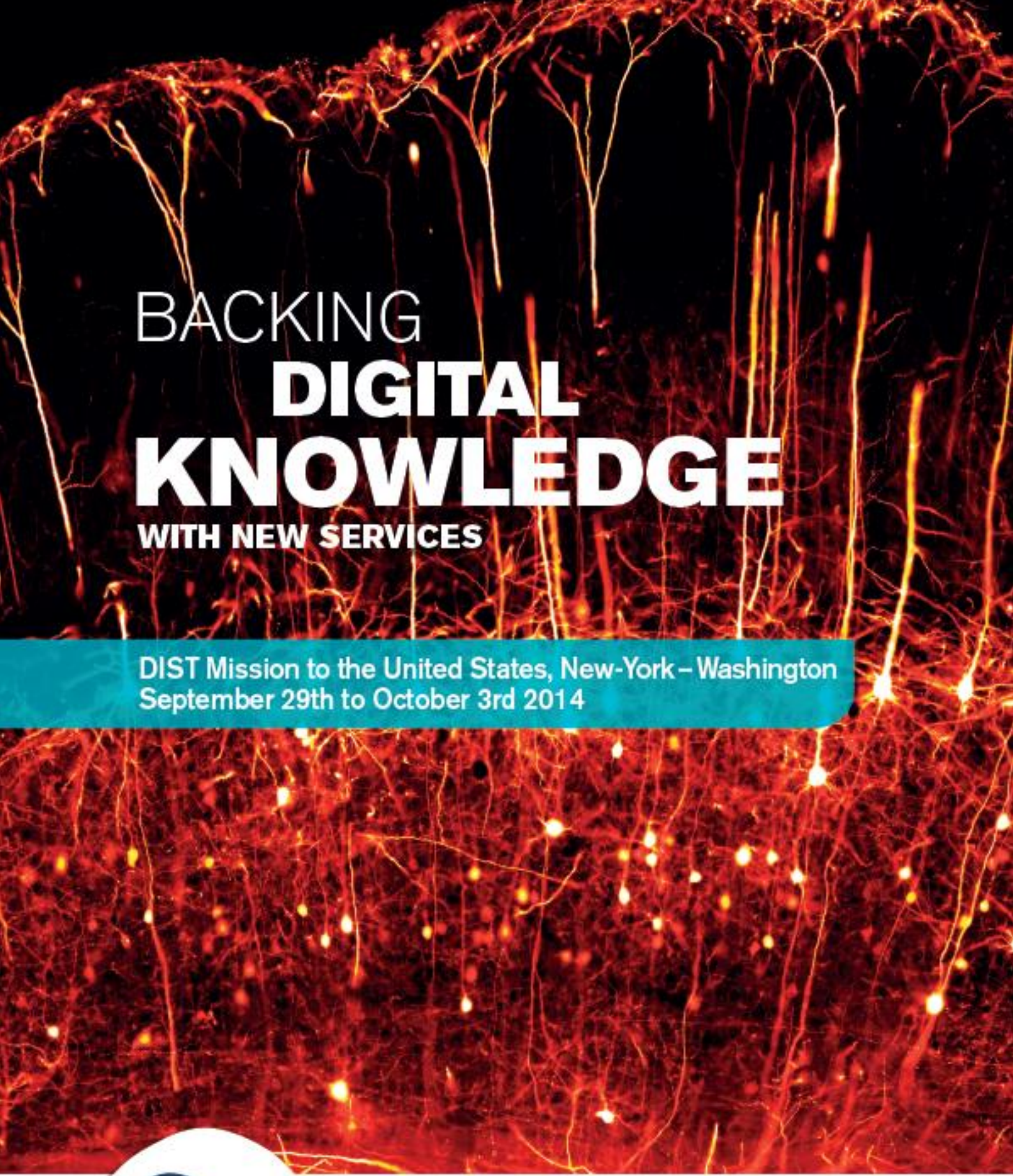


# BACKING DIGITAL KNOWLEDGE

## WITH NEW SERVICES

DIST Mission to the United States, New-York – Washington
September 29th to October 3rd 2014

www.cnrs.fr

Direction de l'information scientifique et technique

# "Backing Digital Knowledge with New Services"

# DIST Mission to the United States

*New York, Washington from September 29th to October 3rd 2014*

This report on the "**Backing Digital Knowledge with New Services**" Mission presents the results of follow-up work on initiatives started by the CNRS's Scientific and Technical Information Department (DIST) in April 2014. These initiatives have been enriched and defined in greater detail to serve as the basis for several working meetings which will take place in the United States in the second half of 2015. The overall subject of these meetings is of particular strategic interest for the future of the CNRS's STI projects - by emphasizing the importance of digital support services for research work, these meetings will provide more in-depth insight into answers to the question: "How do researchers do their research?" It will thus be possible to share a certain number of markers concerning the new generation of support services for research.

I would like to thank the CNRS Office for North America as well as our friends and partners from the American organizations and associations for the quality of their welcome and of the proposals they made to the DIST mission team.

