



## D6.2 EUBrazilOpenBio Joint Action Plan

**Contract number:** FP7-288754 / CNPq 590052/2011-0

**Start Date of Project:** 1 June 2011

**Duration of Project:** 28 months

**Work Package:** WP6 – Policy Dialogue and Exploitation including a sustainable socio-economic and Impact Study between Europe and Brazil

**Due Date:** M28 – 30/09/2013

**Submission Date:** 18/10/2013

**Partner Responsible for the Deliverable:** TRUST-IT (EU) / CRIA (BR)

**Dissemination Level:** PU

**Nature:** Report

**Author(s):** Stephanie Parker, Sara Pittonet, Silvana Muscella (TRUST-IT), Vanderlei Canhos, Renato de Giovanni (CRIA), William A. Gray (SP2000), Donatella Castelli (CNR)

**Reviewer(s):** Ignacio Blanquer (UPVLC)



**CNPq**  
Conselho Nacional de Desenvolvimento  
Científico e Tecnológico  
**60 ANOS**



e-infrastructure



## Change Log

Version	Date	Description	Author(s)
v1.0	12/05/2013	Initial ToC and Draft	Silvana Muscella, Stephanie Parker, Sara Pittonet (Trust-IT); Vanderlei Canhos (CRIA)
v2.0	12/06/2013	Elaboration of the first draft with feedback from the SAB	Stephanie Parker, Sara Pittonet (Trust-IT)
v3.0	25/06/2013	Elaboration of the scenarios with further feedback from the SAB	Stephanie Parker, Silvana Muscella (Trust-IT)
v4.0	12/07/2013	Edits to the foreword, revision of project achievements section, revision of exploitation and synergies, insertion of partner contributions	Stephanie Parker, Sara Pittonet (Trust-IT), Vanderlei Canhos (CRIA), Willian A. Gray (SP2000)
v5.0	24/07/2013	Final draft taking on the review comments	Stephanie Parker (Trust-IT), Vanderlei Canhos (CRIA)
v6.0	19/08/2013	Updated content and re-structuring for the public consultation version for publication in booklet format	Stephanie Parker (Trust-IT), Vanderlei Canhos (CRIA)
v7.0	10/09/2013	Integration of additional partner content and selected position papers from BIH2013	Donatella Castelli (CNR), Wouter Los, Jacco Konijn, (Univ of Amsterdam), Peter Schalk, William Gray (SP2000), Wouter Addink (Naturalis)
v8.0	13/09/2013	Edits to the EUBrazilOpen assets section	Renato De Giovanni (CRIA)
v9.0	25/09/2013	Integration of key trends emerging at EGI Technical Forum and exploitation opportunities.	Stephanie Parker (Trust-IT)
v10	01/10/2013	Final editing	Stephanie Parker (Trust-IT)
V11	09/10/2013	Reviewer's contributions and feedback	Ignacio Blanquer (UPVLC)
V12	14/10/2013	Response to reviewer's feedback	Stephanie Parker (Trust-IT)

## Document Review

Review	Version	Date	Reviewers	Comments
Draft	v10.0	06/10/2013	Rosa M. Badia (BSC)	Minor comments and corrections
QA	v10.0	09/10/2013	Ignacio Blanquer (UPVLC)	
GB	V12	16/10/2013	GB	

## Disclaimer

EUBrazilOpenBio - Open Data and Cloud Computing e-Infrastructure for Biodiversity (2011-2013) is a Small or medium-scale focused research project (STREP) funded by the European Commission under the Cooperation Programme, Framework Programme Seven (FP7) Objective FP7-ICT-2011-EU-Brazil Research and Development cooperation, and the National Council for Scientific and Technological Development of Brazil (CNPq) of the Brazilian Ministry of Science and Technology (MCT) under the corresponding matching Brazilian Call for proposals MCT/CNPq 066/2010.

This document contains information on core activities, findings, and outcomes of EUBrazilOpenBio project, and in some instances, distinguished experts forming part of the project's Strategic Advisory Board. Any references to content in both website content and documents should clearly indicate the authors, source, organization and date of publication.

The document has been produced with the co-funding of the European Commission and the National Council for Scientific and Technological Development of Brazil. The content of this publication is the sole responsibility of the EUBrazilOpenBio Consortium and its experts and cannot be considered to reflect the views of the European Commission nor the National Council for Scientific and Technological Development of Brazil.

---

## Table of Contents

<b>Disclaimer</b> .....	<b>4</b>
<b>1 Where we stand today</b> .....	<b>8</b>
1.1 Europe and Brazil Perspectives on ICT and Cloud Computing .....	8
<b>2 Spotlight on EUBrazilOpenBio – an open access approach to e-infrastructure for biodiversity research</b> .....	<b>10</b>
2.1 EUBrazilOpenBio: Exploitation and Future Synergies .....	13
<b>3 Facing up to the challenges</b> .....	<b>17</b>
3.1 e-Science in Horizon 2020 .....	17
3.2 Challenges for e-infrastructure .....	18
3.3 Biodiversity: challenges and priorities .....	20
<b>4 EUBrazilOpenBio Joint Action Plan</b> .....	<b>24</b>

## Foreword

Horizon 2020 is a time line. The overriding goal of the European Commission is to make every researcher digital by this date through the development and usage of e-infrastructure across an increasingly wider range of disciplines while also meeting the overall needs of society.

2020 is also a time line for international biodiversity targets. The goal of the European Union is to “halt the loss of biodiversity and ecosystem services in the EU by 2020 and restore them insofar as possible, stepping up the EU’s contribution to averting global biodiversity loss”.

Research and innovation are crucial to addressing these challenges. The services provided by the e-Infrastructures for research are essential for progress in any type of research. Such e-Infrastructures include high-performance and high-throughput computing, high-end storage, advanced networking, services like authentication and authorisation, and services to support research workflows.

No single country can tackle the big issues alone. No single institution or initiative can meet the growing demands around big data, compute and storage to make research more cost effective and accelerate discovery. Further, modern research is international by nature. It requires integrated services and interoperable infrastructures across Europe and the world. To implement such an e-infrastructure commons, a high degree of collaboration and standardisation is required.

Open data and open science are essential to boost innovation and demonstrate the return of investments by national governments to society at large. Open research enables the sharing of methods, software and other variants of virtual infrastructure. With more investments to incentivise openness and reward contributors, Europe and other world regions will be able to take huge steps forward in making research data more reproducible, gaining transparency and easing collaboration on existing research data.

EUBrazilOpenBio is a 2-year scoping analysis aimed at understanding how to address biodiversity challenges more effectively by creating a joint data infrastructure for Brazil and Europe. It has demonstrated the benefits of Open Access in all its core aspects, by focusing on the integration of open data, software and services. It promotes the concept of openness for scientific research by leveraging the achievements, components and infrastructures developed in other projects, so that both regions can capitalise on earlier investments and bring to the table experiences on user-centric approaches. It has also demonstrated the benefits of small-scale funding to enable this open integration, overcoming different approaches by focusing attention on the sharing of smarter approaches. The project and its Board have conducted an in-depth analysis of the biodiversity informatics and cloud computing landscapes. This analysis also draws on studies by several partner, including the position paper presented by William A. Gray, Species 2000, as well as the views and visions of external experts.

The EUBrazilOpenBio Joint Action Plan presents a summary of these findings, highlighting priorities for the further development of e-infrastructure in Europe and beyond, and the enabling role e-infrastructure can play in implementing the 2020 and 2050 vision for biodiversity and ecosystem services. It proposes potential scenarios for future collaborative work between Europe and Brazil, not only to address complex biodiversity challenges but also to build stronger links between business communities to contribute to sustainable development.

Our analysis of cloud computing needs in Brazil and the present maturing market of cloud offering in Europe has highlighted the many different opportunities that exist for joint co-operation between Europe and Brazil. It is therefore important to build on the successful exercise of the EUBrazilOpenBio project and ensure that the wider research community benefits from easy to access virtual computing infrastructures such as cloud computing to tackle the rising tide of large scientific data. These opportunities range from new multi-disciplinary approaches to biodiversity, including

internationally recognised post-graduate qualifications and joint facilities, to small and medium-sized businesses creating value-add services around open data. We expect that the many fruitful EU-Brazil discussions will continue in the near future to help realise the vision we are now shaping thanks to the pioneering approach of EUBrazilOpenBio.

The Joint Action Plan is an initial step towards demonstrating that new and creative approaches to scientific discovery will be possible by mastering the main technical and data challenges that lie ahead. Brazil and Europe have much to contribute to the creation of better services for researchers at all the different levels. By embracing the diversity of talent that exists in all fields of research and informatics, future international co-operation can help make collaborative research better, more open, and multidisciplinary.

### **Acknowledgements**

Members of the EUBrazilOpenBio Strategic Advisory Board: Malcolm Atkinson, Director of Edinburgh Data-intensive research and UK e-Science Envoy; Fabrizio Gagliardi, Independent Consultant; Nelson Simões, Director General of Brazil's National Research and Education Network, RNP

William A. Gray, Species 2000

Jacco Konijn, University of Amsterdam and CReATIVE-B

Wouter Los, University of Amsterdam and LifeWatch

Peter Schalk, Acting Executive Secretary Species 2000

Wouter Addink, Naturalis

## 1 Where we stand today

### 1.1 Europe and Brazil Perspectives on ICT and Cloud Computing

There have been significant investments in both Brazil and Europe to develop, deploy and foster the uptake of cloud and virtualised infrastructures to make available a virtual environment for a distributed community, gaining capacity and enhancing collaborative work. Virtualisation (at storage, server, operating system, network and application level) has developed significantly in recent years, reducing IT expenses while boosting efficiency and agility.

On the European side, e-infrastructures have evolved to encompass advances for cloud and grid middleware and applications, including efforts supporting open-standards based federation. EC-funded initiatives include VENUS-C (2010-2012)<sup>1</sup>, StratusLab (2010-2012)<sup>2</sup> and EGI-InSPIRE (2010-2014)<sup>3</sup>. In particular, the VENUS-C project demonstrated that cloud is an effective paradigm to provide computing power, not only to the research community but also to small companies for which HPC systems are not economically affordable. The project adopted a user-centric approach to evaluate the benefits of Platform-as-a-Service style cloud computing in diverse research disciplines underpinned by software that is built around two frameworks: Generic Worker and COMP superscalar (COMPSs). It has therefore contributed to the deployment of easy-to-use platforms that can facilitate access to a wide user base.

