IDCC 2018 - conference proceeding

Fantin Reichler

Eliane Blumer

# 1st Keynote : No (Open) Science without Data Curation: Five lessons from the study of Data Journeys (Sabina Leonelli)

## –> Open Research as an opportunity, including scientific infrastructures, governance and how this should be credited and disseminated:   https://www.datastudies.eu/publications https://icsu.org/cms/2017/04/open-data-in-big-data-world\_long.pdf

Open Research on three aspects:

1. Global Scope
2. Systemic Reach
3. Local Implementation

FAIR data improves your research at many level. BUT their are requirement to make data FAIR, such as :

* coordination of data infrastructures
* making data accessible on many platforms
* etc.

Awareness of Open Science and its tools is still very low in the scientific community (EU Working Group on Educaiton and Skills under Open Science, 2017)

<https://www.garnetcommunity.org.uk/sites/default/files/GARNet_Paper_nplants201786-1.pdf>

<https://www.datastudies.eu/publications>

It is important for researcher to have a bit of knowledge of the tools/methods to make their data FAIR. The most important thing, is that some people - us - can share with them an expertise about these tools/methods, and help ease the confusion that the researchers might feel while putting in practice FAIR.

Focus on qualitative data:

1. Databases, example of plant science
2. Data Re-use cases

Data journey example : TAIR (not FAIR) <https://www.arabidopsis.org/>

1. preparing specimens
2. preparin gand performing imaging
3. data storage dissemination
4. ….
5. …
6. Analysis

Epistemic troubles :

- RD collected represent highly selected data types

- selection basesd on political-economic conditions of sharing

- peer reviews structure unclear

- misalignement between it and research need

- no sustainable plans for maintenance

- ….

Lessons Learnt on a general field

1. Context specific data curaiton is key to data re-use
2. Long-term maintenance is key to trustowrthiness (update, LT Policy)
3. Which data and why?
4. data & materials (connect digital data with data in the physical world)
5. Role of ethics, humanities & social sciences in data management (increase quality and reusability)

<http://press.uchicago.edu/ucp/books/book/chicago/D/bo24957334.html>

————————————————–

————————————————-

# PLENARY : RESEARCH PAPERS

## 

## Measuring FAIR Principles to Inform Fitness for Use

## Carolyn Hank from University of Tennessee

past paper on  <http://datacurationprofiles.org/> => 10.1002/pra2.2016.14505301046

“Fitness for use” > focus on the “reusable” aspect of FAIR

Method : interview

Job-related demographics with questions such as, ‘what is your current job title?’, ‘how many years have you work in this instition?, ’how many have you been work in the discipline?’ etc.

Findability >’how did you find the data?’, ‘DOI’, ‘metadata?’

Accessibility > ‘How did you access the data?’ ‘Open format?’ ‘was the data free?’ ‘was the metadata accessible?’

Interoperability>’was the data in a useable format’ ‘encoded?’ ‘machine-actionnable?’

Reusability>’were the metatadata sufficient ?’ etc.

Potential implications : *data can be FAI, but R requires more research*

=> create ne knowledge of how scientists access and use data

=> producing a framework to enable re-use

—————————————-

## Giving datasets context : a comparison study of institutional repositories that apply varying degrees of curaiton

## (Amy Koshoffer, Cincinnati, USA)

questions :

1. How do the metadata vary for each insittution?

2. completeness of metadata

3. curated datasets do have more documentation

4. DOIs more with curated datasets

5. keywords

What is curation?

- appraisal/selection

- check/run files : include clode review, review sensitive information, merde elle parle trop vite!

4 universities : one repository per institution

20 datasets per repository. Comparaison with mandatory mData / unmandatory for each university

Results

Question 1:

- all universites use title in metadata (for instance), but all of them understand something different

- all datasets had above the minimum metadata required

Question 2:

- 53% completedness, but different for every institution she looked at

- use of the Mann-Whitney U test : <https://fr.wikipedia.org/wiki/Test_de_Wilcoxon-Mann-Whitney>

- no optional use of supplementary metadata in curated and not curated repositories

–> Does curation really have an impact then? note sure if the curation service does.