Renaud FABRE

Director of Scientific and Technical Information

CNRS

## Contents

**Summary Table** ..... **5**

**Monday September 29th 2014** ..... **7**

Libraries Information Services (Columbia University) ..... 7

Facebook ..... 14

**Tuesday September 30th 2014** ..... **17**

New York University (NYU) ..... 17

French Embassy in the United States ..... 20

**Wednesday October 1st 2014** ..... **23**

American Chemical Society (ACS) ..... 23

**Thursday October 2nd 2014** ..... **25**

National Endowment for Humanities (NEH) ..... 25

**Friday October 3rd 2014** ..... **28**

Corporation for National Research Initiatives (CNRI) ..... 28

| Summary Table | | |
|---|---|---|
| **NEW YORK** | | |
| **Partners** | **Proposals** | **Planned dates** |
| **Columbia University Libraries** | Following Rebecca KENINSON's meeting with Phil Bourne, Associate Director for Data Science (ADDS) at the NIH, she will report back to the DIST on their discussion.<br>Amy NURNBERGER is a member of the *Research Data Alliance* (RDA) and will attend the 2015 conference in San Diego. The DIST has been taking part in the RDA since its creation, represented by Francis ANDRE, Project Officer for Research Data. They will be introduced at the conference.<br>The Columbia *Libraries Information Services* has been invited to take part in the autumn 2015 seminar to be organized by the CNRS-DIST and NYU in New York. | **March 9th to 11th 2015**<br><br>**Autumn 2015** |
| **Data Science Center (Yann LECUN – Facebook)** | Yann LECUN restated the Data Science Center's interest in setting up a joint seminar between the CNRS (DIST) and New York University. Juliana FREIRE, NYU professor specialized in Computer Science and Engineering will be the DIST's contact for setting up this seminar.<br>Other institutions may become involved with the seminar such as the SLOAN Foundation or the MOORE Foundation. | **Autumn 2015** |
| **New York University** | The DIST and New York University (*Data Science Center*) will organize a joint seminar which may also involve the SLOAN Foundation, the MOORE Foundation and the Universities of Columbia and Washington.<br>Around fifteen experts will be invited to take part in the seminar.<br>The DIST will send a provisional seminar program to Juliana FREIRE who will then obtain a list. All logistical matters will be managed jointly by the CNRS-DIST and New York University. | **Autumn 2015** |

| Summary Table | | |
|---|---|---|
| **WASHINGTON** | | |
| **Partners** | **Proposals** | **Planned dates** |
| **French Embassy in the United States** | Marc DAUMAS will invite the DIST to take part in one of the NIST's annual data conferences. The Embassy has restated its interest in getting to know better the directors of organizations and their areas of work and interest to be able to represent them more effectively. | |
| **American Chemical Society** | Organization of a joint CNRS/ACS seminar. The ACS will re-contact the CNRS – DIST to set up the necessary exchanges for the organization of this seminar. The program and expert participants (about fifteen) will be proposed by the ACS and the CNRS.<br><br>The Washington office proposes to host this meeting of experts at the Embassy and also perhaps the following *social dinner*. | **Autumn 2015** |
| **National Endowment for Humanities (NEH)** | The CNRS could take part in the "Digging into Data Challenge" program.<br><br>Marin DACOS director of Open Edition (CNRS) has been invited to meet the NEH's teams.<br><br>The American teams working on Maurice GODELIER's "Origin of the State and diversity of its forms" project will get back in touch with the NEH. | |
| **Corporation for National Research Initiatives** | Meeting with P.BAPTISTE, DGDS CNRS<br>A French member will be designated for the DONA | **November 2014** |

## Monday September 29th 2014

### ❖ Libraries Information Services (Columbia University)

COLUMBIA UNIVERSITY LIBRARIES

http://library.columbia.edu/index.html

Meeting with:

- **Robert CARTOLANO**, Associate Vice President, Digital Programs and Technology Services
- **Rebecca KENNISON**, Director, Center for Digital Research and Scholarship
- **Amy L.NURNBERGER**, Research Data Manager, Center for Digital Research and Scholarship
- **Mark NEWTON**, Production Manager, Center for Digital Research and Scholarship
- **Simone SACCHI,** Research and Scholarship Initiatives Manager, Center for Digital Research and Scholarship
- **Leyla WILLIAMS,** Communications Coordinator, Center for Digital Research and Scholarship

| Contacts | |
|---|---|
| **Robert CARTOLANO**<br>Associate Vice President<br>Digital Programs and Technology Services<br>Libraries/Information Services<br>Email: rtc@columbia.edu mailto:<br>Address:<br>508 Butler Library<br>535 West 114th Street,<br>New York, NY 10027 | **Rebecca KENNISON**<br>Director<br>Center for Digital Research and Scholarship<br>Libraries/Information Services<br>Email: rkennison@columbia.edu<br>Address:<br>201 Lehman Library, International Affairs Building<br>420 West 118th Street, New York, NY 10027 |

| | |
|---|---|
| **Amy L.NURNBERGER**<br>Research Data Manager<br>Center for Digital Research and Scholarship<br>Libraries/Information Services<br>Email: ANurnberger@columbia.edu<br>Address:<br>201 Lehman Library, International Affairs Building<br>420 West 118th Street<br>New York, NY 10027 | **Mark NEWTON**<br>Production Manager<br>Center for Digital Research and Scholarship<br>Libraries/Information Services<br>Email: mnewton@columbia.edu<br>Address:<br>201 Lehman Library, International Affairs Building<br>420 West 118th Street<br>New York, NY 10027 |
| **Leyla WILLIAMS**<br>Communications Coordinator<br>Center for Digital Research and Scholarship<br>Libraries/Information Services<br>Email: lwilliams@columbia.edu<br>Address:<br>201 Lehman Library, International Affairs Building<br>420 West 118th Street<br>New York, NY 10027 | **Simone SACCHI**<br>Research and Scholarship Initiatives Manager<br>Center for Digital Research and Scholarship<br>Libraries/Information Services<br>Email: ssachi@columbia.edu<br>Address:<br>201 Lehman Library, International Affairs Building<br>420 West 118th Street<br>New York, NY 10027 |

