A BETTER SHARING OF KNOWLEDGE

Mission DIST aux États-Unis
New-York / Washington, 24 mars - 4 avril 2014
“A Better Sharing of Knowledge”

DIST Mission to the United States

New York, Washington

March 24th to April 4th 2014
This report relates the objectives of the DIST (CNRS Scientific and Technical Information Department) Mission and provides a summary of observations and proposals along with reports on the DIST delegation's fifteen meetings with Universities, Organizations and Research Centers.

These meetings were efficiently prepared and accompanied by the CNRS Office for North America. The Mission would like to greatly thank Xavier MORISE, director of the Office and his team (Sarah MONDET and Jean THEVES) along with Minh-Hà PHAM, scientific advisor for science and technology at the French Embassy in the United States for the quality and energy of her welcome.

An attachment reports on STI in the United States and provides biographies of the CNRS's interlocutors in the US.
Index

A. OBJECTIVES: UNDERSTANDING THE EVOLUTION OF STI PLATFORMS IN THE UNITED STATES ...... 5
  1. What are the medium-term objectives for STI platforms ........................................ 6
  2. Which management approaches? .............................................................................. 6
  3. Which innovative projects for the analysis and promotion of STI? ............................. 6

B. SUMMARY OF PROPOSALS: A FAVORABLE CONTEXT FOR SHARED STI ................................................. 7
  1. The future of platforms: medium-term objectives .................................................. 8
  2. STI management approaches ............................................................................... 9
  3. Innovative projects for STI analysis ....................................................................... 10
Summary table .................................................................................................................. 12

C. THE CNRS MISSION'S MEETINGS AND PROJECTS .................................................................................. 13
Meeting and project 1: Tuesday March 25th 2014 ......................................................... 13
   The American Assembly (Columbia University) ......................................................... 13
Meeting and project 2: Wednesday March 26th and Friday March 28th 2014 ................. 16
   NYU (Center for data science) .................................................................................... 16
Meeting and project 3: Thursday March 27th 2014 ...................................................... 19
   The Andrew W. Mellon Foundation ........................................................................ 19
Meeting and project 4: Friday March 28th 2014 ........................................................... 21
   Alfred P. SLOAN Foundation .................................................................................... 21
Meeting and project 5: Monday March 31st 2014 .......................................................... 24
   National Center for Biotechnology Information ........................................................ 24
Meeting and project 6: Monday March 31st 2014 .......................................................... 27
   National Library of Medicine (NLM), National Institutes of Health (NIH) ............... 27
Meeting and project 7: Tuesday April 1st 2014 .............................................................. 29
   National Science Foundation .................................................................................... 29
Meeting and project 8: Wednesday April 2\textsuperscript{nd} 2014 .......................................................... 31
  \begin{itemize}
  \item American Chemical Society .......................................................... 31
  \end{itemize}

Meeting and project 9: Wednesday April 2\textsuperscript{nd} 2014 .......................................................... 34
  \begin{itemize}
  \item Virginia Bioinformatics Institute .......................................................... 34
  \end{itemize}

Meeting and project 10: Thursday April 3\textsuperscript{rd} 2014 .......................................................... 36
  \begin{itemize}
  \item Corporation for National Research Initiatives .......................................................... 36
  \end{itemize}

Meeting and project 11: Thursday April 3\textsuperscript{rd} 2014 .......................................................... 38
  \begin{itemize}
  \item American Association for the Advancement of Science (AAAS) .......................................................... 38
  \end{itemize}

Meeting and project 12: Thursday April 3\textsuperscript{rd} 2014 .......................................................... 40
  \begin{itemize}
  \item Coalition for Networked Information (CNI) .......................................................... 40
  \end{itemize}

Meeting and project 13: Thursday April 3\textsuperscript{rd} 2014 .......................................................... 43
  \begin{itemize}
  \item National Institute of Standards and Technology (NIST) .......................................................... 43
  \end{itemize}

Meeting and project 14: Friday April 4\textsuperscript{th} 2014 .......................................................... 45
  \begin{itemize}
  \item Smithsonian Institution, Encyclopaedia of Life, National Museum of Natural History (Washington) .......................................................... 45
  \end{itemize}

Meeting and project 15: Friday April 4\textsuperscript{th} 2014 .......................................................... 47
  \begin{itemize}
  \item National Endowment for the Humanities (NEH), Office of Digital Humanities .......................................................... 47
  \end{itemize}

ANNEXE 1: STI IN THE UNITED STATES .......................................................... 49

ANNEXE 2: Contacts, places and projects .......................................................... 54
A. OBJECTIVES: UNDERSTANDING THE EVOLUTION OF STI PLATFORMS IN THE UNITED STATES

The STI strategy of the CNRS (SIST) included the observation that global STI sharing is accelerating in conditions which remain fairly undefined and liable to evolve. A great deal of digital STI is currently distributed by multi-core platforms which house data, publications or analysis while rapidly developing new STI "objects" and networks (analysis services, nano-publications, automatic publications, semantic libraries, researchers' social networks, etc.).

Since the advent of the major global platforms (SCOPUS, WOS, Pub Med, etc.), multi-core platforms redistribute digital STI down to the local level of laboratories. The information industry works continually on these networks of platforms at all levels of production and develops many services to promote and disseminate STI to all user categories and beneficiaries of public science.

By exploring the complex landscape of the new players in STI in its American context, the mission's aim was to detect projects, resources and ideas developed by the STI departments of Universities, organizations and research centres to enhance the CNRS's own STI strategy. The English translation of this strategy built a bridge towards those the mission met. Understanding the evolution of STI platforms in the United States was therefore our main working goal.

On this theme, the DIST opted to combine the approaches of the four leaders of the CNRS Shared Action Plans (Plans d'Actions Partagés or PAPs) namely;

- Mr. Raymond BÉRARD¹, Director of the Institute of Scientific and Technological Information (INIST) - PAP 1: "Obtaining information";
- Christoph SORGER, Director of the National Institute for Mathematical Sciences (INSMI) - PAP 2: "Publishing";
- Bruno DAVID, "Chargé de mission" (Project Officer) at the Institute of Ecology and Environment (INEE) - PAP 3: "Analyzing and promoting information";
- Patrice BOURDELAIS, Director of the Institute for Humanities and Social Sciences (INSHS) - PAP 4: "Supporting and promoting STI".

The mission was prepared and managed by the DIST (Renaud FABRE, Director of the DIST, Francis ANDRE, Project Officer for "Research Data").

¹ In the end Mr. Bérard was unable to go on the mission
Three main questions were defined to guide our approach and the contacts to be made:

1. **What are the medium-term objectives for STI platforms?**

   The aim was to understand the new ways in which STI systems work with and manage ongoing digital changes;

2. **Which management approaches?**

   Here the objective was to find out more about the organization of STI networks used by the communities we were to meet and the ability of those networks to successfully respond to researchers' requirements;

3. **Which innovative projects for the analysis and promotion of STI?**

   On these three themes, the mission aimed to take note of best practices which could be included in the strategy being developed at the CNRS by gathering information and looking for future collaboration opportunities. This approach followed on from the “STI Innovation and Governance in Higher Education and Research” conference (Meudon, France, March 18th and 19th 2014) whose mission was to continue and go further with analysis.
Two overall characteristics prevail in this summary of STI in the United States:

- **the vigour of STI innovation and analysis**: many major innovative projects could be monitored in more depth or shared with the CNRS;
- **the search for overall regulations for world public science**: debate is underway between the public sector (particularly the OSTP) and the information industry which is attempting to define the limits for its involvement.

These two characteristics are predominant and correlated in North American practices.

