

Horizon 2020 Call: H2020-MSCA-RISE-2015 Topic: MSCA-RISE-2015 Action: MSCA-RISE Proposal Number: 690998 Proposal Acronym: *MetaPlat* 

# Table of contents

Section	Title	Action
1	General information	
2	Participants & contacts	
3	Budget	
4	Ethics	
5	Call-specific questions	

### How to fill in the forms?

The administrative forms must be filled in for each proposal using the templates available in the submission system. Some data fields in the administrative forms are pre-filled based on the previous steps in the submission wizard.

Page 1 of 27



Proposal ID 690998

MetaPlat Acronym

# - General information

Topic	MSCA-RISE-2015	Type of action	MSCA-RISE
Call identifier	H2020-MSCA-RISE-2015	Acronym	MetaPlat
Proposal title	Development of an Easy-to-use Metagenomic	cs Platform for Agricultural S	cience
N b	lote that for technical reasons, the following cha e removed: < > " &	aracters are not accepted in	the Proposal Title and will
Duration in months	42		
Panel	LIF - Life Sciences		
Please select up to 5 descriptors (and at least 1) that best characterise the subject of your proposal, in descending order of relevance. Note that descriptors will be used to support REA services in identifying the best qualified evaluators for your proposal.			
Descriptor 1	Genomics, comparative genomics, function	nal genomics	dd
Free keywords	Bioinformatics, metagenomics, microbiome		

#### Abstract

P 0 y

> The aim of this project is to bring together experts from the academic and non-academic sectors and to create an easy-touse integrated hardware and software platform. This will enable the rapid analysis of large metagenomic datasets. It will provide actionable insights into probiotic supplement usage, methane production and feed conversion efficiency in cattle. In the recent years, the number of projects or studies producing very large guantities of seguencing data – analysing microbial communities make-up and their interactions with the environment - has increased. Yet, the depth of analysis done is very superficial and represents an inefficient use of available information and financial resources. This project aims to address these deficiencies and will study the change within microbial communities, under various conditions in cattle guts and impacting probiotic supplement usage, methane production and feed conversion efficiency in cattle. To succeed, we propose to develop faster and more accurate analytic platforms in order to fully utilise the datasets generated. By focusing on better hardware and software platforms, better expertise and training, this project will pave the way for a more optimal usage of metagenomic datasets, thus reducing the number of animals necessary. This will ensure better and more economic animal welfare.

The Meta-Plat project objective is a mixture of innovative research, focused application and commercial awareness. The core objectives being pursued are:

Sample gut collection, from cattle, for sequencing;

Collection of publically available databases - to create a new classification of previously unclassified sequences, using machine learning algorithms;

- Development of accurate classification algorithms;
- Real-time or time-efficient comparison analyses;
- Production of statistical and visual representations, conveying more useful information;
- Platform integration;
- Provide insights into probiotic supplement usage, methan

Remaining characters

0

Page 2 of 27



Proposal ID 690998

Acronym MetaPlat

Has this proposal (or a very similar one) been submitted in the past 2 years in response to a call for proposals under the 7th Framework Programme, Horizon 2020 or any other EU programme(s)?

Declarations

respect.

 1) The coordinator declares to have the explicit consent of all applicants on their participation and on the content of this proposal.
 Image: Content of the content of th

4) The coordinator confirms:

 to have carried out the self-check of the financial capacity of the organisation on https://ec.europa.eu/research/participants/portal/desktop/en/organisations/lfv.html or to be covered by a financial viability check in an EU project for the last closed financial year. Where the result was "weak" or "insufficient", the coordinator confirms being aware of the measures that may be imposed in accordance with the H2020 Grants Manual (Chapter on Financial capacity check); or
 is exempt from the financial capacity check being a public body including international organisations, higher or secondary education establishment or a legal entity, whose viability is guaranteed by a Member State or associated country, as defined in the H2020 Grants Manual (Chapter on Financial capacity check); or
 as sole participant in the proposal is exempt from the financial capacity check.

5) The coordinator hereby declares that each applicant has confirmed:

- they are fully eligible in accordance with the criteria set out in the specific call for proposals; and	$\boxtimes$	
- they have the financial and operational capacity to carry out the proposed action.	$\boxtimes$	
The coordinator is only responsible for the correctness of the information relating to his/her own organisation. Each applicant remains responsible for the correctness of the information related to him and declared above. Where the proposal to be retained for EU funding, the coordinator and each beneficiary applicant will be required to present a formal declaration in this		

H2020-MSCA-RISE.pdf - Ver 1.78 20150427

Page 3 of 27



Proposal ID 690998

Acronym MetaPlat

According to Article 131 of the Financial Regulation of 25 October 2012 on the financial rules applicable to the general budget of the Union (Official Journal L 298 of 26.10.2012, p. 1) and Article 145 of its Rules of Application (Official Journal L 362, 31.12.2012, p. 1) applicants found guilty of misrepresentation may be subject to administrative and financial penalties under certain conditions.

#### Personal data protection

Your reply to the grant application will involve the recording and processing of personal data (such as your name, address and CV), which will be processed pursuant to Regulation (EC) No 45/2001 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data. Unless indicated otherwise, your replies to the questions in this form and any personal data requested are required to assess your grant application in accordance with the specifications of the call for proposals and will be processed solely for that purpose. Details concerning the processing of your personal data are available on the privacy statement. Applicants may lodge a complaint about the processing of their personal data with the European Data Protection Supervisor at any time.

Your personal data may be registered in the Early Warning System (EWS) only or both in the EWS and Central Exclusion Database (CED) by the Accounting Officer of the Commission, should you be in one of the situations mentioned in: -the Commission Decision 2008/969 of 16.12.2008 on the Early Warning System (for more information see the <u>Privacy Statement</u>), or

-the Commission Regulation 2008/1302 of 17.12.2008 on the Central Exclusion Database (for more information see the Privacy Statement).



Proposal ID 690998

Acronym MetaPlat

# 2 - Administrative data of participating organisations

# Coordinator

PIC	Legal name
999885313	UNIVERSITY OF ULSTER ROYAL CHARTER

### Short name: ULster

W

#### Address of the organisation

Street	CROMORE ROAD
Town	COLERAINE
ostcode	BT52 1SA
Country	United Kingdom
/ebpage	www.ulster.ac.uk

### Legal Status of your organisation

#### Research and Innovation legal statuses

Public body	yes
Non-profit	yes
International organisation	no
International organisation of European interest	no
Secondary or Higher education establishment	yes
Research organisation	yes

Legal person .....yes
Academic Sector ....yes

Enterprise Data

SME self-declared status	2010 - no
SME self-assesment	unknown
SME validation sme	unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Nace code 853 -

H2020-MSCA-RISE.pdf - Ver 1.78 20150427

Page 5 of 27



Proposal ID 690998

Acronym MetaPlat

# Department(s) carrying out the proposed work

#### Department 1

Department name	Computer Science Research Institute	
	Same as organisation address	
Street	CROMORE ROAD	
Town	COLERAINE	
Postcode	BT52 1SA	
Country	United Kingdom	

## Dependencies with other proposal participants

Character of dependence	Participant	



Proposal ID 690998

Acronym MetaPlat

#### Contact address of the Host Institution and contact person

The name and e-mail of Host Institution contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and contact details of Host Institution, please save and close this form, then go back to Step 4 of the submission wizard and save the changes. Please note that the submission is blocked without a contact person and e-mail address for the Host Institution.

Organisation Legal Name	UNIVERSITY OF ULSTER ROYAL CHARTER				
First name*	Hiayang	Last r	name*	WANG	
Gender	Male     Female				
E-Mail*	haiyang.wangrise@gmail.co	m			
Position in org.	Principal Investigator				]
Department	Computer Science Research I	nstitute			]
	Same as organisation addre	ess			
Street	CROMORE ROAD				]
Town	COLERAINE			Postcode	BT52 1SA
Country	United Kingdom				
Phone	+442890368908	Phone2/Mobile	+XXXX		XX



Proposal ID 690998

Acronym MetaPlat

# Participant

PIC	Legal name
999974941	THE UNIVERSITY OF EDINBURGH

### Short name: UEDIN

#### Address of the organisation

Street	OLD COLLEGE, SOUTH BRIDGE	
Town	EDINBURGH	
Postcode	EH8 9YL	
Country	United Kingdom	
Webpage	www.ed.ac.uk	
Legal Status of your organisation		

### Research and Innovation legal statuses

Public bodyy	/es	Legal persony	es
Non-profity	/es	Academic Sectory	res
International organisationr	10		
International organisation of European interestr	10		
Secondary or Higher education establishmenty	ves		
Research organisationy	res		

Enterprise Data

SME self-declared status	2007 - no
SME self-assesment	unknown
SME validation sme	2007 - no

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Nace code 853 -

Page 8 of 27



Proposal ID 690998

Acronym MetaPlat

# Department(s) carrying out the proposed work

#### Department 1

Department name	Parallel Computing Centre
	$\boxtimes$ Same as organisation address
Street	OLD COLLEGE, SOUTH BRIDGE
Town	EDINBURGH
Postcode	EH8 9YL
Country	United Kingdom

# Dependencies with other proposal participants

Character of dependence	Participant	



Proposal ID 690998

Acronym MetaPlat

#### Contact address of the participant and main contact

The name and e-mail of Partner Organisation contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and contact details of Partner Organisation, please save and close this form, then go back to Step 4 of the submission wizard and save the changes. The contact person needs to be added as 'Main Contact' for the Partner Organisation.

Organisation Legal Name	THE UNIVERSITY OF EDINB	URGH			
First name*	Terry	Last r	name*	Sloan	
Gender	Male     Female				
E-Mail*	tms@epcc.ed.ac.uk				
Position in org.	Principal Investigator				
Department	Parallel Computing Centre				
	Same as organisation addr	ess			
Street	OLD COLLEGE, SOUTH BRI	DGE			
Town	EDINBURGH			Postcode	EH8 9YL
Country	United Kingdom				
Phone	+441316505030	Phone2/Mobile	+XXXX	XXXXXXXXXXX	XX



Proposal ID 690998

Acronym MetaPlat

# Participant

PIC	Legal name
953569753	NSILICO LIFE SCIENCE LIMITED

Short name: NSILICO LIFE SCIENCE LTD

#### Address of the organisation

Street	GRANGE ERIN LODGE GRANGE ROAD

Town DOUGLAS

Postcode

Country Ireland

Webpage www.nsilico.com

# Legal Status of your organisation

#### Research and Innovation legal statuses

Public bodyno	Legal person	yes
Non-profitno	Academic Sector	no
International organisationno		
International organisation of European interest no		
Secondary or Higher education establishment no		
Research organisationno		

Enterprise Data

SME self-declared status	. 2012 - yes
SME self-assesment	. unknown
SME validation sme	.2012 - yes

Based on the above details of the Beneficiary Registry the organisation is an SME (small- and medium-sized enterprise) for the call.

Nace code 721 -



Proposal ID 690998

Acronym MetaPlat

### Department(s) carrying out the proposed work

#### Department 1

Department name	Development
	Same as organisation address
Street	Grange Erin Lodge, Grange Road
Town	Douglas, Cord
Postcode	0000
Country	Ireland

# Dependencies with other proposal participants

Character of dependence	Participant	



### Contact address of the participant and main contact

The name and e-mail of Partner Organisation contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and contact details of Partner Organisation, please save and close this form, then go back to Step 4 of the submission wizard and save the changes. The contact person needs to be added as 'Main Contact' for the Partner Organisation.

Organisation Legal Name	NSILICO LIFE SCIENCE LIMITED	
First name*	Paul Last name* Walsh	
Gender	Male     Female	
E-Mail*	paul.walsh@nsilico.com	
Position in org.	СТО	
Department	Development	
	Same as organisation address	
Street	Grange Erin Lodge, Grange Road	
Town	Douglas, Cork Postcode 0000	
Country	Ireland	
Phone	353214335572 Phone2/Mobile +xxxx xxxxxxxxx	



Proposal ID 690998

Acronym MetaPlat

# Participant

PIC	Legal name
991821994	FTK FORSCHUNGSINSTITUT FUR TELEKOMMUNIKATION UND KOOPERATION EV

### Short name: FTK

#### Address of the organisation

Street	MARTIN-SCHMEISSER WEG 4
Town	DORTMUND
Postcode	44227
Country	Germany
Webpage	www.ftk.de

### Legal Status of your organisation

#### Research and Innovation legal statuses

Public bodyno	Legal personyes	
Non-profit	Academic Sectoryes	
International organisationno		
International organisation of European interestno		
Secondary or Higher education establishment no		
Research organisationyes		

Enterprise Data

SME self-declared status	2012 - no
SME self-assesment	unknown
SME validation sme	unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Nace code 93 - Other service activities

H2020-MSCA-RISE.pdf - Ver 1.78 20150427

Page 14 of 27



Proposal ID 690998

Acronym MetaPlat

# Department(s) carrying out the proposed work

#### Department 1

Department name	Research		
	Same as organisation address		
Street	MARTIN-SCHMEISSER WEG 4		
Town	DORTMUND		
Postcode	44227		
Country	Germany		

## Dependencies with other proposal participants

Character of dependence	Participant	



#### Contact address of the participant and main contact

The name and e-mail of Partner Organisation contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and contact details of Partner Organisation, please save and close this form, then go back to Step 4 of the submission wizard and save the changes. The contact person needs to be added as 'Main Contact' for the Partner Organisation.

Organisation Legal Name FTK FORSCHUNGSINSTITUT FUR TELEKOMMUNIKATION UND KOOPERATION EV

First name*	Matthias		Last	name*	Hemmje	
Gender	• Male	○ Female				
E-Mail*	mhemmje@ftk	.de				
Position in org.	Professor					
Department	Research					
	Same as or	ganisation addre	ess			
Street	MARTIN-SCHI	MEISSER WEG	4			
Town	DORTMUND				Postcode	44227
Country	Germany					
Phone	+49231975056	60	Phone2/Mobile	+XXXX	XXXXXXXXXXX	XX

Page 16 of 27



Proposal ID 690998

Acronym MetaPlat

# Participant

### Short name: SRUC

#### Address of the organisation

Street	WEST MAINS ROAD
Town	EDINBURGH
Postcode	EH9 3JG
Country	United Kingdom
Webpage	www.sruc.ac.uk

# Legal Status of your organisation

#### Research and Innovation legal statuses

Public bodyno	Legal person
Non-profityes	Academic Sector
International organisationno	
International organisation of European interest no	
Secondary or Higher education establishment yes	
Research organisationyes	

#### Enterprise Data

SME self-declared status	2012 - no
SME self-assesment	unknown
SME validation sme	unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Nace code - Not applicable

H2020-MSCA-RISE.pdf - Ver 1.78 20150427

Page 17 of 27

.....yes



Proposal ID 690998

Acronym MetaPlat

# Department(s) carrying out the proposed work

#### Department 1

Department name	Research & Development		
	$\boxtimes$ Same as organisation address		
Street	WEST MAINS ROAD		
Town	EDINBURGH		
Postcode	EH9 3JG		
Country	United Kingdom		

# Dependencies with other proposal participants

Character of dependence	Participant	



#### Contact address of the participant and main contact

The name and e-mail of Partner Organisation contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and contact details of Partner Organisation, please save and close this form, then go back to Step 4 of the submission wizard and save the changes. The contact person needs to be added as 'Main Contact' for the Partner Organisation.

