

**Dr GIUSEPPE INSANA – 墨白****Bioinformatics, data science, business and linguistics expert***Ph.D. (Cambridge), M.Sc., B.Sc.*

📍 Cambridge, UK

✉ [insana@insana.net](mailto:insana@insana.net)🔑 GPG key [0xCD84801E](#)💻 [www.insana.net](http://www.insana.net)🌐 [linkedin.com/in/insana](https://linkedin.com/in/insana)🐙 [github.com/g-insana](https://github.com/g-insana)🆔 [orcid.org iD](https://orcid.org/iD)🏆 [credly.com badges](https://credly.com/badges)**CORE SKILLS:****INNOVATION AND CREATIVITY**

- ◆ Highly-motivated self-starter who can deliver results without close supervision
- ◆ Ability to rapidly learn an extremely wide range of skills and subjects
- ◆ Innovator and early adopter of new technologies
- ◆ Able to predict trends and ideas, often with independent discovery and foresight
- ◆ Strong sense of ownership, urgency, and drive

**ORGANIZATIONAL MANAGEMENT**

- ◆ Ability to prioritise work and meet deadlines
- ◆ Sharp analytical skills and attention to detail
- ◆ Great team player, able to work effectively with others as well as individually
- ◆ Setup of businesses, overlooking and managing all aspects, from idea to maturity, through design,

**SUMMARY**

With over two decades of cross-disciplinary expertise, I am a results-oriented creative professional known for incredible problem-solving and rapid learning. My skills include exceptional communication, organisational proficiency and managerial acumen, cultivated in both academic and business environments.

As a seasoned data scientist, I blend a scientific mindset, marked by rigorous experimentation and fluid curiosity, with a hands-on, passionate approach to software and technology. Holding a Ph.D. in Bioinformatics from the University of Cambridge and an M.Sc. in Biotechnology, I bring extensive knowledge spanning the fields of biology, IT, linguistics, big data, ML and cognitive science.

My professional journey in bioinformatics research began in 1998, expanding into data science, machine learning, and artificial intelligence in 2000. Notable roles include Senior Bioinformatics Scientist at EMBL-EBI, where I enhanced all aspects related to the production of UniProt.

My experience also encompasses business management and consulting roles, adding a practical dimension to my skill set and the ability to focus on efficiency for actionable insights and appropriate team management.

**EDUCATION**✧ **Ph.D. degree in Bioinformatics**[University of Cambridge](#) – Cambridge, UK

01/2000–02/2004 (thesis defence 12.02.2004) (official duration: 3.5–4 years)

Research at the [European Bioinformatics Institute](#) in Genomic analysis and Human variation under the supervision of M. Ashburner, L. Holm and H. Lehtväslaiho.

Developed online & offline tools for variant effect prediction, for sequence retrieval and for gene coordinate mapping; wrote new BioPerl modules; performed codon usage analyses across genomes, across protein families and in HGT detection; used and developed data analysis, data visualisation and unsupervised machine learning techniques

Thesis: *"DNA phonology: investigating the codon space"*

Note: awarded a fully funded scholarship from EMBL ([European Molecular Biology Institute](#))

✧ **M.Sc. + B.Sc. degree ('laurea') in Biotechnology**[University of Verona](#) – Verona, Italy

09/1993–07/1999 (thesis defence 22.7.1999) (official duration: 5 years)

Research in bioinformatics and protein structure modelling, under the supervision of F. Fogolari and H. Molinari, using simulated annealing and co-variation

Thesis: *"Molecular modelling of plant light-harvesting proteins"*

Final grade: **110** (top score) / Completed 30 courses with distinction.

Note: one of the first in Italy to get a Biotechnology degree and to do research in Bioinformatics

✧ **B.Sc. degree in Economics and Business Administration**[University of Verona](#) – Verona, Italy

01/2009–04/2011 (thesis defence 20.04.2011) (official duration: 3 years)

Research in organizational theory and crowdsourcing under the supervision of L. Mola.

Note: finished the degree in only 2 years.

Thesis: *"Crowdsourcing: a critical analysis of a new paradigm in business management"*

Final grade: **107** (top score 110) / Completed 29 courses with distinction.