EGI is the world's largest, most powerful and most comprehensive distributed computing e-infrastructure supporting world-class research conducted by over 22,000 scientists and researchers across 50 countries<sup>4</sup>. The EGI community is a federation of independent national and community resource providers, whose resources support specific research communities and international collaborators both within Europe and worldwide. EGI has developed a cloud infrastructure platform as part of its strategic vision to support the wider scientific community. This transition to the cloud is led by the Federated Cloud Task to integrate existing communities' resources and make them interoperable based on the use of open standards. This work is paving the way towards a highly flexible European model to deliver innovative cloud services for the international research community.

This goal is also shared by the Helix Nebula<sup>5</sup> initiative, which is setting up a European cloud for science. Based on the experience gathered from "proof of concept" deployments, Helix Nebula's architecture group, led by a series of cloud-savvy SMEs, have defined a standards based federated cloud architecture to enable an open platform for science innovation. EGI.eu is contributing to the development of the architecture so that the EGI publicly funded e-infrastructure can be interfaced with Helix Nebula. Flagship applications from more research disciplines that will stretch the functionality and impact of Helix Nebula have been identified for deployment during 2013.

Early investments by the Brazilian government include cloud computing pilots, focusing on the federation of computing resources, middleware development for service management and PaaS for scientific applications, as well as upgrading the backbone of its National Research and Education

---

<sup>1</sup> VENUS-C, <http://www.venus-c.eu>. VENUS-C in the Digital Agenda for Europe: Bringing open, user-centric cloud infrastructure to research communities, <https://ec.europa.eu/digital-agenda/en/news/bringing-open-user-centric-cloud-infrastructure-research-communities> and SMEs empowered by a pioneering approach to cloud computing, <http://ec.europa.eu/digital-agenda/en/news/smes-empowered-pioneering-approach-cloud-computing>.

<sup>2</sup> StratusLab, <http://stratuslab.eu/>.

<sup>3</sup> EGI, <http://www.egi.eu/>; EGI Federated Cloud Task, <https://wiki.egi.eu/wiki/Fedcloud-tf:FederatedCloudsTaskForce>.

<sup>4</sup> Figures retrieved September 2013, [http://www.egi.eu/infrastructure/operations/egi\\_in\\_numbers/](http://www.egi.eu/infrastructure/operations/egi_in_numbers/).

<sup>5</sup> Helix Nebula, <http://helix-nebula.eu/>.

Network, RNP<sup>6</sup>. Examples of early initiatives include the SINAPAD network of HPC centres<sup>7</sup> and the OurGrid Community<sup>8</sup> complemented by RNP initiatives, such as the Just in Time (JiT) Cloud project<sup>9</sup> and the Working Group MyScientificCloud (GT mc2) within RNP. These initiatives have not only facilitated the development of e-science activities but also substantially contributed to raising awareness of e-science in the country, as highlighted by the Brazilian Ministry of Science, Technology and Innovation<sup>10</sup>.

RNP is driving migration to cloud services, investigating feasibility and the potential for a community cloud approach (e.g. sharing resources and contracts), as well as the implications of buying versus outsourcing services. The Cloud for Science Working Group (CNC) within RNP has deployed a cloud storage service prototype based on the OpenStack Swift object storage<sup>11</sup> with on-going improvements to enable users to store their files in the cloud and transform the prototype into a cloud storage service pilot. RNP also coordinates the deployment of special ICT projects commissioned by the Brazilian government, like the Brazilian Biodiversity Information System (SiBBr)<sup>12</sup> and the Shared Data Center project (CDC). Future strategic plans of RNP include the deployment of a test bed for Future Internet research, upgrading part of its backbone to 100Gbps and creating two new points of international connectivity, at Rio de Janeiro and Fortaleza (currently the international links are connected from Sao Paulo).

Finally, RNP plays a leading role in Latin America, committing considerable financial resources and expertise to RedCLARA. Internationally, it is contributing to the pioneering work of the NREN Global CEO Forum, which is focused on addressing global challenges such as a well-defined and inclusive network architecture interconnecting NRENs on a global scale; global federated identity management; global real-time communications exchange; and global service delivery<sup>13</sup>. Its involvement in the forum, which includes AARNet (Australia), CANARIE (Canada), CERNET (China), CUDI (Mexico), DFN (Germany), Internet2 (USA), Janet (UK), NORDUnet (European Nordics), REANNZ (New Zealand), RedCLARA (Latin America), RENATER (France), SURFnet (Netherlands), clearly illustrates its significance in the international arena.

Initiatives like EELA<sup>14</sup> (I, II and, GISELA)<sup>15</sup>, EUBrazilOpenBio<sup>16</sup>, and the CHAIN-REDS project with the participation of RNP and RedClara<sup>17</sup> have played an important part in fostering international co-operation between Europe and Brazil over the last decade.

---

<sup>6</sup> Brazilian National Research and Education Network, RNP (Rede Nacional de Ensino e Pesquisa), <http://www.rnp.br/>.

<sup>7</sup> SINAPAD, <https://www.lncc.br/sinapad/>.

<sup>8</sup> OurGrid, <http://www.ourgrid.org>.

<sup>9</sup> Just In Time Cloud, [http://www.lsd.ufcg.edu.br/relatorios\\_tecnicos/TR-3.pdf](http://www.lsd.ufcg.edu.br/relatorios_tecnicos/TR-3.pdf).

<sup>10</sup> Rafael H. R. Moreira, director of ICT policy of the Ministry of Science, Technology and Innovation in Brazil, cloudConf LATAM 2012: [http://cloudconf.com.br/arquivos/CloudConf2012\\_Rafael\\_Moreira.pdf](http://cloudconf.com.br/arquivos/CloudConf2012_Rafael_Moreira.pdf).

<sup>11</sup> OpenStack SwiftStack, <http://swiftstack.com/openstack-swift/>.

<sup>12</sup> SiBBr, <http://www.sibbr.gov.br/>.

<sup>13</sup> G. Cattaneo, M. Claps, S. Conway IDC, S. Musella, S. Parker, N. Ferguson, Trust-IT, *Cloud Computing for e-Science and Public Authorities – Final Report* (SMART 0055), pp. 21-23, <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/final-report-clouds-study.pdf>.

<sup>14</sup> EELA, <http://www.eu-eela.eu>.

<sup>15</sup> GISELA, <http://www.gisela-grid.eu>.

<sup>16</sup> EUBrazilOpenBio, <http://www.eubrazilopenbio.eu>.

<sup>17</sup> <http://www.chain-project.eu/>.

## 2 Spotlight on EUBrazilOpenBio – an open access approach to e-infrastructure for biodiversity research

*“EU-BrazilOpenBio has successfully supported co-operation between Brazil and Europe on the key theme of biodiversity. Its focus has been on demonstrating the benefits of integrating and sharing data, tools and services in an open way to facilitate biodiversity research communities across borders rather than on technological innovation. Part of its lasting legacy will be the cultural and environmental bridges between the two regions that it has successfully built.”*

**Malcolm Atkinson, e-Science Institute and National e-Science Centre, UK**

*“At Biodiversity Informatics Horizons 2013, EUBrazilOpenBio presented an impressive array of facilities.”*

**Dave Roberts, Natural History Museum, UK**

Rather than creating a new monolithic infrastructure, EUBrazilOpenBio has focused on demonstrating the benefits of an infrastructure model that is capable of harnessing existing disparate resources to provide a coherent research environment for scientists. The EUBrazilOpenBio infrastructure is the result of the collaborative aggregation of data, tools, services and computational resources into a coherent and integrated research environment for the benefit of the biodiversity community. EUBrazilOpenBio supports the open access movement, promoting the concept of openness for scientific research by leveraging the achievements, components and infrastructures developed in other projects, so that both regions can capitalise on earlier investments and bring to the table experiences on user-centric approaches.

As a hybrid data infrastructure, EUBrazilOpenBio provides a delivery model for data management capability, in which computing, storage, data and software are available “as-a-Service”<sup>18</sup>. It builds on the cloud paradigm offering ‘computing as a utility’, introducing the elasticity of resources and infinite capacity as key features with the aim of making data and data management services available on demand. The platform has been designed based on the evaluation of user requirements. The underlying architecture comprises a number of interacting services underpinned by a service-oriented infrastructure approach, with a focus on re-use and interoperability.

Its aggregative nature enables the creation of a “system of systems”, drawing on a range of synergies with Brazilian and European initiatives to leverage some of the most relevant resources, e.g. textual publications, data sets, maps, taxonomies, tools, services, computing and storage capabilities. Integration spans Species2000/ITIS Catalogue of Life<sup>19</sup>, the Virtual Herbarium of Flora and Fungi<sup>20</sup>, Global Biodiversity Information Facility (GBIF)<sup>21</sup>; speciesLink<sup>22</sup>; List of Species of the Brazilian

---

<sup>18</sup> On the underlying technology paradigm for e-infrastructure adopted by EUBrazilOpenBio, see L. Candela, D. Castelli, P. Pagano, 'gCube: a service-oriented application framework on the grid', ERCIM News, vol. 72 pp. 48 - 48. ERCIM, 2008, <http://ercim-news.ercim.eu/en72/rd/gcube-a-service-oriented-application-framework-on-the-grid>. gCube is a key enabling technology of the iMarine hybrid data infrastructure, [www.imarine.eu](http://www.imarine.eu). See also, ‘the iMarine e-infrastructure: integrating technologies to unlock knowledge on marine living resources’, November 2012, <http://www.isgtw.org/announcement/imarine-e-infrastructure-integrating-technologies-unlock-knowledge-marine-living-resour>, and ‘Surfing the marine data wave’, September 2013, <http://www.isgtw.org/feature/surfing-marine-data-wave>.

<sup>19</sup> Species2000/ITIS Catalogue of Life, <http://www.catalogueoflife.org/>.

<sup>20</sup> Virtual Herbarium of Flora and Fungi, Virtual Herbarium of Flora and Fungi, <http://inct.florabrasil.net/noticias/lista-de-especies-da-flora-do-brasil-nova-versao/>.

