

Meeting Notes: Data Fabrics - Digital Objects and Linked Data

Thursday, April 22 2021,

Breakout 8

This document:

<https://docs.google.com/document/d/11WNC3PqN5k1NLaHH3gY2cJn1uaNTHfGBT6qR0r3GPBg/edit>

Meeting agenda:

(Slides will be available at <https://www.rd-alliance.org/group/data-fabric-ig/wiki/monthly-meetings>)

- 10 Minute Introduction
- 10 Minute Overview of DO Approach
- 15 Minute Discussion of DO Approach
- 10 Minute Overview of LDP Approach
<https://www.youtube.com/watch?v=BqJFPmEo9PA>
- 15 Minute Discussion of LDP Approach
- 25 Minutes of moderated discussion about interoperation in an environment with both Approaches
- 5 Minutes wrap and action items

Attendance list:

Please sign in:

Name	Affiliation	Email
Rainer Stotzka	Karlsruhe Institute of Technology, Germany	rainer.stotzka@kit.edu
Stian Soiland-Reyes	The University of Manchester; University of Amsterdam	
Maggie Hellström	ICOS Carbon Portal and Lund University	margareta.hellstrom@nateko.lu.se
Larry Lannom	CNRI	llannom@cnri.reston.va.us
Ray Plante	NIST	raymond.plante@nist.gov
Hanna Koivula	CSC - IT Center for Science	hanna.koivula@csc.fi
Zachary Trautt	NIST	zachary.trautt@nist.gov

Anton Pirogov	Forschungszentrum Jülich, Germany	a.pirogov@fz-juelich.de
Alex Hardisty	Cardiff University / DiSSCo	hardistyar@cardiff.ac.uk
Robert Ulrich	re3data / Karlsruhe Institute of Technology, Germany	robert.ulrich@kit.edu
Ulrich Schwarzmann	GWDG Goettingen, ePIC	uschar1@gwdg.de
Soizick Lesteven	CNRS/CDS (France)	soizick.lesteven@astro.unistra.fr
Tom Honeyman	Australian Research Data Commons	tom.honeyman@ardc.edu.au
Xiaolei Xia	Computer network information Center,CAS	xlxiao@cnic
Jonathan Clark	The DOI Foundation	j.clark@doi.org
Ahmad Zainul Ihsan	Forschungszentrum Jülich, Germany	a.ihsan@fz-juelich.de
Peter Pennefather	gDial Inc	p.pennefather@gmail.com
Anders Conrad	Danish e-Infrastructure Cooperation (DeiC)	anders.conrad@deic.dk
Peter Mutschke	GESIS, Germany	peter.mutschke@gesis.org
Christian Pagé	CERFACS, France	christian.page@cerfacs.fr
Constanze Curdt	GEOMAR Helmholtz Centre for Ocean Research, Germany	ccurd@geomar.de
Andreas Pfeil	Karlsruhe Institute of Technology, Germany	andreas.pfeil@kit.edu
Binh T. Nguyen	re3data / Karlsruhe Institute of Technology, Germany	binh.nguyen@kit.edu
Gail Hodge	IIA	ghodge@iiaweb.com

Doug Fils	Ocean Leadership	dfils@oceanleadership.org
Rob Quick	Indiana University	rquick@iu.edu
Yann Le Franc	e-Science Data Factory	ylefranc@esciencefactory.com
Barbara Magagna	Environment Agency Austria	barbara.magagna@umweltbundesamt.at
Joan Damerow	Lawrence Berkeley National Lab	JoanDamerow@lbl.gov
Marina Sánchez	UC3M	marinasa@pa.uc3m.es
Fajar J. Ekaputra	TU Wien	fajar.ekaputra@tuwien.ac.at
Jose Benito Gonzalez	CERN	jose.benito.gonzalez@cern.ch
Ralph Müller-Pfefferkorn	TU Dresden / Germany	ralph.mueller-pfefferkorn@tu-dresden.de
David Elbert	Johns Hopkins University	elbert@jhu.edu
Wouter Addink	Naturalis/ DiSSCo	wouter.addink@naturalis.nl
Mariya Dimitrova	Pensoft/BAS	m.dimitrova@pensoft.net
Emanuel Soeding	GEOMAR / HMC	esoeding@geomar.de
Limor Peer	Yale University	
Alex Henderson	University of Manchester	alex.henderson@manchester.ac.uk

Discussions:

