to summarise and evaluate practices in the RIs, to relate them with state-of-the-art discussions and to find broad agreements where possible
- If you or a collegue of you are interested to participate, please send an email to Zsuzsanna (zsuzsanna.szeredi@gmail.com)

Why Digital Object Topic Group?

- 3 large surveys can‘t be wrong
  - 80 % of experts‘ time in data-intensive projects is wasted with data wrangling
  - CrowdFlower survey on big data projects in industry with same results (no semantics involved)
  - To a large extent inefficiencies are due to bad data organisations
- Consequences are well-known (research and industry)
  - The efforts and costs are huge
  - Many DIS projects simply fail
  - Many experts are excluded from DIS
  - A huge diversity and solutions space (proliferation everywhere)

Data volumes and complexity & data sharing/reuse will increase.

Need to change our practices! But how?

Different approaches towards a stable data domain

- Different approaches
  - In big industry
    - Proprietary platforms to bind customers
    - Interoperability frameworks to bridge between tools, formats, etc.
    - Reference architectures as comprehensive top-down designs
  - In research:
    - Research Infrastructures, eInfrastructures, clusters
    - Initiatives such as CODATA, RDA, CTS (DAS, WDS), FORCE11, GOFAIR, DONA, etc.
- Good first results
  - DONA foundation: global domain of resolvable PIDs (Handles, DOIs, ePIC, etc)
  - FAIR principles – an excellent summary of many discussions – a kind of globally agreed guidance for proper data management/stewardship
  - Various RDA WG results

Comprehensive approach by the EC with EOSC.

Yet no breakthrough for harmonised infrastructure building.

What time scale are we aiming at?

Networking Protocol Development

Networking		**Protocols**
People to time-sharing computers	⟷	TTY
People to Files	⟷	File Transfer based on FTP + TCP/IP
People to People		Email based on SMTP + TCP/IP
People to Info	⟷	HTML based on HTTP + TCP/IP
What next?	⟷	What next?

We are using HTML/HTTP now for everything

- The web is great but
  - not made for everything – not for creating a stable data domain

DOs – a few questions

- Digital Objects – what are they, do we have a good definion?
- Could we imagine a unifying protocol for interfacing with DOs?
- Is there any relationship with Abstract Data Types (and OOP) which made complex software designs possible?
- Are DOs just another computer science invention or do they have a relevance for organising sciences in the digital era?
- What is the relation between DOs and FAIR?

- Finally: do DOs have a potential to revolutionise our data practices and to offer a fundament to build on for the coming decades (or more)?

Objective of this workshop is not to find satisfying answers, but to start the discussion about them and to involve the practitioners.

Agenda 1st GEDE DO Meeting

9.20	Keynote on Digital Object Architecture	Robert Kahn
10.30	Coffee	
11.00	Implementation work in C2CAMP	Tobias Weigel
11.20	Objects, types, collections and operations in DOIP	Ulrich Schwardmann (ePIC)
	GOing FAIR & DOing FAIR	Erik Schultess (GO FAIR)
12.00	Lunch	
13.00	DOs – The Scientific Case	Dimitris Koureas
13.20	Digitals Objects as direct input into the CLARIN Language Resource Switchboard	Twan Gosen, Dieter van Uytvanck (CLARIN)
	How a Digital Object Architecture could help ICOS streamline data service provisioning	Margareta Hellström (ICOS)
	DiSSCo Digital Specimens- Widening access to natural science collections	Alex Hardisty (DISSCO)
	The DO Case in Virtual Atomic and Molecular Data Centre	Carlo Maria Zwölf (VAMDC)
	Digital Object Management for ENES: Challenges and opportunities	Tobias Weigel (ENES)
15.00	Final Discussion	Koenraad de Smedt
16.00	end	