

# Enabling Open Science: Wikidata for Research (Wiki4R)

## EINFRA-9-2015: e-Infrastructures for virtual research environments (VRE)

### Sections 4 to 6

#### Table of Contents

4. Members of the consortium.....	2
4.1. Participants (applicants) .....	2
Partner 1 (Coordination): Museum für Naturkunde (MfN), Germany .....	2
Partner 2: Universidad Politécnica de Madrid (UPM), Spain .....	4
Partner 3: Maastricht University (UM), Netherlands .....	8
Partner 4: Wikimedia Deutschland (WMDE), Germany .....	9
Partner 5: Universitat Oberta de Catalunya (UOC), Spain.....	11
Partner 6: Europeana Foundation (EF), Netherlands .....	13
Partner 7: Université Paris Sud (UPS), France .....	15
4.2. Third parties involved in the project (including use of third party resources) .....	16
4.3 Associate Partners .....	17
5. Ethics and Security .....	19
5.1 Ethics .....	19
5.2 Security .....	19
6. Letters of Support .....	19

## 4. Members of the consortium

### 4.1. Participants (applicants)

#### **Partner 1 (Coordination): Museum für Naturkunde (MfN), Germany**

---

The Museum für Naturkunde Berlin (MfN) – Leibniz Institute for Evolution and Biodiversity Science is a research museum within the Leibniz Association and constituted as a foundation under public law. It is one of the most significant research museums worldwide focusing on biodiversity, evolution and geo-sciences with over 250 staff members. Research at the Museum für Naturkunde is organized in four Science Programs: Evolution and Geoprocesses, Collection Development and Biodiversity Discovery, Digital World and Information Science, and Public Engagement with Science. The collections of the MfN are directly linked to research and comprise more than 30 million specimens relating to zoology, palaeontology, geology and mineralogy. In addition, the MfN houses a unique Animal Sound Archive containing approximately 120,000 animal sound recordings. The library of the Museum für Naturkunde Berlin is one of the most important reference libraries in zoology in the German-speaking world.

#### **Relevant experience and role in project**

MfN has been participating and taking a leading role in several EU-projects such as EU BON, SYNTHESYS III (I3),<sup>1</sup> the European Distributed Institute of Taxonomy (EDIT), European Biodiversity Heritage Library (BHL), OpenUp!, 4D4Life, ViBRANT, pro iBiosphere and Europeana Creative, as well as other international collaborative research programmes (e.g. BIOTA). It hosts the secretariat of the European Citizen Science Association (ECSA) and has extensive experience in citizen science. It is also a member of the Science 2.0 research network of the Leibniz Society. MfN coordinates the Wikidata for research project and provides for professional administration, management of the consortium, as well as targeted implementation of the project goals (WP1). It will also take a leading role in the assessment of the benefits of and barriers to openness (WP3) as well as the tasks related to citizen science and cultural heritage (WP4) as well as in dissemination, communication (WP5).

#### **Profile of the personnel**

**Dr. Gregor Hagedorn (m)** is the head of the research division “Digital World and Information Science” at the MfN. He has extensive experience in developing data standards (three Biodiversity Information Standards/TDWG.org standards), descriptive and trait data, computer aided identification and citizen involvement (German Open Nature Guides, Artenquiz). Current projects are “German Federation for the Curation of Biological Data” (GFBio), pro-iBiosphere (EU Project) and Europeana Creative. He is a member of the CETAF and GBIF-Germany IT-commissions.

**Dr. Daniel Mietchen (m)** is a biophysicist specialized in non-invasive imaging, with complementary research interests ranging from evolutionary aspects of vocal production to semantic integration of biodiversity literature. He has broad experience in disseminating scientific information within and beyond the scientific community through open licenses, e.g. by way of data publishing or reuse in educational contexts like Wikimedia platforms. From 2011-2013, he served as Wikimedian in Residence on Open Science at the Open Knowledge Foundation Germany. As a volunteer, he leads a team operating a software that collects openly licensed audio and video materials from the biomedical literature repository PubMed Central and uploads them to Wikimedia Commons for reuse on Wikimedia platforms. The team was one of the inaugural recipients of the Accelerating

---

<sup>1</sup> <http://www.synthesys.info/>

Science Awards,<sup>2</sup> and that work has recently been expanded to a prototype for full-text upload to Wikisource.<sup>3</sup>

**Falko Glöckler's (m)** expertise is biodiversity informatics, especially data workflows, databases, web developments and biodiversity data standards. He is in charge of data integration in several digitization projects and standardization and mobilization of collection and field data for providing these to international aggregation platforms (GBIF, Europeana). His specialities are data quality management and compliance algorithms for biodiversity data.

**Dr. Hatem Mousselly Sergieh (m)** obtained 2014 a double PhD degree in computer science from the University of Passau – Germany and INSA Lyon – France. He has expertise in data mining, ontology matching and information retrieval and works as data scientist.

**Ralf Thomas Schmitt (m)** is curator for minerals and rocks at the Museum für Naturkunde Berlin since 1994. He is responsible for digitization and the continuous development of the mineralogical database. As part of a SYNTHESYS networking activity, he was involved in the development of the ABCDEFG (Access to Biological Collection Databases Extended for Geosciences) XML-Schema for standardization of mineralogical, geological and palaeontological digitized collection data. This scheme is successfully used for the online publication of geoscientific collection data from the Museum für Naturkunde Berlin via the GEOCASE portal, which is combining collection data of several geoscientific institutions from Europe.

## Publications

1. **Hagedorn G, Mietchen D**, Morris R, Agosti D, Penev L, Berendsohn W, Hobern D 2011. "Creative Commons licenses and the non-commercial condition: Implications for the re-use of biodiversity information". *ZooKeys* 150 (150): 127–149. DOI: 10.3897/zookeys.150.2189.
2. Egloff W, Patterson D, Agosti D, **Hagedorn G** 2014. Open exchange of scientific knowledge and European copyright: The case of biodiversity information. *ZooKeys* 414: 109135. DOI: 10.3897/zookeys.414.7717.
3. **Mietchen D** 2014. "The Transformative Nature of Transparency in Research Funding". *PLoS Biology* 12 (12): e1002027. DOI: 10.1371/journal.pbio.1002027.
4. Penev L, **Hagedorn G, Mietchen D**, Georgiev T, Stoev P, Sautter G, Agosti D, Plank A, Balke M 2011. "Interlinking journal and wiki publications through joint citation: Working examples from ZooKeys and Plazi on Species-ID". *ZooKeys* 90. DOI: 10.3897/zookeys.90.1369.
5. **Glöckler F, Hoffmann J**, Theeten F 2013. The BioCASE Monitor Service - A tool for monitoring progress and quality of data provision through distributed data networks. *Biodiversity Data Journal*, 1: e968. DOI: 10.3897/BDJ.1.e968.