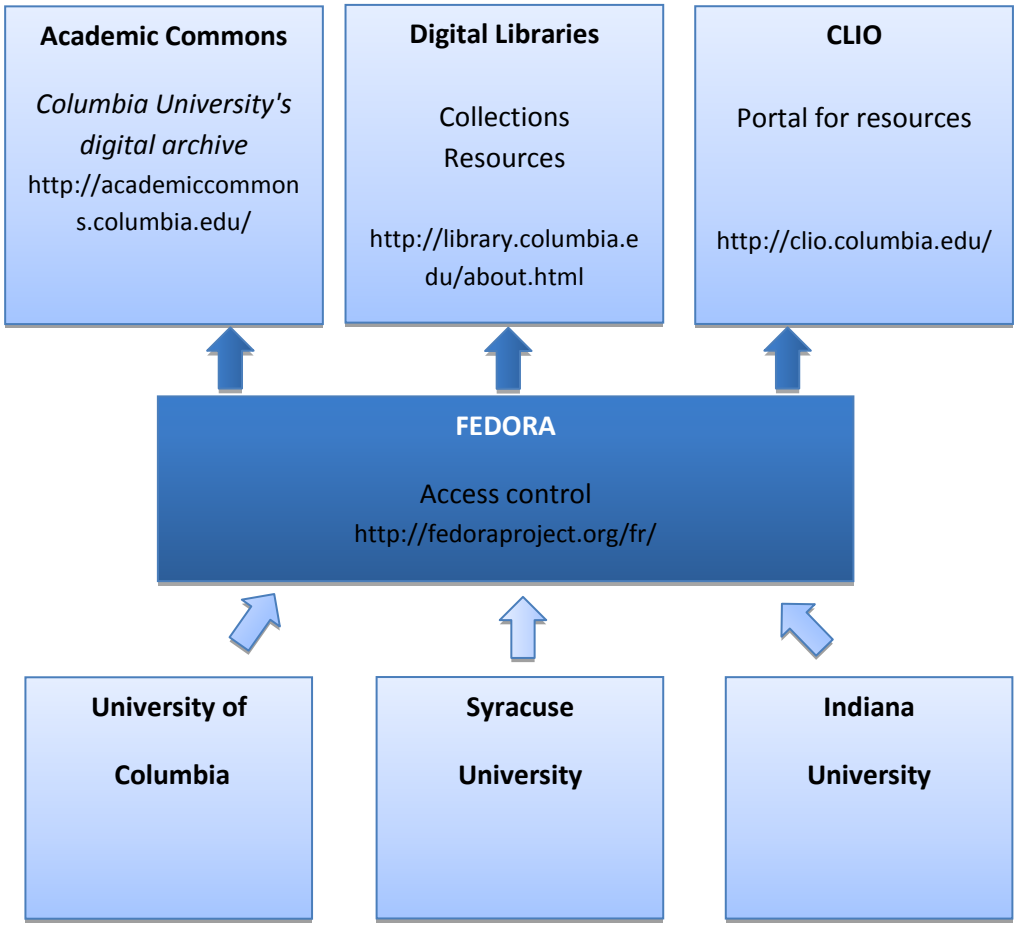

The *Libraries Information Services* of Columbia University provides access to over 12 million volumes, over 160,000 current journals and serials, and a vast collection of electronic resources, manuscripts, rare books, microforms, maps, and graphic and audiovisual materials.

Columbia University manages the digitalization of content and acquisitions to enlarge this collection. The digitalization of print resources or certain audio and video recordings may be carried out by subcontractors or other nearby service providers. The *Libraries Information Services* department enables people outside the University to archive their contents in Open Access research archives or closed *"black"* archives.

For example, it hosts the FORD collection archive for which it has received annual funding of around a million dollars from the Ford Foundation since 2011 in an arrangement which is to run for 7 years[1].

Columbia University also runs the Open Access "*Columbia Undergraduate Science Journal*[2]" which is available on the institutional database.

## Access to Columbia resources




## Evaluation

The evaluation shared by observers today is that there is a growing gap between the quantities of data produced and the means provided to stock, process and archive it. Columbia University finances the acquisition of new resources for the Butler library but data analysis and preservation are financed by subsidies. What should be done when the subsidies end?

[1] https://library.columbia.edu/bts/ford-ifp.html
[2] http://cusj.columbia.edu/about.shtml

Thought about this matter needs to be collective and consortia should be set up so that the work can be carried out on a shared basis.

## Long-term data preservation

Columbia is part of the "Academic preservation trust"[3] project initiated by the University of Virginia which brings together major American universities like Michigan, Stanford, or Columbia who have committed to creating and managing a sustainable environment for the preservation of digital academic and cultural data.

Rapid technological advances (software etc.) have meant that publications with associated data and programs can no longer be reproduced after around a year because of their associated digital environments becoming obsolete. This means we need to define the scope of the environment to be preserved. However no long-term budget has been specifically set aside for "managing research data" and the scale of the issue has yet to be evaluated. Data budgets cover projects and stop at the end of those projects. The potential ensuing loss of information and data could lead to a "digital dark age" as data web specialists are beginning to call it in reference to the fall of the Roman Empire, after which there are very few written historical accounts[4].

## The role of libraries in driving the digital transformation of research institutions

Changes in recent years mean new policies; rules and practices are required to deal with the new challenges.

In scientific communities, researchers lack both the time and knowledge necessary to deal with such problematic subjects. There is a real need for a budget line dedicated to data management including conservation, analysis and archiving. Researchers need to make an effort to document their work and data to thus facilitate interoperability, interdisciplinarity and the exchange of knowledge. Document specialists encourage "the re-use and re-purpose" of data. Certain initiatives will require a legal framework to be defined while others need a technical framework. Libraries have committed to working in these directions and this is the challenge for the next 10 years.