- Talk 0: Here is the slide that was missing from the introduction talk with Guiding Documents and links:
 - Recommendations for Implementing a Virtual Layer for Management of the Complete Life Cycle of Scientific Data (2018/2020)
 - ◦Provides a high-level conceptual framework to support Digital Object management and service development
 - ◦<http://dx.doi.org/10.15497/RDA00026>
 - Consolidated assertions on Persistent Identifiers (2017/2018)
 - ◦Provides a cross-disciplinary collection of statements on PIDs, reflecting commonalities across disciplinary perceptions and use cases

- <http://dx.doi.org/10.15497/RDA00027>
- FAIR Digital Object Forum
- <https://fairdo.org/>
- oSpring 2022 FAIR DO Conference
- Talk 1: What is Digital Object Architecture trying to solve?
 - Overwhelming data amounts
 - APIs, multiple storage, security issues
 - Making data available - but Fragmentation challenges
 - FAIR Guidelines attempt to address - mostly oriented for machine-readability
 - Easy to implement for publications
 - Lab work is much more challenging to FAIRify
 - Add abstraction level for persistently identified digital objects
 - File system abstraction is familiar. For Digital Objects:
 - Globally unique/actionable identifiers
 - Every object is typed
 - Every object has associated metadata
 - Every object has queryable set of operations
 - Every object addressed via Digital Object Access Protocol
 - What are minimal DO requirements?
 - Identifier + Resolution
 - Types, type registry
 - Digital Object repositories (incl. Metadata, repositories)
 - Mapping/brokering systems for existing systems to map to DO
 - Don't assume existing stuff will go away!
 - Protocol for repositories to implement
 - Need to talk common protocol
 - (Showing interactions of DOIP operations)
 - Digital Specimen Object example
 - Digital twin of physical sample
 - Identifier for the digital object different from identifier of physical specimen
 - FAIR Digital Object Forum
 - <https://fairdo.org/>
 - register at <https://fairdo.org/register/>
 - Spring 2022: FDO conference in Europe
 - Combine:
 - FAIR
 - FDO
 - LDP ([Linked Data Platform](#))
 - **Discussion:**
 - URLs from chat:
 - <https://www.dona.net/>
 - <https://www.dona.net/specsandsoftware#block-digitalobjectinterfaceprotocol>
 - https://www.dona.net/sites/default/files/2018-11/DOIPv2Spec_1.pdf
 - <https://www.w3.org/TR/ldp-primer/>

- Stian: Need for a minimal baseline protocol and type operations, e.g. CRUD on HTTP - as needed by LDP. Need for negotiation for support of additional, optional DOIP implementations.
- Laurence: Digital Objects all the way down, types, even methods are identified Digital Objects.
 - (Mint a DOI for <https://tools.ietf.org/html/rfc7231#section-4.3.1> ? :)
-
- One type are variables: for this we would have a proposal: The I-ADOPT [framework](#) see [I-ADOPT Framework 1.0.0 RDA 17](#) and the [related presentation](#)
-
- Luiz Bonino: FAIR Digital Object Framework (FDOF)
 - Luiz video: <https://www.youtube.com/watch?v=BqJFPmEo9PA>
 - Requirements
 - Identifiers, separation of metadata/objects, relations, machine-actionable metadata
 - Informational unit - sequence of bits
 - Current draft: <https://fairdigitalobjectframework.org/>
 - Discussion: Peter points to the FAIR Digital Object Framework definitions & other related materials at <https://github.com/GEDE-RDA-Europe/GEDE/tree/master/FAIR%20Digital%20Objects>
 -
-

Chat Dump

Alex Henderson and 50 others joined the room - 2:01 pm

Stefanie

Welcome! Please add your name to the collab notes document at

<https://docs.google.com/document/d/11WNC3PqN5k1NLaHH3gY2cJn1uaNTHfGBT6qR0r3GPBg/edit>

Stephanie Holmgren joined the room - 2:01 pm

Brian Westra joined the room - 2:03 pm

Robert Szczodruć joined the room - 2:04 pm

Daniel Mohr joined the room - 2:04 pm

Clement Jonquet joined the room - 2:05 pm

Stian

Is HTTP such a common protocol?

