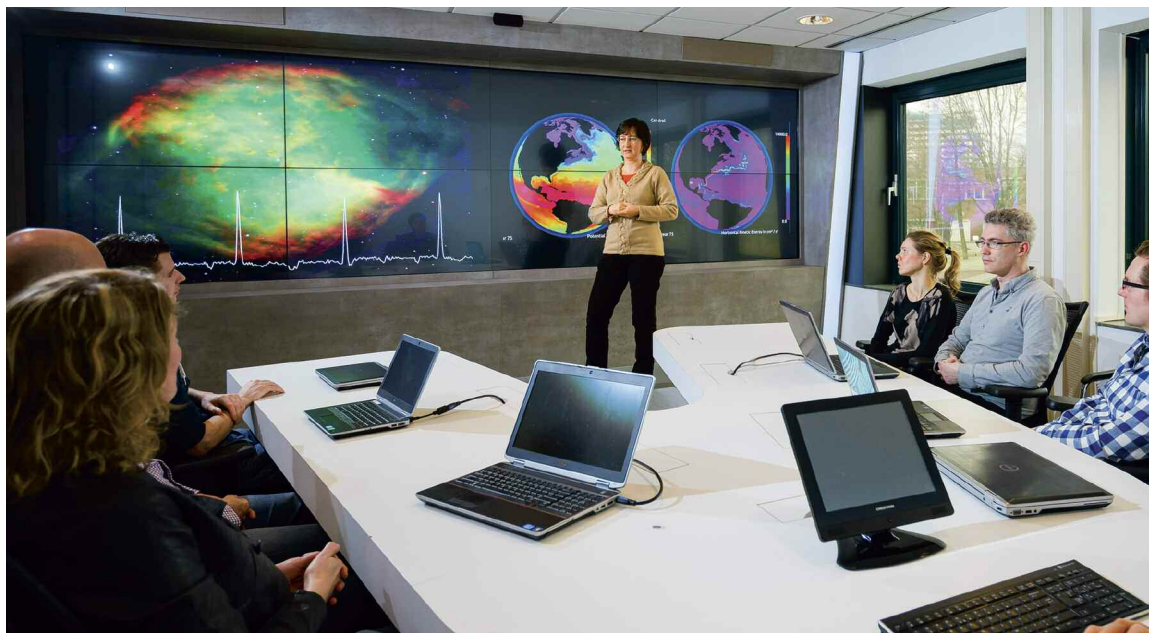

The Collaboratorium *credit esciencecenter.nl/about-the-center/collaboratorium/*

NLeScience deals with scientific challenges

Access to data and the exchange of data across domains is a very important aspect of modern scientific practice. Patrick J.C. Aerts, Director of Strategic Alliances at the Netherlands eScience Center (NLeSC), explains about their focus on the actual deployment of data, software and tools to address complex scientific challenges of the present and the future.

Set up in 2011, NLeSC aims to reinforce and accelerate multi-disciplinary and data-intensive research by developing and applying eScience methods and approaches. eScience ('enhanced science') enables new scientific breakthroughs. NLeSC helps bridging the gap between advanced ICT and e-infrastructures and the scientific disciplines that seek to engage their scientific challenges. Scientists from all disciplines, from humanities to astronomy and from ecology to water management, can count on NLeSC. The particular focus at NLeSC is on excellence in eScience: it develops and reuses tools in expertise domains such as cross-type data integration, data-driven & multi-model simulations, visualization & analytics, high performance (exascale) computing and networking.

Network organization

As a network organization NLeSC collaborates with scientists at universities and maintains a network of eScience Integrators. They are professors in their application domain who act as ambassadors for eScience towards their domain while providing input to NLeSC from their disciplines. Also, NLeSC has established a platform of Dutch eScience/data research organizations (ePLAN) who share a common vision of the role of eScience and Big Data in science and research in both academic and commercial sectors. The importance of software sustainability can be compared to the exponential growth of the web. This growth was enabled by open web codes which could be easily copied and reused allowing every new generation to stand on the shoulders of its predecessors. Today there are libraries and containers with fully fledged solutions to almost everything. By sustaining scientific software a similar track can be followed by reusing software across domains rather than reinventing the wheel.