<sup>21</sup> Global Biodiversity Information Facility (GBIF), <http://www.gbif.org/>.

<sup>22</sup> SpeciesLink, <http://splink.cria.org.br/>.

Flora<sup>23</sup> openModeller<sup>24</sup>, as well as initiatives like BioVeL<sup>25</sup> and i4Life<sup>26</sup> and technologies such as COMP Superscalar (COMPSs)<sup>27</sup> and gCube<sup>28</sup> and cloud infrastructures like VENUS-C.

Standards adoption for the implementation of the interfaces makes it possible not only to interoperate with other infrastructures but also to simplify the adoption of EUBrazilOpenBio by other communities<sup>29</sup>. A good example is the cloud-enabled Ecological Niche Modelling Service developed in EU-BrazilOpenBio and the synergy with BioVeL, integrating the service in the BioVeL workflow and working with the EGI Federated Cloud Task and CloudCaps mini-project to optimise the use case on Ecology.

EUBrazilOpenBio also has the capability to support the creation and operation of virtual research environments (VREs), a web-based collaborative platform connecting researchers irrespective of geographical location. VREs offer transparent and seamless access to a shared set of remote resources, such as data, tools and computing capabilities central to their work. By leveraging the flexible approach of VREs, EUBrazilOpenBio has contributed to an evolving ecosystem for user-configurable e-science environments that help to build communities. This VRE ecosystem and the creation of Virtual Research Communities (VRCs) will continue to be an important feature in Horizon 2020.

### Use Cases

EUBrazilOpenBio enables Brazilian and European research communities to tap into high-quality data sources and tools with easy access to an even greater knowledge base, such as taxonomic intelligence and ecological niche modelling.

**Integration of regional and global taxonomies:** Access to reliable taxonomic information classifying the names of species is of paramount importance for environmental science, for monitoring changes in biodiversity and for effectively addressing challenges related to climate change, both locally and on a global scale. However, the ability to cross-reference between regional and global taxonomic data sets is often masked by complex differences in the taxonomic classification used. EUBrazilOpenBio supports the use of the Catalogue of Life Cross-Mapping Tool, developed by scientists at Cardiff University, to improve taxonomic information relating to Brazil's plant species. The tool cross-maps the regional List of Species of Brazilian Flora, containing around 10,000 species organised under 40,000 plant taxa names (including synonyms), against the global Species2000/ITIS Catalogue of Life, indexing about 150,000 plant species. The Catalogue of Life Cross-mapping Tool helps manage differences between catalogues by detecting, analysing and reporting not only differences between two checklists of species, but also differences in their taxonomic treatment. It then identifies potential resolutions to help researchers resolve these differences.

---

<sup>23</sup> List of Species of the Brazilian Flora, [http://wiki.eubrazilopenbio.eu/index.php/List\\_of\\_Species\\_of\\_the\\_Brazilian\\_Flora](http://wiki.eubrazilopenbio.eu/index.php/List_of_Species_of_the_Brazilian_Flora).

<sup>24</sup> openModeller, <http://openmodeller.sourceforge.net/>.

<sup>25</sup> BioVeL, <http://www.biovel.eu/>.

<sup>26</sup> i4Life – Indexing for Life, <http://www.i4life.eu/>.

<sup>27</sup> COMP Superscalar, Barcelona Supercomputing Center, <http://www.bsc.es/computer-sciences/grid-computing/comp-superscalar>.

<sup>28</sup> gCube Framework, <http://www.gcube-system.org/>.

<sup>29</sup> On key achievements of EUBrazilOpenBio from a technical perspective, see for example, 'EU-Brazil Open Data and Cloud Computing e-Infrastructure for Biodiversity', *Proceedings of the 5<sup>th</sup> International Workshop on Science Gateways (IWGS 2013)*, June 2013, <http://ceur-ws.org/Vol-993/paper7.pdf>. See also, 'Execution of scientific workflows on federated multi-cloud infrastructures', proceedings of the 1<sup>st</sup> International FedICI 2013 Workshop on Federative and Interoperable Cloud Infrastructures, Euro-Par 2013, <http://www.dps.uibk.ac.at/~gabor/FedICI13/program.html>.

The idea is to benefit from the 'taxonomic intelligence' of the tool and support a pilot study that analyses the regional plant catalogue of Brazil against the global index of plants within the Catalogue of Life. This will be used to improve the linkage between the terms in the lists and to enhance both the Catalogue of Life and the Brazilian List of Flora. In turn, this will allow taxonomic and biodiversity specialists to utilise richer data with the potential for the exchange of information in both directions (from the regional Brazilian List of Flora to the Catalogue of Life and vice-versa), thus helping to fill and close gaps between the different taxonomies. This consolidation is likely to be a significant boon to the conservation efforts of scientists who make use of these catalogues in their research and should help provide a fuller picture of biodiversity levels in Brazil

**Ecological Niche Modelling:** EUBrazilOpenBio's co-operation with the Brazilian Virtual Herbarium is also part of the drive towards a more complete and integrated view, identifying knowledge gaps in the biogeography of flora and using herbarium data to generate ecological niche models. Occurrence points are retrieved from speciesLink, a network that integrates data from distributed biological collections, currently serving almost 4 million plant specimen records. The modelling strategy involves generating algorithms in openModeller, helping biodiversity researchers to model ecological niches of plant and animal species across the world. The occurrence points are also retrieved from GBIF with over 380 million points<sup>30</sup>.

Such an approach indicates the ecological requirements for a species to survive and maintain viable populations over time by relating locations where a species is known to occur with environmental variables that are expected to affect its distribution. This makes it possible to predict the impact of climate changes on biodiversity and prevent the spread of invasive species, which is one of the challenges identified in the European biodiversity strategy. Such an approach also helps in conservation planning, identify geographical and ecological aspects of disease transmission, as well as in guiding biodiversity field surveys, among many other uses.

Ecological niche modelling has recently become one of the most popular techniques in macroecology and biogeography. One of the reasons for this trend is the broad range of applications that arise when assessing and projecting the ecological niche of a species in different environmental scenarios and geographical regions. The idea in EUBrazilOpenBio is to investigate and propose efficient ways of generating a large number of ecological niche models through a web service interface that applications like the Brazilian Virtual Herbarium can leverage, integrating a new web-based modelling application in the EUBrazilOpenBio virtual research environment.

### The value of small-scale funding for nurturing an open access culture

In many countries, there are now requirements to make publicly available all data that comes from research funded by the public purse. It is important to find sustainable ways to make this possible. It is equally important to find ways to provide value-add analytical and search services that make using data easier.

EUBrazilOpenBio is a prime mover within the open access movement in terms of building close ties Brazil and Europe that nurture an open access culture. It has gained specific experiences in data exchange across borders, building bridges across different ways of working, different standards and different funding mechanisms. These experiences demonstrate the importance of taking on board cultural, linguistic and scientific differences with regard to data challenges. These experiences also highlight the need to be cautious about applying models developed elsewhere "lock, stock and barrel" to other disciplines or in other parts of the world. It may therefore be appropriate to consider the best approach in each domain in light of current experiences and the long-term experience in bodies concerned with data sharing in individual disciplines or technologies, and generic long-standing international bodies, such as CODATA.

<sup>30</sup> <http://dagendresen.wordpress.com/2012/10/15/gbif-occurrence-data-access-and-download/>.

Further, the experiences of EUBrazilOpenBio demonstrate the effectiveness of small-scale funding to enable the integration of open data, software and services. This open integration process has helped overcome different approaches by focusing attention on mutual benefits stemming from sharing smarter approaches.

The experience of EUBrazilOpenBio clearly demonstrates the value of addressing human challenges when it comes to international co-operation while placing biodiversity as a grand global societal challenge at its centre. While it has focused on two specific demonstration use cases, the infrastructure is designed to meet the needs of a wider range of biodiversity applications. EUBrazilOpenBio has therefore created a new and diverse approach to addressing biodiversity with high potential for social impact. Key challenges ahead for the project are increasing the user base, sustaining the services along with the “glue” that holds them together.

## 2.1 EUBrazilOpenBio: Exploitation and Future Synergies

*“EUBrazilOpenBio has made the first steps towards the goal of supporting key challenges in the area of biodiversity and ecosystems, including at the legal level where open access is a vital prerequisite for interoperability. Such an approach promises well for future co-operation potential with initiatives such as LifeWatch in future funding streams like Horizon 2020.”*

**Jacco Konijn, University of Amsterdam and CReATIVE-B<sup>31</sup>**

EUBrazilOpenBio has dedicated efforts to building synergies across a range of biodiversity initiatives, as well as with EGI as a pan-European cloud-based e-infrastructure that will also leverage federation with the Helix Nebula eScience Cloud. Many of these synergies will be valuable in defining ways to extend the user community and identify potential future collaborations that capitalise on its achievements.

### EGI Cloud Infrastructure Platform

*“The successful collaboration with EUBrazilOpenBio/BioVeL has demonstrated how the BioVeL community could benefit from the full potential of the cloud and help make web services and workflows sustainable for the long term for biodiversity research in Europe.”*

**Nuno Ferreira, User Community Support, EGI.eu**

The EGI Federated Cloud Task is a core activity that is part of a drive to create a platform based on the adoption of open standards in a way that is aligned with the vision and recommendations of the SIENA Roadmap<sup>32</sup>. This open platform is designed to enable resource centres, technology providers and user communities to play an active role and build on the current portfolio of use cases with the aim of boosting innovation through scientific excellence. The federation approach is important to support multi-national deployment that reflects the local environment while driving technical and policy alignment as needed. The ultimate goal is to build on its current user base, particularly through federated cloud services targeting the ‘long tail of science’<sup>33</sup>.

The aim of the EGI Cloud Infrastructure Platform is to provide researchers and research communities with control over deployed applications, elastic resource consumption based on actual

<sup>31</sup> Jacco Konijn, University of Amsterdam, Position Paper for the EUBrazilOpenBio Public Workshop.

<sup>32</sup> Roadmap on Distributed Computing Infrastructure for e-Science and Beyond in Europe, May 2012, <http://www.cloudscapeseries.eu/Content/eLibrary.aspx?id=96&Page=2&Cat=0!19>.