In the past, information was stored in in print versions of books in libraries. Libraries are still the place which houses past, present and future information and need to ensure access is provided to knowledge. Today's knowledge is developing at high speed beyond former physical borders.

---

[3] http://aptrust.org/

[4] http://rt.com/usa/232243-digital-development-documents-lost/

Where should these knowledge databases be constructed?

## Data's legal framework

The digital format brings up the question of author's copyright. The amount of data exchanged and the ease of creation of new digital data are immense and the traceability of data is thus difficult.

Robert CARTOLANO's department works in close collaboration with Columbia University's Copyright Advisory Office to define guidelines and rules for self-archiving articles at Columbia University. The digital work framework is changing so quickly that the Copyright texts written in 2000 now need to be updated.

## The ISTEX project [5]

Columbia University has taken an interest in Version Control on the ISTEX platform. Which data is acquired and is that data relevant? These questions will be answered by the platform's development of services by April 2017 which will enable project beneficiaries (researchers, faculty members etc.) to process the acquired resources using data mining or enrichment techniques for example.

Eventually, the *Libraries Information Services* team would like to see researchers' data stocked on national or international platforms like ISTEX rather than on private platforms or institutional databases.

The question of data succession right planning in the event of the host institution closing is also of interest. The *Libraries Information Services* team is currently setting up its succession plan. Columbia has hailed the Istex initiative which enables long-term access to researchers' published data. Libraries have indeed tended to subcontract publishing to publishers whose mission is not to guarantee long-term storage.

## H2020 recommendations in Europe and OSTP recommendations in the United States

Currently Europe and the United States are faced with the same issues which need solving. The necessary means and structures for data preservation now need to be provided. Currently this is the role of IT institutions but how and when should data be preserved? How should the data be selected?

[5] http://www.istex.fr

Discussions are needed with researchers to help define and document research methods and processes to find the most suitable environment to set up around research communities to enable their data to be preserved.

## Identification of researchers

All researchers who deposit work or interact with a funding institution or agency do so using a "**research identifier"**. To facilitate research mapping, the system of identifiers needs to be normalized and a national identification model set up. Currently two systems co-exist in the United States - the ORCID[6] system which provides researchers with identifiers and the "Authority record[7]" which can be used to create career records for research authors (scientific publications).

The objective underpinning the optimization of research identification is to be able to track both researchers' scientific production and their home institutions so that all institutions could have instant access at any time to an individual's research production.

The **"Linked data for libraries" (LD4L)**[8] project run by Cornell, Harvard and Stanford Universities enables "Authority record" metadata to be linked to libraries' data which thus links researchers to library referentials.

## Tools, projects and studies

### The ODIN[9] Project initiated by *ORCID*[10] *and* the *DataCite Interoperability Network*

This two-year project began in September 2012 and is funded by the European Commission's 7th Framework Programme for the coordination and support of research. It is based on these two successful projects and aims to raise awareness of the importance of using long-term identifiers for authors and objects.

This enables the following to be set up: 1. A reference to a piece of data. 2. Tracking of uses and re-uses. 3 The link between data (as an object and its constituent parts, articles, rights and all those involved in its life cycle (creator, publisher, reviewer, etc).

---

6 http://orcid.org/

7 http://www.loc.gov/marc/uma/

8 wiki.duraspace.org/pages/viewpage.action?pageId=4135402

9 http://odin-project.eu/

10 http://orcid.org/

## SIdora

SIdora is a software program designed to gather all the work produced by researchers at the Smithsonian[11]. Its underlying objective is to actively support the research process by enabling the long-term preservation of all digital content produced by research projects so that it can be curated if necessary.

## Taverna tool kit[12]

Taverna is an open source workflow management system made up of tools which can be used to design, implement and work with scientific workflows. Researchers are increasingly productive in terms of data and depend on using increasingly complex scientific simulation software programs. Incompatible data formats which are produced or used can cause difficulties. Scientific Workflows help solve these interoperability and homogenization issues and give details on which tasks need to be carried out during a given experiment.

### Conclusion

The DIST and Columbia teams will stay in contact to share information on their progress with the subjects discussed, particularly:

- The gap between the data produced and the means available to process, analyse and archive it
- The legal framework, rules and best practices for data
- Data management

Following Rebecca KENINSON's meeting with Phil Bourne, Associate Director for Data Science (ADDS) at the NIH, she will report back to the DIST on their discussion.

Amy NURNBERGER is a member of the Research Data Alliance (RDA) and will attend the 2015 conference in San Diego. The DIST has been taking part in the RDA since its creation, represented by Francis ANDRE, Project Officer for Research Data. They will be introduced at the conference.

The Columbia Libraries Information Services has been invited to take part in the autumn 2015 seminar to be organized by the CNRS-DIST and NYU in New York.

[11] http://www.si.edu/

[12] http://www.taverna.org.uk/download/workbench/2-5/digital-preservation/

❖ **Facebook**


https://www.facebook.com/

Meeting with:

- **Yann LECUN,** Director of AI Research

Yann LECUN is the founder of the ***Data Science Center*** at New York University.

| Contact |
| --- |
| **Yann LECUN**<br>Director of AI Research<br>Email: yann@fb.com<br>Address:<br>770 Broadway, 8th floor<br>New York, NY 10003, USA |

During the first CNRS-DIST mission ("*A Better Sharing of Knowledge*") to the United States from March 24th to April 4th 2014, the team met Yann LECUN who was then the director of the *Data Science Center* at New York University. During this meeting, the organization of a joint seminar for experts on the principles and rules of regulation for science platforms was posited as an interesting idea.

