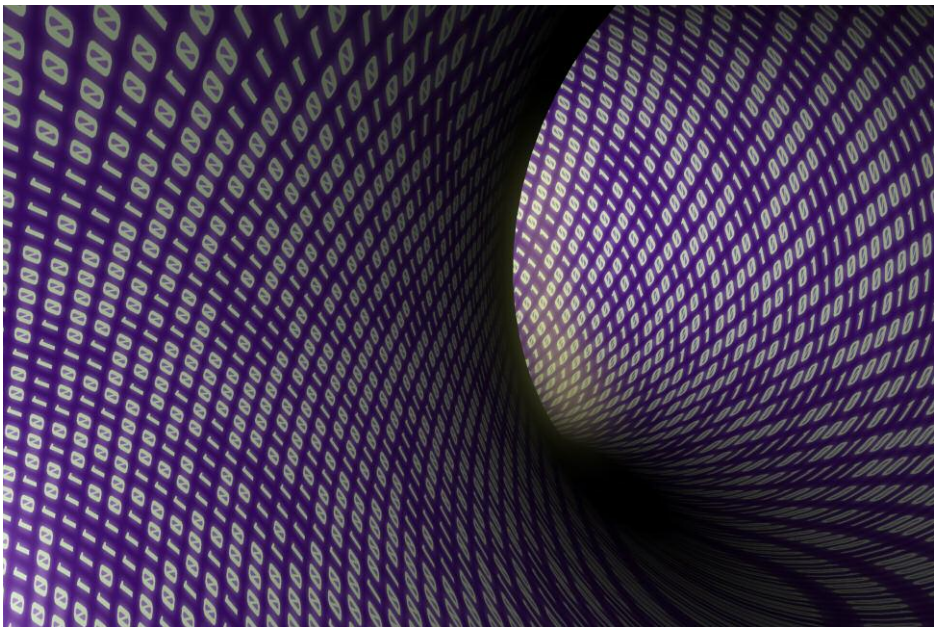
THE PONTIFICAL ACADEMY OF SCIENCES

WORKSHOP ON

BIG DATA AND SCIENCE:

Relevance of Computational Sciences for Data Collection, Data Storage and Data Management in Basic and Applied Scientific Investigations



17-18 November 2015 • Casina Pio IV • Vatican City

Introduction

One of the distinctive features of contemporary scientific research, in both basic and applied sciences, is the large amount of data that is continuously being produced. Quite frequently, data production exceeds the capacity of the available tools for its proper analysis and interpretation, being genomics a paradigmatic example of this situation. Therefore, the use of the most advanced methodologies for data management is of key importance for a successful research enterprise.

In recent years, major advances have taken place in the field of Computational Sciences. They have not only facilitated gaining access to new data in all fields of research, but they have also led to a more efficient processing of the information. These innovations are producing a major impact in the way scientific research is currently conducted. Since collected data becomes readily accessible to potential users, collaborative projects are more common than ever before.

This workshop does not intend to encompass all fields of research in which data management is relevant. Instead, its aim is to present specific examples of appropriate collection, storage and management of data, taken mainly from the Life Sciences, Earth Sciences and Astrophysics. It is expected that sharing experiences among various fields may lead to new opportunities for interdisciplinary research.

This can be expected to provide advice and help for basic and for applied research projects in future scientific investigations. We hope that this workshop provides a unique opportunity for interdisciplinary discussion on big data and sciences.