## WORK EXPERIENCE

### ► 12/2018– ongoing Senior Bioinformatics Scientist



EMBL-[European Bioinformatics Institute](#), Cambridge, UK

Industry sector: Bioinformatics

Notes and skills: database management; workflow and pipeline design, management and optimisation; big data manipulation, analysis and visualisation; deep understanding of omics data

Highlights: Responsible for maintaining and improving the databases, software and data pipelines for the production and release of the [UniProt](#) protein sequence databases; coordinating with teams and researchers across three institutes; performing data science analyses; designing new pipelines with high throughput parallel data processing approaches; improving data quality, availability and ease of access



### ► 02/2020–01/2021 Information security and business strategy consultant



[Geromics](#), Cambridge, UK

Industry sector: Personal genomics

Notes and skills: information security, cryptography

### ► 04/2018–09/2018 Bioinformatics and business consultant



[VanGenes LTD.](#), Xiamen, China

Industry sector: Genomic analysis

Notes and skills: cost benefit analysis, bioinformatics, machine learning, strategy

Highlights: evaluation of viable business growth strategies and of the implementation of a bioinformatics and machine learning reporting and phenotype prediction pipeline for a genomics company currently expanding into D2C WES analysis

### ► 04/2014–10/2018 Cryptocurrency consultant

Notes and skills: Setup and sale of mining rigs, full nodes, wallets and trading accounts for clients

### ► 12/2012–10/2018 Game designer and programmer



[XTeam Software Solutions Srls](#), Rovigo, Italy & [Forgotten Empires](#)

Industry sector: Software development

Notes and skills: design of strategic videogames (three already published, one in the works) and VR software; design and coding of game mechanics, documentation, AI and back-end multiplayer system

Highlights: programmed in JavaScript a full implementation of a strategic board game; designed and programmed its AI with an holistic overseer and fuzzy FSM agents; designed and implemented an online cross-platform multiplayer server; ran two successful Steam greenlight campaigns

### ► 04/2010–11/2018 Business consultant and accounting professional

Chartered Accountants [Insana e Dalla Vecchia](#), Schio, Italy

Industry sector: Accounting, business, finance, auditing, consultancy

Notes and skills: International contracts, tax litigation, business management, investments

Highlights: negotiated joint venture opportunities between major Austrian and Italian companies; resolved high value tax litigation claims with original insight

### ► 05/2001– ongoing Founder and director



[MediaGlyphs Project](#)

Industry sector: Research & development of linguistics services

Notes and skills: Founder and director of an open source project for international communication and multilingual teaching/learning materials. Responsible for general design, technical development, coordination of collaborators... Database management, editing tools, web design, graphic design, Perl+Php+JavaScript+Python coding, automatic language processing, language annotation, ontologies, NLP, MT, semantic relation framework

sourcing, building, training, promotion, up-selling

#### BUSINESS DEVELOPMENT AND

##### STRATEGY

- ◆ Consultant for established companies and start-ups
- ◆ Negotiated international contracts and joint ventures
- ◆ Successfully solved tax litigations with claims of up to 6M euros

##### LEADERSHIP

- ◆ Experience leading both local and international teams, driving operational excellence and best practice
- ◆ Managed businesses and staff in Italy, UK and China
- ◆ Excellent communication abilities to both peers and non-technical people

##### KNOWLEDGE

- ◆ Extensive academic level knowledge in a great number of subjects: biology, computer science, bioinformatics, linguistics, data science, economy, business administration, information security, design.

#### KEY EXPERTISE AREAS:

Data science, ML, AI  
Computational biology  
Linguistics  
Business strategy

► **01/2009–03/2009**      **Bioinformatics consultant**

[Professional Clinical Software GmbH](#), Klagenfurt, Austria

|P|C|S| Industry sector: Research & development of medical software

Notes and skills: Development of data base schemes and data access layer

Highlights: worked as part of Pocemon.eu (FP7 European project)

► **12/2006–11/2008**      **General Manager and Owner**



Sanya G&B Tourism Development LTD, Sanya, China

Industry sector: Hotels and restaurants

Notes and skills: Planning, design, startup setup and full-time management. Customer care, personnel management and training, technical design, planning of promotion, market analysis, marketing, resource management, costing, accounting, relationships with institutions, purchasing, processes' management and optimisation, etc.

Highlights: successfully overseen the design, construction, setup and management of a novel business enterprise with own company of Chinese law

► **04/2005–12/2006**      **Manager and shareholder**

Sanya FS Tourism & Culture LTD, Sanya, China

Industry sector: Hotels and restaurants

Notes and skills: Startup setup and full-time management. Customer care, personnel management and training, resource management, costing, accounting, relationships with institutions, purchasing, etc.