A seminar bringing together the CNRS, the CERN and the DG Connect (European Commission) scheduled for the November 28th 2014[13] in Paris will be the first of a series of 3 seminars to be organised by the DIST of the CNRS on questions related to science platforms. The following seminars will take place in New York and Washington.

During the meeting with Yann LECUN, the questions of data on science platforms which were to be covered at the November seminar in Paris were discussed.

[13] http://www.cnrs.fr/dist/z-tools/documents/CERN_CNRS_DG-CONNECT%20Workshop_nov2014.pdf

## The gap between the production and analysis of Science data

Science platforms process and analyze data. Clear examples of the growth in such data can be seen at the global scale and reveal among other things the reasons for this development and the conditions for storage of results in laboratory archives and reuse of that data. Proof of the growing disparity between the production of data and the capacities and quality of data processing and analysis are gathered through various channels (intensive calculations, visualization, analysis infrastructure etc). The idea is to share ideas on the growing gap between data production and processing based on relevant hypotheses concerning how to share results.

The issue under discussion concerns the management of data flows on science platforms so that data is managed using a process which adds value for Research.

## Sharing knowledge about data analysis

Data and publications are increasingly linked to "digital objects". They are processed using analytical tools of different types (requiring both basic and specific software programs) and are part of the scientific discovery processes which help bring about advances in shared knowledge and excellence. The analytical processes which can be run on advanced digital Science platforms make up a vast landscape including hundreds of high quality tools (publications, data, models, software, etc.)

Today, ideas need to be developed on how European and world approaches can be co-ordinated to promote the emergence "collaboratories" and networks of analytical tools. This aspect is crucial for the development of interdisciplinary science which is not driven solely by the creation of data, but also by the capacity to fully understand and process that data.

The issue under discussion concerns the architectures of state-of-the-art digital platforms for Science in Europe and worldwide and how these can be linked to promote scientific collaboration and knowledge sharing.

## Sharing best practices and rules for data management

Science is undergoing a transformation involving the integration of new paradigms based the intensive use of digital data. However such transformations need to include awareness of heritage in terms of technology, services and practices established in scientific and educational communities.

Research data and its associated tools are used and managed by scientific teams with their own ways of working, ethical rules and capacities for sharing. The rules, practices and laws applied are currently dispersed within a myriad usage pattern.

There is also an important need to map and better understand the main trends and best practices which may develop in Europe to promote and foster the ambition of scientific excellence.

Similarly we need to learn from new experiences in different scientific fields wherein scientists have been able to apprehend the unique dynamics of state-of-the-art digital platforms for Science.

The issue under discussion concerns the new legal frameworks ('soft' or 'hard' law) required to facilitate the development of state-of-the-art digital platforms for Science and thus reduce obstacles to sharing primary information while increasing trust in Science's capacity to share knowledge in a transparent manner.

## Conclusion

Following this discussion, Yann LECUN restated the Data Science Center's interest in setting up a joint seminar between the CNRS (DIST) and New York University. Juliana FREIRE, NYU professor specialized in Computer Science and Engineering will be the DIST's contact for setting up this seminar. Other institutions may become involved with this seminar such as the SLOAN Foundation or the MOORE Foundation.

## Tuesday September 30th 2014

### ❖ New York University (NYU)

NEW YORK UNIVERSITY

http://cds.nyu.edu/

Meeting with:

- **Juliana FREIRE**, Professor, Computer Science and Engineering
-

| Contact |
| --- |
| **Juliana FREIRE**<br>Professor<br>Computer Science and Engineering<br>Email:<br>Juliana.freire@nyu.edu<br>Address:<br>2 MetroTech Center, 10th floor<br>Room # 10.097<br>Brooklyn, NY 11201 |

**Data Science Initiative**

Since 2013, New York University and the Universities of California and Washington have been collaborating on a project with 37 million dollars of funding which is supported by the *Gordon and Betty MOORE Foundation* and the *Alfred P.SLOAN Foundation*. The project's objective is the development of innovative partnerships working on the evolution of support technologies for data management and technical analysis.

## Management of the analysis platforms and data conservation

**Shared observations:**

Both parties noticed:

1. A gulf between the large volumes of data and the capacities for its storage and analysis
2. The necessary elaboration of guides to best practices, ethical charters and a legal framework for platform management
3. The necessity to share the analysis and to invest in data science
4. The necessity of forming trios: the researcher (scientific expert), IT specialist, librarian

There will not be a single answer to these observations, but rather different solutions depending on the scientific domain in question.

New York University (NYU) has included modules devoted to data (management, analysis, storage) in the courses that it proposes. In these modules, the professors communicate to their students the "best practices" for the use of the data produced during the courses. NYU's aim is to put in place for its students, lecturers and researchers a guide to best practices, particularly for data reproducibility. To this end, experts in the scientific domain, computing specialists and experts of databases interact with each other.

To date, it has been possible to install these best practices and tools, for specific domains. However, these processes are difficult to implement at a global and interdisciplinary level.

New York University also wishes to develop a strong policy of data storage in archives (closed, open access, etc.).

In the future, courses devoted to data Science (management, analysis, storage) could be created.

Today, numerous community initiatives exist, such as Github, which is widely used for the codes. Similarly, Git Data[14] is an initiative for data management and exchange. These sites make the API available, allowing access to the reading and writing of Git objects in the users' databases.

[14] https://github.com/

## Center for Urban research Science Program - CUSP[15]

The *Center for Urban Research Science Program*, that was created by NYU under the impetus of Michael BLOOMBERG, mayor of New York, is a unique public-private research center that uses the town of New York as a research laboratory to help towns throughout the world become more productive, livable, equitable, etc. The CUSP observes, analyzes and models towns to optimize the results, provide new solutions, formalize new tools and processes and develop new skills. The aim of the CUSP is to become a world leader in the emerging field of "Urban information technology".