Generally the mission found STI initiatives in the United States to be vigorous and diverse but it also found similar expectations and questions being asked regarding the future of digital STI as those included in the CNRS strategy. As in France, the digital STI developed on American multi-core platforms is faced with the paradox of strong yet uncertain development which does not facilitate the optimization of digital processing of scientific information (calculation, storage or curation capacities, etc.) or the evolution in the ways results are published (authors’ copyright, data, platforms, data sharing ethics, etc.).

Similar expectations to our own are present across the Atlantic namely better traceability of research results and more transparency in the management and value chain of STI, based on more stable ethical rules and regulations which favor public research and its beneficiaries.

This community of expectations was a central guideline for the mission and informed the collaborative projects worked out with CNRS interlocutors. These projects and the observations which influence and drive them will be summarized below in the order of priority of the mission’s work (Perspectives, Management, STI analysis). This summary will be based on the "Meetings and Projects Reports" numbered from 1 to 15 and will be presented in the following section with the context for each proposal.

It should also be noted in these preliminary remarks that modes of production and funding of STI tend to favour the appearance of *structures which drive cooperative*
innovation initiatives. These are examples which should be looked at in depth in the context of the mutualization of STI which is currently materializing.

1. The future of platforms: medium-term objectives

The mission made initial positive contacts with the DATA SCIENCE CENTER at NYU, a federal Center which provides researchers and other professionals with interdisciplinary tools to help them exploit STI from research databases.

The idea of organizing a seminar for experts on the principles and rules for the regulation of science platforms was selected. This seminar which could take place at the end of 2015 - start of 2016 would enable the in-depth study of several questions that have yet to be explored in enough detail - digital law for platforms (the CNRS has invested in this area via the ISTEX project), ethical charters on data sharing and interdisciplinarity, ways of including results on science platforms (particularly through user-generated content) and the architecture of value chains on platforms. The SLOAN Foundation (see below) has stated its readiness to become involved in this joint CNRS - NYU initiative.

This seminar would be a logical extension to the DIST’s work on these themes carried out internally (a CNRS Scientific Board working group will meet in Autumn 2014) and externally (CNRS-CERN seminar and DG CONNECT at the end of 2014 with the SCOAP3 partnership as the starting point).

The mission also studied the diversification of services on the major multi-core platforms coming to maturity (and/or experiencing "growing pains"). Pub Med Central is the biggest global open-access platform and the CNRS has shared the projects of its founder David LIPMAN to control the diversification of its activities and promote Medline’s 23 million citations through the NCBI, (NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION).

The meeting with Robert KAHN, President & CEO of CNRI (CORPORATION FOR NATIONAL RESEARCH INITIATIVES) and co-inventor of TCP/IP with Vint CERF (DARPA) was a highpoint of the mission. He suggested to the CNRS that a French person join the management board of the DONA (Digital Object Numbering Authority), which currently manages rolling out Digital Objects on the Internet worldwide. The mission recommends that the CNRS Research Office (DGDS, Direction Générale Déléguee à la Science) should examine this choice in depth. If the CNRS does not propose someone, the DIST suggests that the role be offered to Pascal GUITTON at the Institute for Research in Computer Science and Control (INRIA).
Platforms also seem to have a future as aggregators, a role the AMERICAN ASSEMBLY has proposed to Columbia University, thus building a bridge between teaching and research. The final objective of this is to eventually unify smaller platforms and incite researchers to deposit their data, while offering an overall regulation service for university publishing, Open Syllabus, which also covers all the tasks of daily student activity and manages classes.

2. STI management approaches

The mission began exchanges which could continue with the NATIONAL SCIENCE FOUNDATION (NSF) on the evaluation of programs and using STI as key "material" for this. It is still too early to know what these exchanges may bring (the first reports are currently being analyzed). It would be useful to send all the information on this contact to the Observatory of Science and Technology (OST) and to the corresponding parliamentary Office.

The management of STI platforms is clearly becoming a new profession of complex dimensions. Subsidies from the A.P. SLOAN Foundation help researchers develop new tools, standards and infrastructures. The Foundation concentrates on projects which encourage scientific access and data sharing, facilitate and promote the development of standards necessary for the interoperability of databases, enable the dissemination of computing research and guarantee the key role played by data management and preservation.

The organization of a seminar for experts by the CNRS and the SLOAN Foundation in collaboration with the CIRHUS at NYU was generally considered an interesting idea. The mission was welcomed positively at the CIRHUS UMI and would like CIRHUS to be involved in organizing the Seminar which could take place at the SLOAN Foundation, at least in part.

The program could partly be run jointly with the NYU CENTER FOR DATA SCIENCE which the SLOAN Foundation collaborates with. The subject selected could cover the "values" of public science in the era of global digital platforms. The experience, studies and tools of the SLOAN Foundation could thus be pooled with those of the Universities and CNRS research centers to run a program aimed at defining STI management's relevant principles and results. The major part of this project involves the Humanities and Social Sciences (HSS) but not entirely. The DIST could liaise with the CNRS Office for North America to send a project preparation mission to the US in September 2014.
Also, the NATIONAL INSTITUTE FOR STANDARDS AND TECHNOLOGY (NIST), a vast federal technical structure (2000 agents) which works on all aspects of standards, including research, is open to the idea of a collaboration project between the CNRS and the DATA SCIENCE PROGRAM. The DIST will send proposals regarding this idea to the counterpart team at the NIST, the aim being to take part from the very start in constructing norms which will define standards for interoperability between databases. The advantages of collaborating on GRID projects with the French Complex Systems Institute (ISC) should also be examined.

The COALITION FOR NETWORKED INFORMATION (CNI) is the equivalent of the COUPERIN Consortium. It is similar to a "technological BSN (Digital Scientific Library)" which supports and accompanies technological and organisational projects working on the shared STI services of the years to come. The mission recommends that the CNRS should subscribe to the CNI which would give access to this centralizing body's wealth of highly representative STI resources. At the European level, the DIST is currently considering joining the equivalent regional structure (Knowledge Exchange) to access first hand information and analysis on key questions linked to the study of the STI value chain.

Joining the CNI community should provide a low cost solution to allow us to mutualize our STI projects particularly via SHARE, to receive a large quantity of information on projects run by the Universities of the United States and to easily share our own national projects.

3. Innovative projects for STI analysis

THE AMERICAN CHEMICAL SOCIETY (ACS) demonstrated clear interest in the CNRS STI strategy and, following the society's suggestion, the mission accepted to look into the organization of a seminar for experts on business models for publishing on platforms and on platforms' analysis tools. ACS is ready to host the meeting on one of its campuses and wishes to develop a new and lasting relationship with the CNRS covering all aspects of STI.