Organisation Legal Name	SRUC
First name*	Richard Last name* Dewhurst
Gender	Male     C Female
E-Mail*	richard.dewhurst@sruc.ac.uk
Position in org.	Professor
Department	Research & Development
	Same as organisation address
Street	WEST MAINS ROAD
Town	EDINBURGH Postcode EH9 3JG
Country	United Kingdom
Phone	+441316517498 Phone2/Mobile +xxxx xxxxxxxxxx

Page 19 of 27



Proposal ID 690998

Acronym MetaPlat

# 3 - Budget

Staff M	lember	Sending Organisation		Seconded to Organisation				Work	Secondment	Duration of Secondment		
ID	Profile	Short Name	Country	Region	Academic Sector	Short Name	Country	Region	Academic Sector	Package Number	Starting Month	(Researcher- Months)
1	ER	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	SRUC	UK	EU/AC	yes	1	1	3
2	ER	SRUC	UK	EU/AC	yes	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	1	12	12
3	ESR	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	SRUC	UK	EU/AC	yes	1	1	12
4	ER	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	ULster	UK	EU/AC	yes	2	1	6
5	ESR	ULster	UK	EU/AC	yes	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	2	6	12
6	ESR	ULster	UK	EU/AC	yes	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	2	18	6
7	ESR	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	ULster	UK	EU/AC	yes	3	36	2
8	ER	ULster	UK	EU/AC	yes	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	3	12	12
9	ESR	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	UEDIN	UK	EU/AC	yes	4	24	12
10	ESR	UEDIN	UK	EU/AC	yes	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	4	12	12
11	ER	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	FTK	DE	EU/AC	yes	5	30	3

# Table A3.1 – List of secondments

H2020-MSCA-RISE.pdf - Ver 1.78 20150427



Proposal ID 690998

Acronym MetaPlat

Staff N	lember	Sending Org	anisation			Seconded to O	rganisatior	I		Work	Secondment	Duration of Secondment
ID	Profile	Short Name	Country	Region	Academic Sector	Short Name	Country	Region	Academic Sector	Number	Month	(Researcher- Months)
12	ESR	ULster	UK	EU/AC	yes	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	5	24	6
13	ESR	SRUC	UK	EU/AC	yes	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	5	24	12
14	ESR	SRUC	UK	EU/AC	yes	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	5	36	6
15	ER	FТК	DE	EU/AC	yes	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	5	30	12
16	ESR	ULster	UK	EU/AC	yes	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	3	24	12
17	MNG	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	SRUC	UK	EU/AC	yes	6	6	1
17	MNG	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	UEDIN	UK	EU/AC	yes	6	30	1
17	MNG	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	FТК	DE	EU/AC	yes	6	32	1
17	MNG	NSILICO LIFE SCIENCE LTD	IE	EU/AC	no	ULster	UK	EU/AC	yes	6	3	1



Proposal ID 690998

Acronym MetaPlat

# Table A3.2 – Summary of secondments per participant (Beneficiaries + Partner Organisations)

						Estimated	d budget support (v	vhole duration of th	e project)	
Participant Number	Organisation Short Name	Country	Academic	Number of secondments	Person-months	Staff member costs	Research, training and networking costs	Management and indirect costs	Total	Requested EU contribution/€
1	ULster	UK	yes	5	48	96000,00	86400,00	33600,00	216000,00	216000,00
2	UEDIN	UK	yes	1	12	24000,00	21600,00	8400,00	54000,00	54000,00
3	NSILICO LIFE SCIENCE LTD	IE	no	10	42	84000,00	75600,00	29400,00	189000,00	189000,00
4	FTK	DE	yes	1	12	24000,00	21600,00	8400,00	54000,00	54000,00
5	SRUC	UK	yes	3	30	60000,00	54000,00	21000,00	135000,00	135000,00
Total				20	144	288000,00	259200,00	100800,00	648000,00	648000,00

# Table A3.3 – Summary of secondments per EU Beneficiary

						Estimated				
Participant Number	Organisation Short Name	Country	Academic	Number of secondments	Person-months	Staff member costs	Research, training and networking costs	Management and indirect costs	Total	Requested EU contribution/€
1	ULster	UK	yes	5	48	96000,00	86400,00	33600,00	216000,00	40500,00
2	UEDIN	UK	yes	1	12	24000,00	21600,00	8400,00	54000,00	58500,00

H2020-MSCA-RISE.pdf - Ver 1.78 20150427

This proposal version was submitted by Hiayang WANG on 28/04/2015 16:11:39 CET. Issued by the Participant Portal Submission Service.



Proposal ID 690998 Acronym MetaPlat Estimated budget support (whole duration of the project) Participant **Requested EU** Number of Country Academic **Organisation Short Name** Person-months Research, Number Staff member contribution/€ secondments Management and training and Total costs indirect costs networking costs 3 NSILICO LIFE SCIENCE LTD ΙE 10 42 no 84000,00 75600,00 29400,00 189000,00 459000,00 DE 4 FTK 1 12 24000,00 21600,00 8400,00 54000,00 18000,00 yes UK 5 SRUC 3 30 60000,00 54000,00 21000,00 135000,00 72000,00 yes Total 20 144 288000,00 259200,00 100800,00 648000,00 648000,00



Proposal ID 690998

Acronym MetaPlat

# 4 - Ethics

1. HUMAN EMBRYOS/FOETUSES		Page
Does your research involve Human Embryonic Stem Cells (hESCs)?	⊖Yes ⊙No	
Does your research involve the use of human embryos?	⊖Yes ⊙No	
Does your research involve the use of human foetal tissues / cells?	⊖Yes ⊙No	
2. HUMANS		Page
Does your research involve human participants?	⊖Yes ⊙No	
Does your research involve physical interventions on the study participants?	⊖Yes ⊙No	
3. HUMAN CELLS / TISSUES		Page
Does your research involve human cells or tissues (other than from Human Embryos/ Foetuses, i.e. section 1)?	⊖Yes ⊙No	
4. <u>PERSONAL DATA</u> (ii)		Page
Does your research involve personal data collection and/or processing?	⊖Yes ⊙No	
Does your research involve further processing of previously collected personal data (secondary use)?	⊖Yes ⊙No	
5. <u>ANIMALS</u> (iii)		Page
Does your research involve animals?	●Yes ○No	4
Are they vertebrates?	⊙Yes ∩No	4
Are they non-human primates?	⊖Yes ⊙No	
Are they genetically modified? (iv) (regulation)	⊖Yes ⊙No	
Are they cloned farm animals?	⊖Yes ⊙No	
Are they endangered species?	⊖Yes ⊙No	
Cattle		

H2020-MSCA-RISE.pdf - Ver 1.78 20150427

Page 24 of 27



Proposal ID 690998

Acronym MetaPlat

6. THIRD COUNTRIES			Page
Does your research involve non-EU countries?	⊖ Yes	No	
Do you plan to use local resources (e.g. animal and/or human tissue samples, genetic material, live animals, human remains, materials of historical value, endangered fauna or flora samples, etc.)? (v)	⊖ Yes	No	
Do you plan to import any material from non-EU countries into the EU? For data imports, please fill in also section 4. For imports concerning human cells or tissues, fill in also section 3.	⊖Yes	⊙ No	
Do you plan to export any material from the EU to non-EU countries? For data exports, please fill in also section 4. For exports concerning human cells or tissues, fill in also section 3.	⊖Yes	⊙ No	
If your research involves low and/or lower middle income countries, are benefits-sharing measures foreseen? (vii)	⊖Yes	⊙ No	
Could the situation in the country put the individuals taking part in the research at risk?	⊖Yes	No	
7. ENVIRONMENT & HEALTH and SAFETY See legal references at the end of the section. (vi)			Page
Does your research involve the use of elements that may cause harm to the environment, to animals or plants? For research involving animal experiments, please fill in also section 5.	⊖ Yes	No	
Does your research deal with endangered fauna and/or flora and/or protected areas?	⊖ Yes	No	
Does your research involve the use of elements that may cause harm to humans, including research staff? For research involving human participants, please fill in also section 2.	⊖ Yes	⊙ No	
8. <u>DUAL USE</u> (vii)			Page
Does your research have the potential for military applications?	⊖ Yes	No	
9. MISUSE			Page
Does your research have the potential for malevolent/criminal/terrorist abuse?	⊖ Yes	• No	
10. OTHER ETHICS ISSUES			Page
Are there any other ethics issues that should be taken into consideration? Please specify	⊖ Yes	No	



Proposal ID 690998

Acronym MetaPlat

I confirm that I have taken into account all ethics issues described above and that, if any ethics issues apply, I will complete the ethics self-assessment and attach the required documents.  $\mathbf{x}$ 

How to Complete your Ethics Self-Assessment



Proposal ID 690998

Acronym MetaPlat

# 5 - Call specific questions

Open Research Data Pilot in Horizon 2020

If selected, all applicants have the possibility to participate in the <u>Pilot on Open Research Data in Horizon 2020</u><sup>1</sup>, which aims to improve and maximise access to and re-use of research data generated by actions. Participating in the Pilot does not necessarily mean opening up all research data. Actions participating in the Pilot will be invited to formulate a Data Management Plan in which they will determine and explain which of the research data they generate will be made open.

We wish to participate in the Pilot on Open Research Data in Horizon 2020 on a voluntary basis (Yes No

Participation in this Pilot does not constitute part of the evaluation process. Proposals will not be evaluated favourably because they are part of the Pilot and will not be penalised for not participating.

<sup>1</sup>According to article 43.2 of Regulation (EU) No 1290/2013 of the European Parliament and of the Council, of 11 December 2013, laying down the rules for participation and dissemination in "Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020)" and repealing Regulation (EC) No 1906/2006.

#### Data management activities

The use of a <u>Data Management Plan (DMP)</u> is required for projects participating in the <u>Open Research Data Pilot in Horizon 2020</u>, in the form of a deliverable in the first 6 months of the project.

All other projects may deliver a DMP on a voluntary basis, if relevant for their research.

Are data management activities relevant for your proposed project? O Yes O No

Meta-Plat RISE

# **START PAGE**

Marie Skłodowska-Curie Actions

# Research and Innovation Staff Exchange (RISE)

# Call: H2020-MSCA-RISE-2015

# PART B

# Development of an Easy-to-use <u>Metagenomics</u> <u>Plat</u>form for Agricultural Science

"MetaPlat"

This proposal version was submitted by Hiayang WANG on 28/04/2015 16:11:39 CET. Issued by the Participant Portal Submission Service.

#### TABLE OF CONTENTS

LIS	ST OF PARTICIPANTS	3
1	SUMMARY	4
2	EXCELLENCE	5
3	ΙΜΡΑCΤ	16
4	IMPLEMENTATION	22
5	REFERENCES	365
6	CAPACITY OF THE PARTICIPATING ORGANISATIONS	37
7	ETHICS ISSUES	44
8	LETTERS OF COMMITMENT OF PARTNER ORGANISATIONS	45

#### Meta-Plat RISE

#### LIST OF PARTICIPANTS

#### Table B1. Participants table

Participant number	Partnership Member	Legal Entity Short Name	Academic (Y/N)	Country
	Beneficiaries			
1	University of Ulster, Computer Science Research Institute	UU	Y	United Kingdom
2	University of Edinburgh, Edinburgh Parallel Computing Centre	EPCC	Y	United Kingdom
3	NSilico Lifescience Ltd.	NSilico	N	Ireland
4	Scotland's Rural College	SRUC	Y	United Kingdom
5	Forschungsinstitut für Telekommunikation und Kooperation	FTK	Y	Germany
	Partner Organisation			
6	Teagasc	Teagasc	Y	Ireland

# **1** Summary

The aim of this project is to bring together experts from the academic and non-academic sectors and to create an easy-to-use integrated hardware and software platform to enable the rapid analysis of large metagenomic datasets. It will provide actionable insights into probiotic supplement usage, methane production and feed conversion efficiency in cattle.

In the recent years, the number of projects or studies producing very large quantities of sequencing data – analysing microbial communities' make-up and their interactions with the environment – has increased. Yet, the depth of analysis done is very superficial and represents an inefficient use of available information and financial resources. This project aims to address these deficiencies and will study the change within microbial communities, under various conditions in cattle guts and will lead to significant impact on probiotic supplement usage, methane production and feed conversion efficiency in cattle. To succeed, we propose to develop faster and more accurate analytic platforms in order to fully utilise the datasets generated. By focusing on better hardware and software platforms, better expertise and training, this project will pave the way for a more optimal usage of metagenomic datasets, thus reducing the number of animals necessary to be engaged in such studies. This will ensure better and more economic animal welfare.

The *MetaPlat* project's objective is a mixture of innovative research, focused application and commercial awareness. The core objectives being pursued are:

- Sample gut collection, from cattle, for sequencing;
- Collection of publically available databases to create a new classification of previously unclassified sequences, using machine learning algorithms;
- Development of accurate classification algorithms;
- Real-time or time-efficient comparison analyses;
- Production of statistical and visual representations, conveying more useful information;
- Platform integration;
- Provide insights into probiotic supplement usage, methane production and feed conversion efficiency in cattle.

# 2 Excellence

# 2.1 Quality, innovative aspects and credibility of the research (including inter/multidisciplinary aspects)

As will be demonstrated below, this project has all the key input elements required to deliver technological and knowledge sharing outputs of the highest quality. First however, some background discussion is merited in order to contextualise what is being proposed:

#### Background

In the recent years, the number of projects or studies producing very large quantities of metagenomic data has increased [2013-2014: 182 publications, 2014-2015: 226 publications], yet the depth of analysis done is very superficial and represents an inefficient use of the available information and financial recourses necessary to obtain samples.

Metagenomics (also referred to as community genomics) is the genomic analysis of microorganisms by direct extraction from an assemblage of microorganisms. The development of metagenomics stemmed from the ineluctable evidence that as-yet-uncultured microorganisms represent the vast majority of organisms on earth. This evidence was derived from analyses of 16S rRNA gene sequences amplified directly from the environment, an approach that avoided the bias imposed by culturing and led to the discovery of vast new lineages of microbial life. Although the portrait of the microbial world is still limited by analysis of 16S rRNA genes only, such studies yielded a phylogenetic description / profiling of community membership.

*Bos taurus* (cow) gut microbiota has a paramount role in cattle performance, productivity, health and immunity. Understanding the topological difference in gut microbial community composition is crucial to provide knowledge on the functions of each member of microbiota to the physiological maintenance of the host. Metagenomics based on high-throughput sequencing offers unparalleled coverage and depth in determining microbial gut dynamics as long as the analytic resources are available.

*B. taurus* is a member of the ruminant, a group of mammals which also include sheep and goats. The value of domesticated ruminants comes from their ability to convert forages into high quality, high protein products for human consumption through rumen fermentation. The rumen is the first chamber of the ruminant stomach, and it contains symbiotic microorganisms that breakdown ingested food. These microorganisms provide nutrients, such as volatile fatty acids and bacterial protein to the host animal. Many studies have investigated the symbiotic microorganisms in the rumen because of their link to economically or environmentally important traits such as feed conversion efficiency, methane production, and more recently discovery of microbes and enzymes that enable fermentation of biomass for biofuel production. A key challenge is identifying rumen microbial profiles, which are associated, and potentially predictive of these traits.

Thus, in order to investigate *B. taurus* gut microbiota in the context of probiotic supplement usage, methane production and feed conversion efficiency, the careful development of a software platform to analyse fully the data generated is paramount.

Currently, one generic sequence database (DDBJ, EMBL-EBI, NCBI) provides the only source reference sequences to identify and classify sequences. However despite the general effort to curate the sequences, no organised project has tried to classify un-classified groups or reduce the level of redundancy produced. A related problem is the lack of new algorithms and approaches to deal with the huge number of generated sequences. Assigning a sequence to an organism is done by sequence similarity, which is not optimal if the sequences diverge greatly or not enough: *e.g.* if more than one sequence are identical, the assignment is often given by the first sequence name encountered by the algorithm, which can be ambiguous and even missing, producing very biased and non-reproducible results and statistics.

To tackle specifically the gut microbiota role in cattle performance, productivity, health and immunity, new tools to be developed for the overall goal of this project but also better analyses and practice must be developed for future and current studies.