<sup>33</sup> S. Newhouse, ‘Federated Clouds and the role of standards’, Cloudscape V, 27-28 February 2013, Brussels, <http://www.cloudscapeseries.eu/Content/CloudscapeSeries.aspx?id=247&Page=3&Cat=0!12!6>.

needs, service performance scaled with elastic resource consumption. This approach significantly reduces waiting time through immediate processing of workloads.

Ecological niche modelling is being increasingly used by the biodiversity research community, which can greatly benefit from cloud computing technologies, especially in complex experiments involving multiple species, algorithms and environmental scenarios. Over the last two years, BioVeL and EUBrazilOpenBio have joined forces with the EGI Federated Cloud Task Force, with the aim of making openModeller ready for cloud deployment and on optimising the software<sup>34</sup>. The successful outcomes of this collaboration means that the BioVeL community can use the EGI Federated Cloud to its fullest potential, demonstrating how cloud technologies could ultimately make web services and workflows sustainable for the long term for biodiversity research in Europe<sup>35</sup>.

## LifeWatch

*“EUBrazilOpenBio has developed easy-to-use technologies for different purposes, while fostering cross-Atlantic cooperation. This is possibly the most important legacy of the project. The project has paved the ground for bringing together different communities with technological and sociological processes. It is an example for scaling up further initiatives in support of biodiversity system research in co-operation with related initiatives. Scaling up with new sustained cooperation models of distributed initiatives is crucial to enter a new research area. This will also promote better valorisation of new developments with public and private partners.”*

**Wouter Los, LifeWatch Project Coordinator**

The ESFRI initiative, LifeWatch<sup>36</sup>, aims to build an infrastructure for biodiversity and ecosystem research. Its plans include integrating smart sensor networks from different locations, metadata standards, communication capabilities. This integration requires facilitating and standardising data gathering to provide pre-processed and interoperable data to virtual labs and sensor data portals. Virtual Lab integration in the ICT-core will leverage the involvement of LifeWatch in EUDAT<sup>37</sup> through LTER-Europe (European monitoring network for long-term ecological research)<sup>38</sup>, which coordinates and manages large amounts of monitoring data.

There is increasing consensus that our planetary environmental systems are strongly interconnected. Biodiversity and ecosystem processes are not only affected by - but also buffer - changes in the climate system, oceanic and atmospheric processes, as well as by the lithosphere and hydrosphere<sup>39</sup>. Priority areas for biodiversity research will need to focus on ecosystem services, on the impact of biodiversity loss, and adaptation/mitigation strategies, which have to be considered in the framework of Earth as a single complex and coupled system. Model and scenario development may assist in decision support for evaluating the effect of potential interventions. It is worthwhile considering how the growing portfolio of enabling technologies may promote targeted data production from diverse organisations to support large-scale analysis and modelling.

---

<sup>34</sup> D. Lezzi, ‘The EUBrazilOpenBio-BioVeL use case in EGI’, EGI Cloud Infrastructure Platform: State of the art and Demo, EGI Technical Forum, 16-20 September 2013, Madrid.

<sup>35</sup> [http://www.egi.eu/news-and-media/newsfeed/news\\_2013\\_0038.html](http://www.egi.eu/news-and-media/newsfeed/news_2013_0038.html).

<sup>36</sup> LifeWatch, e-Science European Research Infrastructure for Biodiversity and Ecosystem Research, <http://www.lifewatch.eu/>.

<sup>37</sup> EUDAT, European Data Infrastructure, <http://www.eudat.eu/>.

<sup>38</sup> LTER-Europe, European Long-term Ecosystem Research Network, <http://www.lter-europe.net/>.

<sup>39</sup> W. Los, Position Paper for EUBrazilOpenBio Public Workshop, September 2013.

Spain is one of the countries participating in LifeWatch with the aim of providing services to applications, especially from the biodiversity domain. The government has mandated the Spanish Research Council (CSIC) to technically articulate LifeWatch, which includes the participation of the Barcelona Supercomputing Center and the 13M department at the Technical University of Valencia. One of the goals is to build an infrastructure in Seville and prototype a virtual lab with integrated real-time sensors in field. LifeWatch stakeholders in Spain have manifested interest in the results of the use cases developed in EUBrazilOpenBio. The project therefore expects to transfer the services developed and tailor them to the purposes of LifeWatch.

### EU Brazil Cloud Connect

EU Brazil Cloud Connect<sup>40</sup> (EUBrazilCC) is aimed at creating a federated research e-infrastructure based on a user-centric approach. It will adapt existing applications to tackle new scenarios in areas of mutual interest to Brazil and Europe with high social impact and innovation: neglected diseases, climate change and heart simulation. It will integrate frameworks and programming models for scientific gateways and complex workflows that meet not only the requirements of three specific use cases but potentially also a much larger user community.

It will drive advances in several research areas, such as virtualised resource federation using clouds that promote sustainability, in programming frameworks in the cloud, including big data analysis, as well as the demand for high capability computing and data in the cloud. The project will pursue interoperability and integration of existing systems, such as coordinating technical interoperability with Helix Nebula, incorporating relevant standards, and extending links with EGI.

The targeted user community extends beyond the project boundaries. EU Brazil Cloud Connect has already received interest from:

- The SINAPAD network and Brazilian National Institutes of Science and Technology: INCT of Innovation in Neglected Diseases, INCT Virtual Herbarium of Flora and Fungi, INCT on Scientific Computing in Healthcare Applications (with the direct participation of LNCC), INCT for Climate Change.
- The LifeWatch-ESFRI initiative and the European Network for Earth System Modelling (ENES).
- It will also build close ties with CloudwatchHUB.eu, with the aim of educating communities on standards to ensure interoperability and broader choice.
- It will introduce in Brazil the Cloudscapeseries.eu, which is now a self-sustained event in Europe.

Two of the three EUBrazilCC use cases relate closely to biodiversity, leveraging several key outcomes of EUBrazilOpenBio, that is, the Leishmania Virtual Lab and research on Biodiversity and Climate Change. The EUBrazilOpenBio cloud-based ecological niche modelling service, to be released as open source, will be used in these use cases to compute, test and project models for predicting the distribution of species, targeting vectors and plants as indicators of climate change.

### CloudWATCH

As an EU cloud observatory, CloudWATCH (cloudwatchhub.eu)<sup>41</sup> aims to accelerate the adoption of cloud computing for business, government and research by focusing on the real issues, practical, independent guides and success stories. It will support the development of common standards profiles to support user community exploitation of cloud computing in the three areas of focus which

<sup>40</sup> EU Brazil Cloud Connect – EU Brazil Cloud Computing for Science, [www.eubrazilcloudconnect.eu](http://www.eubrazilcloudconnect.eu).

<sup>41</sup> [www.cloudwatchhub.eu](http://www.cloudwatchhub.eu).

can be endorsed by standards bodies. The project will also create a long-term sustainable, digital platform to showcase the results of European cloud computing research and development projects and, more generally, offer key information for service providers and users in Europe for supply to meet demand.

Through its partnership with EGI.eu, it will define an outreach plan to disseminate the outcomes of the EGI federated use cases and support strategies aimed at ensuring the long-term sustainability of all value-add cloud resources coming from public funding. It will also monitor open collaboration models such as OpenStack in Europe<sup>42</sup>. It will assess pioneering national strategies driving federated approaches to research infrastructure leveraging secure cloud deployments as enabling platforms for the digital researcher 2020<sup>43</sup>.

### International potential for integration and synergies

The COOPEUS project<sup>44</sup>, which is co-funded by the EC under the Research Infrastructures action of the 7<sup>th</sup> Framework Programme and National Science Foundation (NSF, U.S.), brings together scientists and users from major environmental related research infrastructures in Europe and the US. The aim is to interlink these activities to create new synergies driving the creation of a global integration of existing infrastructures.

With regard to biodiversity challenges, COOPEUS, EUBrazilOpenBio, ENVRI<sup>45</sup>, ViBRANT<sup>46</sup>, EUBON<sup>47</sup>, the Global Biodiversity Information Facility (GBIF), LTER, EGI, EUDAT and LifeWatch started to explore international collaborative efforts at the EGI Technical Forum in September 2013. The aim was to understand how to share progress and define the drivers for future co-operation. In particular, the discussions looked into common biodiversity data services, the potential for common layers, infrastructures, tools, reference models and identity, assessing also technical challenges, such as real-time monitoring, genomic pipelines, and access to taxonomic data, as well as modelling, including earth and satellite data and modelling niches. As a result, EUBrazilOpenBio has been invited to contribute knowledge of mutual interest.

---

<sup>42</sup> For examples of OpenStack private cloud deployments for research, see S. Maffioletti, Elasticcluster: Provisioning computing clusters in the cloud with Python and E. Fernandez, Federating clouds in OpenStack, EGI Technical Forum, 16-20 September 2013, Madrid,

<https://indico.egi.eu/indico/materialDisplay.py?contribId=17&sessionId=34&materialId=slides&confId=1417>

and <https://indico.egi.eu/indico/getFile.py/access?contribId=133&sessionId=34&resId=0&materialId=slides&confId=1417>.

<sup>43</sup> See, for example, P. Kunszt, S. Maffioletti, D. Flanders, M. Eurich, T. Bohnert, A. Edmonds, H. Stockinger, S. Haug, A. Jamakovic-Kapic, P. Flury, S. Leinen, E. Schiller, Towards a Swiss National Research Infrastructure, the 1st International FedICI'2013 workshop: Federative and interoperable cloud infrastructures, 26 August 2013, Aachen, <http://www.dps.uibk.ac.at/~gabor/FedICI13/program.html>.

<sup>44</sup> COOPEUS, <http://www.coopeus.eu/>.

<sup>45</sup> ENVRI, supporting common operations of ESFRI environmental research infrastructures, <http://envri.eu/>.

<sup>46</sup> ViBRANT, Virtual Biodiversity Research and Access Network for Taxonomy, <http://vbrant.eu/>.

<sup>47</sup> EU BON, Building the European Biodiversity Observation Network, <http://www.eubon.eu/>.