## Relevant previous projects or activities

1. EU BON (FP7), Building the European Biodiversity Observation Network<sup>4</sup>
2. Open-Up (EU ICT-PSP), Opening up the Natural History Heritage for Europeana<sup>5</sup>
3. GBIF D, Global Biodiversity Information Facility – Germany<sup>6</sup>
4. EU FP7 4D4Life, Distributed Dynamic Diversity Databases for Life<sup>7</sup>
5. ECSA - European Citizen Science Association<sup>8</sup>

---

<sup>2</sup> <http://blogs.plos.org/plos/2013/10/announcing-the-recipients-for-the-accelerating-science-award-program/>

<sup>3</sup> [https://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Open\\_Access/Signalling\\_OA-ness](https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Open_Access/Signalling_OA-ness)

<sup>4</sup> <http://www.eubon.eu/>

<sup>5</sup> <http://open-up.eu/>

<sup>6</sup> <http://www.gbif.de/entomologie>

<sup>7</sup> <http://www.4d4life.eu/>

<sup>8</sup> <http://ecsa.biodiv.naturkundemuseum-berlin.de>

## **Partner 2: Universidad Politécnica de Madrid (UPM), Spain**

---

The **Universidad Politécnica de Madrid (UPM)**<sup>9</sup> is the largest Spanish technological university. With two recognitions as Campus of International Excellence, it is outstanding in its research activity together with its training of highly-qualified professionals, competitive at an international level. More than 2,400 researchers carry out their activity at the UPM, grouped in 216 Research Groups, 10 Research Centers and 55 Laboratories, all of them committed to transforming the knowledge generated into advances applied to the production sector.

The intense collaboration with governmental bodies and industry guarantees that research at the UPM offers real solutions to real-world problems. The dynamism of research and development/innovation activity at UPM, together with the transfer of knowledge to society, is among its lines of strategy. These two commitments place it among the Spanish universities with the greatest research activity and first in the capture of external resources in a competitive regime. UPM heads the Spanish Universities' participation in the 7<sup>th</sup> European Framework Program with more than 280 projects and more than €80M funding. Moreover, every year, UPM applies for around 40 patents and receives a similar number of concessions demonstrating a high commitment to innovation. UPM is leader in business creation, having generated around 140 businesses. Its support and backing of the business sector is very close and it annually signs around 600 contracts with private businesses.

UPM is an institution **committed to the transfer of knowledge generated through its research structures to society, and its transformation into advances and technological developments applied to the productive sector.**

### **Relevant experience and role in project**

The **Ontology Engineering Group (OEG)**, led by Prof. Dr. Gómez-Pérez, has been working on the provision of **ontologies, semantic infrastructures, multilingualism, sensor data and linked data** since 1995. It is composed of 30 researchers with a consolidated reputation in the fields of: Ontological Engineering, Data Integration, Linked Open Data, Semantic Web, NLP and Semantic e-Science. The OEG has coordinated seven European projects. Members of the OEG participated in more than 15 EU IST projects. OEG members participate on more than 15 working groups at the Consortium of the World Wide Web. The OEG has participated in 19 research and technology transfer contracts with Spanish and International companies and organizations (e.g. AENOR, OMS, FAO, Fujitsu, Telefonica, Indra, Atos Origin, National Spanish Library) and has collaborated with relevant international research groups in prestigious universities worldwide. The OEG hosts the Spanish chapter of DBpedia.<sup>10</sup>

The group's contributions to the project will leverage its expertise in Semantic Web and Linked Data technologies and focus on the property profiles and semantic mapping (WP2), multilingual matters as well as quality assurance (WP3) and providing Wikidata as Linked Open Data (WP4).

### **Profile of the personnel**

**Prof. Dr. Asunción Gómez-Pérez (f)** is Full Professor at UPM, director of the Artificial Intelligence department, director of the OEG and PhD in Computer Science (1993). Before joining UPM, she was visiting (1994-1995) the Knowledge Systems Laboratory at Stanford University. She also was the Executive Director (1995-1998) of the AI Laboratory at the School of Computer Science. She has coordinated SEALS, SemSorGrid4Env and Ontogrid and now she is coordinating LIDER. She has participated in more than 15 EU (Admire, Esperonto, etc.) and Spanish R&D projects (CENITS mIO!, España Virtual, Buscamedia, myBigData, GeoBuddies). Her main research interests are **ontologies, Linked Data, Multilingual Linked Data and the Semantic Web**. She has published more than 150 papers and two books on Ontological Engineering. Her works on Ontological Engineering about Methontology and the NeON methodology are known worldwide. Her works on

---

<sup>9</sup> <http://www.upm.es/internacional>

<sup>10</sup> <http://es.dbpedia.org/>

Ontological Engineering about Methontology and the NeON methodology are worldwide known.<sup>11</sup> She is primarily responsible for UPM task.

**Dr. Jorge Gracia del Rio (m)** is a postdoctoral researcher at the Artificial Intelligence Department at Universidad Politécnica de Madrid (UPM). He obtained his degree in Physical Science and his PhD in Computer Science at University of Zaragoza. His research experience has focused on **semantic measures, ontology matching, and disambiguation techniques** in the field of the Semantic Web. He has been research visitor at Knowledge Media Institute (Open University, United Kingdom) and INRIA (France). He has worked in several European projects: Dynalearn, Monnet, and currently in LIDER (where he develops the role of Quality Assurance Coordinator). His current research topic is **multilingualism on the Web of Data**. He co-chairs the W3C Best Practises on Multilingual Linked Open Data (BPMLOD) community group, and collaborates actively in the W3C Ontology Lexica (ONTOLEX) and Linked Data for Languages Technologies (LD4LT) community groups. You can find out more about Jorge Gracia on his website.<sup>12</sup> He is primarily responsible for carrying out the proposed research related to multilinguality.

**Dr. Mariano Rico (m)** is a physicist (master degree) and computer scientist (PhD) with 10 years of experience in private IT companies and 10 years of academic experience. He is member of the DBpedia Internationalization committee and the person responsible for the Spanish chapter of DBpedia. He has organized three local workshops to disseminate the technologies related to the Semantic Web and Linked data in the context of the Spanish DBpedia. He has been researching in collaborative (wiki) technologies, virtual environments and the relation between linked data and multilingual natural language processing. He is primarily responsible for carrying out the proposed tools related to linked data generation.