Organizers

Werner Arber (PAS President)
Marcelo Sánchez Sorondo (PAS Chancellor)
Rafael Vicuña (PAS Academician)
Takashi Gojobori (PAS Academician)

Programme

Tuesday 17 November 2015	
Session 1-0 **GREETINGS AND OVERVIEW** *Chair*: **Rafael Vicuña (PAS Academician)**	
9:00	*Welcome Speech* **Werner Arber (PAS President)**
9:10	*Welcome Greetings* **Marcelo Sánchez Sorondo (PAS Chancellor)**
9:20	*Outline of the entire Workshop* **Takashi Gojobori (PAS Academician, KAUST, KSA/NIG, Japan) and Rafael Vicuña (PAS Academician)**
Session 1-1 **GENOMICS** *Chair*: **Rafael Vicuña (PAS Academician)**	
9:30	*Data Analysis of Cyanobactrial Genomes* **John Archer (Principal Research Scientist, KAUST, KSA)**
9:50	Discussion
10:00	Break
10:30	*The Architecture of Human Chromosomes: A New Insight Into Genome Organization* **Giorgio Bernardi (Professor at Rome 3 University in Rome, Italy)**
10:50	Discussion
11:00	*Big Data in the Study of Early Microbial Life* **William F. Martin (Professor at Heinrich-Heine-Universität in Düsseldorf, Germany)**
11:20	Discussion
11:30	General Discussion
12:00	Lunch
Session 1-2 **EARTH SCIENCE AND DISASTER** *Chair*: **Marcelo Sánchez Sorondo (PAS Chancellor)**	
14:00	*Big Data: Big Opportunities, ICT Security and Protection of Privacy* **Tomihiro Taniguchi (Former Vice Director-General of IAEA, and Visiting Professor, TIT, Tokyo, Japan)**
14:20	Discussion
14:30	*Reponses to the Data Revolution: Policy, Data Science and Capacity Building* **Simon Hodson (CODATA Executive Director, CODATA/ICSU, Paris)**
14:50	Discussion
15:00	*Big Data for Unlearning of Nuclear Issues After Fukushima Daiichi Accident* **Shuichi Iwata (Professor-Emeritus at Tokyo University, Tokyo, Japan)**
15:20	Discussion

Session 1-3
BIODIVERSITY AND EVOLUTION
Chair: **Takashi Gojobori (PAS Academician, KAUST, KSA/NIG, Japan)**

15:30	*Big Eukaryote Genomes Contain Many Regions for Gene Expression Control* **Naruya Saitou (National Institute of Genetics, Japan)**
15:50	Discussion
16:00	Break
16:30	*DNA Sequence Comparison Can Provide Evidence for Horizontal Gene Transfer* **Werner Arber (PAS President)**
16:50	Discussion
17:00	*Experimental and Theoretical Approaches to Epistatic Interactions* **Shozo Yokoyama (Professor, Emory University, US)**
17:20	Discussion
17:30	General Discussion
18:30	Dinner

Wednesday 18 November 2015

Session 2-1
HUMAN HEALTH
Chair: **Werner Arber (PAS President)**

9:00	*Analysis of Large-Scale Genomic Data, Exemplified by the Search for Landscapes of Cancer Genomes* **Partha P. Majumder (Director, NIBB, India)**
9:20	Discussion
9:30	*Human Cells Returning to Uni-Cellularity – What do We Learn From the Big Data of Cancer Genomics?* **Chung-I Wu (Director, GIB, China)**
9:50	Discussion
10:00	Break
10:30	*Finding Useful Information in Big Biomedical Data* **Vladimir B. Bajic (Director, KAUST, KSA)**
10:50	Discussion
11:00	General Discussion
12:00	Lunch

Session 2-2
SYSTEM
Chair: **Partha P. Majumder (Director, NIBB, India)**

14:00 *Data-Centric Science*
Genshiro Kitagawa (President of Research Organization of Information & System in Tokyo, Japan)

14:20 Discussion

14:30 *Big Data in Robotics*
Shuji Hashimoto (Vice President of Waseda University, Tokyo, Japan)

14:50 Discussion

15:00 *Big Data in Gene Expression Network System*
Wen-Hsiung Li (Director, Institute of Biodiversity, Academia Sinica, Taiwan)

15:20 Discussion

Session 2-3
DATABASE
Chair: **Vladimir B. Bajic (KAUST, KSA)**

15:30 *Bioinformatics Goes Translational: Infrastructural Challenges and Societal Opportunities*
Rolf Apweiler (Director of EBI, UK)