Highlights: successfully overseen the establishment, setup and management of the first international restaurant of Sanya

► **10/2004–06/2005**      **English university lecturer**



Sanya Zhuoda tourism college & [Haikou University of Economics](#), Hainan, China

Industry sector: Public university education

► **03/2004–01/2005**      **Visiting assistant professor**



[Chinese academy of tropical agricultural sciences](#), Hainan, China

Industry sector: Public university education at master level

Notes and skills: I taught the two post-graduate courses "Molecular Biology" and "Bioinformatics" to M.Sc. students, giving bilingual (Chinese and English) lectures, designing the courses and creating the teaching material

Highlights: one of the first foreign bioinformatics professors in China, some of my students are now holding important positions in BGI Shenzhen and Chinese universities

► **05/1997–11/1999**      **IT consultant, sys-admin and web designer**



[A.T. Informatica sas](#), Schio, Italy

Industry sector: Internet outsourcing and web design

Notes and skills: System installation & management, web design, CGI programming

## PLACEMENTS WHILE ATTENDING UNIVERSITY

► **11/2003–12/2003**      **Manager of promotional activities**

Sheraton Resort / Recreation department, Sanya, China

Notes and skills: Management of the setup and promotion of new scuba diving services. Marketing, multicultural communication, customer service, cultural mediation

► **02/2002–12/2002**      **Scientific consultant and programmer**

Sonic Genes Project, Wellcome Trust Science on Stage & Screen, UK

Highlights: mentioned in an article on the Economist (see *Publications and Mentions*, below)

► **02/2001–06/2001**      **Chess teacher**

St Paul's primary school, Cambridge, UK

► **10/2000–06/2002**      **Cognitive science research**

Highlights: analysed, solved and designed Bongard Problems; assisted Harry Foundalis (Douglas Hofstadter group) on his PhD

research using machine learning approach to teach computers to solve BP; discovered together with him whole classes of *meta Bongard problems*

#### ► 09/1994–04/1996 **HP-UX system manager**

Faculty of science, University of Verona – Verona, Italy

Notes and skills: maintaining the OS, protecting against penetration, solving users' issues, creating online services.

Highlights: I learned Unix in my spare time while at university – without any previous knowledge of it – to a level of proficiency such that in less than a year I became the System Administrator of the computer laboratories of the Biotechnology faculty. Developed one of the first online environments for teleconference meetings for Italian biologists

## COMPUTER SKILLS

- **Programming:** **Python, Perl, Php, Bash, Julia, JavaScript**; started coding at the age of 9
- **Bioinformatics** tools & techniques; biological and linguistic data annotation, manipulation and analysis
- **Machine learning**, pattern recognition, data cleaning, modelling, mining and visualisation, large language models
- **Data science** libraries (pandas, scipy, sklearn, keras/tensorflow, seaborn, plotly...)
- **System design, installation and administration** (designed, setup and deployed several custom cloud solutions, servers and micro-services; encrypted cross-platform syncing, device positional tracking, distributed backup; virtualization and containerization)
- **Front-end & back-end** development, web design, database design, graphic design
- **Cryptocurrency**: experience trading and mining cryptocurrencies since 2011
- **Computer graphics**, scalable vector graphics, raytracing, 3d modelling, font and logo design
- **OS**: DOS, Windows, Unix/Linux (Slackware, Debian, Ubuntu & Alpine), Android and MacOS
- **Databases**: flat-file (own custom backends, nosql), MySQL, Oracle
- **Accounting, reporting, interactive computing** (jupyter), **job scheduling** (lsf, slurm) & **version control software** (cvs, git)
- **Word processing & IDE**: vim, Word, LibreOffice, Eclipse

## LINGUISTICS SKILLS

Languages known:        

▪ **Italian**: mother tongue

▪ **English**: excellent

(February 2000 computer-based TOEFL score: 283, equivalent to 657 paper-based)

I studied English since I was 5 years old. I lived, studied and worked in England for many years, attended UK university courses, worked in international research environment, wrote my Ph.D. thesis in English

▪ **Mandarin Chinese**: fluent

I can use the language for all types of work situations. I lived in China for five years, taught scientific subjects at university using both Chinese and English; I also managed two businesses and my own company, always using Chinese for communication with staff, government, customers and suppliers

▪ **Spanish & French**: advanced

▪ **Japanese**: intermediate

(July 2022, Elementary 2; March 2023, Intermediate 1 – Cambridge University Language Programmes) Equiv CERF B2 JLPT N4 (January 2022, Beginner 1; February 2022, Beginner 2 – Saint Petersburg State University online courses)