The mission noted several robust projects in diverse fields where the dynamic of STI platforms directly influences scientific work. These projects all involve interest in the "simple" reprocessing of data which reveals rich and hitherto unknown correlations. The mission noted that this major change in the basis and modalities of research work is developing alongside a strong critical spirit and pragmatism which both represent a departure from "technological futurism". These are the values of the VIRGINIA BIOINFORMATICS INSTITUTE, the MELLON Foundation or the NATIONAL ENDOWMENT FOR HUMANITIES, alliances, organizations or laboratories which are
jointly driving the construction of foundations for new approaches to science. The mission took particular care in its proposals to create the conditions for a long-lasting relationship with the equivalent research teams or counterpart figures in France.
<table>
<thead>
<tr>
<th>Partners</th>
<th>Proposals</th>
<th>Prospective dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>/</td>
<td>Preparatory mission</td>
<td>September 2014</td>
</tr>
<tr>
<td>SLOAN Foundation / CIRHUS UMI and DATA SCIENCE CENTER at NYU</td>
<td>Seminar for experts on &quot;values of public science in the era of global digital platforms &quot;</td>
<td>Start of 2015</td>
</tr>
<tr>
<td>AMERICAN CHEMICAL SOCIETY (ACS),</td>
<td>Seminar for experts &quot;business models for publishing on platforms and on platforms' analysis tools&quot;.</td>
<td>Spring 2015</td>
</tr>
<tr>
<td>ANDREW W. MELLON Foundation</td>
<td>Organization of joint HSS event at Marseille</td>
<td>/</td>
</tr>
<tr>
<td>NCBI (National Center for Biotechnology Information)</td>
<td>Invitation to France for David Lipman and cycle of conferences with the French Institute of Biological Sciences on the &quot;global relationship between the information industry and Open Access in Molecular Biology&quot;</td>
<td>/</td>
</tr>
<tr>
<td>DONA (Digital Object Numbering Authority)</td>
<td>A French member of the CNRS or INRIA (Institute for Research in Computer Science and Control) could join the DONA management board.</td>
<td>/</td>
</tr>
<tr>
<td>NATIONAL SCIENCE FOUNDATION (NSF)</td>
<td>Continue exchanges on the evaluation of programs and use of STI as a key &quot;material&quot; for this evaluation</td>
<td>/</td>
</tr>
<tr>
<td>NATIONAL INSTITUTE FOR STANDARDS AND TECHNOLOGY (NIST)</td>
<td>Build a partnership with the DATA SCIENCE PROGRAM in the short term</td>
<td>/</td>
</tr>
<tr>
<td>COALITION FOR NETWORKED INFORMATION</td>
<td>The CNRS should join this consortium</td>
<td>/</td>
</tr>
<tr>
<td>National Library of Medicine (NLM), National Institutes of Health (NIH)</td>
<td>Create contact with the INIST based on &quot;STI promotion applications for the general public&quot;</td>
<td>/</td>
</tr>
<tr>
<td>SMITHSONIAN</td>
<td>Develop the relationship and look into possible collaborative projects with the INIST</td>
<td>/</td>
</tr>
<tr>
<td>National Endowment for the Humanities (NEH)</td>
<td>Develop the relationship with Huma-Num</td>
<td>/</td>
</tr>
</tbody>
</table>
C. THE CNRS MISSION'S MEETINGS AND PROJECTS

Meeting and project 1: Tuesday March 25th 2014

- The American Assembly (Columbia University)

http://americanassembly.org/

The American Assembly was founded by Dwight D. Eisenhower in 1950 and is part of Columbia University. This Foundation is a national forum working on access to information and scientific publications which provides input to help steer public policy-making and intervenes in the academic sphere.

Meeting with Joe KARAGANIS, Vice-president.

Karaganis joined the American Assembly as Vice President in 2010. His work focuses on the relationship between digital convergence and academic production and has recently included research on broadband adoption, data policy and intellectual property.

Contact:
Jk2991@columbia.edu
475 Riverside Drive
Suite 456
New York, NY 10115
Tel: (+1) 212 870 3500

Perspectives for STI:

Columbia University is made up of a large network of schools whose digital policy is relatively unclear according to the AMERICAN ASSEMBLY: data curation is uneven and databases are often under-used.
The American Assembly's aim is to set up a coherent architecture for databases and to develop social networks for the scientific communities based around a central STI platform. The main field concerned is clearly the Humanities and Social Sciences.

The American Assembly also works as an integrator. Its aim in this capacity is to eventually unify smaller platforms on its own platform and incite researchers to deposit their data thereon without making this in any way an obligation. The platform could also eventually house grey literature (study or research reports, conference proceedings, theses, patents, etc.).

No prior data approval system is planned for this platform and free usage is the rule. User-generated content (UGC) will be managed on a pragmatic basis according to the usage restrictions imposed by publishers.

Providing a link between research and effective access to Higher Education is the basic aim. Currently the cost of access to platforms and databases in the field of Education often involves paying a flat fee and this must not prove dissuasive. To attain the goal of an overall open system, the conditions of access to platforms need to be reworked and those providing funding need to be persuaded to modify their demands. Publishers still dominate Open Access policies even though in terms of "added value" their contribution is less significant (around 6% of publications). Discussions are underway with publishers to set up new boundary lines in the common interest.

**STI Management:**

**Open Syllabus** is a knowledge management system which provides a link between teaching, research and administration. It is a platform which is open to all the restrictions and the management of all forms of intellectual property rights (Open source, author's copyright and ancillary rights, etc.).

The Open Syllabus Project (OSP) is the first large-scale online database of university course syllabi to act as a platform for the development of new research, teaching and administrative tools ([http://opensyllabusproject.org](http://opensyllabusproject.org)). It is based on shared Open Knowledge tools like GitHUb or Zotero.

The **Open Syllabus** project is based on the construction of a vast network of partners involving researchers and personnel from the Universities of Columbia and North Carolina and the New School in New York. Tools will be provided on the platform to
enable users to extract text and citations. The project is also supported by the Alfred P. SLOAN Foundation.

The project will provide new ways of understanding higher education teaching, how disciplines are evolving and initiation to research, changes in teaching methods and cursus management. It will provide more transparency on university courses for students, teachers and administrative staff and thus enable better coordination of classes and the development of study courses.

The project will provide access to new publication parameters and new viewpoints on Scientific and Technical Information practices. The project will also enable monitoring of the usage of open access publications (frequency of teaching, contexts, works taught).

**Analysis of STI:**

Systems are currently being developed that will enable users to map scientific and technical information search behaviour. This will provide those working on academic analysis with their own specific data and even cover interactions between researchers on their dedicated social networks. A project is being developed to map such processes.

For STI analysis, the American Assembly was inspired by the "Academia.edu" platform, a social network for young scientists which gathers and elucidates the links between academic publications' analysis data with a view to promoting the development of common practices among young researchers (they have 9 million registered users).

**PROPOSAL:** To promote Columbia's original integrated platform model (teaching, research, publication, analysis) with the aim of encouraging symmetrical collaborative projects of a similar nature in French universities, organisations etc. working in the same disciplinary fields.
Meeting and project 2: Wednesday March 26th and Friday March 28th 2014

NYU (Center for data science)

http://cds.nyu.edu/

The Center for Data Science aims to exploit the potential of NYU to provide the American academic sphere with a sharing space for emerging knowledge on Research Data Sciences. Their approach is open to all data producers and users (Information Industry, Public Research, Users and Networks, etc.).

Meeting with Yann LE CUN, Founding Director of the NYU Center for Data Science and Director of AI Research, Facebook.

Contact:
yann@cs.nyu.edu
yann@fb.com
NYU, Room 1220
715 Broadway
New York, NY 10003, USA
Tel: (+1) 212 998 3283

Perspectives for STI:

The NYU Data Science Center aims to provide researchers and other professionals with interdisciplinary STI exploitation tools from research databases. Currently these are rarely used more widely than in the laboratory which produced them and often their usage is limited by the lab to immediate verification requirements concerning a limited amount of local hypotheses. The digital format of data now means such databases can be explored using an "overall cross-fertilization" of STI data although such usage is only just beginning. NYU's stated objective is to develop “the country's leading data science research and training facilities".
The United States President's Office for Science and Technology Policy (OSTP) supports the NYU Data Science Center's work which clearly follows the national Open Science policy initiated by the White House since February 2013 (last document: May 2014 on Open Data).