15:50 Discussion

16:00 Break

16:30 *Big Data Challenges in Health and Life Science*
Ron Appel (Director, SIB, Switzerland)

16:50 Discussion

17:00 General Discussion and Conclusions

18:30 Dinner

List of Participants

Speakers

APPEL Ron
APWEILER Rolf
ARBER Werner
ARCHER John A.C.
BAJIC Vladimir B.
BERNARDI Giorgio
GOJOBORI Takashi
HASHIMOTO Shuji
HODSON Simon
IWATA Shuichi
KITAGAWA Genshiro
LE DOUARIN Nicole
LI Wen-Hsiung
MAJUMDER Partha P.
MARTIN William F.
SAITOU Naruya
SÁNCHEZ SORONDO Marcelo
TANIGUCHI Tomihiro
VICUÑA Rafael
WU Chung-I
YOKOYAMA Shozo

Observers

COLETTI Silvano
KAJITA Mami
KAJITA Seiji

Biographies

Ron D. Appel graduated in Computer Sciences at the University of Geneva, where he also received his Ph.D. in 1987, followed by a postdoctoral fellowship at Harvard School of Public Health. He was co-founder in 1998 of the SIB Swiss Institute of Bioinformatics, where he was leading the Proteome Informatics Group, pioneering the development of proteomics software and databases. Ron Appel was also the initiator of ExPASy, the world's first Web site dedicated to life science in 1993. He is scientific co-founder of two biotechnology companies in Geneva, including one in bioinformatics – Geneva Bioinformatics (GeneBio) SA, and a founding member of the Council and of the Executive Board of the Health On the Net Foundation (HON). Ron Appel is currently the Executive Director of the SIB Swiss Institute of Bioinformatics and professor of bioinformatics at the University of Geneva.

Dr. **Rolf Apweiler** is Joint Director of the European Bioinformatics Institute (EMBL-EBI), part of the European Molecular Biology Laboratory (EMBL), where he has strategic oversight of all EMBL-EBI research and service activities. Prior to taking on this position he co-led all EMBL-EBI service activities. Earlier he led the protein resources at EMBL-EBI, including the UniProt Consortium, InterPro and EMBL-EBI's contribution to the Gene Ontology. Rolf received his PhD from the University of Heidelberg in 1994, and has been at EMBL since 1987.

Dr. **John A.C. Archer** is Principal Research Scientist at the Computational Biosciences Research Center (CBRC) of the King Abdullah University of Science and Technology (KAUST). He served in the Office of the President (KAUST) 2009-2011 and now leads its Data Generation, Validation and Engineering research theme. From 1991 to 2009 he served as Principal Investigator at the University of Cambridge. His research interests include microbiology, biotechnology, functional genomics and synthetic biology, in particular, cell factories and their industrial applications.

Prof. **Vladimir B. Bajic** is the director of the Computational Bioscience Research Center and a professor in the Computer, Electrical and Mathematical Sciences and Engineering Division at KAUST. He was professor of bioinformatics and acting and deputy director of the South African National Bioinformatics Institute (SANBI) at the University of the Western Cape. He worked in industry and several academic and research institutions across the globe, including Vinca Nuclear Science Institute in Yugoslavia and the Institute for Infocomm Research in Singapore, where he was head of the Knowledge Extraction Laboratory. His work in modeling and artificial intelligence has resulted in several promoter recognition tools and a knowledge discovery platform that found commercial applications. His primary interest is in facilitating biological discoveries using bioinformatics systems combined with data modeling with emphasis is on the inference of new information not explicitly present in biological data, development of systems and tools with such capabilities on high-performance computing systems and their industrial applications. Prof. Bajic is an elected member of the Academy of Nonlinear Sciences in Russia, and in 2002, he was elected a role model of the Institute for Infocomm Research in Singapore.