Question 3:

- curation does have an impact on documentation

Question 4:

- all support DOIs, but in different ways. They might be other factors to take into account than curation process

Conclusion :

- Curation process may have had a measurable impact, BUT more factors may be impactful

- Curation > more documentation & more readme.txt

———————————————————————————

## Complexities of digital preservation in a virutal reality environment, the case of virtual Bethel

## University of Indianpolis,  Angela Murillo

“CHUUUUUUUUURCH!”

*E. Blumer, 2018*

Creation of a VR space for a churche > the question is : how do we preserve this VR space ?

-  at the time, there was an archive (docuemnts/physical objects) but recently the building was sold

- 3D virtual space of the church + learning space (history of the building)

<https://comet.soic.iupui.edu/bethel/>

Preservation challenges :

- nature of 3D data

- VR operation

Types of data : pre-prod / prod / post-prod + files that make links between the three phase

Updates : 40gb to 60gb (mainly for the creation of learning spaces)

The problem is that until now, there are no VR object preservation framework

Use of NDSA Standard for Levels of Digital Preservation

<http://ndsa.org/activities/levels-of-digital-preservation/>

Essentialy : Work in progress …. progress … progress…. progress… progress…

————————————————————————-

————————————————————————–

# PARALLEL SPEAKS

# ENABLING AND MEASURING FAIR (Fantin)

## Are research data sets FAIR in the long run -

## Dennis Wehrle, Freiburg

Spoiler alert : there’s no definite answer to this question

Pick 10 public rep. through the 1800+ Re3data.org

For each of these 10 rep. , selection of 10 datasets

Limitation for the datasets :

- Open

- etc. (too fast)

Use of Havard’s File Information Tool Set (FITS), which contains 12 analysis tools.

Test dataset : 237 GB to analyse, which represent 5h20 of processing : it represent 85 days of processing for the whole sample (100 datasets), so they took shortcuts (too fast to note)

FITS result : no result / single result / conflicting result / unkown result

Aggregation of identical named format

Unification of “unknown result”, still there were 28 conflicts (2150 files) to post-process

In the end : app. 145 formats identified (lower estimation) - a few files were still unidentified

images : png/jpg

Text encoded format : CSV, XML, RTF, HTML

Script/source code : readable with text editor, base64-encoded in XML, JavaScript in HTML, refereence to external dat in (X)html

Problematic “text files” : unown binaries, matlab, SPSS, OCtet Stream

SUSTAINABILITY :

Formats division :

- high probability (plain text/pdfa)

- medium probability (open formats such as OPEN Office)

- low prob (.doc, prioritary formats)

Applied from data format to datasets : Most of the datasets had LOW PROB (3/4 approx.)

Advice to datasets creators : change their format

Result : single file format migration may not be sufficient. As a matter of fact, most of datasets are heterogeneous.

Lesson learnt :

-Data service shouldn’t refuse “bad file format” (poorly ranked one), but help researchers create workflow to embed them in LT preservation process. Involvment of datasets creator is necesserary.

- tools mentionned (FITS) have weak support

## Enabling FAIR Data in the Earth and Space Sciences

## Shelley Stall, American Geophysical Union

Agu position of data tends to respect FAIR principle

Survey was taken, the top for issues are the following without surprise :

- data complexity

- findingrelevent existing data

- TOO

- FAST

Storytelling : a student had his computer stollen, the data  was only on it. Later the publication was retracted because of this (because of the fact the data weren’t deposited anywhere)

A new funder’s grant is taking place : to get it, your data has to be FAIR

AGU service :

- streamline data policies

-help researchers find support

- dmp support

- etc.

Data Management Traing Clearinghous : [bit.ly/DMTC\_events](http://bit.ly/DMTC_events) :  <http://dmtclearinghouse.esipfed.org/> (not AGU project, but communitary project) => online learning resources.