**STI Management:**

Whatever the size or complexity of today's research databases and whether they are shared or not, they bring up questions regarding promotion, dissemination, access, processing, curation and sharing but above all content analysis and interaction. Currently there are very few transversal approaches to the many questions of all kinds about sharing knowledge acquired by processing digital data. Several major American universities are currently developing such transversal approaches to knowledge engineering and strategies which need to be implemented.

Our rich meetings provided an opportunity to discuss an analysis of the relationships between research and the information industry (The Center most notably works with Elsevier and Thomson Reuters but also Google, Microsoft, Facebook, etc.) particularly by studying the scope of current aims regarding access to STI through the control of corpus exploration tools provided by multi-core science platforms (Text and Data Mining, publishers' Application Programming Interfaces (API), etc.).

**Analysis of STI:**

The 30 research projects currently being developed cover a broad range of scientific fields and approaches (see site). The overall policy is centered on the large-scale architecture of science and its fields of application (biology, artificial intelligence, quantitative methods for the social sciences). Yann Le Cun is also interested in knowledge representation and is taking part in the International Conference on Learning Representations.

The Center is working on the "authorship" of ideas and new forms of publication and production of scientific results along with a set of rules aimed at optimizing the dissemination and sharing of scientific information between its producers and users.

**PROPOSAL:** The idea of organizing a seminar for experts on the principles and rules for the regulation of science platforms was considered of interest. The DIST could liaise with the CNRS Office for North America to send a project preparation mission to the US in September 2014. This seminar could take place at the end of 2015 - start of 2016 and
would enable in-depth study of several questions that have yet to be explored in enough detail - digital law for platforms (the CNRS has invested in this area via the ISTEX project), ethical charters on data sharing and interdisciplinarity, ways of including results on science platforms (particularly through user-generated content) and the architecture of value chains on platforms. The SLOAN Foundation (see below) has stated its readiness to get involved in this joint CNRS - NYU initiative.
The Andrew W. Mellon Foundation

http://www.mellon.org/

The Andrew W. Mellon Foundation was founded on June 30th 1969 and is a non-profit association. It currently subsidizes programs in four main areas:

- Higher Education and study grants
- Scientific and technical information
- History of Art, conservation and Museums
- The performing arts.

Meeting with Don WATERS, Program Officer for Scholarly Communications and Information Technology.

Donald J. WATERS is head of the Foundation's Scientific and Technical Information program. Previously he had been the first Director of the Digital Library Federation (1997-1999) and a librarian at the University of Yale (1993-1997).

Contact:
dwj@mellon.org
140 East 62nd Street
New York, NY 10065
Tel.: (+1) 212 838 8400

Perspectives for STI:

The Foundation's Scientific and Technical Information Program was set up in 2010 and now covers a wide range of activities including gathering, organizing, evaluating, interpreting and preserving information along with the publication and dissemination of scientific research.
The Foundation's objectives are to support libraries and archives in their work preserving and providing access to resources of major cultural and scientific importance (through multi-core platforms) and also to help researchers with the development of specialist resources which influence the start or improvement of HSS research.

The Foundation wishes to work in the very heart of the Humanities and encourage:

- tracking and cataloguing of primary sources;
- the preservation of these primary sources, publications or documents uploaded to the web. This involves two main questions - how to ensure the collection and long-term preservation of data;
- digitalization of collections which are hard to access and the open access dissemination thereof;
- work on the level of relevant aggregation of documents or publications as collections;
- a synergy between library networks;
- the development of new tools for digital resources;
- organizing libraries so that they provide a digital environment corresponding to technological changes and researchers' expectations.

**STI Management:**

This program includes the design, development, implementation and support of IT systems for the conversion, storage, protection, processing, recovery and transmission of digital information.

The Foundation's STI management work is based on:

- working with libraries to encourage coherency between collections and practices
- networks which encourage integration and coherency
- shared platforms with the aim of organising deposits of interoperable documents
- training for young Humanities researchers working on primary sources (*data curation*)
- training programs for young librarians

**PROPOSAL:** To promote and increase awareness of this major Foundation in the field of HSS. The organization of a joint event in Marseille was evoked by the Director of the Institute for Humanities and Social Sciences (INSHS). The DIST will follow up on this proposal during its September 2014 mission in collaboration with the CNRS Office for North America.
Meeting and project 4: Friday March 28th 2014

- Alfred P. SLOAN Foundation

[Logo]

http://www.SLOAN.org/

Founded in 1934 by Alfred P. SLOAN Jr., president, chairman and CEO of General Motors, the Foundation no longer has links with General Motors and provides subsidies to support scientific and technological research, education and economic performances. The overall aim is to improve the quality of life in the United States.

Meeting with Joshua GREENBERG, Director of the Alfred P. SLOAN Foundation’s Digital Information Technology Program, in collaboration with Randall WHITE – Director of the CIRHUS (Center for International Research in the Humanities and Social Sciences) and professor at NYU and Nicolas GUILHOT – deputy director of the CIRHUS and CNRS researcher.

Joshua M. GREENBERG is the director of the Alfred P. SLOAN Foundation's "digital information technology" program.

One of the foundation's departments works in HSS on questions like the differential access of socio-ethnic groups (African Americans, Hispanic women, etc.), gender or higher education, etc.

Contact:
Greenberg@sloan.org
Locaux de l'UMI CIRHUS
4 Washington Square North
New York, NY 10003
Tel.: (+1) 212 649 1633
**Perspectives for STI:**

At the Foundation, young HSS researchers learn on STI courses which are open to *Data Sets* of varied characteristics and size.

Their highly innovative five-year course covers:

- digital data and research on large networks and the capacity to generate data (examples are given involving Google and Ebay auctions)
- empirical social research by students who join Google and others at least on a part-time basis
- study on the question of the privatisation of knowledge and data
- data management platforms in the field of research
- digital data and sharing thereof. How to link articles to their data (including use of *Run my code*)
- speeding up publication and rotation of results in HSS research;
- the organization of a better circulation of "*working papers*"
- data curation
- dissemination processes: networks on the attention economy, etc.
- the difference between "*public access*" and "*open access*",
- the analysis of the Information Industries' new "*business models*" and strategies
- the "*Non profit model*" (definitions, characteristics and limits; competition in the knowledge resource industries)

The generation gap was also evoked by the Sloan Foundation which highlighted issues like inertia in representation and attitudes in journals' editorial boards or conference organization committees. The need to optimize work on the organization of the most useful networks for researchers and for innovation was also stressed.

How should the need to create jobs be dealt with? *Data Scientists* could become a new job category. The questions of *Data Science Initiatives* and *Trading Zones* were also discussed.

**STI Management:**

Subsidies from the A.P. SLOAN Foundation help researchers develop new tools, standards and infrastructures. The Foundation concentrates on projects which encourage scientific access and data sharing, facilitate and promote the development of standards necessary for the interoperability of databases, enable the dissemination of
computing research and guarantee the key role played by data management and preservation.

The Foundation supports the digitalization of scientific and cultural resources.

Examples of projects supported by the Foundation:
- Internet Archive: https://archive.org/
- Open Content Alliance: http://www.opencontentalliance.org/
- Open Syllabus (see The American Assembly): http://opensyllabusproject.org/
- Digital Public Library of America: http://dp.la/

**Analysis of STI:**

The Foundation publishes the results of its most important research in Open Access (for example Sloan Digital Sky Survey).

**PROPOSAL:** The organization by the CNRS and the SLOAN Foundation of a seminar for experts in collaboration with CIRHUS at NYU was generally considered an interesting idea. The mission was welcomed positively at the CIRHUS UMI and would like it to be involved in organizing the Seminar which could take place at the SLOAN Foundation, at least in part. The SLOAN foundation already has a solid working relationship with the CIRHUS UMI.