## 3 Facing up to the challenges

### 3.1 e-Science in Horizon 2020

The overriding goal of the European Commission is to make every researcher digital by 2020 through the development and usage of e-infrastructure across an increasingly wider range of disciplines while also meeting the overall needs of society<sup>48</sup>. In the past, competition in science has been beneficial in stimulating progress. However, current economic constraints call for a renewed focus on research investments that give short-term returns on investment while minimising overlap with regard to large-scale funding. Cost and energy efficiencies need to move much higher up the list of priorities, for example, by leveraging the efficiencies that virtualisation and cloud can help deliver<sup>49</sup>.

It is also important to improve co-ordination of research e-infrastructure at the level of national research centres, promoting the role of centres of expertise and optimising governance. Taking on the grand global challenges will require better co-ordination among stakeholder communities, identifying synergies across different research communities.

The 2030 vision of the High Level Group on Scientific Data is a “*scientific e-infrastructure that supports access, use and re-use, and trust of data. In a sense, the physical and technical infrastructure becomes invisible and the data themselves become the infrastructure – a valuable asset, on which science, technology, the economy and society can advance*”<sup>50</sup>

The GRDI2020 Roadmap is a vision to ensure that global research data infrastructures harness the growing amounts of scientific data and knowledge and optimise data movement across disciplines<sup>51</sup>. The aim is to enable multi- and inter-disciplinary research by defining better data models and query language, developing advanced data tools and infrastructure services, supporting open linked data spaces, supporting interoperability between scientific data and literature, and creating new professions such as data scientists.

In addition to ensuring sustainability, the following pillars are all central to e-infrastructure in Horizon 2020:

- **Scientific data:** wider range of disciplines leveraging data; problems of scale and complexity; importance of sharing and federating data.
- **Computing facilities:** software and instruments, including virtual research environments and virtual research communities.
- **Networking:** global reach and connectivity “at the speed of light”.

---

<sup>48</sup> eScienceTalk Briefing, Horizon 2020, April 2013, <http://www.e-sciencetalk.org/briefings/EST-Briefing-26-HorizonWeb.pdf>.

<sup>49</sup> On the role of cloud computing in contributing to GHG emission reductions, see P. Thomond, *The Enabling Technologies of a low-carbon economy: A focus on cloud computing*, February 2013, commissioned by the Global e-Sustainability Initiative and Microsoft, [http://gesi.org/assets/js/lib/tinymce/jscripts/tiny\\_mce/plugins/ajaxfilemanager/uploaded/Cloud%20Study%20-%20FINAL%20report\\_1.pdf](http://gesi.org/assets/js/lib/tinymce/jscripts/tiny_mce/plugins/ajaxfilemanager/uploaded/Cloud%20Study%20-%20FINAL%20report_1.pdf).

<sup>50</sup> *Riding the Wave – How Europe can gain from the rising tide of scientific data*, Final Report of the High Level Group on Scientific Data, A submission to the European Commission, October 2010, <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>.

<sup>51</sup> C. Thanos, CNR, H. Hanahoe, Trust-IT, Global Research Data Infrastructures – The GRDI2020 Vision, March 2012, [http://www.grdi2020.eu/Pages/SelectedDocument.aspx?id\\_documento=6bdc07fb-b21d-4b90-81d4-d909fdb96b87](http://www.grdi2020.eu/Pages/SelectedDocument.aspx?id_documento=6bdc07fb-b21d-4b90-81d4-d909fdb96b87).

The requirement that publicly funded research results be made available in open access also means that we need to be more vigilant in demonstrating the value of funding research. The European Commission is driving open science as a priority in future project proposals under Horizon 2020. The open science approach is also in accordance with the Nagoya Protocol<sup>52</sup>, adopted at the 10th meeting of the Parties to the Convention on Biological Diversity (October 2010).

The open sharing of data will help researchers to focus their efforts more effectively when working to tackle some of the biggest challenges of the 21st century. The successful development and deployment of the OpenAIRE framework<sup>53</sup> allows researchers to deposit their publications online where they can be freely accessed by other researchers, demonstrating that open access marks an important step towards an open data culture. OpenAIRE, LIBER and COAR strongly support the development of an open, interoperable e-infrastructure for scientific data through the engagement of the relevant actors, including libraries and repositories supporting researchers in their scientific endeavour<sup>54</sup>. The international Research Data Alliance has been set up push for the convergence of standards on how data is stored and categorised<sup>55</sup>.

In May 2013, Neelie Kroes, Vice President of the European Commission, highlighted the economic and social benefits of big data - the fuel driving the knowledge economy<sup>56</sup>. In the public sector, better data allows services that are more efficient, transparent and personalised. Open access to results and associated data allow new ways to share, compare, and discover: permitting whole new fields of research. For scientists and citizens, data is the key to more information and empowerment, and to new services and applications. At the policy level, the EU revised Directive is aimed at making it easier to use and re-use public data, with lower charges and without complicated conditions for re-use.

## 3.2 Challenges for e-infrastructure

In order to ensure the future sustainable development of e-infrastructure and drive international co-operation on grand global challenges, it is important to address the following challenges.

### Lack of service integration and interoperability of research e-Infrastructures

The e-Infrastructure Reflection Group (e-IRG) sees the integration of services and the interoperability of research e-Infrastructures as crucial for paving the way towards a general-purpose European e-Infrastructure. Its White Paper 2013 identifies a number of the current barriers to realising this vision<sup>57</sup>. According to e-IRG, the challenges are mainly organisational and structural rather than technical. Diverse ways of representing and funding key actors exist with insufficient cohesion between them. Better co-ordination is therefore needed among the different e-infrastructure pillars, components and services, e.g. networking, high throughput and high-performance computing, data infrastructures, software/middleware, including authentication and authorisation infrastructures and virtual research environments that are to be used by international research communities. Additional challenges include legal issues and hurdles to accessing and using e-infrastructures.

<sup>52</sup> Nagoya convention <http://www.cbd.int/abs/>.

<sup>53</sup> OpenAIRE, Open Access Infrastructure for Research in Europe, <http://www.openaire.eu/>.

<sup>54</sup> DG Connect's Proposal for a Framework for Action addressing the area of Research Data e-infrastructure – Responses from LIBER, COAR and OpenAIRE, <http://www.openaire.eu/en/component/content/article/9/453>.

<sup>55</sup> Research Data Alliance, <https://rd-alliance.org/node> and RDA Europe, <https://europe.rd-alliance.org/Pages/Home.aspx>.

<sup>56</sup> The Economic and social benefits of big data, SPEECH/13/450, 23/05/2013, [http://europa.eu/rapid/press-release\\_SPEECH-13-450\\_en.htm](http://europa.eu/rapid/press-release_SPEECH-13-450_en.htm).

<sup>57</sup> e-IRG White Paper 2013, [http://www.e-irg.eu/images/stories/dissemination/e-irg\\_white\\_paper\\_2013\\_short\\_version.pdf](http://www.e-irg.eu/images/stories/dissemination/e-irg_white_paper_2013_short_version.pdf); Long version, [http://www.e-irg.eu/images/stories/dissemination/e-irg\\_white\\_paper\\_2013\\_final\\_version.pdf](http://www.e-irg.eu/images/stories/dissemination/e-irg_white_paper_2013_final_version.pdf).

‘A vision towards scientific communication infrastructure’ highlights the importance of integrating research digital libraries and scientific data centres in order to ensure access to the results of complementary phases in the scientific research process<sup>58</sup>. These two pillars of modern scientific communication support researchers in storing, curating, sharing and discovering the data and publications they produce. However, lack of integration hampers the very objective of modern scientific communication, which is publishing, interlinking and discovery of all the outcomes of the research process, from the experimental and observational datasets to the final paper. In order to bridge this gap, the paper proposes the development of scientific communication infrastructures. Such infrastructures should be capable of facilitating interoperability between data centres and research digital libraries and of providing services that simplify the implementation of the large variety of modern scientific communication patterns.

### Business and technical challenges

The EC Tender Study, *Cloud computing for e-Science and Public Authorities*, identifies challenges at the business and technical levels with regard to the future evolution of e-infrastructure.

The need to develop **new business and funding models** for e-infrastructure, where users start paying for the services they consume and providers compete for innovation money by generating revenue from their customer base. New funding models should be easy to manage, because otherwise the cost of fee for service mechanisms outweighs the benefits. It is also important to **move beyond the CAPEX funding model** in research funding in that cloud computing does not fit well with the typical science grant budgets. There are two main challenges to be overcome: fostering a change in funding policies and raising awareness of the many opportunities of using cloud computing for research. Finally, SMART 0055 proposes the **creation of a data and e-research marketplace** of scientific and social science applications to help support wider uptake. One possible approach could be the creation of application store-like services and possibly charging a small fee, payable to the data/software provider with a high flux of transactions, plus a consultation fee. From an EU and Brazilian collaboration perspective linking the market place and services across borders could ultimately boost opportunities for micro, small and medium-sized businesses in both areas.

The need for **interoperable connectivity and cost-effective services**. Many NRENs in Europe and across the globe are facing financial constraints at a time when their user communities are demanding new services. Despite several pioneering activities<sup>59</sup>, most NRENs do not have concrete plans to deploy cloud-based plans<sup>60</sup>, with the risk of fragmentation of services and unclear business models. The *DANTE Strategy 2012-2015: More for all*<sup>61</sup> aims to ensure that all participating NRENs play their part in ensuring long-term sustainability and in defining their own strategic plan to better serve their communities in an increasingly global research environment. However, the DANTE strategy stops short of proposing a “one-size-fits-all” approach as new service delivery should respect national circumstances. The wider Latin American region is also encountering difficulties in embracing cloud computing, where plans for migrating to the cloud are failing to take off due to the lack of adequate resource (both financial and human). These issues have wider, international

<sup>58</sup> D. Castelli, P. Manghi, C.Thanos, ‘A vision towards scientific communication infrastructure – On bridging the realms of digital libraries and scientific data centers’, Springer, July 2013, vol. 13, pp. 155-169 <http://link.springer.com/article/10.1007%2Fs00799-013-0106-7>.