# 2.1.1 Specific objectives and the relevance of the research and innovation project to the scope of the call and in relation to the "state of art"

The proposed project tackles the important area of metagenomics analyses in the agricultural sector. It will to do so by addressing the subject in a holistic way by creating a platform that handles large metagenomics data and produces in-depth analyses and comparisons thereby allowing researchers to make full use of the data generated for each sample. This requires a number of disciplines and skills to be brought together in order to achieve success. Furthermore it requires a mixture of creative research, focused application and commercial awareness, which together will lead to the development of such a platform. The development has been broken down into the following **core objectives**:

- Sample collection preparation, and sequencing
- Curation of the reference databases (phylogeny-aware new classification and previously unclassified sequences using machine learning)
- Development of accurate classification algorithms
- Real-time or time-efficient comparison analyses
- Production of statistical and visual representations conveying more useful information.
- Platform Integration
- Provide insights into probiotic supplement usage, methane production and feed conversion efficiency in cattle

With the foregoing in mind, Meta-Plat brings together five core participants (and one additional partner organisation, Teagasc, who will provide NGS sequencing services and also where necessary provide samples) with the complementary skills, expertise and infrastructure required to address the individual inter- and multi-disciplinary elements of our proposed platform development. SRUC come to this project with a wealth of domain expertise in livestock science. Partner organisation Teagasc has a similar wealth of expertise and in the context of this consortium provides sample processing and sequencing which will be invaluable for insights into probiotic supplement usage, methane production and feed conversion efficiency in cattle. The expertise provided by these consortium members in the agricultural domain is complemented by the expertise, skills and infrastructure provided by the other consortium members in bioinformatics, the computing and IT domains. From the computing perspective EPCC provide the High Performance Computing architecture necessary to process large data sets in short timeframes. At the same time UU have considerable machine learning capabilities and will form a vital component in the overall platform. Similarly, how the platform actually manages these big metagenomic datasets will be informed by the skills and expertise of FTK who are experts in this area. Finally, NSilico provides the ability to develop an easy-to-use front-end software interface as well as the commercial acumen to bring this platform to the market.

The projects objectives will involve significant inter-sectoral collaboration and knowledge transfer between NSilico and the academic partners. Essentially, **NSilico will act as a centralised "hub" to and from which complementary knowledge and skills get transferred between all the partners**. Thus for example the semantic technologies developed by NSilico and UU will be optimised for parallel execution on architectures devised by EPCC's high performance computing centre, embedded at an object level with security protocols devised by FTK and all shall be integrated in a comprehensive health informatics solution developed by NSilico. The samples and genomic use cases provided by SRUC and Teagasc will be integrated in tight cooperation between end users, bioinformaticians and software developers will be achieved to meet domain-specific scientific requirements. The interaction between the partners will be on-going throughout the project due to the highly collaborative nature of the project plan. An agile approach will be taken in many of the project tasks, involving an iterative development of work package deliverables. Moreover, the complementary knowledge, skill sets and experience of the partners are perfectly suited to addressing the objectives of the research programme.

The development has been broken down into five distinct technical Work Packages with two additional work packages focussed on Management and Dissemination/Communication, run over a total of 42 months as summarised in Table B3.

#### Table B3. Work Package List

WP No.	WP Title	Activity Type ( <i>e.g.</i> Research, Training, Management)	Number of person months involved	Start month	End month
1	Metabiome Collection &	Research	27	1	24
	Extraction				
2	Big Data Machine Learning	Research	24	1	24
3	Metagenomics Algorithms	Research	26	12	38
4	High Performance Computing	Research	24	12	36
5	Platform Integration	Research	39	24	42
6	Project Management	Management	42*	1	42
7	Dissemination/Communication	Dissemination/Communication	42*	1	42

\*Note: These are not secondment months

Detailed descriptions of these Work Packages (WP's) is given in Section 4 (Implementation) of this proposal. In working together on these WP's, the partners will acquire specific skills and knowledge from each other (specific details of which are given later in this proposal). Furthermore they will reap the benefits, which accrue from research across the academic/non-academic sectors rather than research within one sector alone – benefits which yield new products and services of economic impact and everyday relevance. In taking this approach, the academic partners will learn how NSilico approach problems with a commercial mind-set and how resources and time are allocated to certain activities in a focussed and applied manner rather than taking a more "blue-sky" approach, common in academia. Researchers from these institutions will also get experience of the demands of real world business practice, and training on practical skills which can be applied to future projects, or make them more appealing to future employers. At the same time, NSilico who have a breadth of knowledge of software development for the human clinical area, have no meaningful experience of parallel computing or the agricultural sector. However through collaborating with the academic partners, NSilico effectively enhance and add depth to their core capabilities, and in doing so build upon their reputation for innovation. As a result, the consortium addresses many of the core principles cited in the Innovation Union initiative including "capitalising on Europe's creative potential" and "pooling forces to achieve breakthroughs", all of which fundamentally underpins Europe's ability to remain competitive. Furthermore, it addresses a core aspect of the RISE programme, which seeks to "align different cultures and expectations, with a view to a more effective contribution of research and innovation to Europe's knowledge economy and society". Moreover, through its intersectoral exchange programme it promotes knowledge transfer between the academic and non-academic sectors fundamentally addresses one of the core principles of the European Charter for Researchers.

#### **Current State of the Art**

Papers by Brulc et al. [1] and Hess et al. [2] were seminal in the development of rumen metagenomics, but involved extremely low numbers of animals. Sequencing costs are reducing rapidly and the ability to sequence either multiple 16S libraries or whole rumen metagenomes has advanced to the point where data analysis capability and acquisition of phenotypes are the major constraint rather than the cost of sequencing. Currently one generic sequence database (DDBJ, EMBL-EBI, NCBI) provides the only source reference sequences and three 16S rRNA gene project provide partially curated dataset based on the DDBJ/EMBL-EBI/NCBI sequences (Table B4). However despite the general effort to curate the sequences and identify and classify NGS data, no organised project has undertaken to classify unknown groups or reduce the use level of redundancy produced.

Databases	Reference	Access	Comments
GreenGenes	[3]	http://greengenes.lbl.gov/	Specific, curated, redundant
SILVA	[4]	http://www.arb-silva.de/	Specific, curated, redundant
Ribosomal Database Project	[5]	http://rdp.cme.msu.edu/	Specific, curated, redundant
DDBJ, EMBL-EBI, NCBI		<pre>http://www.insdc.org/ http://www.ebi.ac.uk/ http://www.ncbi.nlm.nih.gov/</pre>	Generic, no curation, redundant

#### Table B4. Metagenomic databases

Almost all tools currently available (Table B5) are assigning a sequence to an organism by sequence similarity and based on sequence alignment. None are optimised for parallel or cloud execution and most

of them rely on scripts (*e.g.* Perl). This lack of new algorithms and optimised approaches to deal with the huge number of generated sequences is producing biases and non-reproducible results and statistics.

Software	Reference	Access	Comments
RDP suit	[5]	http://rdp.cme.msu.edu/	Single sequence request, blast/alignment based
PhyloSift	[6]	<pre>https://phylosift.wordpress. com/</pre>	NGS data, Phylogeny-driven classification
MGTAXA		http://mgtaxa.jcvi.org/	NGS data, blast/alignment based
PhymmBL	[7]	<pre>http://www.cbcb.umd.edu/soft ware/phymm/</pre>	Discontinued, blast/alignment based
MG-RAST	[8]	http://metagenomics.anl.gov/	Online only, NGS data, blast/alignment based
Metaxa2	[9]	<pre>http://microbiology.se/softw are/metaxa2/</pre>	NGS data, Perl scripts, blast/alignment based
MetaPhlAn	[10]	<pre>http://huttenhower.sph.harva rd.edu/metaphlan/</pre>	NGS data, blast/alignment based
Parallel-META	[11]	<pre>http://www.computationalbioe nergy.org/parallel-meta.html</pre>	NGS data, blast/alignment based
MetaPhyl		http://www.cs.ucr.edu/~tanas eio/metaphyl.htm	Required training data set, k-mer based
CLcommunity		http://www.chunlab.com/softw are_clcommunity_about/	Licenced, NGS data, blast/alignment based

Table B5. Metagenomic software tools available

Finally, representing the results for dissemination is also an untold challenge. Today, only one tool uses by almost all studies, Krona [12], which is a valuable interactive generic hierarchical data browser, but does not convey all degrees of information.

#### Innovative and Original aspects of this project

Our approach aims to marry complementary innovation activities to address the limitations in the current state of the art described above, and thus underline the overall quality of the proposed research programme. In particular, the following innovations are noteworthy:

- Development of non-redundant sequence databases that will encapsulate all current available metadata (species and sub-species/strain, number of sequence) as well as all previously unclassified sequences using machine learning. This will includes a procedure to update the data with newly released information and a public dissemination of the resulting database.
- Development of accurate and reproducible classification algorithms, optimised for parallel and cloud execution. Code will be open source.
- Real-time or **time-efficient comparisons** and **analyses** on large datasets producing phylogenyaware classification but also **quantitative** and **functional analyses** (when possible).
- Production of statistical and **visual representations** maximising information used and conveying more meaning that current visualisations.
- Readily perform high quality analyses and comparisons on **public databases** as well as **user-own dataset**, in an integrated environment (*e.g.* Galaxy [13] web-based platform).

#### 2.1.2 Methodology

Our methodology will interlink diverse research elements: biological samples, big data management, machine learning, search algorithm development, HPC platforms and integrated software development.

Initially, the samples collection and sequencing will occur in parallel to the reference databases preparation and algorithms development.

Rumen fluid samples will be collected using a naso-gastric sample tube system at SRUC. DNA isolation will use the bead-beating procedure of Morrison and Yu [14]; 16S amplicons will be prepared using Caparoasa primers, and sequencing of both 16S amplicons (in parallel) and total genomic DNA accomplished using Illumina sequencers (MiSeq; HiSeq).

A formal software development approach will be used, starting with careful analysis requirements by engaging with researchers. The core effort will be to develop thread safe extensible bioinformatics algorithms for HPC architectures using ISO/IEC 12207 an international standard for software life-cycle processes, which defines best practices for developing and maintaining software. A process will be put in place to identify the most scalable parallel computing architectures and the appropriate functional partitioning strategies to achieve the most optimal solution by maximising data locality and minimising inter-process communication. This activity will be highly inter-disciplinary, as the parallel computing

algorithms derived by EPCC will be complemented in NSilico by the implementation of these algorithms in a platform integration system with the help of UU. This synergistic activity will result in a highly optimised software system that brings high performance computing. NSilico will make this cutting edge technology practical, accessible and useable. The multi-disciplinary consortium will also utilise and align with data formats and ontological models of the field, so as to create compatible and re-usable tools and services.

All of the above methods will be rigorously benchmarked and evaluated in a series of carefully designed experiments.

We will also employ a science-based quality risk management methodology, which allows computer system developers to focus on crucial facets and to develop contingencies to mitigate emergent risks. Software components to analyse and integrate metagenomic analysis will be parallelised across a multi-processor topology to allow large scale "big data" processing while minimising privacy risks. This implementation will be iterative across work packages and will follow the principles of agile development resulting in numerous prototype versions of the parallelised software services.

The efficacy of the output of the project will be established during a last phase of the project, during the assessment phase, as this work will be completed by NSilico and SRUC with the samples collected and generated by SRUC and Teagasc.

#### 2.1.3 Inter-/multi-disciplinarily aspects

The consortium members have been carefully selected so that the partners tightly complement each other in achieving the stated objectives. The disciplines central to achieving the objectives are high performance computation (EPCC), functional software development (NSilico), Next Generation Sequencing (Teagasc), Machine Learning (UU), Big Data Management (FTK) and livestock science (SRUC). Each partner has a highly credible track record in its cognate discipline:

- EPCC hosts the UK's leading supercomputer, making it a world leader in high performance computing.
- NSilico is the developer of web based clinical and genomic software applications and focuses in particular on easy-to-use interfaces, which enable non-bioinformaticians to rapidly and easily process raw genomic data.
- Teagasc is the Irish national body providing integrated research, advisory and training services to the agriculture and food industry and rural communities. It is particularly strong in the area of Next Generation Sequencing (NGS) and although not a beneficiary in this project has agreed to provide this support to the consortium (see letter in Section 8 of this application).
- SRUC has extensive experience in animal science procedures and is one of the UK's leading providers of education and business support for land-based activities.
- UU provides expertise on data analytics including data integration, data mining, knowledge discovery with application to omics data and other healthcare data. It also provides a platform of conducting research and development on the applications of connected health at a scale that is industrially and clinically-relevant and compatible with the types of design, production and manufacturing.
- FTK is as an interdisciplinary research organisation closely related to the FernUniversität in Hagen (FUH). The work of FTK Research Institute for Telecommunication and Cooperation concentrates on the area of Information and Communication Technologies (ICT) for research data management and collaboration. These services have originally been developed in the big data SMART VORTEX project.

At its core, *Meta-Plat* represents an opportunity to build a number of relationships between the consortium members – with different disciplinary expertises. This will be done on both a bi-lateral and a multi-lateral basis. The inter-and multi- disciplinary nature of this project means that the researchers and their respective organisations can attain new and complementary skills and expertise which add to their current capabilities and which open up new avenues of research and innovation. This will in turn lead to further opportunities for commercialisation and collaboration with partners not in this particular consortium.
**2.2 Clarity & quality of knowledge sharing among participants in light of research/ innovation objectives** In order to achieve the research & innovation objectives set out above, significant contributions from each partner are required in terms of the expertise, know-how and skills to be shared. With this in mind, a clear and robust methodology is required to ensure successful exchange of knowledge. This is described in detail below:

#### 2.2.1 Approach & methodology for knowledge sharing in light of the research & innovation objectives

Before considering *how* the management of knowledge sharing will work, it is important to be clear on the overall objectives of the knowledge sharing. This project sets out the following knowledge-sharing objectives:

- Create an environment in which researchers, through pursuing the technical project objectives, develop their careers by building upon their existing skills, facilitating them to acquire new skills and work practices and gaining multi-disciplinary, inter-disciplinary and inter-sectoral expertise.
- Increase mobility of researchers between industry and academia.
- Provide academics an opportunity to learn best practice from the commercial sector and similarly to enable researchers in the commercial sector to get first-hand experience of working in a cutting-edge research environment with world-leading scientists.
- Exploit complementary expertise in each of the consortium partners to deliver the technical objectives of this project whilst also building future research and commercial opportunities.
- Boost the reputation of each of the consortium partners by building upon their existing knowledge base and use this increased reputation to attract new researchers to join their teams, who will in turn acquire and bring new knowledge.
- Cement a relationship between partners, which facilitates them to access each other's extended network of collaborators who could provide significant opportunity for further knowledge exchange in the future.
- Improve innovation and competitiveness at a European level by creating a globally applicable platform which will be achieved through the knowledge sharing required for the delivery of the technical objectives described above.

The above knowledge sharing objectives will be achieved primarily through secondment of experienced and early stage researchers between consortium members.

Specifically, with regard to the partners from each sector, NSilico (the non-academic partner in this consortium) is a life-science software company that is focused on the development of advanced software systems for molecular biology and NGS whole genome assembly pipelines. NSilico already provides a software product (*Simplicity*<sup>™</sup>) as an easy-to-use, cloud-based high performance system for automatic genome assembly and annotation pipelines. NSilico has a strong focus on developing powerful systems by leveraging the scalability of cloud computing technology, while at the same time providing rich, ubiquitous and seamless interfaces that are usable by non-computing specialists. Moreover, NSilico is officially partnered with three major technology corporations, namely Dell, Microsoft and IBM.