▪ **Russian & Bulgarian**: simple communication

### Linguistics competences:

Phonetics, phonology, morphology, syntax, computational linguistics, generative grammars, NLP, NLU, practical applied knowledge, inter-linguistics, language design of conlang and auxlang, intercultural mediation, design of language learning material, language teaching




**Frameworks for NLP**: Spacy, Link grammar, NLTK

**Cognitive science**, semantic relation frameworks, knowledge representation, ontology engineering, taxonomies, inference reasoning engines, hyper-graphs

## OTHER INTERESTS

Photography (exhibited in UK and in Italy), cryptography, design, philosophy, drums, violin, sea sports, board sports, sword sports, Go (囲碁/围棋)

## OFFICIAL QUALIFICATIONS & CREDENTIALS

06/2023	<a href="#">Whole Genome Sequencing of bacterial genomes</a>	
05/2023	<a href="#">Natural Language Processing specialization</a> (four courses)	
04/2023	<a href="#">Deep Learning specialization</a> (five courses)	
03/2022	<a href="#">IBM Machine Learning professional certificate</a> (six courses)	
11/2021	Stanford Online's <a href="#">Machine Learning</a>	
01/2015	Registered to the <a href="#">Italian national Registry of Statutory Auditors</a>	
11/2013	Registered with the <a href="#">Order of Chartered Accountants of Vicenza</a> , Italy	
10/2013	<a href="#">State exam (EQF lev 7)</a> of habilitation to professional body of Accounting Experts; U. of Verona, Italy	

## PUBLICATIONS & MENTIONS

- (with Ignatchenko, A.; Martin, M.J.; Bateman A.) "[MBDBMetrics: An online metrics tool to measure the impact of biological data resources](#)"; Bioinformatics Advances, 2023, 3(1), vbad180
- (with Uniprot Consortium) "[UniProt: the universal protein knowledgebase in 2023](#)"; Nucleic acids research, 2022 Nov gkac1052
- (with Martin, M.J. and Pearson, W.R.) "Improved selection of canonical proteins for reference proteomes"; Poster at [QfO workshop of ECCB](#) Conference, Sitges, Spain 2022
- (with Uniprot Consortium) "[UniProt: the universal protein knowledgebase in 2021](#)"; Nucleic acids research, 2021, 49(D1), D480–D489
- (cited) "[A crowdsourcing open platform for literature curation in UniProt](#)"; PLoS Biology, 2021, 19(12), e3001464
- (with Uniprot Consortium) "[UniRule: a unified rule resource for automatic annotation in the UniProt Knowledgebase](#)"; Bioinformatics, 2020, 36(17), 4643–4648
- (cited) "[Challenges in the annotation of pseudoenzymes in databases: the UniProtKB approach](#)"; FEBS J., 2020, 287(19), 4114–4127
- (cited) "[Caenorhabditis elegans phosphatase complexes in UniProtKB and Complex Portal](#)"; FEBS J., 2020, 287(13), 2664–2684
- (with Uniprot Consortium) "[FAIR adoption, assessment and challenges at UniProt](#)"; Sci Data, 2019, 6(1), 175
- "[Crowdsourcing: a critical analysis of a new paradigm in business management](#)"; **B.Sc. thesis**, Verona 2011
- (cited) "[Phaeaco: A Cognitive Architecture Inspired by Bongard's Problems](#)"; Ph.D. thesis by Harry E. Foundalis, Indiana University 2006
- (cited) "[A Sense of Place: The Global and the Local in Mobile Communication](#)"; Book by János Kristóf Nyíri, Passagen Ed., 2005
- (cited) "[Ramblings of a Madman](#)"; Book by Frank Bua, 2005
- "[DNA phonology: investigating the codon space](#)"; **Ph.D. thesis**, Cambridge University 2003
- (cited) "[The art of DNA; Back to bases](#)"; Article on The Economist (print edition), April 24th 2003
- (cited) [The Bioperl toolkit: Perl modules for the life sciences](#); Genome Res., 2002, 12 (10), 1611–1618
- "Unexpected similarities in codon composition between retroviral genomes and human protein families revealed by a novel method of codon usage analysis"; Poster at [HUGO HGM](#) Conference, Shanghai 2002
- (with Lehväsliho, H.) "Mutable gene model – Computational mutation expression toolkit"; Poster at [ISMB](#) Conference, Copenhagen 2001
- (cited) "[Understanding the Energy Transfer Function of LHCII, the Major Light-Harvesting Complex of Green Plants](#)"; J. Phys. Chem. B, 2001, 105 (3), 604–617
- "[Molecular modelling of plant light-harvesting proteins](#)"; **M.Sc. thesis**, Verona University 1999