**Ms.C. Nandana Mihindukulasooriya (m)** is a Software Engineer who is working for the Center for Open Middleware and the Ontology Engineering Group, Universidad Politécnica de Madrid. He is a member of the Ontology Engineering Group since 2011 and he has participated in the SEALS FP7 project and in the ALM iStack project. His main interests are Linked Data, RESTful design, application integration and transaction processing. He is a member of W3C Linked Data Platform (LDP) WG and he is the editor of the LDP Primer. He was awarded a prestigious two-year Erasmus Mundus scholarship from the European Union during his double-degree M.Sc. from BTH (Sweden) and UPM (Spain). Nandana successfully completed three Google Summer of Code (GSoC) projects in 2006, 2007, and 2010. Previously, he worked as a Technical Lead at WSO2 Inc. where he played a leading role in the design and implementation of WSO2 Web Services and Cloud middleware platforms for which he was awarded the “Outstanding Contributor of the Year Award” in 2008. He was also a member of Web Services Secure Exchange (WS-SX) and Web Services Federation (WS-FED) technical committees in OASIS during that time. Nandana is an active contributor to open source and currently serves as a committer, Project Management Committee member and a mentor for several Web Services and Semantic Web related projects in Apache Software Foundation and was a speaker at ApacheCon USA 2009 and ApacheCon Europe 2012. He is primarily responsible for carrying out the proposed tools related to LDP.

## Publications

1. **Gómez-Pérez A., Fernández-López M**, Corcho O (2003) Ontological Engineering. Springer-Verlag. November 2003.
2. **Fernández-López M**, Gómez-Pérez A, Suárez-Figueroa MC (2013) Methodological guidelines for reusing general ontologies. Data & Knowledge Engineering. Editorial Elsevier. ISSN: 0169-023X. July 2013.
3. Gracia J, Montiel-Ponsoda E, Cimiano P, **Gómez-Pérez A**, Buitelaar P, McCrae J (2012) Challenges for the Multilingual Web of Data. Journal of Web Semantics, 11, pp. 63-71. ISSN 1570-8268

---

<sup>11</sup> <http://www.dia.fi.upm.es/?q=en/node/140>

<sup>12</sup> <http://delicias.dia.fi.upm.es/%7Ejgracia>

4. Gracia J, Mena E (2011) Semantic Heterogeneity Issues on the Web. IEEE Internet Computing, ISSN 1089-7801, DOI 10.1109/MIC.2011.129, volume 16, number 5, pp. 60-67, September-October 2012. Available online at IEEE Xplore digital library.
5. **Rico M, Gómez-Perez A** (ongoing) The Spanish chapter of the DBpedia. Text version: <http://es.dbpedia.org>. SPARQL endpoint: <http://es.dbpedia.org/sparql>. Dataset information: <http://datahub.io/dataset/dbpedia-es>

### Relevant previous projects or activities

1. <http://es.dbpedia.org/> The OEG-UPM serves world-wide the Spanish Chapter of DBpedia, which is relevant for the project. This server stores 200 million RDF triples that received 22 million SPARQL queries in year 2013, and the web site had around 6000 different visitors (8000 visits), 20% with non-Spanish browsers.
2. Members of the OEG have participated at the **W3C Linked Data Platform Working Group and have developed LDP4j** (<http://www.ldp4j.org/>). LDP aims at describing a set of best practices and simple approach for a read-write Linked Data architecture, based on HTTP access to web resources that describe their state using the RDF data model. OEG members are active in the working group and contribute as editors of the LDP Test Cases, LDP Primer, and LDP Best Practices and Guidelines.
3. **Datos.bne.es** (<http://datos.bne.es>): This linked data project started by the end of 2011 has already published 10 million bibliographic and authority records as Linked Open Data under a Public Domain license (Creative Commons CC0), generating more than 60 million triples and more that 1 million links. Furthermore, it provides a portal provides full-text, entity-oriented and faceted search and navigation for end-users using the latest web technologies and with a focus on user experience.
4. **LIDER** (Project coordinated by UPM – A. Gómez-Pérez - *Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe* FP7-610782, support action). LIDER's goal is to define best practices and a reference architecture, develop a roadmap, as well to create a community that can support the release of **language resources as Linguistic Linked Open Data**, to foster the discoverability of such datasets as well as their exploitation by linked-data-aware natural language processing and content analytics services. LIDER is seeking to network with projects in the area of Natural Language Processing, Linked Data and Big Data Analytics. Years: 2013-2015, Value: €693,775
5. The previous most connected project to the subject of this proposal in which UPM was involved is the **NeOn project** (Lifecycle support for networked ontologies - FP6-027595) aimed to advance the state of the art in **Ontology Engineering** and Semantic Web technologies. The main goal was to provide effective methodological and tool support for developing a new breed of semantic applications, able to exploit effectively the large amounts of information and data, which are now available on the web.<sup>13</sup> Years, 2006-2010, Value: €1,135,822.
6. **Monnet** (Multilingual Ontologies for Networked Knowledge FP7-248458). This project provided a **semantics-based solution for integrated information access across language barriers**. As outcome, Monnet developed semi-automatic approaches and methodologies allowing cost effective ontology localization as a basis for implementing an integrated solution to providing semantic-level access to information across languages. One of the results was the lemon model for modelling lexicon and machine-readable dictionaries, linked to the Semantic Web and the Linked Data cloud. Years: 2010-2013, Value: €362,456.

### Significant infrastructure

Thanks to the extensive trajectory in developing and managing R&D projects, the Ontology Engineering Group may provide a wide set of tools and methods that may benefit the project from a scientific and technical point of view.

---

<sup>13</sup> <http://www.neon-project.org/>

- A system administrators' team supports the project technical team in acquisition, configuration and installation of required hardware equipment, IT services or software licenses.
- A professional project management office supports project managers in the financial reporting and auditing processes that take place during the project.
- A set of collaborative tools is at the disposal of project team for communication (online meeting tool, mailing lists manager), for management (budget/payments monitoring spreadsheets, project plan and quality control methods), for IPR management (IPR control spreadsheet) or for software development (software configuration management, change management, continuous integration and delivery, quality control).
- In-house virtualized servers hosting software that support a modern software development lifecycle (SDLC), including but not limited to:
  - Software configuration management (Git, Mercurial, and Subversion)
  - Software project management and collaboration system (Redmine)
  - Software delivery (Nexus)
  - Various state-of-the-art tools and the related expertise for implementing automated testing (unit, integration, functional, stress/stability/scalability)
  - VMs for hosting complex development and testing environments
- Hosted (SaaS) services related to the SDLC, including but not limited to:
  - Software configuration management and related collaboration services (simple wiki, simple issue tracking, social collaboration features) like GitHub and BitBucket
  - Modern software development collaboration environment based on Jira and Confluence
  - Virtual machines used for testing software systems
- Software libraries and modules that can provide the infrastructure for implementing more complex software artefacts and applications for the proposed project.
- A technical library of paper and electronic books related to technical areas relevant to the proposed work.

The OEG infrastructure in a nutshell:

- 19 servers and 8 PCs acting as a server.
- SEALS infrastructure: 7 servers and 1 SAN (storage). 3 servers for virtualization using VMware vSphere 4.1, and 4 servers with different roles (1 frontend server and 3 backend servers). Roles: DHCP, DNS, RRAS, DFS, fibre channel nodes, terminal services, balancer, preproduction/beta test/final services.
- Linked Data infrastructure: 4 servers. SPARQL endpoints with Virtuoso, pubby and some CMS (Joomla, Drupal).
- OEG dedicated: 7 servers with internal/external services (email with Postfix, Mediawikis, Jira, Bamboo, FishEye & Crucible, dedicated virtualization server, FTP, databases and file server. Other server/PCs are fully dedicated to other projects or initiatives.
- We received in the first semester of 2015 **two high performance computers to enhance the Spanish chapter of DBpedia**, These machines and a high reliability data storage system will be hosted in a controlled technical room to allow 7x24 services to the scientific community and to perform high-power computations.