Through the exchange of researchers to and from NSilico, the other consortium partners will gain the following specific expertise and knowledge:

- Expertise in development of complete and robust software products that are secure and provide traceability and data provenance;
- Expertise in developing integrated bioinformatics pipelines that perform analysis on integrated data sets;
- Expertise in usability and user centred software design so that its partners can up skill in the development of software products that are usable by non-computing specialists;
- Expertise in cross platform application development;
- Expertise in highly scalable computing platforms that can meet the high resource demand presented by the computational challenges of the objective.

Big Data and Machine Learning expertise is lead by Dr. Haiying Wang. He has extensive research and development experience for advanced machine Learning applications in various application domains. His research expertise spans integrative data analysis in systems biology, pattern recognition and machine learning, gene ontology, data representation and complex network analysis. By exchanging researchers to

#### Page **10** of **46**

and from Dr. Wang's team, NSilico (and indirectly the other partners) will gain the following specific expertise:

- An integrative, multiplex network based approach to extracting disease signatures from heterogonous biological and medical data;
- Bayesian approaches to data integration;
- Identification of functional modules from biological networks;
- Semantic and ontology analysis;
- Predictive modal and pattern recognition.

The Beef and Sheep Research Centre headed by Prof. Richard Dewhurst has strong skills in animal science procedures – e.g. measuring feed conversion efficiency, methane emissions and gut health. They are very experienced in digesta and tissue sampling and storage, DNA/RNA isolation, quality control and animal genetic and genomics. By exchanging researchers to and from Prof. Dewhurst's team, NSilico (and indirectly the other partners) in the consortium will gain the following specific expertise:

- Issues and questions about regulation and consequences of differences in the intestinal microbiome of commercially important livestock;
- Sampling strategy;
- DNA/RNA isolation;
- Animal nutrition and the microbiome;
- Genetic influence of the host (animal) on the microbiome;
- Prediction of methane based on the Archaea/Bacteria ratio;
- Prediction of feed efficiency and animal health using the microbiome.

EPCC is one of Europe's top supercomputing centres and provides world-class computing facilities for scientists and collaborates with researchers in wide ranging areas. The EPCC HECTOR project at the University of Edinburgh provides supercomputing services using a 360-teraflop Cray XE6 system, the fastest supercomputer in the UK. Their work not only benefits scientific research, but generates new techniques and approaches that are taken up by industry. EPCC provide companies of all sizes with access to novel computing expertise and specialist computing facilities: from enterprise-level servers to world-scale high performance computers. EPCC also offers training for the next generation of computational science professionals. EPCC has led a European Commission-funded visitor programme called TRACS, through which over 400 researchers were able to use HPC systems to further their research. EPCC, which specialises in **hosting** such researchers, is an ideal computing hub for the project, in which semantically aware high performance software can be designed, implemented and tested on state of the art infrastructure. Through exchange of researchers to and from EPCC, NSilico (and indirectly the other consortium partners) will gain specific expertise in:

- Scalable parallel computing architectures;
- Functional partitioning strategies;
- Solution design that minimises inter-process communication;
- Mapping strategies that take full advantage of target architectures;
- HPC Implementation of software pipelines;
- HPC Implementation of ontology networks.

FTK was foundation as an interdisciplinary research organisation closely related to the Fern Universität in Hagen (FUH) in 1991, the work of FTK Research Institute for Telecommunication and Cooperation concentrates on the area of Information and Communication Technologies (ICT). An interdisciplinary team with well-founded research background paired with an expanded network to all important stakeholders in the field of Information and Communication Technologies makes FTK a leading institution for the development and implementation of innovation strategies for business and administration in the ICT area. The Institute is organized as a non-profit membership organisation and is pursuing objectives in the public interest. FTK is related to Fern Universität as a so-called "Aninstitut", an associated but independently run institute. The FTK will bring in services for collaboration, authentication, authorisation and Big data management for metagenomics. These services have originally been developed in the big data SMART VORTEX project. In both, SHAMAN and SMART VORTEX FTK has worked extensively with the engineering industry implementing collaboration and preservation solutions in complex cross-organizational scenarios.

#### Page **11** of **46**

Teagasc is the Irish national body providing integrated research, advisory and training services to the agriculture and food industry and rural communities and will be used as a partner organisation (i.e. not directly a beneficiary) in this project to provide NGS (Next Generation Sequencing )services which will be essential to the overall methodology of the project.

This will be enhanced further through the provision of a series of technical and general initiatives at both a consortium and local level (described below) designed to facilitate knowledge sharing on a collective basis across the lifetime of the project. In accordance with the main thrust of the RISE programme, these initiatives are designed to promote "international and inter-sectorial collaboration... from research to market (and vice versa) for the advancement of science and the development of innovation." These initiatives will be undertaken in the context of an overall knowledge sharing & training plan detailed below.

It is noteworthy that although the RISE programme does not fund secondments between the academic partners, the what the consortium is constructed, with NSilico acting like a centralized hub, will still enable significant direct face-to-face interactions between the academic partners as there is considerable overlap of secondments from different academic partners into NSilico at various stages during the project.

#### 2.2.2 Knowledge Sharing & Training Plan

**1.** Knowledge-Sharing Oversight Group. There will be a sub-committee of the Project Management Board (see section 4 – Implementation) who will form a Knowledge-Sharing Oversight Group to ensure that the requisite knowledge, skills and know-how are successfully transferred from, and acquired by, the seconded researchers and subsequently assimilated by both the sending and host organisations. The Knowledge-Sharing Oversight Group will consist of one academic (Dr. Richard Dewhurst, SRUC) and one non-academic (Dr. Paul Walsh, NSilico) and will be responsible for execution and supervision of the various initiatives described below.

**2. Personalised Knowledge Sharing Plan for Each Researcher.** When any new secondee starts work, a Knowledge Sharing Plan will be drawn up which will clearly outline the key <u>skills to be transferred</u> by the researcher to the host <u>and to be acquired</u> by the researcher from the host, thus ensuring one of the core objectives of the RISE programme, namely that each secondment will "provide benefit to both the seconded staff member ... and the host organisation". The plan will also specify what practical steps can be undertaken to ensure that each researcher and their host have fully acquired/transferred the envisioned knowledge/skills. From the researcher's perspective, such steps may include showing their ability to teach someone else the skills that they have acquired. From the host's perspective, it may include a demonstration of their ability to carry out a specific operation, which prior to the secondment of the researcher they would not have been able to do. The progress of each researcher against their own personalised knowledge-sharing plan will be monitored by the Knowledge Sharing Oversight Group to ensure that the researchers are meeting their own objectives and that these objectives fit with their career plans.

3. Workshops. The Meta-Plat consortium partners will convene workshops consisting of all the consortium members. These will occur every 6 months with the venue rotating between all partners. These workshops will enhance efficient knowledge transfer by facilitating partners to meet collectively rather than just bilaterally via the secondments. These workshops will also provide opportunities to discuss other potential areas of collaboration. In addition to discussing the specifics of the Meta-Plat project, each partner organisation will give a technical module related to keys skills/expertise relevant to the technical deliverables of this project. This will ensure that even those researchers who are not directly exposed to these technical elements through secondments will benefit and expand their own technical horizons. In addition, NSilico will give modules on transferrable skills such as technology commercialisation, project management, licensing, marketing, and communication of technical information to non-technical audiences. These will complement and build upon the technical skills development which the researcher will experience, thus promoting "the right mix of skills... promoting entrepreneurship, self employment and innovation" as called for in the Europe 2020 Flagship Initiative. The specific technical and transferrable skills short courses modules delivered at any given workshop will be dictated in advance by the researchers attending - taking into consideration which ones are best suited to their own individual continuous professional development and career plans. Therefore, the workshops will enable the researchers to

benefit to the maximum from being involved in this project with respect to sharing and transferring knowledge and skills to and from one another and thus impacting positively on their respective careers. In order to further supplement the benefit of the secondments and enabling the researchers to be equally comfortable in the research and innovation environments of both academic and non-academic sectors (thereby meeting the European Charter for Researchers stipulation that "employers and/or funders must recognise the value of... mobility between the public and private sector as an important means of enhancing scientific knowledge and professional development at any stage of a researcher's career"), these workshops will explore the differences in work-practices, project management and decision making and how this can impact on the success or otherwise of inter-sectoral projects. In this regard, previous Marie Curie fellows involved in inter-sectoral research (there are 4 in NSilico, and 2 in EPCC, 1 in UU, 5 in Teagasc, 3 in FTK and 2 in the group at SRUC) will be invited to attend these workshops and give the benefit of their experience to the researchers involved in this project.

**4. Reintegration Plan.** When a seconded researcher leaves their "host" organisation it will be important for them to assimilate and embed the knowledge and skills, which they acquired in the host organisation back into their "home" organisation. To ensure this, they will provide a written report of their time spent on secondment, which details the key points of learning. Furthermore, in order to ensure that the skills and knowledge are embedded and disseminated through their home organisation, they will also be required to train and teach individuals in their home institution in the skills and knowledge which they acquired during their secondment. They will also be encouraged to actively seek out new areas of research and collaboration which can be undertaken with other academics/companies using their new found knowledge and skills-sets.

5. Knowledge exchange and sharing with external parties. In addition to networking events within the consortium, it is planned that the researchers and their host organisations will acquire and transfer knowledge to/from parties who are not in this consortium. This will be achieved in two main ways. Firstly by organising a International Symposium in Cork, to which leading experts (both commercial and academic) in the field of Metagenomics will be invited to participate. In this context it is notable that Paul Walsh from NSilico has organised and run a number of Symposia of this type – both in FP7 (ClouDx-i) and also the Conference on European Research Collaboration (CERC). Furthermore, large animal feed companies such as Gain Feed (part of Glanbia plc) have operations nearby, will be requested to sponsor and promote the event. Symposia of this type give each of the researchers the opportunity to interact and learn from invited guests including industrial, academic and clinical practitioners who will have different points of view and can share the benefits of their experience and expertise with the researcher. This will enable new ideas and knowledge to infuse the consortium as well as providing an opportunity for the consortium partners to transfer knowledge outwards and in doing so to create new opportunities for collaboration and knowledge transfer in the future. It is planned that the International Symposium would take place in the last month of the project when there are sufficient results and outputs of the project as a whole to be presented.

Secondly, with respect to knowledge sharing with external parties, the researchers will be encouraged to attend international conferences such as:

- RECOMB (Annual International Conference on Research in Computational Molecular Biology)
- IWBBIO (International Work-Conference on Bioinformatics and Biomedical Engineering)
- NGS (Applications and Data Analysis: Conference & Workshop)
- BBC (Workshop on Biomedical and Bioinformatics Challenges for Computer Science)
- PBio (IEEE ISPA Workshop)
- CIBB (Computational Intelligence methods for Bioinformatics and Biostatistic)
- EMBL Symposium: New approaches and concepts in microbiology
- EMBL Conference: The Human Microbiome
- Wellcome Trust: Next Generation Sequencing Conference

These will provide opportunities to interact with leading experts from around the world and to stimulate discussion, exchange ideas, seed collaborative opportunities for the future.

Page 13 of 46

#### Quality of the interaction between the participating organisations 2.3

The quality of interaction and exchange and transfer of knowledge, skills and expertise between partners is crucial to the attainment of the ambitious objectives of this project. The Meta-Plat consortium is keenly aware of this and below we describe how we will address this key issue.

	UU transfers	SRUC transfers	NSilico transfers	FTK transfers	EPCC transfers
to UU			Bioinformatics/Software		
to SRUC			Bioinformatics/Software		
to NSilico	Machine Learning	Animal models		Data Management	Parallel Computing
to FTK			Bioinformatics/Software		
to EPCC			Bioinformatics/software		

Table B6. Overview of key elements of direct knowledge exchange between partners

#### Contribution of each participant in the activities planned, including the participants' interactions 2.3.1 in terms of content and expertise provided to reach the project's objectives

At the core of this project is the desire to create a platform that enables the processing in real-time of large datasets and to produce phylogeny-aware classification, quantitative and functional analyses for the project, but also reusable algorithms and methodology for future studies. It is not really conceivable that the diverse areas of knowledge and expertise needed to do this would be available in a single research group and thus in order to deliver upon the project objectives it is necessary for close interaction to transfer these complementary expertise between the partners. NSilico, EPCC, UU, SRUC, FTK and Teagasc have between them a spectrum of skills and expertise which perfectly complement each other and when combined create synergies which will ensure that the goals of the research programme are met. A number of complementary approaches to sharing of knowledge between the partners have been planned. As mentioned in the previous section, each of the consortium members will play an integral part in knowledge and expertise sharing through the initiatives detailed in the Knowledge Sharing and Training plan. However, the most effective and sustained knowledge, skills and experience sharing will be achieved by means of secondments.

Using secondment of researchers between partners as the primary "knowledge sharing vector", the "host" organisations acquire and embed new skills, knowledge and experience which they do not currently have. Furthermore, the seconded researchers will acquire expertise, knowledge and skills from their "hosts" which they can bring back to their "home" institutions and assimilate. Therefore, the two-way exchange of knowledge and skills is of significant benefit to both the individual researchers but also their "host" and "home" organisations.

There will be 15 intersectoral exchanges in total between the academic and non-academic sector. As described in detail in the table below, these exchanges are required for the successful completion of WP's 1 to 5. In the context of the interaction between all the participants, the table below details the "fit" between the researchers for secondment and the expertise required.

WP1 Metabiome collection & Extraction								
Core Objective	Sample collection preparation, and sequencing							
n order to plan the experimental sampling, expertise in bioinformatics and animal models of methane production and								
food production are needed	d. SRUC have expertise in animal models but do not have expertise in bioinformatics and							
therefore SRUC will exchange	ge one researcher each with NSilico who have expertise in bioinformatics to learn these							
skills in bioinformatics whil	e the researcher from NSilico will learn about the "wet" aspects of these activities i.e.							
sample collection and proce	sample collection and processing with a view to completing this WP. NSilico will also send a bioinformatician in the							
early stage of the experimental design, in order to learn about metabiome extraction and ensure good integration								
between experimental const	traints and bioinformatic analyses throughout the project.							
WP 2	Big Data Machine Learning							

Table B7. Specific knowledge sharing objectives required to deliver the core technical objectives					
WP1	Metabiome collection & Extraction				

In order to organise and to distribute the reference data expertise in machine learning and bioinformatics and software development is needed. UU fellows have expertise in machine learning but do not have expertise in commercial bioinformatics and therefore UU will send two researchers to NSilico who have expertise in commercial bioinformatics and NSilico will send one researcher to learn about machine learning with a view to completing this WP.

WP3 Metagenomics Algorithms								
Core Objective	Development of accurate classification algorithms							
In order to develop efficient and effective metagenomics algorithms, expertise in bioinformatics and machine learning								
is needed. NSilico have expe	ertise in bioinformatics but do not have expertise in machine learning. NSilico will host a							
researcher from UU who ha	s expertise machine learning and will send a researcher to UU who has expertise in high							
performance computing to	earn these skills with a view to completing this WP and to transfer skills and knowledge							
between parties.								
WP4	High performance computing							
Core Objective	Real-time or time-efficient comparison analyses							

Production of statistical and visual representations conveying more useful information In order to optimise the software and algorithms for high performance computing, a combined expertise in high performance computing and metagenomic analyses is needed. EPCC have expertise in high performance computing but do not have expertise in metagenomic analyses. EPCC will send a researcher to NSilico to learn these skills with a view to completing this WP.

WP5	Platform Integration
Core Objective	Integration of Meta-Plat platform components
	Provide insights into probiotic supplement usage, methane production and feed
	conversion efficiency in cattle

In order to integrate and manage the data and resources generated, expertise in software integration, big data management, animal models and bioinformatics is needed. NSilico have expertise in software integration and bioinformatics but do not have expertise in big data management or animal models. NSilico will host a researcher from FTK who has expertise in big data management to learn these skills with a view to completing this WP. SRUC will send two researchers to NSilico to integrate animal models into the platform and UU will send a researcher to NSilico that has expertise in biostatistical analyses.