Giorgio Bernardi received an MD degree from the University of Padua (Italy) in 1952 and a degree in Physics from the University of Strasbourg (France) in 1967. He spent most of his scientific career with the Centre National de la Recherche Scientifique (CNRS), first at the Center for Research on Macromolecules in Strasbourg (1959-1969), then at the Jacques Monod Institute in Paris. Between 1998 and 2007, Dr. Bernardi was the President of the Stazione Zoologica Anton Dohrn in Naples and Head of the Laboratory of Molecular Evolution. In 2010, he joined the Science Department of Roma Tre University, Rome, as Visiting Professor. Dr. Bernardi obtained honorary degrees from the Engelhardt Institute of Molecular Biology, Moscow (Russia), from the University of Ancona (Italy), and from the University of Montevideo (Uruguay). Dr. Bernardi is a Member of the European Molecular Biology Organization, a Fellow of the American Association for the Advancement of Sciences, a Foreign Honorary Member of the Genetics Society of Japan, a Member of the Academia Europaea, and of several National Scientific Societies and Academies. He was the Editor-in-Chief of Gene and the founder of Marine Genomics as well as a Member of the Editorial Board of other Journals. He was the Chairman, amongst the others, of the Scientific Council of the UNESCO Office in Venice (2002-2007) and of the International Union of Biological Sciences (2009-2012).

Shuji Hashimoto received the B.S., M.S. and Dr. Eng. Degrees in applied physics from Waseda University, Tokyo, Japan, in 1970, 1973, and 1977, respectively. He is a Professor with the Department of Applied Physics. Since 2000, he was the Director of the Humanoid Robotics Institute, Waseda University for ten years. He was the Dean of the Faculty of Science and Engineering from 2006-2010. He is currently the Senior Vice President and Provost of Waseda University since 2010. He published over 400 technical publications, proceedings, editorials, and books. His research interests include "KANSEI" engineering, signal processing, artificial intelligence and robotics.

Simon Hodson has been Executive Director of CODATA since August 2013. CODATA exists to promote international collaboration for Open scientific data and aims to strengthen international science for the benefit of society by promoting improved scientific and technical data management and use. He sits on numerous project Steering Boards and strategic Working Groups in the area of research data, including the GEO Data Sharing Working Group, the Scientific Advisory Board of CESSDA (the European Data Infrastructure for the Social Sciences) and is a member of the Board of Directors of the Dryad data repository. Simon is also co-chair of the RDA-WDS Data Publication Working Group on Cost Recovery for Data Centres; co-chair of the CODATA-RDA Data Working Group for Data Science Summer Schools in Developing Countries, which is building in the first instance on CODATA partnerships in Kenya; and co-chair of the RDA Working Group to build a 'BioSharing Registry: connecting data policies, standards & databases in life sciences'. He is an

expert on data policy issues and research data management and has recently contributed to reports on data policy issues for the Danish e-Infrastructure Group and on research data management for a consortium of UK research institutions. Simon has a strong research background, as well as considerable project and programme management experience: from 2009 to 2013, as Programme Manager, he led two successive phases of Jisc's innovative Managing Research Data programme in the UK.

Shuichi Iwata is Emeritus Professor at The University of Tokyo; Professor at The Graduate School of Project Design; Editor-in-Chief at the Data Science Journal; Member of EAJ and SCJ; Program Director of Decommissioning Project of MEXT; Program Officer Doctor of Engineering; and Program Officer of Initiative for Strategic Scientific Research on Nuclear Fields of MEXT. Shuichi Iwata works to make data on science and technology available for everyone through materials design, design science and data science. Since 1967 he has been inspired by a set of systematic approaches to develop materials for nuclear applications and he has set his research objective in developing design methodologies on materials. As maps to explore new materials, he and his colleagues have designed and developed an information infrastructure of phase diagrams, electronic-, atomistic-, crystallographic- and microscopic- structures and properties, that makes it feasible to strategically design materials by integrating direct (deductive) and inverse (inductive and adductive) approaches rather than traditional trial and error approaches. He received the Honda Memorial Young Researcher Award, the Iketani Science Foundation Award, the Promotion of Science and Technology Information Award, the Japan Science and Technology Agency Paper Award and the Japan Institute of Metals/GIW Best Paper Award.