<sup>59</sup> See, for example, Advanced North Atlantic 100G Pilot, <http://www.surfnet.nl/en/nieuws/Pages/Internet2,NORDUnet,ESnet,SURFnet,CANARIE,andG%C3%89ANT.aspx> and the Enlighten Your Research Global initiative, <http://www.hpcwire.com/hpcwire/2013-06-20/five-leading-national-research-and-education-networks-launch-competition.html>.

<sup>60</sup> ASPIRE – The Future Roles of NRENs, September 2012, <http://www.terena.org/activities/aspire/docs/ASPIRE-future-of-nrens.pdf>.

<sup>61</sup> [http://www.dante.net/Media\\_Centre/Publications/Pages/DANTE\\_Strategy\\_2012-15.aspx](http://www.dante.net/Media_Centre/Publications/Pages/DANTE_Strategy_2012-15.aspx).

implications and call for specific actions to assess, support and co-ordinate the adoption of new technologies.

### 3.3 Biodiversity: challenges and priorities

Modern society depends on data-driven models to understand complex systems and to coordinate most areas of life and of planning, from climate to medicine and from economics to transport. Biodiversity represents one of the major aspects that remains hard to integrate into these systems, even though the human race has been observing and measuring biodiversity and its patterns for centuries, and even though we recognise the importance of such understanding not only for conservation but for all aspects of land planning, primary industry and health.

**Coordination of programmes.** One of the most important lessons learnt from the EU-BrazilOpenBio project is the very different approaches adopted by biodiversity initiatives in Brazil and Europe despite the many common challenges. One of the main differences is the competitive environment in which initiatives have been funded at EU level. A vast array of initiatives exist. Many of these projects directly address the challenges of deploying e-infrastructures for biodiversity science. While they have similar characteristics, they differ substantially in their architectures and technology approaches. These different approaches reveal the lack of a common understanding on how best to deploy e-infrastructures for biodiversity and ecosystem research. As a result, there are overlaps, dead-ends, and often, complete lack of mainstream industrial involvement.

Integration of existing tools and technologies is important within the broader framework of GEOSS<sup>62</sup>, where GEO BON<sup>63</sup> aims to deliver a decentralised and distributed informatics infrastructure. The system will be based on a service oriented architecture approach, building mainly on contributing systems developed at regional, national and multi-national scales, as a “system of systems”. The planned ESFRI LifeWatch<sup>64</sup> research infrastructure, EBONE<sup>65</sup> and EU BON<sup>66</sup> will ultimately form the European contribution.

In order to address these issues, the white paper, *A decadal view of biodiversity informatics: challenges and priorities* (April 2013)<sup>67</sup> makes **12 recommendations** focus on the following strategic areas for biodiversity informatics. Strategic area one: reducing duplication, improving collaboration, facilitating the creation of new knowledge by synthesis using the data and tools generated. Strategic area two: enhancing usability and better deployment of existing technologies. Strategic area three: developing new structures to facilitate the integration of foundational technologies by exploiting technologies in new ways. Strategic area four: New technologies, such as observatories employing novel sensors are delivering data in unprecedented volumes, especially molecular data, as the Genomics Observatories Network has emphasised. This will require the development of new technologies, or adaptations of technologies from related fields, new information systems, and platforms offering overviews of detectors and experimental set-ups for biodiversity research to facilitate exploitation of the opportunities presented.

The White Paper also recommends that future proposals leverage completed and existing funded projects to gain the maximum benefit for the future biodiversity infrastructure. Proposals should explain how they have taken earlier and current project results into account and demonstrate that they are building on them rather than offering incompatible alternatives. Letters of Support from

<sup>62</sup> <http://www.earthobservations.org/geoss.shtml>.

<sup>63</sup> <http://www.earthobservations.org/geobon.shtml>.

<sup>64</sup> <http://www.lifewatch.eu/web/guest/home>.

<sup>65</sup> <http://www.wageningenur.nl/en/Expertise-Services/Research-Institutes/alterra/Projects/EBONE-2.htm>.

<sup>66</sup> <http://www.eubon.eu/>.

<sup>67</sup> A. Hardisty, D. Roberts, *A Decadal View of biodiversity informatics: challenges and priorities*, April 2013, BioMed Central, <http://www.biomedcentral.com/1472-6785/13/16>.

other projects should be used to demonstrate that community-wide discussion and acceptance of proposals has taken place prior to submission for funding, and thus contributing to the community's decadal vision.

The White Paper summarises the opinion of a large part of the biodiversity informatics community with respect to their views on future developments in terms of systems, information needs and global sustainability. Enormous progress has been made in establishing a framework for sharing collection and observation data, such as the Global Biodiversity Information Facility (GBIF)<sup>68</sup> and its Global Biodiversity Informatics Outlook (GBIO)<sup>69</sup>, at the species level by the Encyclopedia of Life (EoL)<sup>70</sup> and with respect to genetic information by the Consortium for the Barcode of Life (CBOL)<sup>71</sup>. An integrated systems approach, which moves significantly beyond taxonomy and species observations, is needed to truly understand biodiversity.

The Catalogue of Life (CoL) is a global effort that is instrumental in connecting the thousands of resources and making them interoperable in a meaningful way. This authoritative synonymic index is a true core-product of the taxonomic community, unifying expert views and opinions in a single database resource. European funded initiatives like 4D4Life and i4Life have laid important foundations, facilitating a virtual taxonomic community and underpinning an iterative cycle of matching the CoL against other sources with taxonomic components that will result in an enriched CoL (more names) plus a system with 'unplacable names' that need at a certain moment to be removed from circulation. The EUBrazilOpenBio project has provided a more versatile cross-mapper web service working in combination with a partly automated piping tool funneling data back to the global species database sources. Species 2000 will include the cross-mapping tools and piping mechanism in the regular CoL services.

Further developments in biodiversity informatics should address key challenges for the future with respect to the Catalogue of Life.

**Proliferation of digital sources for biodiversity data on the web.** The growing number of (thematic) aggregator sites harvesting from these has resulted in a complex environment for users to locate properly validated up-to-date taxonomic information. The community should consider a more collaborative approach to integrating services and reduce the number of portals and sites.

**Increasing number of unauthoritative sources.** The growing number of web sources with old names, misspellings and misapplied names hampers interoperability and the reliable recombination of information sources. It further confuses the taxonomic realm. There is a pressing need for a "cleaning up" process. The biodiversity community should come together to address this challenge in ways such as those facilitated by recent CoL initiatives.

**Vulnerable funding basis and individual specialists working on a shoestring budget.** An integrated approach to a shared ICT infrastructure to store and treat data, and a more robust funding model for data processing would help address these issues.

**Further automation of data processing in taxonomy to ensure reliable, up-to-date biodiversity information services.** Priority areas to further enhance the usefulness of the CoL for the biodiversity community include filling the gaps in the CoL, such as continuing efforts on fossil taxa. There is also a pressing need for stable identifiers.

---

<sup>68</sup> <http://www.gbif.org/>.

<sup>69</sup> Global Biodiversity Informatics Outlook: Delivering Biodiversity Knowledge in the Information Age, September 2013, [http://www.gbif.org/orc/?doc\\_id=5353](http://www.gbif.org/orc/?doc_id=5353).

<sup>70</sup> <http://eol.org/>.

<sup>71</sup> <http://www.barcodeoflife.org/>.

**Overcoming IPR issues to ensure the full openness of the CoL.** The community as a whole should contemplate a full open data model to enhance usage and stimulate feedback. Moving towards Creative Common licenses underpinning data providing and use is therefore recommended.

**Sustainability of biodiversity information services.** Larger taxonomic facilities might consider adopting data services generated by subsidized projects or grants and integrating them in their core business to ensure continuity. It is important to offer such services as part of the GBIF mission on sharing data.

**Further integration of basic taxonomic services.** Integrating these services into (automated) workflows and research environments of other disciplines has several financial benefits, such as a more cost-effective use of resources, reducing duplication of effort.

At the international level, the Global Biodiversity Informatics Outlook (GBIO)<sup>72</sup> offers a framework for coordinating and planning our informatics activity to address these critical needs. It identifies four major areas requiring global investment and collaboration. First and foremost, changes in culture and infrastructure are required to ensure that all data can be managed in stable, carefully-curated repositories that intelligently support use and analysis. Secondly, all past and future observations and measurements of biodiversity need to be delivered in digital forms that can be managed in this infrastructure. Thirdly, a range of comprehensive discovery and access tools are required to deliver these data in the form that researchers need. Lastly, investment is required to build and optimise models that exploit all available resources and offer the kind of high-level syntheses and indicators required to deliver a rich understanding of biodiversity and its interrelationships with ecosystems and society.

### Biodiversity Challenges and the Aichi Targets

Nearly all the **Aichi Targets**<sup>73</sup> depend on achieving **true interoperability** and **accessibility between biodiversity data sources** in order to further science and deploy the knowledge for policy and management<sup>74</sup>. The science community needs proper maintainable ICT tools to support this interoperability. A large ICT-tool-basis is available as a result of a great number of biodiversity data sharing projects in the EU. Now it is important to further refine and deploy those tools for achieving synergy between the information resources and systems.

Future developments in biodiversity informatics must focus on **completing a validated taxonomic index of all known living organisms**. Such a complete synonymic index is crucial to achieve interoperability between the various multidisciplinary data sources in biodiversity research from taxonomy to ecology, from molecular data to physiology, from conservation to economic value of species. The importance of a unified global register of species based on the opinion in the taxonomic community cannot be stressed enough. Automation for generating and continuously updating such a central taxonomic register from an array of quality sources is a condition for success. A sturdy cross mapping mechanism for continuous checking names in the central register against other sources, and the automatic funnelling of unlisted names via piping tools towards the authoritative resources in the

<sup>72</sup> Global Biodiversity Informatics Outlook: Delivering Biodiversity Knowledge in the Information Age, September 2013, [http://www.gbif.org/orc/?doc\\_id=5353](http://www.gbif.org/orc/?doc_id=5353).

<sup>73</sup> <http://www.cbd.int/sp/targets/>. Aichi Targets, <http://www.cbd.int/sp/targets/>. The EU target for 2020 is: “to halt the loss of biodiversity and ecosystem services in the EU by 2020 and restore them insofar as possible, and step up the EU’s contribution to averting global biodiversity loss”. The EU vision for 2050 is: “Biodiversity and ecosystem services - the world’s natural capital – are preserved, valued and, insofar as possible, restored for their intrinsic value and so that they can continue to support economic prosperity and human well-being as well as avert catastrophic changes linked to biodiversity loss”, <http://ec.europa.eu/environment/nature/biodiversity/policy>.