## A legal framework around the data

NYU recognizes the importance of installing a legal framework (charters of best practices, rules) around the use of the data. Thought processes concerning these aspects are ongoing. These questions raise issues of private property. Who does the data belong to?

### Conclusion

The DIST and New York University (*Data Science Center*) will organize a joint seminar in autumn 2015, which may also involve the SLOAN *foundation* and the MOORE foundation, as well as Columbia and Washington Universities.

Around fifteen experts will be invited to take part in the seminar.

The DIST will send a provisional seminar program to Juliana FREIRE, who will then obtain a list of experts, including for example Josh GREENBERG (director at the Sloan of the Digital Information Technology program and one of the founders of the ZOTERO program). The logistical questions will be managed jointly by the CNRS-DIST and New York University.

[15] http://cusp.nyu.edu/about/

### ❖ French Embassy in the United States

http://france-science.org

Meeting with:

- **Marc DAUMAS,** Attaché for Science and Technology, French Embassy in the United States

Presentation of the French Scientific Mission to the United States and of the Scientific Service (CNRS).

| Contact |
| --- |
| **Marc DAUMAS**<br>University Professor<br>Attaché for Science and Technology<br>Email :<br>marc.daumas@diplomatie.gouv.fr<br>Address: French Embassy in the United States<br>4101 Reservoir Road NW<br>Washington, DC 20007 |

## Their programs, missions and events

- STEM Chateaubriand Fellowship[16]

The *STEM Chateaubriand Fellowship* is a research program for about 30 students per year, for 4 to 9 months. This program allows links to be forged between American PhD students (in their 3rd or 4th year of thesis – a thesis lasts 5 years in the United States) and French laboratories.

[16] http://stem.chateaubriand-fellowship.org/

The program has existed for about fifteen years and it allows co-supervisions to be set-up, as well as reinforcing partnerships between American and French researchers. The NSF has a similar program destined to promote collaborations with Northern Europe.

- New Technology Venture Accelerator (NETVA)[17]

This program is destined for French companies who want to establish themselves in the United States or to work in collaboration with the United States. It is a minor program that allows immersion of the CEOs of businesses into American companies for a week. This program concerns innovation and small companies. The cost is approximately 1700€. The program extends to Canada.

- Young Entrepreneur Initiative (YEI)[18]

In contrast to NETVA, this program is for American entrepreneurs who wish to establish themselves in France.

Other missions / programs / events are directed by the French Scientific Mission to the United States, such as the scientific cafés, France Atlanta, *The French American Innovation Day,* etc.

### Horizon2020

Horizon2020 has requested the French Scientific Mission to the United States to set up co-operative projects with the United States. However, in the context of these projects, only the European partners are financed and not the American groups. Bi-financed partnerships for European/United States do not yet exist in the framework of Horizon2020.

On the other hand, the NIH is prepared to finance the American partner in European / United States projects from the American defense budget.

### National Institute of Standards and Technology (NIST)

The *National Institute of Standards and Technology* organizes open events all year long, particularly on the subject of BIG DATA: http://bigdatawg.nist.gov/home.php.

[17] http://www.netvafrance.com/

[18] http://www.yeifrance.com/

The French Scientific Mission to the United States is in contact with the NIST, which has constituted, for example, the working group "Data Science". A joint thought process is under way concerning strategies around the data. The documents related to this working group are accessible on-line at http://www.nist.gov/itl/iad/data-science-symposium-2014.cfm

**Conclusion**

Marc DAUMAS will re-contact the DIST to invite it to take part in one of the NIST's annual data conferences.

The Embassy has restated its interest in getting to know better the directors of organizations, their fields of action and interests, to be able to represent them better.

## Wednesday October 1st 2014

### ❖ American Chemical Society (ACS)

http://www.acs.org/international

Meeting with:

- **Bradley D.MILLER**, Director, Office of International Activities
- **Brandon NORDIN**, Vice president , Sales Marketing & Web Strategy-Publications

| Contacts | |
|---|---|
| **Bradley D.MILLER, Phd**<br>Director<br>Office of International activities<br>Email :<br>b_miller@acs.org<br>Address:<br>American Chemical Society<br>1155 Sixteenth Street, NW<br>Washington, DC 20036<br>USA | **Brandon NORDIN**<br>Vice President<br>Sales, Marketing & Web Strategy Publications<br>Email :<br>b_nordin@acs.org<br>Address:<br>American Chemical Society<br>1155 Sixteenth Street, NW<br>Washington, DC 20036<br>USA |

**Joint ACS/CNRS Seminar**

During the first DIST Mission to the United States from March 24th to April 4th 2014, the organization of a joint CNRS / ACS seminar about economic models of publication on chemistry platforms had been envisaged.

Effectively, both parties noticed:

1. The gulf between large volumes of data and the capacities for their storage and analysis
2. The necessary development of guides to best practices, ethical charters and a legal framework for platform management
3. The necessity to share the analysis and to invest in data science

The modalities of this seminar were defined during this follow-up Mission:

- Themes to be addressed
- Setting up exchanges for the organization
- Date

The seminar with the ACS will be based on the same model as the seminar held in Paris with the participation of the CNRS, the CERN and the DG Connect (European Commission). The definitive program of this latter seminar, along with the acts, will be transmitted to the ACS at the beginning of 2015 to determine the joint themes to be approached during the Washington seminar

http://www.cnrs.fr/dist/z-outils/documents/CERN_CNRS_DG-CONNECT%20Workshop_nov2014.pdf.