Genshiro Kitagawa is Director General of the Institute of Statistical Mathematics, President of the Research Organization of Information and Systems (Tokio) and Professor of Statistical Science at the Graduate University for Advanced Study. His primary interests are in time series analysis, non-Gaussian nonlinear filtering, statistical modeling and discovery science. He is the executive editor of the Annals of the Institute of Statistical Mathematics, co-author of Smoothness Priors Analysis of Time Series, Akaike Information Criterion Statistics, Information Criteria and Statistical Modeling, and several Japanese books. He was awarded the Japan Statistical Society Prize in 1997 and Ishikawa Prize in 1999, and is a Fellow of the American Statistical Association and an elected member of the International Statistical Institute.

Wen-Hsiung Li has been Director and Distinguished Research Fellow of the Biodiversity Research Center at the Academia Sinica (Taiwan) since 2008. He was James Watson Chair Professor (2004-2012), George Beadle Chair Professor at the Department of Ecology & Evolution of the University of Chicago (1999-2004). He is a member of the Editorial Boards of the journals Scientific Report (Nature series) and Genome Reserch. He is also Editor of the journal Molecular Phylogenetics and Evolution. He was awarded with the Balzan Prize for Genetics and Evolution (2003), the Horace Mann Medal (2004), the HUGO 2008 Chen Award for Distinguished Academic Achievement in Human Genetic and Genomic Research, and the Mendel Lecture and Medal (2009). He is Member of the National Academy of Sciences since 2003, Fellow of the American Academy of Arts and Sciences since 1999 and member of the Academia Sinica, Taiwan since 2008. His research interests are Evolutionary genomics, Molecular Evolution, Bioinformatics, Population Genetics, and Human Genetics.

Partha P. Majumder is the founding Director of the National Institute of Biomedical Genomics, and concurrently a Professor of the Indian Statistical Institute, Kolkata. His contributions to human and population genetics span a vast range, from structure and evolution of human populations to genomics of diseases and development of statistical methods for genomic analyses. He is an elected Fellow of all the three science academies of India, of The World Academy of Sciences and the International Statistical Institute. He is a Council Member of the Human Genome Organisation. He is the Indian National Coordinator on the International Cancer Genome Consortium. He is a recipient of many awards and medals, including the TWAS Biology Prize – 2009, G.D. Birla Award for Scientific Research – 2002, Ranbaxy Research Award in Applied Medical Sciences – 2000, and the New Millennium Science Medal, Government of India, 2000.

William F. Martin has been a professor at the University of Düsseldorf since 1999, where he has taught Ecological Plant Physiology (1999-2011) and Molecular Evolution at the since 2011. Since 2013, he has been Visiting Professor at the Instituto de Tecnologia Química e Biológica, Oeiras. He is Elected Member of EMBO (European Molecular Biology Organisation) and of the Nordrhein-Westfälische Akademie der Wissenschaften; in 2006 he was Elected Fellow of the American Academy for Microbiology. Since 2009 he is Editor-in-Chief of the journal *Genome Biology and Evolution.* William Martin was Editor-in-Chief of *Molecular Biology and Evolution* (2004-2007) and Editorial Board Member for 11 different journals in microbiology and molecular evolution from 1995 to 2008. His research interests include Endosymbiosis and cell evolution (Bioinformatics), Anaerobic energy metabolism in eukaryotes (Biochemistry), Hydrothermal vents, anaerobic autotrophs, and the origin of life.

Naruya Saitou. Born in 1957 in Japan. 1975-1981, B.S. and M.S. from University of Tokyo, majoring in physical anthropology. 1982-1986, Ph.D. from Graduate School of Biomedical Sciences, University of Texas Health Science Center at Houston. One chapter of Ph.D. thesis proposed neighbor-joining method, which became one of standard methods for making phylogenetic trees. 1989-1991, assistant professor at Department of Biology, University of Tokyo. 1991-2002, associate professor at Division of Evolutionary Genetics, National Institute of Genetics, Japan. From 2002, Professor at Division of Population Genetics, National Institute of Genetics. 1991-2011, involved in DNA Data Bank of Japan activity. 2005-2014, Council member, Science Council of Japan. Concurrent positions: Professor, Department of Genetics, Graduate University for Advanced Studies; Professor, Department of Biological Sciences, University of Tokyo. Saitou's research interest is evolution of human and other organisms using genome wide big data.