Face 2 face meeting : rd-alliance.org

Include your organization sstall@agu.org

———————————————————

———————————————————

# PARALLEL SPEAKS

# Cross-institutional and national data services (Eliane)

**Lisa R. Johnston - Data Curation Network: A cross Institutional Staffing Model for curating research data**

- Building the data curation network

- all universities in USA

Idea: collaboratively sharing data curation staff

- How would we deal with conflicting policy issues?

- What do researchers actually need our help with? Will they care if curation is distributed?

- Can I trust someone else to curate our data? What about quality control?

Start with: 9 institutions (all of them contributing to the curators) , 19 data curators, 1 project cooridnator, 1 program director, 8 DCN representatives, 2 admin leads

Day 1: Business Meeting

Day 2-3 : Curator Training/Network

Process: Ingest, Appraise and Select, DCN, Facilitate Access, Preserve Long-term

DCN: Review, Assign, CURATE, Mediate, Approve

- Check files and metadata

- Understand and run files

- Request missing information

- Augment metadata

- Transform file formats

- Evalute for FAIRness

= CURATE

Assessment:

* Is a network approach to curate research data more efficient? Indicators: number of datasets, frequency, variey, efficiency
* Are Curated data more valuableIndicators: track reuse indicators, implement a DCN registry, apply badges and metadata to signal that data sets curated by the DCN are FAIR

**Making everything available: British Library Research Services and research Data Strategy : Rachel Kotorski, British Library**

- New department : everything available “research services”

- change management portfolio

Research data strategy: make research data business as usual (this is not the case at the moment), users will be able to use reserach data via tools

<http://blogs.bl.uk/digital-scholarship/2017/08/announcing-the-new-british-library-research-data-strategy.html>

Four themes:

* data management: documented data management processes, british library data management plans, data management plan engagement, data management training
* data creation: generating data at the library, advising on data creation, carify approach to data collection, engaging and linking with others –> **idée for EPFL: rajouter quelques sources de données sur le site web des BL?** [**https://data.bl.uk/**](https://data.bl.uk/) **: also digitized content**
* data archiving and preservation: preserving library data, sharing data preservation expertise, data preservation services for third parties, digital shared storage
* data discovery, access and reuse : discovery for library data, third-party data discovery, new models of data access, tools and skills for data exploration, datacite UK

This cannot be done alonE!!!

I**nternaitonal Reserach Infrastructure - funder or partner ? Angeletta Miranda Leggio (ANDS)**

Working together with other existing groups

A lot of collaborations on local levels

Funded projects like : Open Access to Marine Data, Open River, PetaJakarta

Do you see ANDS as funder, provider or partner?

Too fluffy ….I do not reall now what to do out of it

——————————————————————————————-

LUNCH. It was really good.