#### 2.3.2 Justification of the main networking activities

The networking activities described above will lead to a number of highly impactful outcomes which will be felt far beyond the immediate technical scope of this project. In particular the comprehensive knowledge sharing and training plan described above will:

- Develop the careers of the participant researchers by building their inter- and multidisciplinary skills, and facilitating work practices which are equally applicable in both industry and academia.
- Increase researcher mobility between industry and academia, and facilitate better understanding of the various issues at play in each sector in the area of metagenomics in general.
- Increase commercial acumen of academic researchers & enable commercially based researchers to get first-hand experience of working in a creative, cutting-edge research environment.
- Help build future research and commercial opportunities by exploiting complementary expertise in each of the consortium partners.
- Enhance each of the consortium partners' reputations by building upon their existing knowledge base, and use this increased reputation to attract new researchers who will in turn acquire and bring new knowledge.
- Solidify a mutually beneficial collaborative relationship between the partners, which allows them to reach into each other's extended network of contacts and collaborators who can also provide future opportunities for collaboration and knowledge transfer.

All of the above bullet points address key issues which have been cited in one form or another within The European Charter for Researchers, the European 2020 Flagship Initiatives – "Innovation Union", "Agenda for new skills and jobs" and "Youth on the Move" and are thus wholly justified. Also it to be noted in the context of justifying the networking activities, is that in order to achieve breakthroughs and significant developments in personalised medicine and health informatics, the type of inter-sectoral and interdisciplinary collaboration facilitated by the networking activities described herein is crucial.

Europe's ability to sustain its preeminent position as a leader in R&D is increasingly dependent upon socalled *Open Innovation* and close synergistic relationships between researchers in the public and private sectors. This interdependence brings benefits to both parties. From an industry perspective, partnering with academia facilitates product portfolio diversification into unmet needs without risking critical levels of capital. It also enables the opportunity to work collaboratively with leading academic researchers and to perform research in a pre-competitive environment of open data exchange. Meanwhile, the academic partner benefits from the expertise, resources and development capabilities of industry. The proposed networking activities thus encourage this type of Open Innovation environment.

### 3 Impact

- **3.1** Enhancing research- and innovation-related human resources, skills, and working conditions to realise the potential of individuals and to provide new career perspectives
- **3.1.1** The research and/or innovation project contribution to realising the potential of individuals providing new skills and career perspectives

Metagenomics has the potential to provide unprecedented insight into the form and function of heterogeneous communities of microorganisms and their vast biodiversity, without the need for isolation and lab culture of particular organisms. Microbial communities affect human and animal health, support the growth of plants, are critical components of all terrestrial and aquatic ecosystems and can be exploited to produce fuels or chemicals. Metagenomics, thus pervades a number of hugely important industries central to economic growth and employment. Although this project focuses primarily on animal health, many of the principles learned can be applied across this range of industries and so the project will enable the maximum impact on immediate and long-term career prospects of the researchers – whether they pursue career paths in academia or the private sector, or indeed if they decide to become entrepreneurs. It does so through a coherent blend of (i) participation in a cutting-edge R&D project in which there is interand multi-disciplinary research and knowledge exchange achieved through international secondment, with (ii) formal training in transferrable skills and competences. It thereby addresses key European policy objectives mentioned previously, including the Agenda for New Skills & Jobs, which amongst other things cites the need to:

- Improve flexibility and security in the labour market ("flexicurity");
- Equip people with the right skills for the jobs of today and tomorrow;
- Enhance geographical mobility throughout the EU;
- Promote entrepreneurship, self-employment and innovation.

Through its inter-national, inter-sectoral and inter-/multi-disciplinary strands, *Meta-Plat* will help develop and add to the pool of researchers and innovators who are equally "at home" in the commercial or academic realms. Moreover, in enabling the researchers to address the project's ambitious research and innovation objectives and participate in training and knowledge exchange of the highest quality, Meta-Plat decisively rises to the challenge posed in the European Charter for Researchers which says that the shortage of researchers of the requisite calibre "will pose a serious threat to EU's innovative strength, knowledge capacity and productivity growth in the near future and may hamper the attainment of the Lisbon and Barcelona objectives". Furthermore, the inter-sectoral and international secondments planned in this project play to the objectives of both the "New Skills and Jobs"<sup>1</sup> and the "Youth on the Move"<sup>2</sup> flagship initiatives of the Europe 2020 Agenda<sup>3</sup> which focus on "…strong development of transnational learning and employment mobility for young people…" and underscore the need for "researcher mobility across sectors countries and disciplines…" and seek "…stronger links between education, research and innovation…" and "…equip researchers with relevant skills that will match both public and private sector needs…"

The European Council has also repeatedly emphasised the need to "...ensure a better transfer of information on labour market needs and to provide a better match between those needs and the

<sup>&</sup>lt;sup>1</sup> http://ec.europa.eu/social/main.jsp?langId=en&catId=958

<sup>&</sup>lt;sup>2</sup> http://ec.europa.eu/youthonthemove/

<sup>&</sup>lt;sup>3</sup> http://ec.europa.eu/social/main.jsp?catId=956&langId=en

*development of knowledge, skills and competences...*"<sup>4</sup> whilst the RISE guide for applicants cites the need for RISE projects to promote *"sharing of knowledge and ideas from research to market (and vice versa)"*. These very issues are addressed *via* the knowledge and skills exchange in this project where partners in both sectors are kept abreast of the latest market and technological trends.

The *Meta-Plat* consortium is perfectly suited to addressing these issues by facilitating partners to work together closely. It is also noteworthy that this project is at the cutting-edge of metagenomics at a time when metagenomics is really beginning to gain very significant traction in the scientific and commercial communities. Not only does this put the individual researchers in this project at a distinct advantage, it also significantly benefits their organisations by enabling industry partners to serve as a conduit of information to the academic partners for the market's current and future needs in this area, and in turn helps the academic partners develop a coterie of graduates with the requisite skills to meet these needs.

The training in transferrable skills given at the project workshops will also significantly enhance the capabilities and career potential of the researchers involved in this project. This will be further cemented by the day-to-day exposure to research and innovation in a different environment than the researchers normally operate in. For example, those researchers seconded to the academic partners will be exposed to research and innovation within creative cutting-edge environments with world-class infrastructure at their disposal. At the same time, those researchers seconded to the industry partners will immerse themselves in the entrepreneurial culture of working in an innovative SME seeking to address global markets. These complementary experiences will inculcate in both cohorts a new sense of creativity, entrepreneurialism and appreciation of the issues and work practices pertaining to each sector and thus give them an inherent flexibility which will sustain them in whichever direction their career leads them.

It is also planned during this project to publish a number of peer-reviewed publications. The highly innovative nature of the planned research means that many of these publications will be in high impact journals. These will add significant "weight" to each Researcher's *Curriculum Vitae* when seeking future employment.

### 3.2 To develop new & lasting research collaborations, to achieve transfer of knowledge between

# research institutions & to improve research & innovation potential at the European & global levels 3.2.1 Development of new and lasting research collaborations resulting from the inter-sectoral and/or international secondments and the networking activities implemented

It is notable that there are some existing bilateral relationships between NSilico and each of the other parties, but none of others have previously worked together. In coming together collectively under the umbrella of this consortium, this project gives all the partners the forum to work collaboratively and forge new and lasting relationships formed on *inter*-sectoral, *inter*-disciplinary and *inter*-national linkages across Europe. Thus it serves as a means for each consortium member to tap into the extended network of the other parties to build upon existing collaborations and create new ones. This in turn positively impacts the long-term sustainability of the partnership, described in section 3.2.2 (below). Notwithstanding the fact that this is a newly formed consortium, it represents the first step in an ongoing sustainable set of beneficial inter-relationships between participants. Meta-Plat will have a significantly beneficial impact upon the culture of research in both the public and private sector partner organisations and will underscore the advantages of a more collaborative innovative approach to undertake and exploit research activities an approach which avoids researchers from each sector working in a silo, but rather promotes working collaboratively taking the best practices from public and private sector in order to develop skills which are transferable. Traditionally academics just collaborated with other academics and true collaboration with industry partners was a rarity. This often resulted in solutions being developed for problems which did not really exist or technologies for which there was no real commercial potential. As a result, both the academic and industry sector lost out. This can largely be attributed to a lack of real understanding (from both sides) of the trends, culture and work practices in operation in the other sector. But it has had a significant negative impact insofar as it has meant that significant opportunities for job creation have been lost and Europe lagged behind other economies like the US and Japan in terms of commercialising research

<sup>&</sup>lt;sup>4</sup> 3164th EDUCATION, YOUTH, CULTURE and SPORT Council meeting Brussels, 10 and 11 May 2012.

activities. In turn, this places significant pressure upon funding agencies to justify continued expenditure on Research and Innovation when (particularly during an economic downturn) there are other competing priorities for tax payers' money.

However, the culture change fostered and built upon by this project through the planned secondments and networking, training, dissemination and outreach activities will help to further bring down these barriers which previously inhibited collaboration and commercialisation, and in doing so, addresses a key action point of the Innovation Union agenda, namely *"removing obstacles to innovation"*. It will thus help foster *"openness and capitalise on Europe's creative potential"* thereby ensuring the maximum benefit to the researchers themselves, their host institutions and the European economy as a whole.

#### 3.2.2 Self-sustainability of the partnership after the end of the project

As described later, a dedicated Project Manager will be appointed (see section 4 – Implementation) and amongst their core duties will be to actively seek out and coordinate further funding opportunities as well as opportunities to expand the consortium to include other members for future funding calls and bring this information to the attention of the Project Management Board (PMB). The Coordinator will also regularly prompt all the consortium members to actively seek opportunities for further collaboration. "Sustainability" will be a constant agenda item at all PMB meetings with individual members of the PMB tasked with tapping into their respective networks to seek collaboration and funding opportunities. In this regard the consortium have already discussed applying for further Horizon 2020 calls and are actively working towards a submission in 2016 – as well as identifying previously funded H2020 and FP7 consortia related to this area to whom it will reach out and create connections. *e.g.* MAMBA a consortium investigating marine metagenomics; METAEXPLORE a consortium investigating the metagenomes of microbial communities of selected agricultural and aquatic habitats. It is also worth noting the fact that the coming together of the partners for this RISE application is testament to their determination to build upon the current bilateral relationships, which exist and thus underscores their willingness and ability to meaningfully address the issue of sustainability.

# **3.2.3** Contribution of the project to the improvement of the research and innovation potential within Europe and/or worldwide

The market for metagenomics solutions falls within the *global* bioinformatics market which was estimated to be worth over  $\in 2$  billion in 2012 (and is forecast to grow to  $\in 6$  billion by 2017) and the Next Generation Sequencing market which estimated to be worth  $\in 2$  billion in 2014 (and is forecast to grow to over  $\in 8$  billion by 2020). However this does not paint a full picture, as more powerful bioinformatics tools will accelerate discovery and market deployment which in turn will drive the development/commercialisation of new products. This in turn will lead to creation of new biotech, diagnostic, software and other solution and service providing companies and will drive demand to providers of software, hardware, storage and other service providers as they seek to meet the growing demand for design and analysis of metagenomic-focussed products and services.

EU economic growth and the sustainability of European-based world class companies and academic institutions will be directly impacted by the quality of research being performed within the Member States. In turn, this depends upon excellent and long lasting intra- and inter-sectoral and international collaborative links. In bringing together the various partners from both the academic and non-academic sectors, this consortium will play a significant role in bettering the integration of European science and thus yielding a long-term economic impact. By leveraging the inherent and differentiating benefits of the platform developed in this project, NSilico (who will lead the commercialisation of the project outputs) will be able to significantly add to Europe's competitiveness in bioinformatics and to take and to create and maintain a leading global position in the niche area of metagenomics for agricultural applications.

SME's have a huge importance to the European economy – as underlined by Commitment #7 of the Innovation Union agenda document which states that *"the Commission will design future EU research and innovation programmes to ensure simple access and stronger involvement of SME's particularly those with a high growth potential"*. With this in mind, it is clear therefore that the overall impact of this project is enhanced by the fact that NSilico is an SME – working in the rapidly growing bioinformatics and software industries. Furthermore, the training and experience the researchers seconded to NSilico will get will be

#### Page 18 of 46

hugely valuable in terms of equipping them with the knowledge and skills and insights into commercialisation of academic research. This will help address a distinct need that currently exists – which is the need to commercialise (and thus demonstrate an economic return from) the large amount of publicly funded research in academic institutions. Currently, across Europe there is a great lack of people with the requisite skills to translate academic research into monetary return to the exchequer. In addition, through the dissemination and outreach activities outlined below, this project will serve as an excellent example of industry-academic collaboration. This can then serve to inspire others to replicate the model and build similar collaborations across Europe, and therefore significantly enhance the innovation potential and economic benefit to Europe.

#### **3.3** Effectiveness of the proposed measures for communication and results dissemination

In line with the various agendas of the Europe 2020 Flagship Initiative, we have devised a strategy for communication and dissemination which:

- Promotes open access to research results (Innovation Union)
- Promotes ideas getting quickly to market (Innovation Union)
- Promotes the attractiveness of skills development for the knowledge economy (*Youth on the Move*)
- Promotes entrepreneurship and innovation (Agenda for new skills and jobs)
- Brings awareness of societal benefits (Innovation Union)
- Directly engages with the public to increase engagement and understanding of science (*European Charter for Researchers*)

# **3.3.1** Dissemination strategy - targeted at scientists, potential users and to the wider research and innovation community - to achieve the potential impact of the project

The following dissemination activities are planned by the *Meta-Plat* partners. The designated leaders of each of these activities are indicated within the square brackets.

**Conferences.** [Individual Researchers] The researchers will attend and present at national and international conferences. Conferences such as the Annual International Conference on Research in Computational Molecular Biology, EMBL Conferences, Computational Intelligence methods for Bioinformatics and Biostatistic, IEEE Conference on Bioinformatics and Biomedicine and International Conference on High Performance Computing & Simulation, the ACM International Symposium on High-Performance Parallel and Distributed Computing, International Conference on Parallel Distributed and network-based Processing and MTSR will be targeted.

**Trade Fairs** [Individual Researchers] Trade fairs are also seen as an excellent way of disseminating project results and at the same time adding to the commercial knowledge of the researchers with a view to giving them further opportunities to network and explore opportunities for alternative applications of their research. Fairs such as the European Federation on Biotechnology<sup>5</sup>, the Congress of European Microbiologists<sup>6</sup> and the European Forum on Farm Animal Breeders<sup>7</sup> will be targeted.

**Organisation of an International Symposium.** [Paul Walsh & Project Manager] The partners have agreed to host a scientific meeting that will take place over two days and to which leading industry and academic researchers will be invited. This will help boost the profile of the partners and give the researchers an opportunity to network with a wider scientific audience outside of the consortium partners.

**Website.** [Project Manager] A *Meta-Plat* website will be launched at the beginning of the project. In this regard it is notable that the domain name meta-plat.eu is currently available. The website will report on project progress, will be the main host for dissemination of the resources generated (public databases and open source tools) and will have links to the individual partners websites as well as updates and links to other relevant websites, recent articles and publications etc. Search Engine Optimisation will be undertaken in order to ensure the *Meta-Plat* website will appear prominently in search results.