The program could, in part, be run jointly with the NYU CENTER FOR DATA SCIENCE which the SLOAN Foundation collaborates with.

The subject selected could cover the "values" of public science in the era of global digital platforms. The experience, studies and tools of the SLOAN Foundation could thus be pooled with those of the Universities and CNRS research centers to run a program aimed at defining the STI management's relevant principles and results with a view to developing Open Science.

The SLOAN Foundation clearly possesses in-depth expertise in all dimensions of the relationship between publishing and services (Open Access and Open Process). The major part of this project involves the Humanities and Social Sciences (HSS) but not entirely. The DIST could liaise with the CNRS Office for North America to send a project preparation mission to the US in September 2014.
Meeting and project 5: Monday March 31st 2014

- National Center for Biotechnology Information


Senator Claude PEPPER created The National Center for Biotechnology Information (NCBI) on November 4th 1988 as a division of the National Library of Medicine (NLM) at the National Institutes of Health (NIH).

Meeting with David LIPMAN, director of The National Center for Biotechnology Information.

David LIPMAN has directed the NCBI for 25 years and created Pub Med Central. He is an authority on research and STI in molecular biology.

Contact:
lipman@ncbi.nlm.gov
Building 38A, Room N807
8600 Rockville Pike MSC
6075
Bethesda, MD 20894-6075
Tel: (+1) 301-496-2475

Perspectives for STI:

The NCBI is part of the Pub Med Central project. This platform offers 23 million citations from MEDLINE along with on-line life sciences journals and books. These citations may also include direct links to full text articles from the PubMed Central platform or publishers’ own sites.

David LIPMAN has selected three main directions for the future of the NCBI and the databases:
- usage analysis to help implement a robust "service-oriented" direction;

- API development to enable use by external applications;

- strategic choices on precise subjects need to be made because everything seemingly cannot be managed in parallel.

TDM was also discussed at this meeting.

According to D. LIPMAN, if researchers wish to publish in Open Access, this needs to be imposed at the call for projects stage. Recommendations in the framework of H2020 do impose (pre)-publication in an open archive but this does not apply to actual publication (in an OA or journal or not). The time scale for OA publication also remains to be decided (6 months or 1 year).

**STI Management:**

According to D. LIPMAN, customarily in biomedical research (pre)-publication is the norm so researchers publish to obtain later financing. There is a lack of selectivity in publication which causes a tricky bibliometrics problem.

It is currently necessary to set up improved project monitoring to be able to evaluate the impact of calls for projects on the corresponding editorial production which makes more sense than calculating the "impact factor" of publications derived from a given project. Strategies are complex - authors often first attempt to obtain a high "impact factor" but make do with a lower "impact factor" later to avoid extremely long publication delays. This idea seems to be shared by the NSF.

The NCBI databases:

- The NCBI's C++ manual: A comprehensive manual on the C++ language used by the NCBI along with a C++ indexing library, examples of software, demos, FAQs and notes. The manual can be consulted on-line and downloaded as a series of PDF documents.

- A static database of the pages of the NCBI site: This database is a search tool for the whole of the NCBI site except the FTP repertoires. These pages included a specialized range of sequence analysis tools, old editions of information sheets, descriptions of resources’ heritage sequence, examples of code and various other resources.
PROPOSAL: An invitation to France for David LIPMAN could be set up along with a cycle of conferences with the French Institute of Biological Sciences on the global relationship between the information industry and Open Access in Molecular Biology. More generally, David LIPMAN's unique experience in digital scientific publishing could be extremely useful for understanding the overall usage of contemporary digital scientific publishing, its expectations and dead ends. He could be asked to provide international expertise.
Meeting and project 6: Monday March 31st 2014

❖ National Library of Medicine (NLM), National Institutes of Health (NIH)

The National Institute of Health (NIH)

http://www.nih.gov/

This NIH is under the authority of the United States Department of Health and Human Services and is the US medical research agency.

National Library of Medicine (NLM)

https://www.nlm.nih.gov/

The NLM was created in 1836 and is today the world's most important biomedical library. It keeps and makes available a vast collection of journals and produces digital resources on a wide range of subjects. It also coordinates a national network of 6000 members of medical libraries to promote and enable access to information on health in communities all over the United States.

Meeting with Pertti Juhani HAKKINEN, Acting Head of the Office of Clinical Toxicology (study of the impact for health of different elements which are present in our environment).

Pertti Juhani HAKKINEN is a specialist on databases and other online information in the health field.

Contact:
Pertti.hakkinen@nih.gov
Division of Specialized Information Services (NIH/NLM),
6707 Democracy Blvd, Suite 440
North Bethesda, MD 20817
Tel: (+1) 301 827 4222
The Special Information Services division of the NLM develops databases on toxicology for researchers, students and the general public.

At our meeting Pertti Juhani HAKKINEN spoke about free services and projects for public users of the National Library of Medicine (NLM)

**STI Management:**

To attain its objectives the NLM provides access to biomedical information and works on creating high quality information services in the fields of toxicology, environmental health, health services and public health.

These include specialist databases from the *Specialized Information Service Division* (Lactmed, Toxnet, ALTBIB) based on secondary information extracted from various publications and sites. These databases can be accessed on tablets and telephones.

The general philosophy is to satisfy user requirements and particular care has gone into creating the access interfaces (CHEMM for example)

The NLM facilitates interoperability through the development of subsets of terminological data and other resources. It supports the development, enhancement and distribution of specific clinical vocabularies aimed at facilitating the exchange of clinical data and improving information recovering in the health field.

Examples of projects managed by the NLM:

- *PubMed Central* (see NCBI)
- *MedlinePlus* is an information website in the field of health for private individuals.

**Analysis of STI:**

It is currently difficult to predict how these services will develop because we have no information on the real usage of such secondary information services. However at another level, the sites developed by the INIST (*Sidasciences, Psycho Temoins, Recherches Arctiques*) are in a similar vein. These services require a great deal of resources and it is essential to evaluate the relevance of such specialized services.

**PROPOSAL:** Put the NLM in contact with the INIST (if this is not already the case) and particularly ensure exchanges on "STI promotion applications for the general public" especially using mobile phones.
The NSF is an independent federal organization created by Congress in 1950 to promote scientific progress. The NSF is the Federal Government's main source of intervention in many fields like mathematics, computing and the social sciences, etc.

Meeting with Alejandra MEDINA-BORJA, Program director, Evaluation and Assessment. She is implementing a cutting-edge program at the NSF for the evaluation and control of STI in research programs.

Contact: amedinab@nsf.gov
NSF
4201 Wilson Boulevard, suite 505N,
Arlington
VA 22230, USA
Tel: (+01) 703 292 9013

Perspectives for STI:

The NSF's expressed policy (on the governmental level and for all the agencies) is to move towards a funding system which is more efficient, effective and open and therefore better evaluated.

STI Management:

Currently the evaluation of programs is used as a decision-making tool aimed at increasing innovation but researchers have difficulty understanding how the mechanism works.
This evaluation mechanism is rolled out through all phases of a research program - from *early stage design* to studies of the impact of the science produced on society.

**Analysis of STI:**

The NSF has clearly developed a complete and attractive process with which it evaluates the programs it funds. Their model for planning projects which correspond to the various evaluation objectives could be a source of innovation.