### **Partner 3: Maastricht University (UM), Netherlands**

---

Maastricht University is the youngest Dutch university ranking in the world top 20 of universities under 50 years. The Department of Bioinformatics, BiGCaT, is a 12 year old group using bioinformatics for data integration of omics using statistical approaches and pathway analysis using pathway databases.

#### **Relevant experience and role in project**

UM develops systems biology solutions to biological questions. Central roles here are set aside for pathway databases and semantic technologies to link experimental omics data. The group has extensive experience with setting up international, multi-disciplinary data platforms, including ISATab, ToxBank, DiXA, Open PHACTS, and eNanoMapper. Experimental data is linked up into systems biology approaches using Open databases software including WikiPathways, PathVisio, BridgeDb, the Chemistry Development Kit, Bioclipse, and others. UM will focus on correctly capturing the chemical aspects of the profiles (WP2), generally contributing Semantic Web expertise, support the incorporation of two CC0 data sets (WP3), and coordination of the training events (WP5), in addition to other dissemination and communication activities (WP5).

#### **Profile of the personnel**

**Dr Egon Willighagen** (m) is a senior post-doc with a PhD degree in chemometrics on the computer representation and statistical analysis of chemical entities, with experience at various European institutes including the Karolinska Institutet and Cambridge University. He leads the cheminformatics toolkit the Chemistry Development Kit (CDK) and has been Invited Expert to the W3C's Health Care and Life Sciences interest group on using semantic web technologies to the life sciences. He has worked on the FP7 projects ToxBank and Open PHACTS, has been advisory board member of the compound database ChemSpider and the FP7 project OpenTox. Most relevant research to this project is the nanoQSAR platform he developed including a wiki-based knowledgebase and matching ontology, containing more than 300 ENMs and a CDK-based computational library for nanoQSAR descriptor calculation and statistical modelling building. He also collaborates with the NIH/NCBI on a semantic version of PubChem.

**Prof Dr Chris Evelo** (m) is the head of the Department of Bioinformatics – BiGCaT, which he founded in 2001. He was trained in Toxicology and Bioinformatics and has over 25 years of research experience. He is renowned for collaborative, open source development of bioinformatics and integrative systems biology approaches (e.g. WikiPathways, PathVisio and in Open PHACTS) and development of community standards and practical applications of those in biology. Dr Evelo coordinates the FP7 IRSES project Microgennet, he is a steering committee member of the IMI project Open PHACTS, he is a council member and coordinator for bioinformatics and systems biology activities in the NuGO foundation for nutrigenomics and its predecessor the FP6 project NuGO and the current NuGO, he is a work package leader in the FP7 systems biology programme for toxicogenomics DiXA, PI in the Dutch Consortium for Systems Biology (NCSB) and a participant in the ESFRI project EuroDish, he is an NBIC faculty and think tank and DISC coordination team member and he is a recipient of an Agilent Thought Leader Grant.

#### **Publications**

1. The Chemistry Development Kit (CDK): an open-source Java library for Chemo- and Bioinformatics. Steinbeck et al. Journal of chemical information and computer sciences 2003 43 (2), 493-500.
2. The Blue Obelisk - Interoperability in chemical informatics, Guha et al. In Journal of Chemical Information and Modeling 46 (3) 991-8
3. WikiPathways: pathway editing for the people. Pico AR et al. PLoS Biol. 2008 6(7): e184.
4. Presenting and exploring biological pathways with PathVisio. Van Iersel MP et al. BMC bioinformatics 2008, 9(1), 399.
5. Open PHACTS: Semantic interoperability for drug discovery. Williams AJ et al. Drug Discovery Today 2012, 71: 21/22. 1188-98 Nov.



6. Scientific lenses to support multiple views over linked chemistry data. Batchelor C et al. The Semantic Web – ISWC 2014. Vol. 8796 of Lecture Notes in Computer Science. Springer International Publishing, pp. 98-113.

### **Relevant previous projects or activities**

1. eNanoMapper<sup>14</sup> - EC project that services a cluster of more than 30 FP7 and H2020 projects in the area of safety of nanomaterials. It develops an ontology and database platform to support the ongoing academic and industrial research and development.
2. Open PHACTS<sup>15</sup> - IMI-funded project to develop a Semantic Web platform to support research in the pharmaceutical industry.

### **Partner 4: Wikimedia Deutschland (WMDE), Germany**

---

Wikimedia Deutschland – Gesellschaft zur Förderung Freien Wissens – is a charitable non-profit organisation under German Law. Founded in 2004, it is the first and largest chapter of the Wikimedia Movement, as strong partner of the US-based Wikimedia Foundation and a leader in the free knowledge movement with a membership of 22000+, 70 employees and an annual budget around €5M. WMDE supports Wikimedia projects such as the German-speaking Wikipedia, Wikimedia Commons, Wikisource and others.

### **Relevant experience and role in the project**

In 2012, WMDE began developing Wikidata, a free, collaborative, multilingual, secondary database, collecting structured data to provide support for Wikipedia, Wikimedia Commons, and other Wikimedia projects. Wikidata has quickly evolved into a structured data repository with many uses beyond the Wikimedia projects.

Since 2012, Wikidata has quickly gathered a community of more than 1.5 million editors who have so far contributed about 190 million edits to Wikidata, resulting in the creation of about 13 million data items. In addition, the potential of beyond-Wikimedia uses has attracted the attention of many individuals, groups and stakeholders, including the research and science communities. The “Gene Wiki” and the Wikidata community recently announced<sup>16</sup> that every human gene (according to the United States National Center for Biotechnology Information) now has a representative entity on Wikidata. Next steps to support life sciences data in Wikidata are to establish bots that populate Wikidata with entities representative of two other key classes: diseases and drugs; and to expand the scope of these bots to include the addition of statements that link these entities together into a valuable network of knowledge.

In November of 2014, Wikidata received the Open Data Institute’s Open Data Publisher Award<sup>17</sup> - celebrating “high publishing standards and use of challenging data“. In December of 2014, Google announced<sup>18</sup> that all the data in Freebase will be migrated to Wikidata and that Freebase will be wound down: “We believe strongly in a robust community-driven effort to collect and curate structured knowledge about the world, but we now think we can serve that goal best by supporting Wikidata -- they’re growing fast, have an active community, and are better-suited to lead an open collaborative knowledge base.” With this, Wikidata has become the premier structured data base project, uniquely positioned to serve as a repository for research data as well.