<sup>74</sup> Position Paper presented by William A. Gray, Species2000.

taxonomic community for processing is a necessity to clean up the millions of misspelled and incorrect names in our resources.

In biodiversity there is a large range of data and software services that are required to conduct any research successfully. Also the data and the services are growing and evolving all the time and it is impossible to define rigid standards for these. It is therefore important that a **federated approach** is used which enables data prepared in diverse ways and services created under different systems to interoperate seamlessly. This can be achieved by storing services and data in a heterogeneous environment where new services and data can be added to the environment in a way that is simple and allows them to evolve. This can be achieved by using an e-science approach where the services are provided through portals that hide the underpinning details of the software and data. It is important in these types of environment that new services can be created that extend the use of the data. For the **Catalogue of Life** this would mean providing the data in an open manner with the basic services provided but extending these services with tools that allow the catalogue to be capturing additional information that increases its coverage of the world species and their alternative names.

Global initiatives such as the Catalogue of life are of great importance but prove hard to fund and sustain. It is important that global initiatives are built upon, but also form part of, regional programs and activities. This promotes a better local embedding, shared responsibility and ownership, and a more flexible funding strategy. A global versus regional hub structure will serve this purpose well.

It is essential to work on deployment of technology that is feasible, affordable and maintainable. Some tools may be open-source other can be commercial. We must recognise an infrastructure of data, and an infrastructure of tools.

## 4 EUBrazilOpenBio Joint Action Plan

Together with its Strategic Advisory Board, EUBrazilOpenBio has analysed the current biodiversity informatics, cloud computing and policy landscape to identify possible future synergies that can help tackle key challenges within the context of Horizon 2020 and beyond. The analysis spans biodiversity strategies, global community efforts on biodiversity such as the Global Biodiversity Informatics Outlook (GBIO-GBIF), the Brazil-EU high-level policy dialogue, including the current Joint Action Plan 2012-2014<sup>75</sup>, and the VI Brazil-EU Summit (January 2013)<sup>76</sup>, the EU Biodiversity Strategy 2020<sup>77</sup> as well as the outcome document of Rio +20 (June 2012)<sup>78</sup>.

The emphasis on sustainable development through the implementation of the green economy is one of the primary objectives of the Rio +20 outcome document, *The Future we Want*. The January 2013 VI Brazil-EU Summit reiterates the importance of sustainable development through continued co-operation between Brazil and Europe, highlighting the role of ICT, research and development in tackling grand global challenges. The Summit Statement also stresses the need to strengthen contacts between business communities, promoting both investments and innovation. In particular, it calls for actions increasing support to micro, small and medium enterprises.

EUBrazilOpenBio and its Strategic Advisory Board propose potential future scenarios to address the global challenges through joint collaboration. These scenarios highlight co-operation opportunities not only for multi-disciplinary research but also for building stronger ties between the business communities, and supporting sustainable development through the green economy. Most of the proposed scenarios require relatively small-scale investments.

Central to the proposed future collaborative scenarios is the ability to harness existing technologies with a view to contributing to a “system of systems”, by adopting a service-oriented approach to addressing the challenges posed by biodiversity loss. Future collaborative work on biodiversity and climate change should therefore take on board the recommendations of the White Paper, *A decadal view of biodiversity informatics: challenges and priorities*, as well as recommendations formulated by other relevant reports cited in the paper<sup>79</sup>.

As more and more scientific disciplines become data hungry, and a rising number of social scientists (economists, e-humanities researchers) follow suit, governments and international bodies are starting to commit to developing frameworks that allow convenient, secure and intelligent access and exchange of data. Open government data is presenting new opportunities to create new value-add services, including the creation of new start-ups to push services to market<sup>80</sup>. For example, current

<sup>75</sup> V European Union-Brazil Summit Statement and Joint Action Plan, pp. 5-28, [http://www.consilium.europa.eu/uedocs/cms\\_Data/docs/pressdata/EN/foraff/124878.pdf](http://www.consilium.europa.eu/uedocs/cms_Data/docs/pressdata/EN/foraff/124878.pdf).

<sup>76</sup> VI Brazil-EU Summit Joint Statement, 24 January 2013, Brasilia, [http://eeas.europa.eu/delegations/brazil/documents/eu\\_brazil/declaration\\_summit\\_january2013\\_en.pdf](http://eeas.europa.eu/delegations/brazil/documents/eu_brazil/declaration_summit_january2013_en.pdf).

<sup>77</sup> European Parliament Resolution (2011/2307(INI)), April 2012 and Our life insurance, our natural capital: an EU Biodiversity Strategy to 2020, (COM(2011)0244),

[http://ec.europa.eu/environment/nature/biodiversity/comm2006/pdf/2020/1\\_EN\\_ACT\\_part1\\_v7%5B1%5D.pdf](http://ec.europa.eu/environment/nature/biodiversity/comm2006/pdf/2020/1_EN_ACT_part1_v7%5B1%5D.pdf)

<sup>78</sup> “The Future we want”, in Report of the United Nations Conference on Sustainable Development (June 2012), <http://www.unctd2012.org/content/documents/814UNCSD%20REPORT%20final%20revs.pdf>. See also the follow-up of Rio+20, <http://sustainabledevelopment.un.org/rio20.html>. On the EU’s recent proposal to tackle poverty eradication and sustainable development, see [http://ec.europa.eu/commission\\_2010-2014/piebalgs/headlines/news/2013/07/20130716\\_en.htm](http://ec.europa.eu/commission_2010-2014/piebalgs/headlines/news/2013/07/20130716_en.htm).

<sup>79</sup> A. Hardisty, D. Roberts, op cit. See also The LERU Biodiversity Working Group: Challenges for biodiversity research in Europe, *Procedia Social and Behavioral Sciences* 2011, 13:83-100, [http://ac.els-cdn.com/S1877042811001777/1-s2.0-S1877042811001777-main.pdf?\\_tid=6dd1c724-08dd-11e3-b46f-00000aab0f6b&acdnat=1376923467\\_f3b7667bf49902d2507f02ad7d8a5732](http://ac.els-cdn.com/S1877042811001777/1-s2.0-S1877042811001777-main.pdf?_tid=6dd1c724-08dd-11e3-b46f-00000aab0f6b&acdnat=1376923467_f3b7667bf49902d2507f02ad7d8a5732).

<sup>80</sup> Brazil’s open government initiative is dados.gov.br. Sixteen EU member states have currently implemented an open government data initiative: Austria, Belgium, Denmark, Estonia, Finland, France, Germany, Greece,

case studies in the health sector are demonstrating cost savings through shared analytics and services<sup>81</sup>. The future collaborative scenarios also centre on the priority of nurturing an open data culture, leveraging new opportunities at every level of society.

### *Research data e-infrastructure supporting multidisciplinary science, analytics and international cooperation*

Biodiversity is inherently connected to fields such as ecology, genetics, epidemiology, climate science, economics, social anthropology and theoretical modelling. Research approaches therefore need to become increasingly multi-disciplinary and trans-boundary.

The biodiversity informatics community has recently made significant step towards realising the Commission's wish that consortia formation becomes more collaborative and less competitive<sup>82</sup>. There is now consensus on a common goal, which is to deliver predictive modelling for biodiversity. The community also recognises that this is a multi-decadal ambition. The first steps will be to conduct a scoping study, equivalent to the design studies with which the astronomers start a new telescope.

Goals for the short term include:

- Clarity of vision with greater focus on end-goals in the bigger picture.
- Good, simple tools with syntactic operability.
- Community identity.
- Links, both networking and links between people, links between machines holding data resources and links with other communities, especially ecology and policy.

An immediate challenge is to question the need to digitise all sources of information and to start asking specific questions about what the end-goals of an activity that requires considerable effort. Answering this question will dictate what further data is needed and can also dictate the data quality required.

The 3 key words are: **integration; co-operation; promotion.**

EUBrazilOpenBio and the synergies it has established at various levels propose the following action lines for future co-operation

**Usability and better deployment.** A user-friendly, user-centric approach is crucial for successful uptake by the wider research community, from biologists to statisticians and ecologists, who have much to contribute to cross-disciplinary science but who typically lack advanced IT skills. The same is true for biodiversity researchers and professionals working at natural history museums and botanical gardens, who have much to gain from the current achievements, for example, EUBrazilOpenBio. Sufficient resources dedicated to marketing service(s), particularly outside the project community and ideally in significant numbers.

---

Ireland, Italy, Netherlands, Portugal, Slovak Republic, Spain, Sweden, and the UK. Norway and Switzerland are the two associate countries with an implemented initiative.

<sup>81</sup> See, for example, use cases incubated by Open Data Institute (ODI, UK), <http://www.theodi.org/>, such as OpenCorporates. With a small staff of just two persons, the company managed to collate data from 51 million companies with the aim of creating an open database for the corporate world (similar to OpenStreetMap) connecting and adding clarity to corporate data. See also the science|business webinar on Big Data with Neelie Kroes, VP of the European Commission, 23 May 2013, <http://www.sciencebusiness.net/events/SmarterData/> and the NIST workshop on Cloud Computing and Big Data, January 2013 with examples of new developments in healthcare data, <http://www.nist.gov/itl/cloud/nist-joint-cloud-and-big-data-workshop-webcast.cfm>.

<sup>82</sup> Biodiversity Informatics Horizons 2013, 3-6 September 2013, Rome, <http://conference.lifewatch.unisalento.it/index.php/EBIC/BIH2013>.

**New knowledge creation and integration of foundational technologies.** Use cases should focus on the creation of new knowledge by synthesis activities using the data and tools thus generated, including open data and data encoding, and developing new structures using existing technologies, such as building a repository for classifications and a single portal for currently accepted names.

Overall, the e-infrastructure landscape needs to be part of a wider ecosystem of cloud-services that help create economies of scale. For example, NREN jointly procured services could enable better links between research and other public institutions to open up new collaboration opportunities.