**PROPOSAL:** The mission began exchanges which could continue with the NSF on the evaluation of programs and using STI as key "material" for this. It is still too early to know what these exchanges may bring forth (the first reports are currently being analyzed). It would be useful to send all the information on this contact to the Observatory of Science and Technology (OST) and to the corresponding parliamentary Office.
Meeting and project 8: Wednesday April 2nd 2014

- American Chemical Society

[Logo of the American Chemical Society]

http://www.acs.org/content/acs/en.html

*The American Chemical Society* was founded in 1876. It is an independent organisation which is original insofar as it represents professionals with all levels of studies working in all fields of chemistry and chemical sciences. The ACS is the world's biggest learned society.

**Chemical Abstracts Services**

https://www.cas.org/

Meeting with:
Bradley MILLER, Director of the Office of International Activities.

Bradley MILLER joined the ACS in 1999 and currently works on developing programs, products and services aimed at helping chemical science progress thanks to collaborative projects in Africa, Asia, Europe, Latin America and the Middle East.

**Contact:**
b_miller@acs.org
ACS
1155 Sixteenth Street, NW
Washington D.C, 20036
USA
Tel: (+1) 202 872 6317
Brandon NORDIN, Vice president, Sales, Marketing & Web Strategy - Publications.

Brandon NORDIN joined the ACS in 2008 and, among other projects, works on managing change, product line and organisational development strategies, the construction of multiple flows, IPs and content licenses, etc.

Contact:
b_nordin@acs.org
ACS
1155 Sixteenth Street, NW
Washington D.C, 20036
USA
Tel: (+1) 202 872 8063

Michael DENNIS, Vice president, Legal Administration.

Michael DENNIS works in the Chemical Abstracts Service (CAS) a branch of the ACS.

Contact:
mdennis@cas.org
CAS
2540 Olentangy River Road
Columbus, Ohio, 43202-1505, USA
Tel: (+1) 614 447 3613

Perspectives for STI:

The ACS is very up-to-date in the services it provides like the CAS or the journals it publishes. Its strategy is mainly commercial (even though it is a "non profit organisation").

Questions asked: how to ensure the French contract is updated to include the ACS's new journals (as this is apparently not the case).

STI Management:

The presentations reiterated a description of the ACS (Miller) and Chemical Abstracts Services CAS (Dennis). Like most publishers, CAS has developed a range of Open Access services (gold publishing, ACS Central Science megajournal). The discussion
with Brandon NORDIN was highly interesting and he was clearly keen on the proposed collaborative projects.

**PROPOSAL:** The ACS demonstrated clear interest in the CNRS STI strategy and following the society’s suggestion, the mission accepted to look into the organization of a seminar for experts on business models for publishing on chemistry platforms. ACS is ready to host the meeting on one of its campuses and wishes to develop a new and lasting relationship with the CNRS covering all aspects of STI.
Meeting and project 9: Wednesday April 2nd 2014

❖ Virginia Bioinformatics Institute

https://www.vbi.vt.edu/

The VBI opened in 2000 in the Tech Corporate Research Center in Virginia. It is a world-class research institute dedicated to the study of biological sciences.

Meeting with:

Stephanie SHIPP, Deputy Director and Research Professor, Social and Decision Analytics Laboratory.

Contact:
Steph19@vbi.vt.edu
Social and Decision Analytics Laboratory (SDAL)
National Capital Region, 3rd floor
900 North Glebe Road
Arlington, VA 22203
Tel: (+1) 571 858 3123

Sallie KELLER, Professor and Director, Social Decision and Analytics Laboratory (SDAL).

Contact:
sallie41@vbi.vt.edu
Social and Decision Analytics Laboratory (SDAL)
National Capital Region, 3rd floor
900 North Glebe Road
Arlington, VA 22203
Tel: (+1) 571 858 3102
**STI Management:**

This meeting gave an excellent illustration of how well the processing of statistical data can be applied to research in the Social Sciences field, the foundations of the *Digital Humanities*. Cross-referencing heterogeneous data seems fertile for research: illustration with the association of heart attacks and air pollution.

Currently there are difficulties in sorting: *The Signal and the Noise, Nate Silver, 2013*.

According to Stephanie SHIPP and Sallie KELLER, *Big Data* in the most general sense of the term is a true revolution which will transform our lifestyles and way of thinking.
Meeting and project 10: Thursday April 3rd 2014

- Corporation for National Research Initiatives

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

Meeting with Robert KAHN, President & CEO of CNRI.

Robert KAHN became interested very early in Digital Objects Architecture after having invented TCP/IP with Vint CERF (DARPA). He is also the creator of our current Internet insofar as it is based on his first network Arpanet. He is a great figure and the meeting was a great moment for us.

Contact:
rkahn@cnri.reston.va.us
1895 Preston White Drive
Suite 100
Reston, VA 20191 – 5434
Tel: (+1) 703 620 8990

Perspectives for STI:
Accessing information on the internet requires 3 components:

- Repositories

- Metadata registries

- A Handle system and Domain Name System.

The Digital Objects Architecture orchestrates these component parts. One piece of the puzzle is the Digital Object Numbering Authority which attributes numbering prefixes to objects that are created.
These components also enable machines to communicate between themselves to evolve towards a decentralised mechanism which places the resource or document entity at the center of the system (using a handle system which can be updated by the resource owner). The DOI system is the first successful step in this direction.

Robert KAHN proposed that a French person should join the board of directors of the Digital Object Numbering Authority.

**PROPOSAL:** The mission recommends that the CNRS Research Office (DGDS) should examine this choice in depth. If the CNRS does not propose someone, the DIST suggests that this role be offered to Pascal GUITTON at the Institute for Research in Computer Science and Control (INRIA).
Meeting and project 11: Thursday April 3rd 2014

- American Association for the Advancement of Science (AAAS)

http://www.aaas.org/

The AAAS, founded in 1848, is an international non-profit organization dedicated to the promotion of science in the interest of all the world's peoples.

Meeting with Vaughan TUREKIAN, Chief International Officer and Chief Editor of Science & Diplomacy.

Contact:
vturekia@aaas.org
AAAS
1200 New York Avenue, NW
Washington, D.C 20005 USA
Tel: (+1) 202 326 6650

Perspectives for STI:

The AAAS medium term aims are to:

- Improve communication between scientists, engineers and the public
- Promote and defend the integrity of science and how it is used
- Strengthen support to work in science and technology
- Give science a voice on questions of society
- Promote the responsible use of science in public policies
- Strengthen and diversify the working population in the fields of science and technology
- Promote Education for all in science and technology
- Increase public participation in science and technology
- Promote international cooperation in the sciences
**STI Management:**

The idea of an international meeting on the importance and benefits of STI was rapidly evoked. We should also note the existence of the Richard Lounsbery Foundation which finances Franco-American scientific activities (link with the French *Académie des Sciences*).
Meeting and project 12: Thursday April 3rd 2014

Coalition for Networked Information (CNI)

http://www.cni.org/

The CNI was founded in 1990. It is a non-profit association of research libraries and management staff in STI computing centers and is very well represented among the major universities.

Meeting with Joan K. LIPPINCOTT, Associate Executive Director.

Contact:
joan@cni.org
CNI
21 Dupont Circle – Suite 800
Washington D.C 20036
Tel: (+1) 202 296 5098

The CNI is the equivalent of the COUPERIN Consortium and similar to a "technological BSN" which accompanies technological and organisational projects preparing the STI shared services of coming years.

STI Management:

The members of this association meet twice a year to discuss new developments in this field.

The structure is financed by subscriptions from its members (>200) some of whom are European: DEFF[2], JISC[3], DFG[4].

[4] Deutsche Forschungsgemeinschaft - Netherlands
These four European participants are associated in the Knowledge Exchange (http://www.knowledge-exchange.info/) project which has been hoping to collaborate with the CNRS for a while now. The CNRS could consider taking part in the two bodies - Knowledge Exchange for the European side of things and CNI which Joan LIPPINCOTT has invited the CNRS to join.