**Conclusion**

The joint CNRS / ACS seminar will be held in the autumn of 2015.

The ACS will re-contact the DIST-CNRS to set up the necessary exchanges for the organization of this seminar. The program and expert participants (about fifteen) will be proposed by the ACS and the CNRS.

The Washington office proposes to host this meeting of experts at the Embassy and also perhaps the following *social dinner*.

## Thursday October 2nd 2014

### ❖ National Endowment for Humanities (NEH)


http://www.neh.gov/

Meeting with:

- **Brett BOBBLEY**
- **Jennifer SERVENTI**

| Contacts | |
|---|---|
| **Brett BOBBLEY**<br>Chief Information Officer<br>Director, Office of Digital Humanities<br>Email : bbobley@neh.gov<br>Address:<br>Constitution Center<br>400 7th street SW<br>Washington DC 20024 | **Jennifer SERVENTI**<br>Senior Program Officer<br>Office of Digital Humanities<br>Email : jserventi@neh.gov<br>Address:<br>Constitution Center<br>400 7th street SW<br>Washington DC 20024 |

## Open Access

The NEH incites researchers to deposit by Open Access on the platforms that it subsidizes such as "Open Context" or "the Perseus Digital Library" that are increasing in popularity due to the NEH's support.

## Digging into Data Challenges[19]

The *Digging into Data Challenge* is a grant program sponsored by several major international research organizations, which has as its aim to analyze how "Big Data" changes the research panorama in the Humanities and Social Sciences.

Today, the huge databases that are available for research into the Humanities and the Social Sciences are of all imaginable types: digital books, newspapers, music, information generated by Internet activities and portable communications, administrative data from public organizations, client databases from private sector organizations, etc. Faced with all this data and with new research methods, at a time when the world is becoming more and more digitalized, new techniques are necessary to research, analyze and understand these materials.

The *Digging into Data Challenge* incites the research community to develop and accompany the new research infrastructures of the 21st century through financing and grants.

**First wave (2009)**
When it was launched in 2009, the *Digging into Data Challenge* was sponsored by four funding bodies (NEH, NSF, CRSH, JISC) that represented the United States, Canada and Great Britain. Finally, 8 international projects were retained for funding. Several of these projects led to publications in newspapers such as the New York Times. These projects were also the subject of a major research report published by the CLIR (Council on Library and Information Resources). The 8 projects were presented during a conference in Washington in June 2011.

**Second wave (2011)**
Four additional funding partners joined the *Digging into Data Challenge* in 2011. During this second wave, 14 projects chosen by international experts were given awards. The 14 projects were presented during a conference held in Montreal, Canada on the 12th of October 2013.

**Third wave (2013)**
Two more financial partners joined the project in 2013. At this time, 10 countries participated in *Digging into Data Challenge*. At the end of this third wave, 14 projects were recognized.

[19] http://diggingintodata.org/

The digital libraries, archives and museums represent an important part of *Digging into Data Challenge.* These are the organizations that create, re-group, curate and preserve the numerical data studied by researchers. A list of reference documents created during the *Digging into Data Challenge* projects is made available to researchers.

Brett BOBBLEY invited the CNRS to join this program.

## Trans-Atlantic Platform[20]

The *Trans-Atlantic Platform* is a partnership between 15 funding organizations for European and American research. The aim of this platform is to improve transatlantic research collaborations in the key fields of mutual commitments which respond to the challenges of society in the 21st century, involving Humanities and Social Sciences (HSS). The French National Research Agency (ANR) is involved in this collaboration. Each participant contributes 1 million euros to finance its groups over approximately two and a half years.

## Presentation of Maurice GODELIER's project "Origin of the State and diversity of its forms"

The NEH proposes grants for HSS projects such as that put forward by Maurice GODELIER "Origin of the State and diversity of its forms". The NEH is ready to meet with the American members of the consortium proposed to evaluate the funding to which this project could apply.

### Conclusion

The CNRS could take part in the "Digging into Data Challenge" program.

Marin DACOS director of Open Edition (CNRS) is invited to meet the NEH's teams.

The American teams working on Maurice GODELIER's "Origin of the State and diversity of its forms" project will get back in touch with the NEH.

[20] www.transatlanticplatform.com

## Friday October 3rd 2014

### ❖ Corporation for National Research Initiatives (CNRI)


http://www.cnri.reston.va.us/

Meeting with:

- **Robert KAHN**, President, CNRI
- **Patrice LYONS**, General Counsel, CNRI

Robert Kahn, who was met during the first DIST Mission to the United States, indicated the creation of the DONA Foundation (Digital Object Numbering Authority), a non-profitable organization, www.dona.net, in January 2014. This is based in the Geneva municipality and is housed by Geneva University. Christophe BLANCHI was elected Executive Director in July 2014 and will take up his position from January 2015.

#### Its aims

The mission of the DONA is to favor interoperability between heterogeneous information systems. It will allow and extend the use of architecture of digital objects by:

- Supplying **management services**, the **development of software** and other strategic services for the **technical co-ordination**, evolution, application and other uses of public interest around the architecture of digital objects.
- In the framework of these prerogatives, **administering and maintaining the stability** of the **GHR** (Global Handle Registry) function, the GHR being an essential element of the architecture of digitalized objects, and **authorizing and co-ordinating** the GHR administration with the MPA (Multi-Primary Adminstrator).

**Philippe BAPTISTE, DGDS of the CNRS, agreed that France should be represented on the DONA board of directors, if possible by the CNRS.**

**The key figures who are members of this board of directors** must maintain its current diversity and **geographical equilibrium** as far as possible.