David Elbert joined the room - 2:06 pm

Megan Force joined the room - 2:06 pm

David Elbert joined the room - 2:08 pm

Christian Pagé joined the room - 2:08 pm

Stefanie

Welcome! Please add your name to the collab notes document:

<https://docs.google.com/document/d/11WNC3PqN5k1NLaHH3gY2cJn1uaNTHfGBT6qR0r3GPBg/edit>

Christian Pagé joined the room - 2:11 pm

Ulrich

hand

Alex

DOIP spec: https://www.dona.net/sites/default/files/2018-11/DOIPv2Spec_1.pdf

Alex

SDK: <https://www.dona.net/specsandsoftware#block-digitalobjectinterfaceprotocol>

Stian

Combined with LDP would presumably be a HTTP-based implementation

Alex

DONA Foundation: <https://www.dona.net/>

Alex

FAIR Digital forum: <https://fairdo.org>

Stian

Linked DAta Platform <https://www.w3.org/TR/ldp-primer/>

Peter

true Stian - then it would be a duty I guess.

Peter

LDP is relying on the Web protocol stack

Ian and 1 other joined the room - 2:14 pm

Riina Luoma joined the room - 2:15 pm

Stian

with optional other DOIP access. Perhaps there could be a bridge as well. It sounds like some negotiation is needed..

Krishna Patel joined the room - 2:16 pm

Ulrich

but an operation is a special object wrt to just bitstreams, . similar to functions wrt parameters

Barbara

May I say something about this variability?

Peter

sounds interesting Barbara

Peter

please send the link

Barbara

https://docs.google.com/document/d/1NGC5BQ9tKsGWnh_IFtMLBWu-BIT8vpulAanF6zYEr2I/e/dit# and the related presentation

Barbara

The I-ADOPT framework

Barbara

<https://padlet.com/barbaramagagna/sogprgszse1bgd24>

Maria Stoica joined the room - 2:24 pm

Barbara

sorry here the presentation

https://docs.google.com/presentation/d/16IEftQkeUttEvefENTVHDNN2jdwMbj1ClzmLLmIF-04/edit#slide=id.gd0d27f88ae_0_12

Ulrich

what the I-Adopt group did, really looks promising.

Alex Henderson

+1 for I-ADOPT

Nicola Knight joined the room - 2:27 pm

Zachary

can you post the link to youtube?

Christian Pagé joined the room - 2:27 pm

Rainer

Video: <https://www.youtube.com/watch?v=BqJFPmEo9PA>

Limor Peer joined the room - 2:28 pm

Günther Gerrit joined the room - 2:29 pm

Stefanie

collab notes document for those who just joined:

<https://docs.google.com/document/d/11WNC3PqN5k1NLaHH3gY2cJn1uaNTHfGBT6qR0r3GPBg/edit>

Siddeswara Guru joined the room - 2:29 pm

David Elbert joined the room - 2:30 pm

Barbara Magagna joined the room - 2:32 pm

Maria Luiza Campos joined the room - 2:32 pm

David Elbert joined the room - 2:33 pm

Peter Pennefather joined the room - 2:35 pm

Christophe Parisse joined the room - 2:36 pm

David Elbert joined the room - 2:36 pm

Dieter Van Uytvanck joined the room - 2:37 pm

Christian Pagé joined the room - 2:38 pm

Christian Pagé joined the room - 2:38 pm

Christian Pagé joined the room - 2:38 pm

Christian Pagé joined the room - 2:38 pm

Tom Honeyman joined the room - 2:38 pm

Stian

I'll have to go, but have to say how I like the FDOF approach as exactly the kind of "base types", and "base methods"

Stian

and base protocol is then HTTP

Maria Luiza Campos joined the room - 2:39 pm

Krishna Patel joined the room - 2:39 pm

Peter

FDOF basics defined here:

<https://github.com/GEDE-RDA-Europe/GEDE/tree/master/FAIR%20Digital%20Objects>

Christian Pagé joined the room - 2:42 pm

Christian Pagé joined the room - 2:43 pm

Christian Pagé joined the room - 2:43 pm

Alex

\hand

Peter
hand

Bryan Mc Alister joined the room - 2:46 pm

Christian Pagé joined the room - 2:47 pm

Charles Schmitt joined the room - 2:47 pm

Christian and 1 other joined the room - 2:48 pm

Peter

what about DO contents based identifier. The original DO will be owned by someone who could self-curate the persistence

Zachary

RO-Crate was speaker hosted in a monthly meeting:

<https://www.rd-alliance.org/group/data-fabric-ig/wiki/monthly-meetings>

Christian Pagé joined the room - 2:48 pm

Wouter

<https://w3id.org/ro/crate>

Ulrich

at ePIC we have a special kind of short live Handles that have prefixes like 21.Txxxx with T for test or temporarily

Christian Pagé joined the room - 2:49 pm

Christian Pagé joined the room - 2:50 pm

Christian Pagé joined the room - 2:50 pm

Maggie

Of course we still need to support the equivalent of "tombstones", where the DO may be unavailable, but its related metadata live on