The CNI's work themes cover:
- The management of distributed information;
- The development of technologies and infrastructures;
- The transformation of organisations, skills and careers.
These themes are the foundations of the CNRS STI strategy and we really need to develop collaborative projects with the CNI.

For example, cooperative projects led by the CNI deal with Data Management, Data Visualisation, new forms of digital publishing, the development of Data Scientists' skills, etc.

The CNI launched the SHARE initiative (http://www.arl.org/focus-areas/public-access-policies/shared-access-research-ecosystem-share#.U0z603-JSxg).

This research into the shared access ecosystem is a joint initiative involving The Association of Research Libraries (ARL), The Association of American Universities (AAU) and The Association of Public and Land-Grant Universities (Aplu) which began on February 6th 2014. SHARE's detailed action plan includes a series of activities aimed at guaranteeing that scientific research results are constructed and published in such a way as to facilitate and accelerate the research process. The SHARE notification system is a logical part of the permanent Higher Education mission to encourage community-based solutions which improve public access to research and maximise knowledge creation.

The ORCID project:

This project provides a long-term numerical identifier which distinguishes researchers and supports automated links between researchers and their professional activities thus ensuring recognition of their work thanks to its integration into key research workflows like manuscripts and presentations.

The CNI also works on many projects promoting Open Access:
- The SPARC project
- Digital preservation
- Electronic theses and dissertations (ETDs)
- E-Textbooks
PROPOSAL: The mission recommends that the CNRS should subscribe immediately to the CNI as this would give access to this centralizing body's wealth of highly representative STI resources. At the European level, the DIST is currently considering joining an equivalent regional structure (*Knowledge Exchange*) to access first hand information and analysis on key questions linked to the analysis of the STI value chain.
Meeting and project 13: Thursday April 3rd 2014

- National Institute of Standards and Technology (NIST)

http://www.nist.gov/

The NIST is a very large technical structure (2000 agents) which covers all aspects of standards including research.

Meeting with:
Ileana MARTINEZ, International Affairs Advisor, Standards Coordination Office
Ashit TALUKDER, Information Management and NIST Data programs
Vicki PILLITTERI, Cybersecurity
Dean PROCHASKA, Smart Grid
Cuong NGUYEN, Smart Grid

Contact:
NIST, 100 Bureau Drive, Stop 1070
Gaithersburg, MD 20899-1070
Tel: (+1) 301 975 2000

Perspectives for STI:

The Communication Technology Laboratory's work is still in its early stages and no documents could be provided.

STI Management:

The NIST runs Data Science orientated programs particularly based on developing metrics. One example presented was that of "information retrieval tools" in TREC
conferences. A French clone of this has been successfully developed by Annie CORET and Laurent SCHMITT.

**Analysis of STI:**

The NIST *Data Science Program* is based on collaboration with all scientific disciplines and sectors of activities to permit the definition of reference frameworks to evaluate the metrics to be used.

PROPOSAL: Building a collaboration project with the Data Science Program is envisaged in the short term. The DIST will send proposals regarding this idea to the counterpart team at the NIST, the aim being to take part from the very start in constructing norms which will define standards for interoperability between databases.
Meeting and project 14: Friday April 4th 2014

Smithsonian Institution, Encyclopaedia of Life, National Museum of Natural History (Washington)

http://www.si.edu/

Founded in 1846, the Smithsonian is a major museum and research center made up of 19 museums and galleries, a zoological park and 9 research centers.

Meeting with Cynthia PARR, Chief Scientist.

Contact:
Parrc@si.edu
1000 Jefferson Dr SW,
Washington D.C, 20004, USA
Tel: (+1) 202 633 1000

STI Management:

The Smithsonian is developing a major application with its Encyclopedia of Life (EOL). This involves an investment of $25M over 5 years for a collaborative project providing access interfaces adapted to the requirements of users ranging from researchers, students and members of the public interested in biodiversity to the public as whole.

It is an application which constructs itself by harvesting data from external databases under the scientific supervision of "curators" in a spirit of total openness.

EOL requests authorization to use other databases (even if this is not formally necessary as American law protects database structures though not the actual data). EOL received a significant amount of funding when it was set up but is now looking for the right business model. One possibility that was discussed is setting up well-targeted EOL-
specific "portals", for example a portal on certain types of "species" or for specific clients (zoos), and which would be funded by those requesting the portals. This system is currently being set up.

This could be an example of a best practice to adopt if the INIST considers moving towards this type of activity.

PROPOSAL: Provide the INIST with the foundations for a collaboration project as proposed to the mission, verify interest in an STI relationship (INEE) and develop the relationship with the Smithsonian.
Meeting and project 15: Friday April 4th 2014

- National Endowment for the Humanities (NEH), Office of Digital Humanities

http://www.neh.gov/divisions/odh

The NEH is an independent federal organization created in 1965. It is one of the main sources of funding for HSS programs in the United States.

Meeting with Brett BOBLEY, Chief Information Officer.

Contact:
bobbley@neh.gov
National Endowment for the Humanities
Office of Digital Humanities
1100 Pennsylvania Ave., NW Washington,
D.C. 20506
Tel: (+1) 202 606-8401

Perspectives for STI:
The Digital Humanities are well underway - research is no longer carried out on a small number of print documents, researchers now work with a great quantity of digital documents. This fundamental change to research practices has two consequences:

- data sharing
- sharing processing and usage methods for this data (open process)
There has been a true cultural shift with young researchers becoming major data "consumers".

Another approach which generates data and is developing quickly is crowdsourcing, a trend which involves the public in research projects and in mechanisms for the acquisition, annotation or translation of data.
Another project in the pipeline is the *Humanities Commons Open Repository Exchange* or *Humanities CORE*. This project aims to develop a "Humanities Commons" type platform to aim to share the collections of academic societies in the HSS field. This initiative should be closely monitored and could be useful.

**PROPOSAL:** Develop the relationship with Huma Num as quickly as possible (if this is not already the case) and with the CLEO.
ANNEXE 1: STI IN THE UNITED STATES

I- Perspectives for STI
II- STI management
III- STI analysis

I- Perspectives for STI

1. OSTP (Office of Science and Technology Policy)

In the United States, the Office of Science and Technology Policy published various articles about American government STI initiatives on its website on the "Initiatives" page [http://www.whitehouse.gov/administration/eop/ostp]

Several of these initiatives are also on the "Promoting Open Data, Open Science and Open Government" page. Here are the contents:

Open Government
The Obama Administration firmly believes that openness in government strengthens our democracy and promotes a more efficient, effective and accountable government. OSTP is a leader in support of President Obama’s historic Open Government Directive that requires Federal agencies to take specific steps to achieve key milestones in transparency, participation and collaboration.

Open Data
The Obama Administration is committed to responsibly unleashing data from the vaults of government to fuel the innovation that grows the economy while also advancing government efficiency and accountability. On May 9, 2013, President Obama signed an executive order [http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government], making open and machine-readable data the new default for government information, taking historic steps to make government-held data more accessible to the public and to entrepreneurs while appropriately safeguarding sensitive information and rigorously protecting privacy. To build on this landmark effort to open up data across government, the Administration has also launched—through OSTP—several Open Data Initiatives aimed at scaling up open data efforts across the Health, Energy, Education, Finance, Public Safety and Global Development sectors.

