from Open PHACTS to Open Science and back

Barend Mons
Open PHACTS conclusive meeting Vienna 2016
YOU ARE HERE

Open PHACTS

ODEX4all

FORCE11, JDDCP, FAIR

EXCOR

EOSC

ELIXIR


CWA

FDG

FAIRdom, CEDAR, EUCAT, PH

Long-Term Data Stewardship

How to finance ESFRI's and EBI SIB type + infra

Mainly private for reliability

Interoperability Backbones, Standards, Procedures

Mainly H2020 + ESFRI-type domain expertise

95% MS

5% EC

Mandatory for Research Projects H2020 & Member States
Summary of contributions from SAB and Key operational players

- **One Giant Proof of Concept:**
  - Semantic Web is not equal to ‘Nerdy Linked Data’ and Linked Data became ‘mature’
  - A **science question driven**, agile approach with a **CTO** and a **TTF** (no loose sand)
  - **PPID** (multiple identifiers, CW, IMS, IRS) - stroke of brilliance > EOSC
  - Bridging **TWO divides**
    - Engineer/scientist divide (Drive/Build/Manage)
    - Industry/Academy divide
  - An **exemplar** of effective EU projects (for IMI) and **inspiration** for other projects & companies
  - **Instrumental** for ELIXIR development
  - **FAIR concept** inspired by (a.o.) Open PHACTS
  - Industrial IT departments become increasingly **convinced** (use of pre-structured external data)
  - Dealing professionally with IP and licensing of data
ELIXIR Members

On 18 December 2013, ELIXIR became a permanent legal entity following the ratification of the ELIXIR Consortium Agreement (ECA) by EMBL and the first five countries. The countries that have signed the ECA are full members of the ELIXIR Board. Provisional Members (France and Spain) are progressing the ratification of the ECA. They have the same benefits as full members for the duration of their term as Provisional Members and make a full financial contribution to ELIXIR in this period.

<table>
<thead>
<tr>
<th>Country</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMBL (June 2013)</td>
<td></td>
</tr>
<tr>
<td>UK (Sept 2013)</td>
<td></td>
</tr>
<tr>
<td>Sweden (Sept 2013)</td>
<td></td>
</tr>
<tr>
<td>Switzerland (Oct 2013)</td>
<td></td>
</tr>
<tr>
<td>Czech Republic (Nov 2013)</td>
<td></td>
</tr>
<tr>
<td>Estonia (Dec 2013)</td>
<td></td>
</tr>
<tr>
<td>Norway (Jan 2014)</td>
<td></td>
</tr>
<tr>
<td>Netherlands (Feb 2014)</td>
<td></td>
</tr>
<tr>
<td>Denmark (Mar 2014)</td>
<td></td>
</tr>
<tr>
<td>Israel (May 2014)</td>
<td></td>
</tr>
<tr>
<td>Portugal (July 2014)</td>
<td></td>
</tr>
<tr>
<td>Finland (Sept 2014)</td>
<td></td>
</tr>
<tr>
<td>France (Oct 2015 as Provisional Member)</td>
<td></td>
</tr>
<tr>
<td>Spain (Oct 2015 as Provisional Member)</td>
<td></td>
</tr>
<tr>
<td>Belgium (November 2015)</td>
<td></td>
</tr>
<tr>
<td>Italy (January 2016)</td>
<td></td>
</tr>
</tbody>
</table>
The FAIR Guiding Principles for scientific data management and stewardship

Authors

Mark D. Wilkinson¹, Michel Dumontier², IJsbrand Jan Aalbersberg³, Gabrielle Appleton³, Myles Axton⁴, Arie Baak⁵, Niklas Blomberg⁶, Jan-Willem Boiten⁷, Luiz Bonino da Silva Santos⁸, Philip E Bourne⁹, Jildau Bouwman¹⁰, Anthony J Brookes¹¹, Tim Clark¹², Mercè Crosas¹³, Ingrid Dillo¹⁴, Olivier Dumon³, Scott Edmunds¹⁵, Chris T Evelo¹⁶, Richard Finkers¹⁷, Alejandra Gonzalez-Beltran¹⁸, Alasdair J G Gray¹⁹, Paul Groth³, Carole Goble²⁰, Jeffrey S. Grethe²¹, Jaap Heringa²², Peter A.C. ’t Hoen²³, Rob Hooft²⁴, Tobias Kuhn²⁵, Ruben Kok²², Joost Kok²⁶, Scott J. Lusher²⁷, Maryann E. Martone²⁸, Albert Mons²⁹, Abel L. Packer³⁰, Bengt Persson³¹, Philippe Rocca-Serra¹⁸, Marco Roos³², Rene van Schaik³³, Susanna-Assunta Sansone¹⁸, Erik Schultes³⁴, Thierry Sengstag³⁵, Ted Slater³⁶, George Strawn³⁷, Morris A. Swertz³⁸, Mark Thompson³², Johan van der Lei³⁹, Erik van Mulligen³⁹, Jan Velterop⁴⁰, Andra Waagmeester⁴¹, Peter Wittenburg⁴², Katherine Wolstencroft⁴³, Jun Zhao⁴⁴, and Barend Mons⁴⁵
Open Science as a Social Machine

&

the support through

The European Open Science Cloud
Analysis of input documents provided to HLEG ESOC
Provided to HLEG ESOC
Hard lessons learned:

- Limitations of RDF
- Sustainability (not ‘automatic’ that companies step in)
- How many of the > 50 associated partner contribute?
- The Toddler bike in the Formula 1 race syndrome
- Inability to keep excellent people involved in the project and its follow up.
OXS long term sustainability: minimally four synergistic pillars of income