**Peer Reviewed Publications.** [Individual Researchers] Publication of peer-reviewed articles will be pursued. In line with the move toward Open Access, the consortium will explicitly target publications in well-

<sup>&</sup>lt;sup>5</sup> http://www.efb-central.org/index.php/Main/

<sup>&</sup>lt;sup>6</sup> http://fems-microbiology.kenes.com/

<sup>&</sup>lt;sup>7</sup> http://www.effab.org/NewsandEvents.aspx

regarded open access journals (see also Data Management Plan in the Implementation section of this application). With this in mind, it is envisaged that this project will lead to publications in open access journals (or journals which offer the choice of publishing the work as open access) journals such Concurrency and Computation: Practice and Experience (Wiley), Future Generation Computer Systems (Elsevier), EMBO, BMC Genomic, BMC Bioinformatics, BMC microbiology, PLoS Genomics, PLoS Computational Biology and the International Journal of Information Management (IJIM).

**LinkedIn.** NSilico is a member of the following specialist LinkedIn Groups: Metagenomics Network (>300 members) and Bioinformatics Computing (>7,300 members). These forums will be used to disseminate the results amongst various interested stakeholders. It will also serve to build the partners' networks by enabling them to meaningfully engage with other group members about a tangible product which impacts upon the areas of personalised medicine and electronic health records. Dissemination within these groups also boosts the profile of the consortium members and will undoubtedly lead to further collaboration opportunities.

**Twitter.** A twitter account will be set up to disseminate the progress of the project, and link this progress to the latest research articles, conferences etc.

#### 3.3.2 Communication strategy, outreach plan and the activities envisaged engaging the public.

The partners plan to outreach to as wide an audience as possible. Some of this will be achieved through the dissemination activities (described above). However a wider audience including the non-technical general public, national research agencies and policy makers will be engaged through the following series of outreach activities. Each researcher will undertake to lead at least one of these activities during their involvement in the project. The designated leaders of each of these activities are indicated within the square brackets.

**Website.** [Project Manager] As mentioned previously, a project website will be set up. The website will contain non-technical sections in which the general public can read about the partners, the project, its applications in everyday life and its importance with respect to its impact potential on the economy.

Local, National & International Media. [WP Leaders & Project Manager] UU and EPCC/University of Edinburgh have dedicated communications offices. These will be leveraged to the full to ensure that there are regular releases to their local and national media, and that when available (and appropriate) that the opportunities are taken by senior researchers to act as expert commentators in the area of health informatics to these media outlets. Also noteworthy is that the academic partners each have Alumni magazines and newsletters, which are sent to past graduates. Therefore it is also planned to feature this project in these publications.

**Engagement with Schools.** [Individual Researchers] Local schools will be invited to send students to visit the academic partners on specially planned open days. These are planned to coincide with consortium meetings held in the academic partners so as the industry partners can also participate. The researchers will also be given the opportunity to visit local schools and to participate the national and local STEM activities (such as the Big Bang Young Scientists and Engineers Fair) to engage the students in the possibilities, which a career in research presents and help inspire others to follow in their footsteps.

**Facebook.** [Project Manager & Individual Researchers] The power of social media is becoming increasingly apparent. The partners will leverage this by setting up a Facebook page for this project and the individual researchers involved will invite their Facebook friends to connect.

**Piggy-back on other outreach initiatives.** [Project Manager] As evidenced by SRUC's involvement with numerous outreach activities with various agricultural events (e.g. Royal Highland Show), Open Days (e.g. Beef Centre Open Day; SMART Farming Technology Day), as well as events for the public understanding of science (e.g. Midlothian Science Fair)there is significant opportunity to piggy-back on other initiatives which engage the public – either directly through information days, public seminars etc., or indirectly through the policy makers and implementers who fund such initiatives.

**Marie Sklodowska Curie European Researchers Night.** [Project Manager] The partners will seek to participate in their national European Researchers night (which takes place across Europe every September) and seeks to engage the public with research in all its guises.

Page 20 of 46

#### 3.3.3 Expected impact of the proposed measures

As will be deduced from the technical description of the proposal, there is a high probability of significant technical developments resulting from this project. These will be communicated and disseminated by the various means described above and will highlight the partners' credentials as leading exponents in their particular disciplines. The ensuing enhanced reputational impact can then be exploited to the full, in terms of attracting new high quality researchers and collaborative academic and non-academic partners which will in turn ultimately drive the creation and exploitation of new innovations and lead to further employment thus providing tangible impact which can be recognised by both the technical community and the general public.

Specifically, regarding the planned program of dissemination activities, the impact will be widespread and on-going. Conferences at which the project results will be presented will provide useful opportunities to boost the consortium's (and the individual researchers) reputation, and network with experts in the field. Most of these conferences will have a good balance of attendees from both industry and academia and any connections made through networking have the potential to be exploited in the future as opportunities for collaboration. Similar opportunities will be presented at the 2-day scientific symposium (see Section 2 – Excellence), which the consortium will host in Cork. In addition, attendance at trade fairs will further bolster the researcher's network and profile as well as serving to build upon their commercial knowledge. Publication of results in peer-reviewed journals will add to the reputation of consortium members and the individual researchers and will enhance their respective ability to attract new collaboration, avoid duplication of effort, increase research efficiency in this area, accelerate innovation and help involve a much broader range of stakeholders (other scientists, public bodies, general public etc.) through improved transparency of the scientific process.

Regarding the planned communications and outreach program, once again significant impact will ensue. The communications are designed to engage as wide an audience as possible and show how expertise from different disciplines and diverse sectors can come together to achieve an important project outcome which not only advances knowledge but which will also ultimately impact everyday lives through improving patient care and enhancing treatment outcomes. Furthermore clear links between this project and the creation of sustainable high-value jobs will be communicated. At national and European level, at a time when we are still suffering the effects of a global recession and where there are many competing priorities for tax-payers money, from a policy makers perspective the communication and dissemination from this project will provide significant collateral support to underscore the impact that investment in research and innovation can have in terms job creation, competitiveness and productivity.

As mentioned above, it is planned that each researcher will lead at least one of the outreach activities during their involvement in the project. This will help them build their own individual communications capabilities and thus helps to address the notion of training the researchers in skills which are transferrable between the public and private sectors – and thus addresses key elements of the European Charter for Researchers. School engagements and *e.g.* careers nights will serve as excellent means of engagement with younger members of the public and seek to inspire them to become the researchers of the future. Noteworthy is that the researchers from this project who volunteer at these events will have a unique insight into life as a researcher in both the academic and commercial sectors as a result of this project and so they can pass on these insights to their student audience.

#### 3.3.4 Intellectual property rights aspects (if applicable) and exploitation of results

It is envisaged that this project will lead to substantial IP – including know-how and patents being generated. Both NSilico has substantial experience in the commercialisation of such IP. The IP developed will offer significant synergies with the company's existing products. For example NSilico's iCare platform, which is used in the Irish national health care system to manage cancer EHRs will benefit greatly from UBITECH's IT security expertise for this sensitive data, which will lead to significantly increased commercial opportunities for exploitation – which will in turn lead to the creation of new jobs. The unique combination of technologies, skills and infrastructure brought together under the umbrella of this consortium is difficult to replicate and thus provides a significant barrier to entry for other would-be solution providers in this space. This provides a level of competitive advantage for NSilico. There may also be significant opportunities for EPCC to develop high performance informatics software, which might form the basis of a

license, which could be offered commercially. Similarly both SRUC and Teagasc who provide education and business support to land-based industries in the UK and Ireland respectively will benefit greatly by engaging with NSilico and by having access to and exploiting such a metagenomics platform which they can use to advance research within their own organisations but also provide an extended range of value-added services to the business community which they already support. By engaging with NSilico UU, UM and FTK will gain first-hand experience in the commercial application of their respective expertises and technology in "real-life" settings and will be able to use this opportunity to develop further IP that may be commercially licensed.

It is also planned that if any party seeks to exploit the IP generated, it is intended that they will use all reasonable efforts to include the originating researcher(s) in the commercialisation activities where possible and appropriate – thus adding to the impact of training and career development which are available to benefit the researchers who participate in this project and directly addressing The European Charter for Researchers which states that "Employers and/or funders of researchers should recognise it as wholly legitimate, and indeed desirable, that researchers be represented in the relevant information, consultation and decision-making bodies of the institutions for which they work, so as to protect and promote their individual and collective interests as professionals and to actively contribute to the workings of the institution".

### 4 Implementation

# 4.1 Overall coherence and effectiveness of the work plan, including appropriateness of the allocation of tasks and resources

The aim of this project is to create a platform that enables a new era of personalised medicine by semantically integrating genomic big data with electronic patient records by the development of secure and scalable software engineering techniques. The attainment of this aim has been broken down into 7 individual work packages summarised below. Each work package has a clearly planned set of tasks along with associated deliverables and milestones. Specific technical risks associated with each WP have also been identified along with contingency plans which can be activated should the need arise.

# 4.1.1 Consistency & adequacy of the work plan and the activities proposed to reach the project objectives

Each of the WP's is described in detail in the Tables below.

WP Number	1	1							
WP Title	Metabiome	Metabiome Collection & Extraction							
Activity Type	Research								
Participant Short Name	NSilico	SRUC							
Person months per participant 15 12									
Objectives Sample collection preparation, and sequencing									

#### Table B8.Work Package Description

#### Description of Work

This WP will be led by Prof. Dewhurst (SRUC) and involves the collection of animal gut microbiota and recording of measuring feed conversion efficiency, methane emissions and gut health. To assess the effect of management strategies on feed efficiency and growth, we will undertake a major trial in which groups of low and high chronic stress will be compared. The choice of chronic stressor will be made after examining existing SRUC data on husbandry systems from commercial beef farms. Rumen fluid samples will be collected for analysis of the changes in rumen microbial population due to chronic stress. These changes could be used as an indicator of responses to chronic stress among animal groups. Interaction between breeds (or sire progeny groups) and stress treatment groups will be estimated to identify the breed (or sires) with higher susceptibility to competition or stress. There will be 80 animals tested a rotational cross between Aberdeen Angus and Limousin sires, and are referred to as Aberdeen Angus (Limousin) crossbred animal depending on the breed of the purebred sire used. They will be fed one complete diet consisting (g/kg DM basis) of either 480 forage: 520 concentrate. The steers had been on the diet for a minimum of 8 weeks before methane measurements were made. Six indirect open-circuit respiration chambers will be used to measure methane emissions. Rumen fluid will be collected using a naso-gastric sampling device, DNA isolated and either multiplexed sequencing of 16S amplicons or sequencing of total genomic DNA conducted using a next-generation sequencing platform (to be done by partner organisation Teagasc)

**Because animals will be used, full ethical approval will be sought by the appropriate authorities** (UK Home Office). A comprehensive document management system with workflow and audit trail will be set up to ensure full compliance with all ethics regulations.

SRUC will exchange researcher one researcher each with NSilico with a view to completing this WP with NSilico. The researcher sent by NSilico to SRUC will aid with and learn about the "wet" sample collection and processing, after which a researcher from SRUC will be sent to NSilico to ensure a smooth handover to aid and learn the "dry" data processing. Additionally, NSilico will send a bioinformatician in the early stage of the experimental design, in order to ensure the good integration between experimental constraints and bioinformatics analyses throughout the project and to learn about metabiome extraction.

#### Tasks

Task 1.1 Experimental plan

Task 1.2 Experiment preparations (animals, diets, consumables)

Task 1.3 Data collection and recording (feed conversion efficiency, methane emissions and gut health)

Task 1.4 Sample processing and 16S sequencing

Deliverables

Deliverable 1.1 Ethical clearance (**Month 1**)

Deliverable 1.2 Data collection and recording (Month 18)

Deliverable 1.3 Sequencing data (Month 24)

Deliverable 1.4 Submission to EBI ENA database of the raw sequence files (Month 30)

WP Number	2							
WP Title	Big Data Machine Learning							
Activity Type	Research							
Participant Short Name	UU	NSilico						
Person months per participant	18	6						

#### Objectives

Development of accurately generated reference databases via machine learning techniques (phylogeny-aware, new classification and previously unclassified sequences using machine learning)

#### Description of Work

This **WP will be led by Dr. Wang (UU)** and entails the formal analysis of the complete set of reference sequences. It order to reduce the redundancy, a system will be developed to keep track of each sequence origin and redundancy, quality, taxonomical position and associated biological genome (gene set). It will also involve the development of a system allowing automatised update of the databases and the public accessibility of the resulting dataset. The system will be designed to provide homogenised and high quality data, by visualising and analysing in a fully integrated way.

UU fellows have expertise in machine learning but do not have expertise in commercial bioinformatics and therefore UU will send two researchers to NSilico who has expertise in commercial bioinformatics and NSilico will send one research to learn those skills with a view to completing this WP.

#### Tasks

Task 2.1 Characterisation and curation of sequencing data (Month 1-6)

The goal of this task is to have a better understanding of sequencing data collected in WP1. In order to reduce the redundancy, the data will be characterised and preprocessed. A system will be developed to keep track of all the related information associated with each sequence, including its origin, quality, taxonomical position and associated biological genome.

#### Task 2.2 Phylogeny-aware classification (Month 7-12)

The main objective of this task is to develop a new phylogeny-aware algorithm to classify sequencing data. It will make use of phylogenetic information to label each sequence.

#### Task 2.3 Machine learning models (Month 13-24)

In this task, a new machine learning based classification model will be proposed and implemented, which will be used to predict previously unclassified sequences. The data curated using phyogeny-aware classification in Task 2.2 will be used to train, test and evaluate the model.

#### Task 2.4 Reference database distribution (Month 13-24)

A publicly available reference database will be developed and implemented. It will involve the development of a system allowing autotomized update of the databases. The system will be designed to provide homogenised and high quality data, by visualising and analysing in a fully integrated way.

#### Deliverables

Deliverable 2.1 Data model and curation (Month 6)

Deliverable 2.2 Machine learning model (Month 12)

Deliverable 2.3 Database distribution (Month 24)

Deliverable 2.4 Implementation of software platform (Month 24)

Deliverable 2.5 Academic papers (Month 30)

WP Number	3							
WP Title	Metagenomics Algorithms							
Activity Type	Research							
Pre-requirement	WP2 (deliverable 2.1 and 2.2)							
Participant Short Name	icipant Short Name NSilico UU							
Person months per participant 2 24								
<b>Objectives</b> Development of accurate classific	cation algorith	ims	-					

#### Description of Work

This **WP will be led by Dr. Walsh (NSilico)**. Metagenomic samples can contain reads from a huge number of organisms. For example, in a single gram of soil, there can be up to 18,000 different types of organisms, each with its own genome. Classically, binning approaches are used for grouping reads or contigs and assigning them to operational taxonomic units. Based on Mande et al. [15] review publication, we will develop new algorithmic approach, allowing for both accuracy and speed which will subsequently be implemented in WP5 in a HPC infrastructure. We will build on the big classification algorithms developed in WP2 to implement a comprehensive pipeline of algorithms for the integrated analysis of microbial populations.

In order to develop metagenomics algorithms expertise in bioinformatics, machine learning and high performance computing is needed. NSilico have expertise in bioinformatics but do not have expertise in machine learning and limited expertise in high performance computing. NSilico will exchange a researcher with UU who has expertise machine learning and will send a researcher to UU who has expertise in high performance completing to learn these skills with a view to completing this WP.

Tasks

Task 3.1 Current algorithm evolution

Task 3.2 Development of new analyses algorithm

Task 3.3 Development of new classification algorithm

Task 3.4 Software suit development (Linux platform)

#### Deliverables

Deliverable 3.1 Current Algorithm review and evolution – Review/position publication (Month 12)

Deliverable 3.2 Development of new analyses algorithm (Month 18)

Deliverable 3.3 Development of new classification algorism (Month 24)

Deliverable 3.4 Implementation of software platform (Month 32)

Deliverable 3.5 Academic papers and technical reports (Month 38)

WP Number	4						
WP Title	High Performance Computing						
Activity Type	Research						
Pre-requirement	WP2 and WP3						
Participant Short Name	EPCC	NSilico					
Person months per participant	12	12					

#### Objectives

The development of high performance computing algorithms to rapidly annotate and analyses genomic sequences for meta-genomic analysis, increasing the high through put computational capacity of analysis pipelines two orders of magnitude.