Wikimedia Deutschland will be the leader for WP4, Enabling the use of Wikidata in research contexts. WMDE will support the project in developing a SPARQL endpoint to enable easy access

---

<sup>14</sup> <http://enanomapper.net/>

<sup>15</sup> <http://www.openphacts.org/>

<sup>16</sup> <http://blog.wikimedia.de/2014/10/22/establishing-wikidata-as-the-central-hub-for-linked-open-life-science-data/>

<sup>17</sup> <http://theodi.org/news/first-odi-open-data-awards-presented-by-sir-tim-berners-lee-and-sir-nigel-shadbolt>

<sup>18</sup> <https://plus.google.com/109936836907132434202/posts/bu3z2wVqcQc>

to the data in Wikidata for researchers and make Wikidata follow established standard procedures of the semantic web.

### Profile of the personnel

**Lydia Pintscher (f)** is the product manager for Wikidata at Wikimedia Germany. She has extensive experience in Open Source and Open Content projects. She studied computer science at the Karlsruhe Institute of Technology with a focus on open collaboration, language and medicine. After that she joined Wikimedia Germany to manage the community side of the development of Wikidata and later moved on to doing product management for Wikidata. She is a regular contributor to many Open Source projects like KDE, where she is the current president of KDE e.V., the non-profit supporting the KDE community.

**Daniel Kinzler (m)** works as a software developer and system architect of the Wikidata project at Wikimedia Germany. Daniel wrote his diploma thesis in informatics about extracting a multilingual thesaurus from Wikipedia, and has been involved in several projects related to semantic integration and mining of semi-structured data since. During his employment at Wikimedia Deutschland since 2008, Daniel has had several roles ranging from software development to serving in the MediaWiki architecture committee.

**Tobias Gritschacher (m)** is part of the Wikidata project since it started in April 2012. His responsibilities ranged from software development to browser testing and test automation. Currently he is the SCRUM master for the Wikidata team and other software development teams at Wikimedia Deutschland. He studied Software Development and Business Management at Graz University of Technology. He has previously worked at crowd-sourcing and crowd-funding platforms Neurovation.at and [1000x1000.at](http://1000x1000.at) as well as within the Open Source project Catrobat.org which provides free open source educational apps for children and teenagers.

### Publications

1. Erxleben F, Günther M, Krötzsch M, Mendez J, Vrandečić D 2014. Introducing Wikidata to the Linked Data Web. In Proceedings of the 13th International Semantic Web Conference. Springer.<sup>19</sup>
2. Krötzsch M. 2014. Hot to use Wikidata: Things to make and do with 30 Million Statements - Wikimania 2014<sup>20</sup>
3. Wikibase software<sup>21</sup>

### Relevant previous projects or activities

1. EU FP7 RENDER<sup>22</sup> - Reflecting Knowledge Diversity
2. First phase of Wikidata development (funded by Google, AI<sup>2</sup> and the Gordon and Betty Moore Foundation)
3. Second Phase of Wikidata development (funded by Yandex and Wikimedia Deutschland)
4. Wikidata meets Archeology
5. Upload of images of the Bundesarchiv including mapping of data

### Significant infrastructure

Wikimedia Deutschland's offices are located in Berlin, Germany and house 70 employees, three program departments, an executive office, finance department, communications, fundraising and evaluation departments. The organization features all necessary technical and physical infrastructure, as well as programmatic and financial management capacity to effectively participate in the Wiki4R consortium.

---

<sup>19</sup> <http://korrekt.org/papers/Wikidata-RDF-export-2014.pdf>

<sup>20</sup>

[http://wikimania2014.wikimedia.org/wiki/Submissions/How to use Wikidata: Things to make and do with 30 million statements](http://wikimania2014.wikimedia.org/wiki/Submissions/How_to_use_Wikidata:_Things_to_make_and_do_with_30_million_statements)

<sup>21</sup> <http://wikiba.se>

<sup>22</sup> <http://render-project.eu>

The technical infrastructure of Wikidata is provided by the Wikimedia Foundation. This includes hosting of the website itself, bug tracking, source code repositories and more. Office infrastructure for the Wikidata development team is provided by Wikimedia Germany. This includes computers, administration and facilities.

### ***Partner 5: Universitat Oberta de Catalunya (UOC), Spain***

---

The Universitat Oberta de Catalunya (UOC, Open University of Catalonia) is a state-of-the-art technological university with a highly innovative learning model, providing a benchmark for quality in both teaching and R&D. It was created in 1994 as one of the world's very first completely online higher education establishments and currently has more than 50,000 students. The UOC's core goal is to be the university of the knowledge society, promoting innovative education, personalised learning, technological leadership, R&D work on the information society and e-learning and the dissemination of knowledge. The UOC promotes R&D activities via 45 groups linked to a department or to one of the university's research centres: the eLearn Center, devoted to e-learning studies, and the Internet Interdisciplinary Institute (IN3), specialising in the study of the networked society and the knowledge economy, network technologies and specific software development. In total, more than 400 people work in R&D at the UOC.

#### **Relevant experience and role in project**

Over the last five years, the UOC has participated in more than 260 R&D projects, either national or European. What is more, the UOC works to promote knowledge transfer and has, over the last four years, signed more than 1000 agreements to this end. The UOC forms part of more than 30 international networks, including the European University Association (EUA), the International Council for Open and Distance Education (ICDE) and the IMS Global Learning Consortium. The research group involved in this proposal, Open Science and Innovation, belongs to the above-mentioned IN3, and has recently carried out a research project on the use of Wikipedia at universities. At present the group develops another research project on Science & Wikipedia which aims to analyse the scientific contents of Wikipedia and to promote the active contribution to Wikipedia by scientists and researchers. UOC's role in the project will be to develop and organize the training activities aimed at fostering the use of Wikidata for research among different potential users (WP5). This will mainly encompass the design and development of tutorials and course materials able to be used both in e-learning environments and in traditional educational settings, and the organization of training events for different audiences.

#### **Profile of the personnel**

**Eduard Aibar (m)** is an associate professor at the Department of Arts and Humanities and a researcher in the Internet Interdisciplinary Institute, both at the Universitat Oberta de Catalunya (Barcelona, Spain). He teaches Science and Technology Studies and has published several works on the interaction between technological innovations and social change in different arenas. He leads the research group on Open Science and Innovation and has recently finished a research project on the use of Wikipedia by faculty members in higher education and is leading a new project, funded by the Spanish Ministry of Education and Science that aims to analyse the scientific content of Wikipedia.

**Josep Lladós (m)** is an associate professor at the Department of Economics and Business Studies and a researcher in the Internet Interdisciplinary Institute (IN3), both at the Universitat Oberta de Catalunya (Barcelona, Spain). He teaches economic geography and international economics and has published several works on innovation processes. He is a member of the IN3's research group Observatory of the New Economy, focused on the study of the new forms of digital businesses. He has recently participated in a research project on the use of Wikipedia by faculty members in higher education and is taking part in a new project, funded by the Spanish Ministry of Education and Science, which aims to analyse the scientific content of Wikipedia. Currently he is the director of the Internet Interdisciplinary Institute (IN3).