———————————————————————————————

## Minute madness

1. Metro Fun - Train the Trainer R. Schneider (vote 1)
2. Of coooooooooose! The Bible holds the answer to eeeevrything!
3. There goes my time….joggling during RDM trainings.
4. **Data Citation in Social Sciences (à regarder)**
5. FDMentor [www.forschungsdaten.org/index.php/FDMentor](http://www.forschungsdaten.org/index.php/FDMentor) (more than one minute)
6. Maredata - uanc apon a taim, thea wea several (best accent!) - RDM Iberia
7. HODs/rd: research data harvester based on repositories …Gugeell=Google
8. Holistic RDM service Hannover
9. Jisc RDM toolkit for international community (expörts in se field)
10. Grace : exporing the cost and scalability of reserach data management services (Göttingen) [https://www.sub.uni-goettingen.de/en/projects-research/project-details/projekt/grace/  (2 vote)](https://www.sub.uni-goettingen.de/en/projects-research/project-details/projekt/grace/)
11. Building a reserach data management training community (efaluation foorm)
12. How federated reserach data infrastructure work
13. Crosswalk - Resurrecting data back from the dead
14. Long-tail of data
15. Agile data eco-system
16. Data sharing workflow for large datasets with globus (the shortest presenter of the posters)
17. **Dtaa Processing Pipeline - Finnish National Preservation Service (a little bit taller than the speaker before) (à voir)**
18. **Defining Library Capacity for Big Data curation (the tallest presenter from all posters) (à voir)**
19. Research data management courses : overview and gap analysis
20. **scientific data science service at Brown University (à voir)**
21. Springer Nature Research Data Service (äreeund=around) (\*buuuuuuh\*)
22. Curriculum RDM at Toronto University
23. Supporting Open research using KiltHub <https://kilthub.figshare.com/>
24. RDM for phd (icecreeeeeeam!!!!!)
25. Preservation of Canadian reserach data (service model)
26. **scaling up data management services with metadata in gene sequencing (à voir)**
27. **surveying data management practices among neuroimaging researchers (àvoir )**
28. Forsbase (ELLE S’APPELLE ELIANE!!!!!)
29. New online course, deliver RDM services from DCC

——————————————————————————————–

## Demonstrations (Fantin et Eliane vont dans la même session, car DMPOnline on connaît par coeur)

T**he Arctic World Archive**

<https://www.piql.com/arctic-world-archive/>

Piql is a norvegian preservation company

digital vault designed to protet most valuable dat a from wars, cata strophes and cyberattacks.



This is REALLY a vault. It may be study trip suggestion ?

ealare not sht[tps://www.aridhia.com/platform/analytixagility/](https://www.aridhia.com/platform/analytixagility/) inaugurated in March 2017

natural cold conditions without energy

signed by 43 countries in 2018

what can be stored ? Piqlfilms are the storage medium and they can handle every format (digital)

binary data encoding on a 2D barcode > printed on 35mm film but it is digital, not analog.

Guaranteed information recovery in 500 years (ISO standards). I guess we can call this “long term preservation”.

Data is visible

open platform for reading, only light source and sensor are necesary. Metadata physically embedded on the film. m

OAIS model compliant

Offline storage, migratoin free, few people manages data, trustworthy materials, automatic archival system

Conclusion de FAR: Mais, ce n’est pas OPEN!

Réponse d’Eliane: Bein, soit open, soit long-term preserved. Faut faire un choix!

**Data Deposit Recommendations, DANS, P. Doorn**

<https://halshs.archives-ouvertes.fr/halshs-01531337>

Built on re3data.org

“user-friendly service guiding humanities researchers to deposit data”

Tailyoring re3data, which is a well known service. To improve it, re3data needs to be improved.

Voilà, le site: <https://ddrs-dev.dariah.eu/ddrs/>

Helps you find an adequate repository regarding your field of research and your country. Attention : this only concerns HUMANITIES RESEARCH.

As this tool is built upon re3data, the quality of mData will depend on the quality of mData included on re3data.

## DMPRoadmap : new features in DMPonline and DMPTool (finally we had time to see this one too)

## Sarah Jones DCC

<https://dmponline.dcc.ac.uk/> = ancienne version

<https://dmponline-test.dcc.ac.uk/> = version beta de l’update

There’s been an update on the DMPTool

“ask for feedback” functionnality

“invite collaborator” functionnality

## What about the beat (bit) in the middle

A lot of talk about the context of how research data is currently made. but …

What about active data management? because, they told us at the beginning of the presentation that they were definitly talk about active data management (actually, the term is in the subtitle of the talk).

10 minutes later, we got a demo from a product, “AnalytixAgality” (we repeated together 1h in order to be able to pronounce the name correctly)

<https://www.aridhia.com/platform/analytixagility/>

Different kinds of - quiet powerful - visualisation possible.

Possible to code in R inside the software. LateX also enabled in the platform.

Genomic data

——————————————————

# Parallel sessions II

———————————————-

# More than Data (Fantin)

## Incorporating software curation into research data management services: lessons learned

## (F. Rios, Arizona)

Can we treat software different from other generic data ? Yes but “we are more interested in what the software does than what it is”.