Community of practice

eScience has been called a community of practice growing towards a discipline per se. This reflects the fact that eScience in 2020 will be different from eScience today, but will develop organically with the ever increasing growth in ICT potential and e-infra-structure capabilities on the one hand and the ever more demanding challenges in public and private science and research on the other.

p.aerts@esciencecenter.nl
eScienceCenter.nl

The European framework for audit and certification

How can we bring trust to the system of digital preservation and sustainable access? Ingrid Dillo, Policy Director at DANS, has the answers.

Data sharing is far from common practice. Although there are a lot of advantages (transparency, ability to replicate and verify research, reuse of data, higher return on investment), researchers still argue that they cannot trust data produced elsewhere by somebody else. So, how can we build trust into the system of digital repositories?

Three levels of evaluation

In 2010, a European Framework for Audit and Certification of Digital Repositories was set up to provide an answer to the question of trust. The parties involved have come up with three levels of certification that offer increasing reliability:

1 Basic Certification is granted to repositories that obtain Data Seal of Approval (DSA) certification. DSA, initially developed by DANS, ensures that research data can be processed in a high-quality, reliable manner, provided the 16 guidelines for self-assessment are followed.

2 Extended Certification is granted to Basic Certification repositories that perform an additional structured, externally reviewed and publicly available self-audit based on ISO 16363 or DIN 31644. The DIN 31644 standard, an initiative by NESTOR, is a catalogue of 34 criteria that trusted digital repositories should meet. The ISO 16363 standard presents over 100 metrics for different aspects of a digital repository.

3 Formal Certification is granted to repositories which, in addition to Basic Certification, pass a full external audit and certification based on DIN 31644 or ISO 16363.

If you are a researcher you can deposit your data in a trusted digital repository, or you can ask the repository how they make sure data are treated well. If you are a funder of research proposals you can draw the attention of researchers to the importance of depositing their data in a trusted digital repository after completing their research. Or you can make it mandatory for researchers to deposit their data in a trusted digital repository. You can provide researchers with some funding to take care of their data management during research and to archive them afterwards. Alternatively, you can withhold a percentage of the initial funding if they don't. If you work at a research infrastructure, make sure that the resources you offer remain meaningful and usable over time by applying sustainable business models. If you work at a digital repository yourself, make sure it is trustworthy.

The time is now

We all endorse the critical importance of sharing and preserving reliable data produced during the scientific process. Trust is at the very heart of storing and sharing research data. The time is now; together we can make the digital world more reliable.

ingrid.dillo@dans.knaw.nl
trusteddigitalrepository.eu

DSA seal *credit DSA*

COLUMN

Let's take data to Broadway

Dear eScientist, I bet one of your first actions in the morning is switching on your computer(s). Science without computers is inconceivable today. Yet, we make life difficult for them.

From 'reading to form a hypothesis' we move towards 'reading to confirm a hypothesis' that was 'thrown in our faces' by data pattern recognition. For decades we have been integrating data, using Extract-Transform-Load (ETL) into yet another non-interoperable data warehouse for local use. ETL is no longer a viable option in eScience, but we continue to bury our results in narrative, clumsily linked 'supplementary data' and non-interoperable formats, using 'our own standards' which create nightmares for computers which we should be supporting as our main assistants by now.

Semantic technologies enable 'functional interlinking' of datasets for everyone to use and reuse, and they can be processed by machines 'as one', even though the underlying sources are disperse and in different formats. The 'narrative publishers' are also exploring business models that involve exposing the associations buried in their texts in machine-readable format.
Key associations reported in these data sources need to be carefully selected for their KD value by pattern recognition. What's more, those associations need to be made available in so-called machine-readable FAIR format: Findable, Accessible, Interoperable and Reusable. The underlying resources will increase in value, can be found and used more easily and will always remain crucial for 'confirmative reading'. As long as *de novo* data 'speak the same language' immediate patterns can be discerned that were escaping our attention without Functionally Interlinked Data.
All this means that our 'local shows' need to be 'taken to Broadway', which implies Professional Data Publishing with safeguards against all the mistakes made in narrative publishing. Listing the latter would make this column way too long, and we know them all too well.

Barend Mons
photo marietmons.nl