Tomihiro Taniguchi. Immediately after graduating in Nuclear Engineering from the University of Tokyo, Tomihiro Taniguchi joined the Ministry of International Trade and

Industry (MITI) in 1968. During his 30-year-career at the MITI, he served as Director and Deputy Director. In the 90s, he was appointed as Deputy Director General in charge of trade negotiation for establishment of WTO. Subsequently, as DDG for Nuclear Policy, he was the Japanese Government's chief administrative officer on all aspects of commercial use of nuclear energy. In 1998 he was appointed Visiting Professor of the University of Tokyo. In the same period, he served as the Vice Chair of the Intergovernmental Panel on Climate Change (IPCC) for its Third Review. Tomihiro Taniguchi served in a number of important positions at international organizations, including OECD Deputy Director for Science, Technology and Industry and the Director of Oil Market and Emergency Preparedness of the International Energy Agency (IEA). From 2001 to 2010, he worked as DDG and the Head of Department of Nuclear Safety and Security of the International Atomic Energy Agency (IAEA). He was also a part of the IAEA group and the IPCC group, which were awarded the Nobel Peace Prize in 2005 and 2007, respectively. In 2011, Tomihiro Taniguchi was named Executive Director of the Science and Technology in Society (STS) *forum* and also appointed as a professor of the Tokyo Institute of Technology. Since August, 2014, he has been Senior Advisor at the STS *forum*.

Chung-I Wu, PhD, is Professor and Former Chairman of the Department of Ecology and Evolution of the University of Chicago. He is also Professor of the Committee on Evolutionary Biology and of the Committee on Genetics, Genomics, and Systems Biology of the same University. Furthermore, Chung-I Wu is Senior Fellow and Former Director of the Beijing Institute of Genomics, Chinese Academy of Science. He is interested mainly in the molecular and population genetics of species and racial differentiation. His three main areas of research are genomic approaches to genes of species differentiation; gene expression differences between species, focusing on the roles microRNAs play in expression divergence; and the evolution of "cells within multi-cellular organisms", or cancer evolution.

Shozo Yokoyama earned his B.S. and M.S. degrees in Biology in Japan, and a Ph.D. degree in biomathematics in 1977 from the University of Washington. He moved to Emory's Department of Biology in 2003 as an endowed chair, Asa G. Candler Professor of Biology. Previously, Dr. Yokoyama was at Washington University (1978-1987, Assist Professor), the University of Illinois (1987-1991, Assoc. Professor), and Syracuse University (1991-2003). His research focuses on molecular genetics and the evolution of dim-light and color vision. Since 1990, his lab has been conducting genetic analyses of the visual pigments of a diverse range of organisms. His lab was the first to identify the amino acids that control red-green color vision and UV vision in various vertebrate species. These and other analyses on the origin and evolution of color vision produced "the deepest body of knowledge linking differences in specific genes to differences in ecology and to the evolution of species (The Making of The Fittest by Sean Carroll, 2006)." Dr. Yokoyama is currently trying to understand the molecular genetics and evolution of vertebrate and invertebrate vision using quantum chemistry. He is author and coauthor of over 150 articles. He was a Panel member of the Genetics Study Section at National Institutes of Health (1988-1991) and was a President of American Genetics Association (2003). He is an elected Fellow of the American Association for the Advancement of Science.

For the biographies of the other Academicians of the PAS and PASS see www.pas.va and www.pass.va

MEMORANDUM

1) From 17, 18 November the buses for the Academicians and the Speakers will leave the hotels as follows:

Domus Sanctae Marthae-Casina Pio IV-Domus Sanctae Marthae	
Tuesday 17 November	h. 8.30 a.m. h. 8.30 p.m.
Wednesday 18 November	h. 9.00 a.m. h. 8.30 p.m.