We may loose knowledge about : its execution / its relationships (workflows), other aspects (code, rep. int. prop.), attribution (visibility

Software in RDM

- consulting (best practices / funder & publishers reqs. / IP / repository selection / mData / how does it fit in DMP)

- archiving (instituional data repo / dataset curation)

- education (workshops / training)

knowledge-builidnt & planing is linked with the 3 bullet points above.

Planning outcomes :

- expand archiving workflow (treat sofware separately from data / scope : only sofware produced as part of the research => DON’T TRY TO CAPTURE EVERYTHING, DO FULL RE-EXCUTION)

- expand consulting expertise (eleveate team knowledge on research software in RDM)

- training

Sharing sofware in an Archive :

- how archive worflow ?

- how can software be made more visible in the archive ?

- how should id be curated ? => packaged in one entity / break down in multiple module ? => more of the second option (give enough info to replicate / give credit / reuse)

mData is added while uploading THEN controlled by the researcher who created it once uploaded.

Best practices (advices) : version controle / code organization

The team learned a lot about software data while creating educational materials.

Future goals: chose a better mData schema : CodeMeta could be a good solution <https://github.com/codemeta/codemeta>

## Curating scientific workflows for biomolecular nuclear magnetic resonance spectroscopy

## (M. Gryk, Illinois, UCONN, Health, USA)

Workflow : transform raw data (1)  into frewuency (2) / indentify signals (3) / transform it into biophysical claim (4). The process take between months and years.

(2) Shell script => not really reproducible (although NMR specialist say so). They created their own workflow design / execution system (CONNJUR Workflow system). Exports are in XML. But xml is not really human-intellegible. They exchanged a problem with another.

Implementation of PREMIS in those XML => finnally developped a CONNJUR XML sdchema  (not found on github)

Goal of this : workflow should be more understandable to a broader audience.

CWB sits inside NMRbox (VM downloadable)

## Emedded Metadata patterns across web sharing environments

## (S. Thomspon, Houston)

Images mData => it’s quiet chaotic since everybody is allowed to make photographs and upload them (smartphone, cheapers cameras)

Typical metadata :

- exif

- icc profile

-GFIF (not sure)

- IPTC

- too fast

First, they choosed a sample from various location. Then, they developped data collection, tested metadata stability and finally recorded results (shared on which plateform / embedded mdata)

Result : soft. like photoshop / windows media viewer CHANGEd the embedded mData. No use of such softs for now.

Goal : see if embedded mData travel with the image while beeing uploaded on a social media platform. They do, but they’re changed. Some of them don’t. It’s easy to change some fields are editable in standard media viewer.

Some mData schemes are not readable by some social media platforms.

It arises more questions than answers. But it permits to have a clearer overiew on reusability of suche media files.

## Designing and building interactive curation pieplines for natural hazards engineering dta

## (M. Esteva, Texas)

Natural hazards engineering research : large scale experiments / simulation / hybrid simulation / etc.

DesignSafe\_CI : large national initiative. Cyber infrastrucure platform : data management / data analysis / data curation / data publication > RD END-TO-END PLATFORM.

<https://www.designsafe-ci.org/>

Data challenges : changes accross research steps / final results achieved through serveral iterations / large numbers of files, in different formats / multi-relational datasets / involved documentation / user roles : data creators VS re-users > quite conflictual roles

at first : only curation/publishing module. They had to pass through the whole R&D process (interviews/meetings/modelling research workflows + translate those workflows into front-end UI and back-end mechanisms

user-centered design : choice of vocabularies made with researchers. Every type of experiments are made possible in the platform.

For the researches the boundaries between the steps of DLC are quite burry. They need their data AT ANYTIME. They don’t think step by step : the sofware architecture needs to follow this mindset.