- Open PHACTS/IMI Partners (contributions) In Kind & in Cash
- Services (for a fee) Data related
- Third Parties (applications) With revenue sharing
- Data Stewardship Plans in future Data Generating Grants (Pharma, EC, National, ESFRI, etc.)
- Partners
- Services
- Applications
- Grants

OPS
YOU ARE HERE

Open PHACTS

ODEX4all

FORCE11, JDDCP, FAIR

EXCOR

EOSC

ELIXIR

CWA

Data Management Plans
Mandatory for Research Projects H2020 & Member States

Long-Term Data Stewardship
How to finance ESFRI's and EBI SIB type + infra
Mainly private for reliability

Interoperability Backbones, Standards, Procedures
Mainly H2020 + ESFRI-type domain expertise

95% MS

5% EC
The Data Stewardship Cycle

5%
Open PHACTS
Open Pharmacological Space

The Open PHACTS Foundation + elixir Data for Life + Private Partners = DataStewardship for IMI

163.8
Mike Barnes: (coming in today:) on community building and leadership

Open PHACTS didn't have the most auspicious start, and some would argue it was an **arranged marriage**, but the project has gone from strength to strength, in a way that has truly built a far reaching community and some damn useful tools and infrastructure! The success of the Foundation is also a testament to this, at the end of most research collaborations, the partners usually can't wait to get on with the "conscious uncoupling" but these guys are practically renewing their vows!

This is all a credit to the leadership of Open PHACTS for making this a friendly and enthusiastic team for all involved.

**Long may it continue……**
Paul Groth:

I think the biggest impact of Open PHACTS is in its methodologies for getting work done. I've seen over and over again the way of working that was fostered in Open PHACTS is spreading. Whether it's the adoption of *athons, focus on iteration, thinking about sustainability from day 1, iteration, the adoption of APIs, focus on delivery. Also, the embrace of community. Things like the dataset descriptions, provenance, linkset publication, target ontologies, are all a result of Open PHACTS embracing what's already out that. I see numerous projects that point to Open PHACTS as an inspiration.

I've also heard time and again that Open PHACTS has showed that IMI projects can deliver.

Lastly, from the semantic technology point of view. Open PHACTS showed that semantic technologies can be applied in complex settings to deliver robust infrastructure.
Lee Harland:

For me it has been to catalyse and collaborate with other efforts to increase the better availability of public data. Key points:

- Before Open PHACTS there were very few pharma-relevant RDF datasets
- Much better capturing of provenance and how prov standards should be used
- Mappings between databases and infrastructure for this
- How various chemistry databases can be better connected
- Disgenet, the gene-disease axis
- Surechembl

In summary - each of these is not down to Open PHACTS alone, but Open PHACTS has had a major contribution in many areas, so the big impact is that there are more data sets available for drug discovery and these are now much better represented and integratable.
David Searls:

Quite aside from the functionality of the resulting framework as a data integration platform, the Open PHACTS project has set an impressive example of an enterprise that demonstrated successful integration in a number of other ways. It was a transnational project in the true spirit of the European Union, with as little apparent friction as any American project of similar scale. It also bridged the academic/commercial divide deftly, including dealing with complex intellectual property issues very professionally. Finally, it dealt with the biggest cultural divide of all, that between scientist and engineer, to create a very functional product. Open PHACTS has truly shown itself to be an integration project *par excellence*, in the broadest sense, and one of its biggest impacts may well be as a model for other similar efforts.
Janssen (Herman van Vlijmen):

- Proof of concept that semantic data integration is of value to early drug discovery and is as a technology ready for implementation.
- Academic/industry consortia are useful to explore new directions/technologies
- Science: making us work on important and hard questions
- Innovation: providing large amounts of integrated data and new ways to access the data, to answer these questions
- Collaboration: making company and academic teams work together
- Because of the Open PHACTS project we are now convinced that we need to use interoperable databases to access all public, private, and commercial data that we have. We are actually doing that together with people in our IT department.
- The achievements of Open PHACTS and our activities played a large role in changing the perspective of our IT people from relational-database-only, to strongly-appreciating-semantic-databases
Dean Allemang:

The impact I have seen from Open PHACTS ranges from the very basic to the very abstract.

At the basic level, the identity service (which has been incorporated into BridgeDB and thence in to many other services) has been so successful that is it almost taken for granted. While being taken for granted isn't a good thing in a relationship, it is the sign of lasting project impact. What would we do without it? My own example was a hackathon I was in a couple years back. We needed, very quickly (a few hour hackathon), to convert identifiers from one authority in to another. BridgeDB, and specifically the Open PHACTS parts of it, to the rescue. As a sign of the impact, the team member who knew of this and was able to make it happen, had no formal relationship to Open PHACTS.

At the other end of the spectrum, Open PHACTS has inspired other projects with a view to what is possible. I routinely describe Open PHACTS to people who want to share data on an industrial scale, sometimes in closely related areas (agribusiness), other times in very different areas (finance). They are inspired organizationally by the IMI, that it is possible to organize multinational companies in to productive pre-competitive research. Then they are inspired by the technological success of Open PHACTS. Then the discussion moves to "could we do that here?" and "could we take advantage of the Open PHACTS infrastructure in our field?" These projects are still ongoing, but the success of Open PHACTS has provided impetus to convince others that this can be successful.

Another benefit that Open PHACTS could provide beyond its walls is some good, well-tested policies about licensing in a linked open data context.