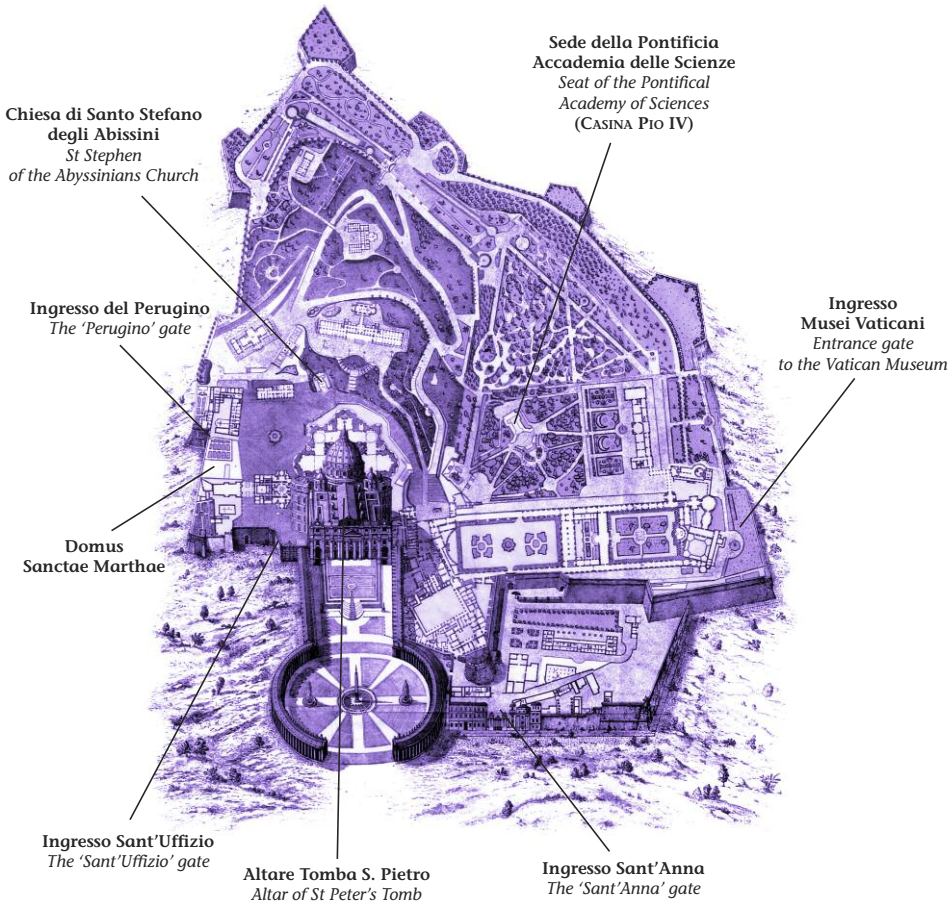
Hotel Columbus-Casina Pio IV-Hotel Columbus	
Tuesday 17 November	h. 8.30 a.m. h. 8.30 p.m.
Wednesday 18 November	h. 9.00 a.m. h. 8.30 p.m.

2) Lunch and dinner for the Speakers and Academicians will be served at the Academy. If you are a vegetarian, please let us know as soon as possible.

3) Wifi is available in the Casina Pio IV's Conference Hall. Please log in to the network called WLAN_PADS using "guest" as the username and "password" as the password.

4) Cable internet access is available at the Domus Sanctae Marthae for 3,00 euros per day.

Note for the **Speakers/Academicians:**
Please give your form for the refunding of expenses to the Secretariat.



THE PONTIFICAL ACADEMY OF SCIENCES • CASINA PIO IV • V-00120 VATICAN CITY
Tel: +39 0669883451 • Fax: +39 0669885218 • Email: pas@pas.va

For further information please visit:
www.pas.va