The White House Open Data Initiatives
Accelerating and expanding efforts to make government information resources more publicly accessible in “computer-readable” form and spurring the use of those data by entrepreneurs as fuel for the creation of new products, services and jobs.

http://www.whitehouse.gov/innovationfellows/open-data-initiatives

Blog - Open Government Data Spurs Entrepreneurship and Jobs
Freely available data from the US Government is an important national resource, serving as fuel for entrepreneurship, innovation, scientific discovery and other public benefits.

http://www.whitehouse.gov/blog/2013/02/04/open-government-data-spurs-entrepreneurship-and-jobs

Open Data Executive Order
Making open and machine-readable the new default for government information.


Open Data Policy

2. Open Access Policy in the United States
In February 2013, an article was published on the nature.com blog announcing the implementation of a new Open Access policy in the United States.

http://blogs.nature.com/news/2013/02/us-white-house-announces-open-access-policy.html

3. Other stakeholders

- The University of New York (NYU)
Several articles have been published about the NYU's STI projects:

“To Collaborate On Multi-Million Dollar Initiative To Harness Potential Of Data Scientists And Big Data”, NYU Center For Data Science, November 2013.


“Initiative In Data Science And Statistics: A Collaborative Effort with Two Distinct Components” NYU Center For Data Science, April 2013.

"NYU Announces Project to Create Arabic Collections Online, An Arabic-Language Digital Public Library” NYU News, April 2013.


“NYU Launches Initiative in Data Science and Statistics to Push Advances in Medicine, Science, Technology and Other Fields” NYU News, February 2013.

UNESCO

In November 2011, a forum took place about UNESCO's Open Access strategies. A PDF of the report is available at:


ELSEVIER

Several videos about STI have been put on-line by the publisher Elsevier:

The article of the future:
http://www.youtube.com/watch?v=1dXkmgkYuEg#t=309

The executable paper
http://www.youtube.com/watch?v=FQ5FaeHcdwo

Open Access
http://www.youtube.com/watch?v=8m8-QYrcewU

Elsevier also has a department dedicated to research data alone:
http://researchdata.elsevier.com/
II - STI management
Various polices about scientific publication and copyright:

1. MIT (Massachusetts Institute of Technology)
MIT Scholarly Publishing policy:
http://libraries.mit.edu/scholarly/publishing/copyright-information-for-mit-faculty/

2. Association of American Publishers
Public Policy and Issues: http://www.publishers.org/issues/

3. Copyright law in the US
http://www.copyright.gov/title17/

4. Social Science Research Network
“How Copyright Drives Innovation in Scholarly Publishing” Adam Mossoff, George Mason University School of Law, April 2013.

5. UNESCO
UNESCO Policy Guideline’s for the development and promotion of Open Access:
http://unesdoc.unesco.org/images/0021/002158/215863e.pdf

III - STI analysis

1. Information Analysis Centers (Vancouver)
http://iac.dtic.mil/iac_mission_goals.html

IAC Missions: Collect, analyze, synthesize, produce and disseminate worldwide Scientific and Technical Information (STI)
Vision: Drive Innovation and technological developments by anticipating and responding to the information needs of the defense and broader community, while enhancing collaboration through integrated Scientific and Technical Information (STI) development and dissemination.

http://iac.dtic.mil/support_docs/strat_plan_10_15.pdf
2. **Society for scholarly Publishing**  
   http://www.sspnet.org/

3. **National Science Foundation**  
   http://www.nsf.gov/  

   The National Center for Science Engineering Statistics is the main supplier of statistical data about science in the United States.
ANNEXE 2: Contacts, places and projects

List of contacts

❖ Lorcan DEMPSEY - Head of the research department at OCLC (Online Computer Library Center) in Dublin [OHIO] – Expert on metadata.

Mail: dempseyl@oclc.org
Blog: orweblog.oclc.org/about.html
OCLC page: http://www.oclc.org/research/people/dempsey.html?url=m=160985

❖ Randall WHITE – Director of the CIRHUS (Center for International Research in the Humanities and Social Sciences) and professor at NYU

Tel: 212 992 7488
Mail: randall.white@nyu.edu
Address: Randall WHITE – Director
CIRHUS – UMI 3199 (CNRS-NYU)
4 Washington Square North
New York, NY 10003
CIRHUS page: http://cirhus.as.nyu.edu/object/cirhus.white

❖ Nicolas GUILHOT – Deputy director of the CIRHUS and CNRS researcher.

Tel: 212 992 7492
Mail: nicolas.guilhot@nyu.edu
Address: Nicolas GUILHOT – Deputy Director
CIRHUS – UMI 3199 (CNRS-NYU)
4 Washington Square North
New York, NY 10003
CIRHUS page: http://cirhus.as.nyu.edu/object/cirhus.guilhot
❖ **Alexandra MEDINA-BORJA** - Interim Program Director for Evaluation and Assessment – National Science Foundation (NSF).

Tel: (703) 292-7557  
Email: amedinab@nsf.gov  

❖ **Robert E. KAHN** – Director of the CNRI (Corporation for National Research Initiatives)

Tel: 703.620.8990  
Address CNRI: Corporation for National Research Initiatives  
1895 Preston White Drive • Reston, Virginia • 20191  
301 East High Street • Charlottesville, Virginia • 22902  
CNRI page: [http://www.cnri.reston.va.us/bios/kahn.html](http://www.cnri.reston.va.us/bios/kahn.html)

❖ **Clifford LYNCH** – Director of the CNI (Coalition for Networked Information)

Mail: clifford@cni.org  
Tel: 202.296.5098  
Fax: 202.872.0884  
Address: Clifford A. Lynch -Executive Director, Coalition for Networked Information  
21 Dupont Circle  
Washington, DC, 20036  
CNI page: [http://www.cni.org/about-cni/staff/clifford-a-lynch/](http://www.cni.org/about-cni/staff/clifford-a-lynch/)

❖ **David LIPMAN** – Director of the NCBI (National Center for Biotechnology Information)

Mail: lipman@ncbi.nlm.nih.gov  
Tel: 301-496-2475  
Fax: 301-480-2288
Address: David J Lipman, MD
Building 38A, Room N807
8600 Rockville Pike MSC 6075
Bethesda, MD 20894-6075


Carole L.PALMER – Director of the CIRSS (Center for Informatics Research in Science and Scholarship)

Mail: clpalmer@uiuc.edu
Tel: 217-244-0653
Address: Carole L.PALMER
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
501 E. Daniel Street
Champaign, IL 61820-6212

Places and Projects

CSHL (Cold Spring Harbor Laboratory) - New York
Cold Spring Harbor Laboratory recently initiated a project to set up an open archive for pre-prints for biologists similar to the service provided for physicists.


Columbia University - Institute for Data Science and Engineering – New York
http://idse.columbia.edu/


CNI: Coalition for Network Information - Washington
http://www.cni.org/program/current-program-plan/2012-2013/

CNRI: Corporation for National Research Initiatives - Washington
http://www.cnri.reston.va.us/

NSF National Science Foundation - Arlington/Washington

**NCBI National Center for Biomedical Information** – Bethesda/Washington

**CIRSS: Center for Informatics Research in Science & Scholarship** – Urbana, south of Chicago, Illinois
http://cirssweb.lis.illinois.edu/index.php

**MIT CSAIL (Computer Science and Artificial Intelligence Laboratory)** – Cambridge, Massachusetts.
http://www.csail.mit.edu/research

**MIT Libraries**
http://libraries.mit.edu/guides/subjects/data-management/
data-management@mit.edu

**Thomson Reuters** - Philadelphia
Thomson Reuters has published documents about its future WOS platform including Open Access (Scielo) and links on the open web for integrated citation searches.


Thomson REUTERS is also working research data citation with the recently launched data citation index.