**Antoni Meseguer-Artola (m)** is an associate professor at the Department of Economics and Business Studies and a researcher in the Internet Interdisciplinary Institute (IN3), both at the Universitat Oberta de Catalunya (Barcelona, Spain). He teaches Statistics, Econometrics and Mathematics and has published several works on price competition on the Internet, driving factors to e-commerce diffusion, consumer behaviour on virtual learning environments, game theory, and e-learning. He is a member of the IN3's research group Observatory of the New Economy, focused on the study of the new forms of digital business, and the relationship with the online consumer. He has recently participated in a research project on the use of Wikipedia by faculty members in higher education and is taking part in a new project, funded by the Spanish Ministry of Education and Science, which aims to analyse the scientific content of Wikipedia.

**Julià Minguillón (m)** received his Ph.D. degree from the Universitat Autònoma de Barcelona (UAB) in September 2002. In January 2001 he joined the Universitat Oberta de Catalunya (UOC) where he is a faculty member of the Computer Science, Multimedia and Telecommunication Studies department. He has developed learning resources for object oriented programming, abstract data types engineering and compiler construction. He is also involved in the integration of e-learning standards in virtual learning environments, such as IEEE LOM, SCORM and IMS LD. His main research interests include the formal description of the learning process by means of ontologies, personalizing the learning process by means of adaptive itineraries based on reusable learning objects, and user modelling in virtual e-learning environments applying web and data mining techniques for improving user experience and usability, accessibility and mobility issues. He is also interested in open educational resources and the uses of social tools such as Wikipedia for teaching and learning.

## Publications

1. Meseguer A, **Aibar E**, Lladós J, **Minguillon J**, Lerga M (forthcoming). "Factors that influence the teaching use of Wikipedia in Higher Education". Journal of the Association for Information Science and Technology.
2. **Aibar E**, **Lladós J**, Minguillon J, **Meseguer A**, Lerga M (forthcoming). "Wikipedia at University: what faculty think and do about it". The Electronic Library. Vol. 33. Issue 4.
3. **Aibar E** 2014. "Lessons from the Digital Divide". In: Antonio López Peláez (ed.). The Robotics Divide. A New Frontier in the 21st Century? New York: Springer; 157-171. ISBN: 978-1-4471-5358-0
4. **Aibar E** 2014. "Ciència oberta, encerclament digital i producció col·laborativa". In: T. Iribarren, O. Gassol and E. Aibar (eds.). Cultura i tecnologia: els reptes de la producció cultural en l'era digital. Lleida: Punctum; 99-120. ISBN: 978-84-9419874-8.
5. **Aibar E** 2013. "Producción colaborativa y ciencia: un estudio empírico sobre las percepciones y prácticas del profesorado universitario respecto la Wikipedia". A: González Alcaide, G., Gómez Ferri, J. i Agulló, V. (eds.). La colaboración científica: una aproximación multidisciplinar. València: Nau Llibres; 381-392. ISBN 978-84-7642-930-3

## Relevant previous projects or activities

1. The use of Internet open content for university education: an empirical study on the perceptions, attitudes and practices of university faculty on Wikipedia. Recercaixa: ACUP/Fundació La Caixa; from 1-1-2012 till 31-12-2013. Ref: 2011ACUP00051
2. Information, culture and knowledge: New citizens' practices, new public policies. A case comparison of Spain and United States (Sinde Law versus SOPA Law). Ministerio de Economía y Competitividad (Spain). Ref: CSO2012-37851. From 1-2-13 till 31-1-2015.
3. Scientific authority in the public sphere in twentieth-century Spain. Ministerio de Economía y Competitividad (Spain). Ref. HAR2012-36204-C02-02. From 1-2-2013 till 31-2-2015.
4. iCity: Linked Open Apps Ecosystem to open up innovation in smart cities. Programme: Competitiveness and Innovation Framework Programme (CIP) - The Information Communication Technologies Policy Support Programme (ICT-PSP). Ref. 297363. From 2013 to 2015.

## **Partner 6: Europeana Foundation (EF), Netherlands**

---

The Stichting Europeana (Europeana Foundation) is a foundation under Dutch law that owns and operates the Europeana Digital Service Infrastructure (DSI). The Europeana Foundation aims to transform the world with culture. Europeana is the initiator of a network, representing more than 2500 cultural heritage organisations and a thousand individuals from these and other walks of life, passionate about bringing Europe's vast wealth of cultural heritage to the world. Doing so will unlock untold economic and societal benefits, transforming lives in the process. Culture unites Europe, and making it more accessible promotes understanding and new economies.

The Europeana Foundation's responsibilities include providing a legal framework for the governance of Europeana DSI, employing staff, bidding for funding and making sure the Europeana DSI is sustainable.

### **Relevant experience and role in project**

Europeana has collected and created structured information (metadata) about the objects held in Europe's combined collections, made available through a single interface.<sup>23</sup> Europeana developed data standards – specifically the Europeana Data Model (EDM) – to make that information interoperable on the web, and shares that information as widely as possible by applying the Creative Commons Public Domain Mark to all of its metadata. At the same time, Europeana has started to engage users very personally in their shared history through collection days across Europe, Europeana 1914-1918<sup>24</sup> (the largest repository of personal stories about the First World War) and Europeana 1989<sup>25</sup> (one of the largest collections on the events of 1989 in Central and Eastern Europe).

Within this project, the Europeana Foundation will both make data available and explore its usage via Wikidata. Two Linked Open Data sets will be made available, Firstly, 30 million metadata records related to digitised cultural heritage across Europe, and secondly a set of 90 million bibliographic records collected by The European Library, a subunit within the Europeana Foundation concentrating on the library sector. Once this data is made available on Wikidata, The Europeana Foundation will work with project partners to explore different angles by which the data can be used both by the digital humanities research community and the cultural heritage domain.

### **Profile of the personnel**

**Jill Cousins (f)** is Executive Director of the Europeana Foundation and Director of The European Library. She created both operational services, The European Library and Europeana. She has a strong web publishing background, having worked for VNU as their European Business Development Director and then transferred the lessons learnt from commercial business-to-business publishing to scholarly publishing working for Blackwell Publishing and several other academic publishers in the UK. Prior to a publishing career, she worked in the online environment for many years, first as a researcher with her own company specialising in providing business information to large corporate companies. After selling this company Jill worked as the Marketing Director for Online information. She has been involved in several international publishing industry bodies such as CrossRef and COUNTER.

**Antoine Isaac (m)** is R&D Manager at Europeana Foundation. He holds a PhD in Computer Science from University Paris-Sorbonne, where he started to work on applying Semantic Web and Linked Data techniques for cultural heritage (then at the French National Audiovisual Institute, INA). He has contributed to various national and European research projects, and has been involved in a number of W3C groups, notably for SKOS and Library Linked Data. He is co-author of the French book "Le Web sémantique en bibliothèque". He is also a guest researcher at the Web & Media group in the Free University Amsterdam.