They must be **involved in the domain of free architecture of digital objects** and possess **the technical know-how and appropriate expertise** to carry out the DONA's aims.

Furthermore, it has been decided that potential candidates should belong to multipartite organizations with a **certain governmental independence**.

The creation of the DONA has allowed parties other than the CNRI to administer the GHR and thus this responsibility is shared between several organizations. At the end of the year, or as soon as the current contracts with the CNRI (previously the only body authorized to administer DONA) end, the establishments, organizations and entities wishing to use the Handle system will be able to approach any one of the members of the DONA board of directors.

## The initial MPAs designated in July 2014

- Corporation for National Research Initiatives (CNRI)
- Coalition for Handle Services -- China (ETIRI, CHC and CDI)
- International Telecommunication Union (ITU- Union internationale des télécommunications)
- Gesellschaft für Wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)

The organizations could once again, as previously, contact the CNRI (based in the United States) or the GWDG (based in Europe). The ONG at the ITU, along with those who do not want to address any of these bodies, can address China.

## The members of the DONA board of directors – Meeting of July 2014

**Dr Robert Kahn,** (President), CNRI
**Dr Stephen Wolff** – Internet2. Internet international co-operations.
**Dr. Peter Wittenburg** - The Language Archive & Max Planck Institute for Psycholinguistics, First to lead the RDA (Research Data Alliance)
**Dr. Norman Paskin** International DOI Foundation
**Dr. Antoine Geissbuhler** Division of eHealth and Telemedicine (standard for medical records and medical information), Geneva, University & Hospitals
**Mr.Stefan Eberhard** (Secrétaire) ABELS Avocats Geneva (legal interface with the government in Switzerland: tax exemption, corporate staff). He will step down from the board as soon as he is not needed.
It could become an international organization:
**Mr. Adama Samassekou** President of CIPSH, President of the MAAYA Network, Previous Executive Secretary of ACALAN BP E 214 Bamako – MALI, he was the president of the 2 world security summit.
**Mr. Gao Xinmin** - Deputy Head of the Internet Society of China.

## Recommendation UIT-T X.1255

The DONA produced recommendation UIT-T X.1255, the aim of which is to define a framework of open architecture where it is possible to find information relative to identity management (IdM).
The International Telecommunication Union (ITU) is a specialized institution of the United Nations in the domain of telecommunications and information and communication technologies (ICT).

This information will necessarily be represented in different ways and will be managed by various trust frameworks, or other IdM systems, using different metadata schemas. For example, this framework will allow entities functioning in the context of an IdM system to precisely solve the identifiers originating from other IdM systems. The users or organizations (or programs exploited on their account) that cannot find this information have no choice other than to ascertain the best way to establish the credibility and verify the authenticity of an adequate identity, whether this be for a user, system resource, information entity, etc.

In the light of this knowledge, it is up to the user or the organization to determine whether or not, all things considered, he can believe in a given trust framework or in another IdM system.

The principal elements of the framework presented in this recommendation are the following:
**1)** a model of digital entity data;
**2)** a protocol for the digital entity interface;
**3)** one or more resolution/identifier system(s);
**4)** one or more metadata registers.
These elements comprise the basis of the open architecture frame.

## The CNRI

Dr. Robert KAHN developed the concept of the architecture of digital objects. This notion supplies a framework for the interoperability of heterogeneous information systems and allows numerous applications such as the Digital Object Identifier (DOI). He is also the co-inventor of the Knowbot programs, mobile software agents in the network environment.

## Among the programs

The CNRI is developing an infrastructure for the **Knowbot programs**. These are mobile agents destined to be used in systems that are widely distributed, such as the Internet.
They supply a simple-to-use architecture for the development of distributed agents' secured systems. The agents can be any type of software and the architecture allows

several programming languages. The agents are clearly defined. They are autonomous entities that interact with their environment, known as "service-stations", according to precise rules.

A Knowbot software program could be used for different applications, such as the extraction of data or the reinforcement of the negotiation of a protocol between geographically disparate systems. Furthermore, it could serve to mediate access to information in a network environment. Target applications for these systems are the digital libraries that contain sensitive documents. The use of a system based on access agents could allow any entity to examine a document in the library, but restrict the modification or suppression of a document to those entities with the required security level. The opening of digital libraries to confidence agents will allow more flexible research on larger data groups.

**The Handle system** is an effective, extendable service which provides a secured payment system for the unique and persistent identifiers of digital objects. It is a component of the CNRI Digital Object Architecture. The architecture of digital objects provides a means of digital information management in the environment of a network. A digital object has an independent structure (machine and platform) that allows it to be identified, accessible and protected, as the case may be. A digital object can integrate, not only elements of information, i.e. a digital version of a document, a video or an audio recording, but also the unique identifier of the digital object and other metadata pertaining to the digital object. The metadata can include restrictions of access to a digital object, ownership notices and identifiers for permit agreements, as the case may be.

The Handle system includes a group of open protocols, a space for names and a reference implementation for protocols. The protocols enable a computing system to be distributed, to stock the identifiers of arbitrary resources, called Handles, and to form these Handles in the information needed to locate, access, contact, authenticate, or carry out any other use of the resources. This information can be changed as required to reflect the current state of the identified resource, without changing its identifier. Thus the persistence of the element's name during changes of place and other connected information is allowed. Several examples of applications that use the identification services and HDL® resolution for the infrastructures are functions of management of rights, registers, deposits of digital objects, institutional conservation of data and archiving.

**Conclusion**

Philippe BAPTISTE, DGDS of the CNRS, will meet Robert KAHN to study the possibility of the CNRS belonging to the DONA board of directors.