Interactive mockups were set up and tested by researchers. Their requirements, for instance : all models should have sensor informations

Transition from Active Management to publication (they should have the choice of “when” they’re able to publish, and what files, etc.)

mData : editable after active management => mdata transformed in DublinCore Pro in back-end

Browzing  interface (categories / relations between data and process / relations btw data and documentation) => They got a DOI at the end through EZID

Evaluation : curration process seems still a bit of a painful process for researchers. Further automations should be implemented

———————–

# Repository Services (Eliane)

**Two libraries using one texas data repository / A. Dabrowoski, Texas**

- ILSSI : provide workflow for dataset submission

- UTCT : visibility of data

Challenges:

- navigating relationships with stakeholders

- integrating with existing practices

- dealing with large quantities of data

- ensuring maintenance and continuation

Lessons learnt:

- labs face complicated social, policy and technical challenges

- furhter software and policy development

- consortial model is an advantage for sharing solutions

- scaling local solutions involves more engagement from liaison librarians

Future work:

- working with liaison librarians

- costmodels for preservation

- …. trop vite!

**From passive to active, from generic to focussed: how can a ninstitutional data archive remain relevant in a rapidly evlovling landscape, M. J. Cruz, Delft University)  [AUTHOREA ME REND FOU]**

- 4TU.Dataservice

- 10 years old <https://data.4tu.nl/repository/>

- DSA Certified

- 15 years long-term preservation ensured

- metadata is checked and improved

How can we remain relevant in the next years?

- repositories need to have a subject or format focus to remain relevant

–> netCDF 90% of data <https://www.unidata.ucar.edu/software/netcdf/> : standard in atmospheric science, climate sciences and geology

The presentation is online: <https://zenodo.org/record/1175238#.WoxCJefjI2w>

Some services which could be provided:

- visualisation serfices, data processing and data mining

- scalable data citation like recommended by : <https://rd-alliance.org/group/data-citation-wg/outcomes/data-citation-recommendation.html>

- training, advice and guidance

- outreach and international collaboration

if researchers want to share datasets in netCDF, TU Delft would be willing to collaborate

**Building open-source digital curation services and repositories at scale, R. Marciano, Maryland, USA**

Digital curation service

Project: Digital Repostiory at scale : layering different frameowkrs together

<http://dcic.umd.edu/10032016-introducing-open-source-platform-dras-tic/>

Example: Developping a cloud-bsed digital curation service : NSF Brown Dog Project

Another example: creation of a testbed of justice, human rights, and cultural heritage collections

—–

—–

—–

DAY 2

—–

—–

—–

# Remediation Data Management Plans : a Tool for Recovering Research Data from Mesy, Messy Projects

# Clar Llebot Lorente (OSU Libraries and Press, Oregon State University, USA)

Environmental project : “experimental forest”

WRC => they had no RDM at all. At the phase of syntesis/analysis of their data, they started to think “hmmm.. maybe we should have implemented some data management planning ” they contacted the OSU Libraries

The library created a “remedial” rDMP

Differennce between classical DMP and rDMP :

1. Audience : researchers for DMP / researchers OR administrators & manager for rDMP

2. Data Inventory was the most difficult because there were so many datafiles : <http://watershedsresearch.org/>

- Relational databases, trask tabular dtaa, fish database, other digital data, physical samples

- First of all: group the data which are together (like climate data and nutrional data)

- a looooot of data: 50 0000 excel files, 520000 images, 334 databases, 9000 documents, 17000 pdfs

- second: collecto metadata about the data : subject, location, responsibilities, manager, versions, formats, documentation, sensitivity, sharing status

- third: implementation is different, focus on priorities and not on guidelines (i.e. priority 1 : clean, document, preserve in controlled datasets, priority 2 : clean, document and preserve data and associated to past publications, priority 3, triage data in shared drive folders

problems: motivation of researchers, what are the priorities, what should be selected

conclusion: DMPs can be used for remediation, but is this true?