---

<sup>23</sup> <http://www.europeana.eu/>

<sup>24</sup> <http://www.europeana1914-1918.eu/>

<sup>25</sup> <http://www.europeana1989.eu/>

**Alastair Dunning (m)** is currently Scientific Coordinator of the Europeana Cloud project, and has several years' experience running and being involved in large-scale projects at a European level, such as Europeana Newspapers and Arrow Plus. He has particular interest in how knowledge from specific academic disciplines is shared with other disciplines. At his previous job for the UK funding Agency JISC, he initiated a portfolio of projects related to harmonising digitised metadata and content from different disciplines to allow for greater cross searching and reuse.

**Victor-Jan Vos (m)** is Head of Programmes, Policy and Research at the Europeana Foundation. He has been working with Europeana Foundation since May 2014; before that he held different positions in collection development, online services and digital preservation at the Koninklijke Bibliotheek, the National Library of the Netherlands. He holds a master in Media Studies from the University of Amsterdam.

### **Publications**

1. Wickett M. K, **Isaac A**, Doerr M, Fenlon K, Palmer C, Meghini C. Representing Cultural Collections in Digital Aggregation and Exchange Environments. D-Lib Magazine, 20(5-6), 2014. doi:10.1045/may2014-wickett
2. Stiller J, Petras VMaria G, **Isaac A**. Automatic Enrichments with Controlled Vocabularies in Europeana: Challenges and Consequences. Proceedings of the 5th International Conference on Cultural Heritage (EuroMed 2014). [http://link.springer.com/chapter/10.1007%2F978-3-319-13695-0\\_23](http://link.springer.com/chapter/10.1007%2F978-3-319-13695-0_23)
3. **Antoine I**, Haslhofer B. Europeana Linked Open Data -- data.europeana.eu. Semantic Web Journal, 4(3):291-297. Wang S, **Isaac A**, Charles V, Koopman R, Agoropoulou Ai, van der Werf T. Hierarchical structuring of Cultural Heritage objects within large aggregations. Proceedings of the 3rd International Conference on Theory and Practice of Digital Libraries (TPDL 2013). <http://arxiv.org/abs/1306.2866>
4. **Alastair D**, Gregory I and Hardie A. Freeing up digital content with text mining: new research means new licences. Serials, 2009, vol. 22, n. 2, pp. 166-173. <http://eprints.rclis.org/18049/>

### **Relevant previous projects or activities**

1. EU ICT-PSP - Europeana Cloud. Europeana Cloud is a three year project to explore cloud infrastructures in cultural heritage field. [pro.europeana.eu/web/europeana-cloud](http://pro.europeana.eu/web/europeana-cloud)
2. EU ICT-PSP - Europeana Creative. Europeana Creative is a European project which enables and promotes greater re-use of cultural heritage resources by creative industries <http://www.pro.europeana.eu/web/europeana-creative>
3. EU ICT-PSP - Europeana Versions 1, 2 and 3 – three projects that have provided funding for the core functions of Europeana

## **Partner 7: Université Paris Sud (UPS), France**

---

Université Paris Sud is one of the largest research universities in France, with a large spectrum of scientific disciplines and fields. It has a leading role in physics and mathematics; driven by its close ties with chemistry and biology, pharmaceutical research at Paris-Sud University focuses primarily on therapeutic innovation, active principles, therapeutic target identification, and drug vectorization. Medical research, be it fundamental, clinical or translational, is behind Paris-Sud University's leading role in key areas such as oncology, immunology and biotherapy, neuroscience and reproductive endocrinology, and public health; research in social sciences include law, economics and management.

Université Paris Sud is a co-founder of the Center for Data Science (CDS) at Paris Saclay. The goal of this initiative is to establish an institutionalized agora in which these scientists can find each other, exchange ideas, initiate and nurture interdisciplinary projects, and share their experience on past data science projects. To foster synergy between data analysts and data producers we propose to provide initial resources for helping collaborations to get off the ground, to mitigate the non-negligible risk taken by researchers venturing into interdisciplinary data science projects, and to encourage the use of unconventional forms of information transmission and dissemination essential in this communication-intensive research area. The CDS fits perfectly in the recent surge of similar initiatives, both at the international and at the national level, and it has the potential to make the University one of the international forerunners of data science.

### **Relevant experience and role in project**

The Data group at CDS develops a Data as a Service platform which goals and technologies are closely related to the project. The future platform will interconnect all the data of laboratories amongst themselves but also with the scientific information on the Web. Wikidata has already become a major focus point to openly share scientific information on the Web and so, it will play a major role in the future platform of CDS.

CDS includes the Machine Learning and Optimization team (AO), a joint team of INRIA/CNRS/University Paris Sud. The team has been participating and taking a leading role in several EU-projects (PASCAL, SYMBRION, CITINES, MASH, EGEE/EGI) as well as in national industry-oriented projects such as TIMCO (Technologies for In-Memory Computing).

Its role in the project will be to experiment the manners to reuse and to complete the information of Wikidata in the tools of laboratories (WP4). The aim is to start to build in the university a unified research area open to the world. Moreover, it will establish links with the ongoing effort of the EGI.eu foundation towards open science as described in the Open Science Commons [whitepaper](#).<sup>26</sup>

### **Profile of the personnel**

**Dr. Cécile Germain (f)** is a full professor of Computer Science. She is a member of the Machine Learning and Optimization team (AO), a joint team of INRIA/CNRS/University Paris Sud. Her research interests are in Machine Learning and its applications to e-science and autonomic computing. As the policy officer of University Paris Sud for Scientific Computing and chair of the Data group in CDS, she has extensive experience with the scientific data systems at all scales, including international scientific collaborations in High Energy Physics. Current relevant projects are the Center for Data Science, EGI.eu and TIMCO. She has initiated and leads the Grid Observatory initiative (<http://www.grid-observatory.org>), a digital curation facility for the digital assets of the EGI flagship European grid for computer science and engineering. She is also involved in the design of interdisciplinary scientific challenges, including the recent HiggsML challenge.

---

<sup>26</sup> <http://go.egi.eu/osc>

**Karima Rafes (f)** is a researcher at the Center for Data Science of Paris-Saclay and she started a thesis related to Linked Data. Her expertise is research about the Semantic Web since 2007. She developed the LinkedWiki tool for MediaWiki; and in the CDS project, she researched pragmatic solutions to help laboratories integrate the latest Cloud and Linked Data technologies in their workflows.