This **WP will be led by Terry Sloan (EPCC)** and involves the development of thread safe extensible semantically integrating genomic big data with electronic health records algorithms for HPC architectures using ISO/IEC 12207, an international standard for software life-cycle processes, which defines best practices for developing and maintaining software. Investigate suitable functional HPC algorithms for sequence analysis pipelines identified in WP2 and WP3 and the benchmarking of their performance against existing standards and develop new standards.

- Identify the most scalable parallel computing architectures for meta genomic sequence analysis.
- Identify the appropriate functional partitioning strategies to achieve the most optimal solution.
- Design a solution that maximises data locality and minimises inter-process communication.
- Architect mapping strategies that take full advantage of target architectures.
- Prove clear commercial value propositions by demonstrating benefits of the system in the field of meta-genomics.

In. EPCC will send a researcher to NSilico who has expertise in metagenomic analyses to learn these skills with a view to completing this WP.There is a dearth of HPC skills in the bioinformatics community so it is crucial for SME development in this space that NSilico staff upskill in this valuable discipline. Tasks

Page 25 of 46

Task 4.1 Analysis Design of HPC architecture and parallelisation strategies (Month 12-18)

Task 4.2 HPC Implementation of assembly algorithms (Month 18-24)

Task 4.3 HPC Implementation of meta-genomic analysis pipeline (Month 18-36)

Task 4.4 Integration of HPC Implementations with animal models (Month 24-36)

#### Deliverables

Deliverable 4.1 HPC platform infrastructure (Month 9)

Deliverable 4.2 Parallelised Assembly for meta-genomics (Month 12)

Deliverable 4.3 HPC Implementation of meta-genomic analysis pipeline (Month 24)

Deliverable 4.4 Academic papers and technical reports (Month 36)

WP Number	5							
WP Title	Platform Integration							
Activity Type	Research							
Pre-requirement	WP1 and WP4							
Participant Short Name	NSilico	FTK	SRUC	UU				
Person months per participant	3	12	18	6				

#### Objectives

Integration of Meta-Plat platform components. Provide insights into probiotic supplement usage, methane production and feed conversion efficiency in cattle.

#### Description of Work

This **WP will be led by Dr. Walsh (NSilico)**. Using the data collected in WP1 (feed conversion efficiency, methane emissions and gut health) and the meta-genomic analysis pipeline In WP4, extensive testing, improvement of the platform will be provided to the other WP during all the development and implementation stage. Finally it will ensure a sanity check of the project integration while using the collected data to analyse the changes in rumen microbial population due to chronic stress and inform best practice in ruminant farming. Population structure will be correlated to feed conversion efficiency, methane emissions and gut health using multiple statistical models using R.

NSilico will exchange a researcher with FTK who has expertise in big data management to learn these skills with a view to completing this WP. SRUC will send two researchers and UU one researcher to NSilico who has expertise in Bioinformatics integration and biostatistical analyses.

#### Tasks

Task 5.1 Testing and reporting of algorithm implemented in WP2

Task 5.2 Testing and reporting of algorithm developed in WP3

Task 5.3 Testing and reporting of implementation of meta-genomic analysis pipeline in WP4

Task 5.4 Analyse of WP1 recode with WP4 meta-genomic analysis pipeline

#### Deliverables

Deliverable 5.1 Technical reports and Platform Integration of WP2 results and databases (Month 14)

Deliverable 5.2 Technical reports and Platform Integration of WP4 pipelines (Month 30)

Deliverable 5.3 Academic papers on changes in rumen microbial population due to chronic stress (**Month 36**) Deliverable 5.4 Public release of the meta-genomic analysis platform (**Month 36**)

**PLEASE NOTE:** The "MANAGEMENT" and "COMMUNICATION/DISSEMINATION" WP's are described in detail Section 4.2 and 3.3 respectively. It is planned that there will be 4 secondment months in total for Management as reflected in Part A3. These 4 months will be split into shorter stays and will facilitate occasional visits by the CEO of NSilico (who will be the main commercialisation vehicle for the outputs of this project to oversee development and at the same time transfer market knowledge to the academic partners).

#### 4.1.2 Credibility and feasibility of the project through the activities proposed

The feasibility of this project requires inter-disciplinary collaborative activity in the domain of high performance computation, semantic search, machine learning, animal science and software development.

#### Page 26 of 46

Research fellows from NSilico who have expertise in software development will be involved in design and implementation of parallel software in EPCC, thereby credibly realising EPCC's parallel algorithm designs by codifying algorithms in a commercial software development framework. Similarly the machine learning processing and semantic technology expertise of UU will be brought to bear on the project when UU research fellows bring their expertise to NSilico where the deliverable of the project will be developed using commercial software development best practices. The IT security aspect of the project will also be highly attainable as the commercial security and big data management expertise of FTK will be leveraged when research fellows from UU are mentored in big data software development.

However, the partners are acutely aware that by their nature, research projects of this type have inherent risks, which will impact upon their feasibility. Thus they have identified the following risks and credible mitigation strategies to counteract them:

In WP1, risk 'Animal death, insufficient data collected' has been identified but SRUC's experience and the large number of animals available through both SRUC and Teagasc should allow collection of enough new data despite the risk.

In WP2 and WP3, risk 'lack of suitable new models' has been identified but if this happens we will be able to use improved current models that apply to the farm animal model scenarios identified in WP1.

In WP4, risk '**software specification is not sufficiently refined'** has been identified but we will mitigate by the use of agile software development practices.

In WP4, risk 'parallel algorithms do not map to hardware' has been identified but we will mitigate this by the use of a parallel programming language that combines functional and procedural techniques, thereby allowing vast flexibility in how code is mapped to parallel architectures. Moreover, we will target an array of architectures ranging from shared memory to distributed memory configurations, thereby allowing code to be retargeted to a number of systems.

In WP2, WP3, and WP4, risk '**software system is unstable**' has been identified but WP5 will be used to evaluate and improve the integrated platform and system along side each WP, we will also mitigate this by the use of agile software development and test driven development practices.

#### 4.1.3 Gender aspects in the planning of the activities

The partners fully understand and commit themselves to tackling the issue of gender balance. Each of the partners already has equal opportunities policies in place and encourage gender balance. It will be ensured that an equal opportunity for secondment is given to male and female researchers within all organisations. Furthermore, any female secondees to NSilico will be encouraged to join WISER (Women in Science & Engineering Research) which as well as offering support for female researchers, also provides excellent networking opportunities. Similarly, those seconded to EPCC, UU or SRUC will be encouraged to join WISE whose mission is to increase the gender balance in the UK's STEM (Science Technology Engineering & Maths) workforce, while those seconded to FTK, will be encouraged to join the European Association for Women in Science, Engineering and Technology as will those seconded to UM. UU is a member of the Athena SWAN Charter and gained its institutional Athena Swan Bronze Award.

				<b>.</b>	<u>.</u>	- !!
Del.	Deliverable Name	WP	Lead Participant	Nature	Dissemination	Delivery
No		No.	Short Name		Level	Month
1.1	Ethical clearance	1	SRUC	Management	CO	1
1.2	Data collection and recording	1	SRUC	Research	CO	18
1.3	Sequencing data	1	SRUC	Research	CO	24
2.1	Data model and curation	2	UU	Research	CO	6
2.2	Machine learning model	2	UU	Research	CO	12
2.3	Database distribution	2	UU	Research	CO	24
2.4	Implementation of software platform	2	UU	Research	PU	24
2.5	Academic papers	2	UU	Dissemination	PU	24
3.1	Current Algorism evolution –	3	NSilico	Dissemination	PU	18
	Review/position publication					
3.2	Development of new analyses algorithm	3	NSilico	Research	CO	24
3.3	Development of new classification	3	NSilico	Research	CO	30
	algorithm					

#### Table B9. Deliverables List

Del.	Deliverable Name	WP	Lead Participant	Nature	Dissemination	Delivery
No		No.	Short Name		Level	Month
3.4	Implementation of software platform	3	NSilico	Research	PU	38
3.5	Academic papers and technical reports	3	NSilico	Dissemination	PU	38
4.1	HPC platform infrastructure	4	EPCC	Research	CO	18
4.2	Parallelised Assembly for meta-genomics	4	EPCC	Research	CO	24
4.3	HPC Implementation of meta-genomic analysis pipeline	4	EPCC	Research	CO	36
4.4	Academic papers and technical reports	4	EPCC	Dissemination	PU	36
5.1	Technical reports and Platform	5	NSilico	Dissemination	PU	24
	Integration of WP2 results and					
	databases					
5.2	Technical reports and Platform	5	NSilico	Dissemination	PU	36
	Integration of WP3 pipelines					
5.3	Technical reports and Platform	5	NSilico	Dissemination	PU	36
	Integration of WP4 pipelines					
5.4	Academic papers on changes in rumen	5	NSilico	Dissemination	PU	36
	microbial population due to chronic					
	stress					
5.5	Submission to EBI ENA database of the	5	NSilico	Dissemination	PU	42
	raw sequence files					
5.6	Public release of the meta-genomic analysis platform	5	NSilico	Dissemination	PU	42

#### Appropriateness of management structure & procedures, incl. quality & risk management

### 4.1.4 Project organisation and management structure, including the financial management strategy, as well as the progress monitoring mechanisms put in place

The management structure for this project is shown below. As can be seen it consists of varying layers of management at the local and consortium level.



#### **Project Management Board (PMB)**

*Meta-Plat* will be managed by the Project Management Board (PMB). There will be one representative from each of the beneficiaries on the PMB, including the Project Coordinator (Haiying Wang) who will chair the Board. The PMB will convene quarterly - either in person (which will be timed to coincide with the project workshops, see below) or *via* video conference/telephone – and will be **responsible for ensuring quality across all aspects of the project** and that the consortium is compliant with the contractual obligations which exist with the European Commission. In this regard it will oversee the distribution of project funds (*via* the Coordinator), and ensure that the project continues to address the objectives outlined herein. The PMB it will also ensure that the Researchers are undertaking the Work Packages as outlined in the work plan. If any changes to this plan need to be made, these will first be ratified by the

#### Page 28 of 46

PMB before being cleared by the REA Project Officer. The PMB will also ensure that the exchange of knowledge and skills between partners is taking place satisfactorily. In this regard it will appoint Dr Paul Walsh (CTO of NSilico) and Dr. Richard Dewhurst (SRUC) as the respective commercial and academic members of a Knowledge Exchange Oversight Group (See Section 2). This subcommittee of the PMB will be responsible for the execution and supervision of the various Knowledge Exchange Plan activities described in Section 2. Furthermore the Knowledge Exchange Oversight Group will interact closely with each researcher and their co-workers to ensure that the knowledge exchange objectives are being met. Decisions to be made by the PMB or the Knowledge Exchange Group will be made where possible on a consensus basis. If a vote is needed a simple majority will suffice with the Coordinator having the deciding vote if required.

#### Coordinator

The Coordinator, Haiying Wang, is very experienced at running intersectoral collaborations of this type (See section 6). He will manage financial matters between the EC and the beneficiaries; In addition he will monitor progress against the planned schedule and to propose such actions to the Commission as are necessary to reach the project objectives; to review progress, periodic and final reports as well as sending the progress and financial reports to REA. Furthermore he will chair all meetings of the PMB and have a casting vote in the event of any tied votes. He will also liaise with the WP leaders who are responsible for ensuring the smooth running of their respective Work Packages. In turn, they will report on results to the Coordinator any organisational issues that may result in a deviation from the time schedule or result in technical or quality problems. Reporting will be carried out according to the Commission rules.

#### **Project Manager**

Using the "Management and indirect costs" allocation of the grant, a part-time Project Manager (PM) will be recruited at the commencement of the project. This has been agreed in advance by the consortium partners. The PM will act as a bridge between the Coordinator, the WP Leaders and the individual researchers. The PM will assist the Coordinator with the majority of the administrative duties and ensure that they are properly undertaken. This will include taking and distributing minutes of consortium meetings; collating consortium documents; maintaining project website; seeking and coordinating future funding applications to ensure sustainability of the consortium. In addition the PM will also be responsible for ensuring that the various communication (described in Section 3) and project monitoring initiatives (described below) are successfully implemented. Because of her/his administrative oversight, the Project Manager will be able to highlight any problems as they arise in the project and seek to either directly address them or to bring them to the attention of the Researchers, WP Leaders, PMB or project officer in the European Commission.

#### **WP Leaders**

The technical supervision and management of the seconded researchers will be undertaken on a day-today basis by their WP Leaders in the host organisation. As the WP Leaders will also be members of the PMB they will have a good understanding on the overall project requirements and context, and can readily brief their PMB colleagues on project progress at the Quarterly PMB meetings or flag any issues to the PMB in a timely manner. Furthermore, the WP Leaders will also act as career mentors to the younger incoming researchers and will be able to advise and steer them on future career options, which may be available to them.

#### **Conflict Resolution**

The philosophy of the *Meta-Plat* consortium is to seek consensus where possible. Where this is not possible at PMB/Sub-Committee level, a majority vote system is in place (as outlined above). Should conflict arise on a personal/individual basis, there are conflict resolutions in place at an institutional level in each of the academic organisations which will be used as appropriate. These procedures cover resolution protocols, which would be used in the case of both (a) conflict between a researcher and his/her supervisor and (b) conflict between two researchers. Similarly NSilico have defined conflict resolution procedures in place in their Health & Safety Handbooks, which cover these eventualities.

#### **Communication & Project Monitoring**

The international and inter-sectoral nature of this collaboration will be challenging, and therefore it is important that the lines of communication are clear so that the project can run smoothly. In this regard, the following measures will be undertaken:

- A secure cloud-based project management tool (e.g. Planbox, Microsoft Project or other similar package) will be used by the *Meta-Plat* consortium. This will facilitate access to project materials regardless of location. The access to information can also be managed in such a way that certain members of the consortium only have access to appropriate material. This tool will also assist the Project Manager and Coordinator in clearly communicating tasks, deadlines, project progress etc.
- Standardised Reporting. The Coordinator and PM will design a standardised reporting format to be
  used by the researchers. There will be standard reports for all planned meetings and events listed
  in Section 2. Having standardised reports will ensure that there is consistency in how information is
  captured and communicated and will help to avoid confusion as well as ensuring smooth transfer
  of knowledge.
- Language. The workings of the *Meta-Plat* consortium will be conducted in English. All of the identified seconded researchers are fluent in English.

#### **Key Meeting Schedule**

The following table summarises the schedule of key meetings, which will ensure good communication across the lifetime of the project

Meeting Participants	Frequency of Meeting
Whole consortium	Every 6 months at consortium workshops
Project Management Board	Quarterly, commencing at kick-off meeting in Month 1
Knowledge Sharing Oversight Group	At beginning and end of each researcher secondment and at consortium workshops
Researcher & Host Supervisor	Formal weekly meeting
Project Manager & WP Leaders	Quarterly visit by PM to each WP leader

#### Table B10. Meeting schedule

#### **Financial Management Strategy**

All of the partners have strong financial management structures in place. EPCC, SRUC, UU, UM and FTK will be supported by their Research and Finance Offices. As described in section 6, all of these institutions have vast experience in the management of projects of this type. In this project it has been agreed that Haiying Wang (as the project co-ordinator) will manage the finances and distribute grant monies accordingly. The Coordinator will work closely with the finance officers in the partner organisations to ensure tight budgetary control and operational discipline. Timely drawdown of the successive tranches of funding from the European Commission will be ensured by the Coordinator who will be responsible for the on-time submission of all reports to the Project Officer in Brussels. The Coordinator will be assisted in this endeavour by the Project Manager.