Christian Pagé joined the room - 2:51 pm

Peter Wittenburg joined the room - 2:51 pm

Christian Pagé joined the room - 2:51 pm

Christian Pagé joined the room - 2:52 pm

Christian Pagé joined the room - 2:52 pm

Christian and 1 other joined the room - 2:52 pm

Peter

Maggie is right: at that moment when people start referencing to an FDO you need to maintain the PID

Christian Pagé joined the room - 2:52 pm

Raymond

Objects that are created as a result of an operation will often have a "short-lived" identifier

Christian Pagé joined the room - 2:53 pm

Peter

it depends on the reuse of such an object

Christian Pagé joined the room - 2:54 pm

Christian Pagé joined the room - 2:54 pm

Christian Pagé joined the room - 2:56 pm

Peter

finally a PID does not cost much - just some bytes, so add a thombstone note

Anders

I think Larry put it right: take everything away to begin with, that doesn't break the architecture

Thomas Jejkal joined the room - 2:56 pm

Maggie

PIDs in themselves are cheap - but maintenance (and associated quality control and quality assurance) of their kernel metadata record may be quite costly!

Rob(Q)

agreed Maggie

Raymond

@Maggie +1

Peter

An FDO should also point to infos saying how the 15 FAIR principles are addressed (eg referenceto the metadata scheme, the usage licence etc)

Maggie

Thus, "sustainability" is a key thing - and we'll need to lobby stakeholders & policy makers to get enough funding!

Barbara

Peter: this again links than to FIPs..which are implementations for each of the FAIR principles for this resource

Peter

@Barbara: Indeed. As I already said at the FAIR Convergence Symposium, I see nice relationships between the two approaches

Peter

Ideally, a FDO should contain a reference to a FIP, to make it more machine-actionable

Barbara

I agree

Alex

I agree we won't resolve today. Because we must think about what it is that conforms to any FDO specification. It's the system.

Peter

Folks - I do not know how FIPs relate in this context. FIPs somehow state the profiles of some provider which is another level - isn't it?

Maggie

So I think we'll need to collect a lot more use cases and examples on how all of this can be concretized

Alex

Rainer/Rob: Can you keep the chat for the notes please?

Stefanie

@Alex we can capture the chat and send it to the chairs

Rob(Q)

yes, stefanie please send it to us and we'll publish it to the group

Peter

@Peter: yes, but you can always break it down to the DO level

Rob(Q)

@Rainer, maybe we can schedule a full presentation from Barbra in an upcoming Monthly meeting?

Peter

ok will have to look again, but we cannot leave the definition of what an FDO is to the individual providers. We need a "plug-in" structure with formal validators that checj compliance

Rainer

@Rob: yes. Having Barbara presenting I-ADOPT in our monthly project shares would be great.

Tibor Kálmán joined the room - 3:12 pm



Maggie

Thanks all for an extremely interesting session, with a great dynamic exchange of views and ideas!

Barbara

thank you

Stefanie

It won't cut off :-)

Stefanie

As long as at least one speaker stays on, it won't close

Paul Stokes joined the room - 3:13 pm

Peter

indeed thanks to Rob and rainer

Alex

<https://fairdo.org>

Maggie

Be FAIR, or be square!

Alex

Thx Rob and Rainer.

Anders

Thanks, all!

Barbara

thanks to all

Ulrich

thanks

Jonathan

thanks - good session

Peter

Thanks all!

Yann

thanks for this great session

Christian

Thanks for this great session!!

Maggie and 1 other joined the room - 3:14 pm

Peter

Great session!

Emanuel

thank you!!

Alex and 1 other joined the room - 3:14 pm

Xiaolei

thanks!

Robert Ulrich joined the room - 3:14 pm

Alex Hardisty joined the room - 3:15 pm

Barbara

thank you happy to join one of your next sessions then

Barbara Magagna joined the room - 3:15 pm

Emanuel Soeding joined the room - 3:15 pm



Q&A Dump (well the end of it anyway, not all questions were on screen)

Laurence

https://www.dona.net/sites/default/files/2018-11/DOIPv2Spec_1.pdf

0 votes

Peter

Those associated with creation of DO have a stake their persistence. If the DO also has links to those who have authorship rights. They can self-curate DOs with content based identifiers that will be uniquely identifiable because they were created at particular time and place

0 votes

Peter

A FDO is by definition created for re-use. But the author of the FDO may want to choose when it is published to be examined a re-used by others. So there will be a time before which the same object becomes FAIR

0 votes