——————

# A landscape survey of #ActiveDMPs

# Sarah Jones for the team which is international

transform static documents in active, machine-actionable DMPs

During IDCC 2017, discussions about:

- interoperability with research systems

- institutional perspectie

- repositry use cases

- evaluation & monitoring

- utilising PIDs

A white book with results of this workshop was published <https://riojournal.com/article/13086/>

common standards

leveraging PIDs

capacity planning

increasing data discvoery & reuse

….. to fast

GITHUB user stories were created : <https://github.com/RDA-DMP-Common/activeDMPs>

A study was made in the framework of OpenAire : “what h2020 dmp users prioritised ?” : <https://zenodo.org/record/1120245#.Wo0_1-fjI2w>

There are still WGs working on that matter :

- during the next RDA face 2 face meeting (March 2018 in Berlin)

- Force11 : Force2018 Montreal <https://www.force11.org/meetings/force2018>

- Australian DMP IG is another group

- Data Management Record in Queensland Australia

- DMPRoadmap <https://github.com/DMPRoadmap/roadmap/wiki>

- UQRDM <https://guides.library.uq.edu.au/for-researchers/research-data-management>

- ReDBox DLC

- Auckland DMP Tool

- Data Stewardship Wizard Elixir

- ezDMP IEDA

- Data planning toolUNINETT Sigma 2

- DMP Service OpenAIRE, EUDAT

To be read: <https://www.slideshare.net/sjDCC/10-simple-rules-for-machineactionable-dmps>

Websitesoon open :<https://activedmps.org/>

Slides will be on: <http://doi.org/10.5821/zenodo.1174283>

**Mandate to integrate DMP in the program for students (idée pour l’EPFL?)**

————————————————–

PARALLEL Sessions

————————————————-

# Data Policy and guidance (Eliane)

## Advancing policy and changes for graduate data management, Virginia Tech, Zhiwu Xie

- policy wordings: encourage DMP, ought to release, as widely as possible, must be provided

- it is necessary to use specific wording

- once data management is routine, it is not necesary to educate on it anymore

- librarians

-          Librarians are not really convincing

-          Read the book « Nudge »

-          Universtiy owns the data, but you can, should, ought to share it, however, if something goes wrong, you are responsible for it

**The impact on authors and editors of introducing data availability statements of nature journals, R. Grant, Springer Nature**

Springer Nature has four policy types

-          Sharing data and data citation is encouraged but not required

-          Data sharing and evidence of data sharing encouraged

-          Data sharing encouraged and statements of data availability required

-          Data sharing not encouraged

Data availability statement (DAS)

-          Templates are available

<https://www.nature.com/news/announcement-where-are-the-data-1.20541>

Editors needed to write down the time they need to write down a DAS

Processing times go up when making data policies mandatory

**Disciplinary data publication guides, Z. Beckles, Bristol UK (for EPFL?)**

**–> à aller voir la présentation**

- very big differences in disciplines

Does the library need to publish guidance ? well, we did it anyway

Methodology:

* Paper type
* Data class
* Data type
* Publication mechanism

Ontology for this different classes

 Transform into a decision tree for it

Conclusion

- by in from academics is critical

- requires a lot of background information

- sustainability, maintannce?

To be found on : <http://10.0.23.196/m9.figshare.5803266> (but it doesn’t work)

**Data Stewardship - adressing disciplinar data management needs, M. Teperek, TU Delft**

<https://zenodo.org/record/1175032#.Wo1Yt-fjI2w>

It is actually what we’ve learned there already, but I will try to take a few notes

Goal of the project: Research integrity

Data stewards:

Data stewards are more like consultants at TU Delft

a few data stewards are still missing

a lot of trainings for data management

- weekly meeting

- one to one meetings with the data stewards per month

- meetings with the faculty superivsor

- sharing with other institutions (blog, open meetings)

34 interviews with researchres, RDM Survey

Challenges:

- prioritize

- how do we know whether the approach we have chosen, is actually working

- Survey on yearly basis could be a very good idea

- is one data steward per faculty enough? or do we need data champions?