## Publications

**Rafes K**, Nauroy J, **Germain C** 2014. TFT, Tests For TripleStores. Semantic Web Challenge 2014.<sup>27</sup>

1. **Rafes K** (ongoing). SPARQL Protocol and RDF Query Language (course on Wikiversity, in French).<sup>28</sup>
2. Feng D, **Germain C**, Glatard T 2013. Efficient distributed monitoring with active collaborative prediction Future Generation Computer Systems, 29(8).<sup>29</sup>
3. Zhang X, Furtlehner C, **Germain C**, Sebag M 2014. Data Stream Clustering with Affinity Propagation. IEEE Transactions on Knowledge and Data Engineering (TKDE), vol. 26, no. 7, pp. 1644-1656.
4. **Germain-Renaud C**, Cady A, Gauron P, Jouvin M, Loomis C, Martyniak J, Nauroy J, Sebag M 2011. The Grid Observatory. 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid) pp 114-123.

## Relevant previous projects or activities

1. EU FP7 NoEs PASCAL and PASCAL II, Pattern Analysis, Statistical Modelling and Computational Learning<sup>30</sup>
2. EU FP7 CP-CSA EGI-inspire, Integrated Sustainable Pan-European Infrastructure for Researchers in Europe<sup>31</sup>
3. Kaggle challenge, Higgs Boson Machine Learning Challenge<sup>32</sup>

## Significant infrastructure

The Virtual Data computing centre located at the Orsay campus of University Paris Sud is a joint facility with other regional institutions. This top-level infrastructure (PUE 1.3) is currently designed for 400KW IT, and will be extended up to 1.5MW IT in the next seven years. It hosts ~ 4000 cores and 500TB of disk storage. It includes a tier2-node of the EGI grid and a 1000 cores cloud using the IaaS technology developed in the FP7 StratusLab project. Extensions are planned in 2015-2016 to include resources dedicated to database operations (project OpenData@UPSud), with 100TB of storage and in-memory computing facilities.

## 4.2. Third parties involved in the project (including use of third party resources)

No third parties are involved.

<b>Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)</b>	<b>Y</b>
The only tasks to be subcontracted referred to the preparation of certificate of financial statements for the Coordinating institution Partner 1. All <b>sub-contracting</b> will follow Horizon 2020 financial guidelines and is related to rules for awarding contracts	

<sup>27</sup> [http://challenge.semanticweb.org/2014/submissions/swc2014\\_submission\\_4.pdf](http://challenge.semanticweb.org/2014/submissions/swc2014_submission_4.pdf)

<sup>28</sup> [http://fr.wikiversity.org/wiki/SPARQL\\_Protocol\\_and\\_RDF\\_Query\\_Language](http://fr.wikiversity.org/wiki/SPARQL_Protocol_and_RDF_Query_Language)

<sup>29</sup> <http://hal.inria.fr/hal-00784038>

<sup>30</sup> <http://www.pascal-network.org>

<sup>31</sup> <https://www.egi.eu/about/egi-inspire/>

<sup>32</sup> <https://www.kaggle.com/c/higgs-boson>



according to the principles of best value for money (best price-quality ratio), transparency, and equal treatment.	
<b>Does the participant envisage that part of its work is performed by linked third parties</b>	N
<i>Not applicable</i>	
<b>Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)</b>	N
Not applicable	

### 4.3 Associate Partners

Table 4.3a List of associates involved as key stakeholders in the project. Their letters of support are attached in Section 6. Further stakeholders will be involved during the project.

Organisation	Country	Contact Person	Role
The Scripps Research Institute <sup>33</sup>	US	Andrew Su, Associate Professor	Code sharing and coordination with Gene Wiki project
Histropedia <sup>34</sup>	UK	Navino Evans, Co-founder	Providing use cases for and visualizations of historical information
MetaBrainz Foundation <sup>35</sup>	US	Robert Kaye	Integration of metadata about music <sup>36</sup>
Universitat Pompeu Fabra <sup>37</sup>	ES	Xavier Serra, Head of Music Technology Group	Advice on semantic modelling of concepts related to music and music technology
Pensoft Publishers <sup>38</sup>	BG	Lyubomir Penev	Advice on publication and taxon-related metadata; integration with RefBank
Interdisciplinary Centre for Mathematical and Computational Modelling <sup>39</sup>	PL	Łukasz Bolikowski, head of Data Analysis	Integrating Common Map of Academia or other ICM data with Wikidata

<sup>33</sup> <http://www.scripps.edu/>

<sup>34</sup> <http://histropedia.com/>

<sup>35</sup> <http://metabrainz.org/>

<sup>36</sup> <https://musicbrainz.org/doc/Wikidata>

<sup>37</sup> <http://www.upf.edu/>

<sup>38</sup> <http://www.pensoft.net/>

<sup>39</sup> <http://www.icm.edu.pl/>

at University of Warsaw (ICM)		Laboratory	
Bern University of Applied Sciences <sup>40</sup>	CH	Beat Estermann	Advice on mapping and integrating research and cultural heritage data
Botanic Garden Meise <sup>41</sup>	BE	Quentin Groom	Advice on data about botanical collectors
Würzburg University <sup>42</sup>	DE	Konrad Förstner	Advice on handling genome and transcriptome information of microbial species
Ruđer Bošković Institute <sup>43</sup>	HR	Alen Vodopijevec	Advice on integrating data from Croatian sources
ETH-Bibliothek Zurich <sup>44</sup>	CH	Franziska Regner	General advice on WP3 on integrating suitable data from ETH-Bibliothek (T3.1) and on WP3 more generally
University of Bologna <sup>45</sup>	IT	Silvio Peroni	Advice on ontologies relating to scholarly publications
Royal Society of Chemistry <sup>46</sup>	UK	Richard Kidd	Advice on integration of chemical information
Plazi <sup>47</sup>	CH	Donat Agosti	Advice on taxon related data, integration of the Plazi treatment repository and Biodiversity Literature Repository
CrossRef <sup>48</sup>	UK	Ed Pentz, Executive Director	Advice on bibliographic metadata and persistent identifiers
Center for Open Science <sup>49</sup> (COS)	US	Andrew Sallans, Partnerships Manager	Integration of Wikidata with workflows in Open Science Framework (OSF)
ContentMine <sup>50</sup>	UK	Peter Murray-Rust	Advice on integration of facts mined from the scholarly literature

<sup>40</sup> <http://www.bfh.ch/>

<sup>41</sup> <http://www.br.fgov.be/>

<sup>42</sup> <http://www.uni-wuerzburg.de/>

<sup>43</sup> <http://www.irb.hr/>

<sup>44</sup> <http://www.library.ethz.ch/>

<sup>45</sup> <http://www.unibo.it/>

<sup>46</sup> <http://www.rsc.org/>

<sup>47</sup> <http://plazi.org/>

<sup>48</sup> <http://www.crossref.org/>

<sup>49</sup> <http://centerforopenscience.org/>

<sup>50</sup> <http://contentmine.org/>

## 5. Ethics and Security

### 5.1 Ethics

There are no specific ethical issues associated with the Wiki4R project.

### 5.2 Security

**Please indicate if your project will involve:**

- Activities or results raising security issues: **NO**
- 'EU-classified information' as background or results: **NO**

## 6. Letters of Support