Short- to mid-term opportunities identified with a view to supporting these future developments include:

- Fostering the development and provision of analytics as a service.
- Supporting an international, inter-university consortium for socio-economic research to help define impact assessment models and related studies of mutual interest to Brazil and Europe.
- Encouraging the involvement of savvy SMEs to support the development of innovative data and science-tailored services around an open and extensible open science cloud stack, including increased access to the cloud through smart devices [61]. The procurement of cloud computing for processing and storage would play a key role in encouraging user uptake, such as adopting an approach similar to the VENUS-C Open Call with free trial usage and technical support. Such an approach was also acknowledged during the EUBrazilOpenBio Public Workshop in September 2013. This workshop also highlighted the need to enlarge the community vision and ambitions by addressing challenges at much larger scale with increased funding to match the level of difficulty involved.
- Providing opportunities for technology transfer, which is high on the Brazil and European ICT agendas. Planning should include the involvement of specialised commercial organisations to facilitate and strengthen enterprise clusters within the cloud ecosystem and encourage a risk-taking culture. A shift of focus towards professional commercial assistance that prepares research results for commercialisation would be an important part of the technology transfer process to help avoid the “valley of death”. This will require changes in the ways academic technology transfer offices typically operate. Use cases could also demonstrate the potential for new uses of existing solutions, for example, cross-mapping tools to address social challenges like criminal records; services that can be scaled out to public authorities (e.g. climate/weather monitoring services) and educational services adapted for smart devices.

A parallel development could be the creation of an **EU-Brazil joint Biodiversity and Business Platform** along the lines of the European model<sup>83</sup>. The platform could promote both commercial offers and relevant exploitable assets from funded initiatives to ensure long-term sustainability, including their commercialisation and internationalisation. Over time, the platform could incorporate products and services in areas such as smart cities, sustainable urban development and sustainable tourism. Such services relate directly to the implementation of the green economy. Commercial services should be based on clearly defined business models and ensure corporate social responsibility.

---

<sup>83</sup> [http://ec.europa.eu/environment/biodiversity/business/index\\_en.html](http://ec.europa.eu/environment/biodiversity/business/index_en.html). EP Resolution, April 2012, Action 47, [http://ec.europa.eu/environment/nature/biodiversity/comm2006/pdf/EP\\_resolution\\_april2012.pdf](http://ec.europa.eu/environment/nature/biodiversity/comm2006/pdf/EP_resolution_april2012.pdf).

### *Supporting Training and eSkills Development*

The experiences of EUBrazilOpenBio have shown that training programmes are crucial to increase the adoption of cloud computing and ICT tools and to ensure that underlying systems are efficient and available in the future. Research e-infrastructures are an important vehicle for achieving results that are vital to creating better knowledge and more informed policy making. Without the availability of high-quality hands-on and remote training facilities, without which it will be hard to foster the uptake of tools and services that have the proven potential to improve biodiversity research. Training is needed not only at the level of the resources but also at the level of the tools, as shown by the EUBrazilOpenBio ecological niche modelling use case. Future directions need to ensure that training on the use case also includes training on niche models as the tools are not useful without the underlying scientific knowledge, which is also a project recommendation to GBIF.

Investments need to respond to these requirements through dedicated educational and training programmes at all levels but individually designed to meet a specific purpose and with clear outputs that can be quantified and certified.

In the field of biodiversity, one possible approach could be to ensure budget allocation for co-located training sessions at community events (e.g. training sessions at events for botanists). In the scientific domain, it is important to foster a “mind set” shift that focuses on real-world examples, with users communicating their progress through a blog or user stories and the involvement of cloud champions sharing best practices and practical advice. Such an approach will enable the shift from definitions and functionalities towards tangible returns on investment.

General educational programmes should focus on the benefits of cloud computing and on real issues and concerns of business leaders, governments and public authorities, such as legal issues, contracts, data protection, trust and security. Educational services also need to communicate the relevance and maturity of standards for interoperability.

The Strategic Advisory Board within EUBrazilOpenBio highlights the following considerations:

- Overcoming skill shortage in specific areas, such as digital research.
- Supporting the recruitment and training of operational system professionals to ensure that underlying systems run as efficiently as possible, now and in the long term.
- Drawing on lessons learnt from current training programmes, proposals should consider the availability, structure of existing training facilities and allocate sufficient budget for promotional campaigns, including tactics to overcome sociological barriers, timeliness and tangible learning outcomes.
- Addressing data sharing issues. Despite increasing recognition of Intellectual Property Rights (IPRs) from different data providers, there is still a need to focus on making, at least partially, data public and sharable to foster better interaction. Commercial usage is not an issue since fostering industry on adopting biodiversity preservation is part of their aims.

### *Supporting large-scale, long-term information processes*

The EU-Brazil Joint Action Plan 2012-2014 highlights the benefits of establishing joint facilities in the field of biodiversity and sustainable tourism. Such facilities could leverage opportunities to explore very large-scale, very long-term information gathering processes in nature. EU-Brazil collaborative projects in the field of biodiversity could combine with others to stimulate an international PhD training college, bringing together the best students from different disciplines, studying under a common programme with common goals, meeting virtually and annually.

The Strategic Advisory Board within EUBrazilOpenBio proposes the following short, mid- and long-term timelines with a view to developing large-scale, long-term information processes:

- **Short term:** enable a team of cross-discipline experts to plan the criteria, curriculum and topics of joint facilities, the scope and a roadmap for their development, including investments needed.
- **Mid-term:** support the planning and delivery of an international European Summer/Brazil Winter School, its objectives, student selection criteria, programme, evaluation criteria, promotion and dissemination of outcomes.
- **Long-term:** establish an international PhD studentship supported by reciprocal recognition in EU and Brazil. The programme committee driving the studentship should ensure that the syllabus design addresses challenges not only linked with biodiversity but also data-intensive science and cloud computing based on a multi-disciplinary approach and that the programme is of sufficiently high standard to serve as a model for similar initiatives elsewhere.

### *Supporting a market for open science and open data services, including opportunities for public-private partnerships: short-term-long-term*

This scenario identifies ways in which the involvement of businesses and public-private partnerships could support open science and open innovation in line with one of the recommendations of the 2013 EU-Brazil Summit is to strengthen ties between the business communities.

One possible way of building closer ties between the Brazilian and European business communities would be to create conditions favourable to generating a data market. For example, new environmental monitoring and healthcare services, which could be cloud-based and provided by new start-ups leveraging the benefits of the cloud. Further, there are increasing opportunities for businesses to contribute to the development of new services, such as for the healthcare sector, smart cities and environmental sciences.

Another opportunity lies in supporting EU-Brazil coordination activities coming together in a single “EU-Brazil Co-operation Hub”, aggregating and promoting products, services and solutions in the private sector and through public-private partnerships. It could offer a directory of EU and Brazilian businesses interested in expanding in the respective regions, career and job opportunities. It could also highlight eSkills shortages that can be addressed through specialised training schemes.

Possible revenue streams resulting from a virtual market for data might include:

- Free of charge usage on a cloud infrastructure for research that is publicly funded.
- A nominal fee for sporadic usage by other research communities.
- Low entry-fee for a start-up using the data generated by a research group; charges to large pharmaceutical and other large enterprises that makes intense use of the data for research and development.

Pathways towards these new approaches could include a small-scale study providing a data-roadmap that defines what is the most valuable data and why, e.g. “core reference data” that is useful many times. It should propose areas in which value creation around data can help unleash the innovation potential and propose ways in which funding might be channelled towards mentoring data-driven start-ups, as well as for training researchers and journalists in data literacy.

### *Coordination activities supporting advances at the network layer and opportunities for public-private partnerships*

Researchers rely on a highly pervasive affordable network, where cloud services are increasingly helping to create a key enabling platform. Without being able to connect researchers around the world “at the speed of light” we cannot take on the grand global challenges of the 21st century.

Leading NRENs are currently involved in advancing the network for tomorrow’s research community, addressing interoperability challenges and defining new business and governance models to gain better efficiencies. However, the pace of technological advances and provision of cloud-based services by NRENs in Europe is diverse and fragmented. While there are examples of well-defined strategies and deployed services, there is a serious risk of fragmentation of services and unclear business models, among other difficulties identified. Similar issues are being faced across Latin America.

It is important to share experiences and new knowledge emerging from pioneering NRENs so that best practices can be adapted to specific circumstances. For example, supporting data-intensive science disciplines, overcoming cross-border interoperable challenges, and launching public-private partnerships with the commercial sector (equipment vendors and carriers).

Joint coordination activities are not only important to increase collaboration between the EU and Brazil but also to ensure that pioneering developments benefit the wider, global open science community. The role of the Brazilian NREN and EU NRENs in addressing major global challenges and providing user-centric services presents new opportunities to build closer links between the R&E sector and other public institutions (e.g. libraries, museums and hospitals). Future joint co-ordination activities through small-scale funding schemes could play a role in promoting leadership and best practices around service delivery, standards and interoperability. It could help identify potential public-private partnerships for the integration of new services, such as for healthcare centres, libraries and museums. Other outputs could include the evaluation of business models and portfolio of services, identifying best practices and new services that could be replicated elsewhere or adapted to suit local or national circumstances.

One possible approach for a future support and coordination action could be to build on current analyses on the adoption of new technologies and business modelling, helping to identify areas for the involvement of both public and private organisations. Closer co-operation with national resource providers (e.g. National Grid Initiatives or NGIs for short) should also be encouraged to help drive federated research infrastructures and ensure the sustained use of digital resources.

Ideally it would provide a roadmap for short-term actions to create an ecosystem of services from NRENs, including joint procurement leading to economies of scale and preferential rates for R&E sector, such as contracts with telecom providers. Actions should focus on ensuring that services are “innovation enablers”, combining in-house solutions to meet specific needs while procuring existing tools and services that researchers find useful to create an ecosystem that effectively supports the digital researcher of 2020.

Such an approach should also focus on success stories and insights for policy makers and the wider public along the lines of the eScienceTalk project and its work with iSGTW. This would also play a key role in helping to change mind-sets as one of the main barriers to driving change, identifying and recruiting champions of cloud and digital tools, new business models and new organisational structures suited to today’s service-oriented approach.