Future plans:

- awareness raising and training

- data champions programme

- and tooo fast…..

# Data creation and re-use studies (Fantin)

## Finding, Accessing, Reusing : Art Making, Digital Curation and Real-World Value

## Laura Molloy (University of Oxford, UK)

Art is strong economic asset - 2 billion pound a year

UK Art Market : different mechanisms and mindset. second biggest art market in the world

How the artist survive > profession increasingly under prossure. Oh surprise : artist are underpaid.

Side jobs necessary, mostly self-employed

The project :

-develop and guidance recommandation that

- support good practice in digital object seeking management and sissemination which will

-benefit visual art practictioners and

-support sustainable careers in visual arts

Approach :

- practitione-centered enquiry

- semi-structured interviews (quantitative data: 47 artists in the UK) > workflow for artists : which task is digital, what knowledge do they need to realise those tasks

- produces audio, textual and visual data

+ workshops to make confortable with this methods

Sample :

- visual artists (not specifically digital art

- currently practicing in the UK

- other req. too fast to not them all

Definition of artforms, categories (eg. sculptures) + data format

Digital objects were created everyday, although no participants received RDm training from their art school.

The practices in backup are lacking, and yet there is distress of thinking of losing these DO.

Their knowledge was quite poor (auto-evulation) but they had an appetite for training

In this case, data skills are valuable

## Experimenting with Citizen Scholarship Our Theatre Royal Nottingham : Its Stories, People and Heritage

## Laura Carletti - University of Oxford

Project still ongoing : “I’m not here to give a lesson today, but to receive your advice”

Performing venue (theatre), archival is not their main role, but they definitly do have an archive

Focus groups led to key themes that would be developped eventually. They had a funding, volunteers were engaged (mix of professionals and amateurs).

Theatre had an analogue archive not accessible to public + a small digital archive

Huge work on the analogue side =>  massive reorganisation

Digital side : creation of a web platform <http://ourtheatreroyal.org/s/default/page/home>

Final notes .

- involoving communities in a research-led process to co-curate local heritage

- citizen scholarship : engagement and training to contribute in arts and humanities research process

## Emerging Roles for Optimising Re-Use of Open Government Data

## Fanghui Xioo (University of Pittsburgh, USA)

Research sample :

- geographical  distribution / the city size / scale

- maturity : 2015 data center for one of the three / some of them are far more older

- familarity : some of them were well-known from the authors, some of them weren’t

Raw data collected of official website => analysis, the raw data could be classified in 5 categories

- availability (make sure the data is open)

- understandability (should be easy to understood)

- User data Literacy Improvement (develop skills and training)

- Technical Help (creation of APIs)

- Social Engagement (interactivity : promote citizens participations in the social archive)

Results :

- what are the common issues of open government data center : engange citizens to use OGD / empowerment use of data / easier to find and uses => IMPROVING THE USE OF DATA

- what user services and supportive tasks are provided by open gov data centers  … too fast => CREATE NEW SUPPORT SERVICE

## Tiny Data : Building a Community of Practice auround Humanities Datasets

## Veronica-Gaia Ikeshoji-Orlati (Vanderbilt University, USA)

How Faire Principles are integrated in Humanities Data ?

Tiny Data : “are at the core of traditional and contemporary hmanistic inquiry, reflecting scholars’ crictical engagmenets with texts,  iamages, sound and performance.”

Still work in progress

make the community self-sustaining : for instance create a “data champion” initiative for humanist

Tiny data documentation for humanists

Rethink what FAIR humanities data look like and the potential impact on humanities researchers’s ideas of repdoucibility in their fields

Teach liaison librarians how to find, and encourage engangement with existing datasets

——————————————————————

## Keynote : Collaborating across communities

<https://www.dataone.org/webinars/dpm-%E2%80%9Cstack%E2%80%9D-management-infrastructure-frame-digital-preservation-parallels-technical>