#### Support to incoming seconded researchers

As described in their profiles, each of the partners has an excellent track record and experience of international collaborations meaning that they are well-placed to contribute to and sustain a collaboration of the highest quality. They are well acquainted with the issues involved in hosting incoming researchers from other countries and in this regard all of the academic partners each have dedicated research and graduate studies offices that are well versed in the support required for incoming researchers. The full support of these offices will be at the disposal of the researchers. The team in NSilico will also offer full support with respect to relocation, accommodation, and other "settling-in" issues. Each of the consortium members is committed to ensuring that the scientific and cultural integration of the researchers is satisfactorily fulfilled. In this regard they will also liaise closely with their local EURAXESS offices for assistance in helping the incoming researchers.

#### **Management of Career Development**

The consortium are committed to the European Charter for Researchers which states that *"Employers and/or funders of researchers should draw up, preferably within the framework of their human resources management, a specific career development strategy for researchers at all stages of their career"*. The Knowledge Exchange Oversight Group will be responsible for ensuring that all the researcher training needs

Page **30** of **46** 

are met and, in consultation with the individual researchers and their host WP Leaders, that the addressing of these needs are integrated into a tailored Career Development Plan for younger researchers (ESRs) and a Continuous Professional Development plan for experienced researchers (ERs). These training elements extend beyond the specific scientific and technical knowledge acquired through the secondments. A broader training will be provided through of a series of non-technical modules outlined in Section 2. This will train the researchers in transferable skills which can be used in both industry and academia and will thus broaden their professional options. Moreover as alluded to previously in this section, the host WP Leaders will be available to advise the younger researchers on how they might best make the transition from academia to industry (or *vice versa*) should that be of interest to them.

#### Intellectual Property Strategy

The partners have already in place a Letter of Intent (LoI) which will form the basis of a Consortium Agreement (CA). The LoI addresses intellectual property issues and how they will be managed. It is planned that the CA will follow the DESCA model for European projects and the associated IP provisions therein which include ownership of the pre-existing knowledge and know-how (background knowledge), foreground knowledge and intellectual property protection, publication and dissemination activities, promotion of the exploitation of the results and IPR issues beyond the end of the Agreement. It is further planned that any novel/patentable/commercialisable developments or inventions will be immediately disclosed to the PMB who will between them agree the most appropriate exploitation strategy. The consortium members have also agreed that where possible, they will seek to involve those researchers who were responsible for any commercialisable Intellectual Property in the exploitation of same. This will have the effect of boosting the training and career development elements of this project and will extend beyond the project lifetime.

#### Data Management Plan

- This will be led by the Coordinator with the assistance of the Project Manager. It is anticipated that all non-confidential data pertaining to this project will be published in open access journals and the raw data to be deposited in public databases. Prior public submission, the data will be handle and backup by NSilico infrastructure with the support of FTK as the partner in charge the big data integration. All partners have agreed to follow guidelines:
- Research data will be managed to the highest standards throughout the research data lifecycle as part of the Project's commitment to research excellence.
- Responsibility for research data management through a sound research data management plan during the research project lies primarily with the WP leaders.
- The project will provide training, support, advice for the research data management.
- The project will provide mechanisms and services for storage, backup, public deposit and retention of research data assets in support of current and future access, during and after completion of the projects.
- Exclusive rights to reuse or publish research data will not be handed over to commercial publishers or agents without retaining the rights to make the data openly available for re-use.

#### **Ethics Management Plan**

The ethics management plan is dealt with comprehensively in section 7.

# 4.1.5 Risks that might endanger reaching the project's objectives and the contingency plans to be put in place should risks occur

In addition to the specific technical risks and contingencies associated with each Work Package, which have already been described earlier, the partners have identified the following risks and contingency plans.

**Data Loss**. A comprehensive data backup and disaster recovery plan will be formulated to ensure against data loss and to provide the ability to continue the project in the event of any eventuality. The plan will consist of methods and timings of regular backups as well as server redundancy and off-site storage facilities.

**Disputes on Intellectual Property Ownership**. IP ownership will be covered in the Consortium Agreement using the DESCA template which clearly sets out the nature and ownership of Background IP and details procedures for deciding ownership and usage of foreground IP (see also section on IP below).

**Researchers leave the project prematurely**. In the case of premature departure of seconded researchers, existing staff members of the "sending" organisation will be substituted where possible.

**Change in personal circumstances of identified secondees**. In consultation with the REA Project Officer, existing staff members of the "sending" organisation will be substituted where possible.

**Confidentiality not being maintained**. Each seconded researcher will be required to sign a non-disclosure agreement on arrival at their host organisation.

**Failure to produce new data (sequencing and sampling issue)**. Historical data and public repository will be used as a replacement, allowing the development of new algorisms and the full training and personal exchange.

#### 4.2 Appropriateness of the institutional environment (infrastructure)

#### 4.2.1 Availability of the expertise and human resources, to carry out the proposed research project

*UU's* contribution will be led by Dr. Haiying Wang. Dr. Wang is one of the lead researchers within the Computer Science Research Institute (CSRI) which comprises four research groups; Artificial Intelligence and Applications (AIARG), Information and Communication Engineering (ICE), Intelligent Systems (ISRC) and Smart Environments (SERG). The CSRI currently comprises 45 academic staff, together with research associates and research fellows, and approximately 100 doctoral research students. Within the CSRI, there is a strong emphasis on research output quality, with frequent publication in leading academic journals and conferences. Much of the research within CSRI is supported by highly competitive awards from external funding bodies. Total external grant income in the current UK REF assessment period exceeds £24M, including over £5.5M from the UK Research Councils indicating the quality of expertise and human resources available within CSRI. The computational biology research interest group, which crosses the Smart Environments Research Group and the Artificial Intelligence Research Group, has expertise on genomics and proteomics data integration and analysis from the past 10 years research experience including FP6 funded projects and a recent TSB funded project. Specific expertise relevant to this project includes knowledge and data engineering, semantic analysis, modelling, optimisations and pattern analysis.

Prof Richard Dewhurst leads **SRUC's contribution** to this project. Richard has for many years contributed at the interfaces between dairy nutrition, milk composition and rumen function - notably modelling of forage composition, dry cow feeding strategies, forages and fatty acids, fatty acids and fertility, and rumen diagnostics. Some work used directly by the animal feed industry – notably feed evaluation systems, feed values and dry cow strategies. Much more of his work has been used by farmers and advisers – notably to develop advice and models to predict and increase forage intakes, improve the utilisation of feed protein (reduce N pollution), improve milk composition, and facilitate the use of forage legumes in conventional and organic farming. Demonstrated practicality and feeding value of baled legume silages, which have been adopted by the industry; Development of 'Acidogenicity Value' concept for feed evaluation, which has been used by French feed manufacturer, COFNA/Evialis; Results from a series of silage-feeding studies incorporated into UK system for predicting silage intake; Results concerning effects of legume silages and organic farming on milk fatty acids used for promotions by UK Organic Milk Supply Company; Results from a long series of dry cow feeding studies incorporated into guidelines on dry cow feeding (UK Milk Development Council). He thus has ideal background and expertise to lead a significant contribution to this project.

**FTK** has been a competent partner for the development and realisation of innovative strategies in private economy and public administration since 1991. Interdisciplinary teams create and develop strategies in order to guarantee an efficient implementation of latest ICT systems, and beyond. FTK consists of a team with broad and profound expertise in various fields such as consulting, research, information, innovation-and project management, event organisation and web development. The institute's independent research premises are located in Dortmund. FTK currently employed a team of 14 full time researchers and non-researchers, 18 PhD research members (incl. external members) and is supervising three habilitation research projects. FTK can provide profound knowledge and expertise related to Cross-Organisational

Collaboration, Digital Preservation in Industrial and Scientific Innovation, Semantic Web Technologies, Information Retrieval and Information Visualisation Support, Digital Libraries and Media Archive Technologies, Knowledge Management Technologies and Methods.

**NSilico** is also a <u>research and innovation driven company</u> of 7 people in total. All of their staff are university graduates (4 PhDs and 3 MSc's) -the personnel being a mix of software developers and life scientists. Of specific relevance to this project, the expertise resident within the company includes bioinformatics, programming, agile software development, easy-to-use UI design and implementation, commercialisation and project management. The expertise NSilico also offers an excellent environment for incoming researchers and is experienced at hosting same, through its previous experience in the FP7 Marie Curie – IAPP programme, in which it hosted two incoming researchers for prolonged periods (>12 months). As a small (but growing) SME it is however mindful of the dangers in over-committing to outward secondments for long periods of time. With this in mind the outward secondments from NSilico have been structured in such a way as that no individual secondment is longer than 3 months.

**EPCC**: The Edinburgh Parallel Computing Centre (EPCC), a University of Edinburgh facility founded in 1990, has built up a unique set of skills, which it has employed in collaboration with hundreds of businesses and academic research groups. Since it foundation, EPCC has worked with companies from blue-chip multinationals such as Cisco and Rolls-Royce to smaller, local companies such as DEM Solutions and OHM Surveys. In 2002 EPCC became lead partner in the HPCx consortium, supporting the national supercomputing service for UK academic research and in 2008, it became the host for HECTOR, a second national service, funded to the tune of £115 million over six years, until 2014. It now hosts the £43m ARCHER national supercomputing service. It also runs the Computational Science and Engineering support service for ARCHER. EPCC has hosted a European Commission-funded visitor programme called TRACS for a decade, through which over 400 researchers were able to use HPC systems to further their research. TRACS was replaced by HPC-Europa 2004-2012, during which period EPCC hosted more than 800 research visits. EPCC, which specialises in hosting researchers, is an ideal computing hub for the project in which semantically aware high performance software can be designed, implemented and tested on state of the art infrastructure. EPCC also offers training courses on a variety of subjects, and has contributed over the years to standards efforts such as MPI. In 2001, EPCC introduced one of the world's first MSc degrees in HPC, which has proved popular and successful.

# 4.2.2 Description of the necessary infrastructures and any major items of technical equipment (if required)

**UU**: CSRI at UU conducts internationally excellent and world-leading research in intelligent systems, assistive technologies, next generation networks, and semantic analytics. Within the last 5 years, CSRI has completed extensive new research facilities, with over £12.6M of infrastructural investment, with a further £8.5M external funding to develop new infrastructure already secured in 2013 to expand and sustain research activity over the next 5-year period. Specific to this project, the SERG and AIA Intelligent Environment is a dedicated research environment developed with funding from the NI DEL Research Capital Investment Fund (RCIF) to support Deployment of Sensing Technology in Connected Health Care. It has a large-scale Intelligent Environment to support research, deployment and evaluation of Connected Health solutions, data acquisition and semantic analysis of user environments.

**EPCC** is one of Europe's major supercomputing centres and hosts and administers a number of facilities for use by researchers on behalf of various EU and UK research funding councils. EPCC has a long history of providing services to research scientists and industry. In 2002 EPCC became lead partner in the HPCx consortium, supporting the national supercomputing service for UK academic research and in 2008, it became the host for HECTOR, a second national service, funded to the tune of £115 million over six years, until 2014. It now also hosts the £43m ARCHER national supercomputing service. EPCC also runs the Computational Science and Engineering support service for ARCHER and is thus more than adequately equipped from an infrastructural perspective for the demands of this project.

**NSilico** is the SME participant in this project. The company's facilities are located in the Rubicon business centre. This centre contains over 3,000m<sup>2</sup> of office, laboratory and incubation space. The company currently has a 50m<sup>2</sup> unit and space within the Rubicon centre can be augmented as needed. Therefore the

#### Page 33 of 46

company can comfortably cope with the demands of hosting research fellows. The company also has in place a very experienced management team who crucially in the context of this project, have experience of managing numerous grant funded inter-sectoral and international collaborations (see section 6). NSilico houses a VRTX Server with over 100 compute cores, 16 To of storage and 512 Go RAM, designed in partnership with Dell. This enables the deployment of high throughput compute instances for production and test environments. This is anticipated to be a perfect development environment for this project as it allows multiple configurations, parallel computing of big metagenomic datasets and cloud deployment simulation. Thus, NSilico is more than adequately equipped to ensure that the objectives are delivered upon and it will do its part in driving the project to a successful conclusion.

**SRUC** has established world-class livestock greenhouse gas (GHG) emission research facilities. These are part of the new SRUC Beef & Sheep Research Centre at the Bush Estate near Edinburgh: Respiration chambers to measure methane output from individual cattle or a cow and calf (or sheep); 44 HOKO feeders enabling individual feed intake in group pens (enabling 130+ cattle to be recorded at a time). HOKO feeders have integrated canopies with 'sniffers' to enable gas sampling from identified cattle whilst they feed. SRUC can provide:

- Methane measurements with '*high*' throughput linked to '*in situ*' measures of feed intake/growth and 'in situ' methane sampling.
- Individual performance monitoring facilities for hundreds of cattle for investigating genetic, nutritional and other factors influencing feed efficiency and methane output and for testing '*proxy*' measures of GHG outputs.
- A suite of ultrasound measurement equipment, and associated capture and analysis tools.
- Meat quality prediction and assessment laboratory; other laboratory facilities (blood and feed analysis, molecular biology).
- Strong links to commercial abattoir facilities for full carcass evaluation and tissue retrieval.

**FTK** is a research participant in this project. The institute is located at the Technologie Park in Dortmund and offers a full research setting with rooms and infrastructure for the exchange of researchers. In relation to Metageomics, FTK will bring in an existing IT infrastructure for data fusion, collaboration, authentication and authorisation and its profound expertise in research data archiving. These services have originally been developed in the big data project Smart Vortex (http://www.smartvortex.eu/) and the long term preservation projects SciDIP-ES (http://www.scidip-es.eu/) and APARSEN (http://www.aparsen.eu/). In Smart Vortex, FTK has worked extensively with the engineering industry implementing cloud based data fusion technologies, collaboration and digital preservation solutions in complex cross-organizational scenarios. This IT infrastructure is running on in-house servers as well as in a virtual server environment that will be deployed in cloud based scenarios to support collaboration across all project partners in a more flexible way. Furthermore FTK was developing progress in formal qualifications of digital long term preservation, building up an online training portal, and interactive map in APARSEN and focusing on data preservation infrastructures and training in SciDIP-ES.

# 4.3 Competences, experience & complementarity of participating organisations & institutional commitment

# 4.3.1 Adequacy of the partnership to carry out the project explaining how participants' synergies & complementarities will be exploited

*Meta-Plat* requires the partners to combine their specific complementary expertise and knowledge in order to successfully develop and commercially exploit a platform that analyse and accurately classify metagenomic 16S data. On an individual basis none of the consortium members have the requisite skills, experience or expertise required to develop such a platform. However by working together to leverage their complementarities and synergies they can meet this objective and furthermore to create a means for collaboration which lasts beyond the lifetime of this project. The specifics of these complementarities and synergies are described in detail below. It should be noted that these could further be with respect to the future development and growth of each of the consortium members once the researchers return after their secondments. This is because the knowledge and skills, which have been acquired as a result of leveraging the synergies/complementarities in this project, will help increase the quality and efficiency of their research and help build and embed improved cultures within the respective research groups.

**Complementarities**. The consortium has been carefully balanced in terms of inter-sectoral and interdisciplinary complementarities with the academic partners transferring knowledge and know how in a series of exchange with NSilico. Moreover, there is a balance in a European dimension with North-western based entities in Ireland (NSilico and partner organisation Teagasc) and UK (EPCC, SCRI and SRUC) linking with a central European partner in Germany (FTK).

The skillsets of the partners in *Meta-Plat* are not sufficiently broad to enable any one of them to undertake the complex inter- and multidisciplinary work proposed herein. Therefore the choice of partners is crucial to ensure that the gaps in expertise and knowledge which exist in one partner are filled by the knowledge and skills provided by another. For example, in order to achieve the main objective of this project (*i.e.* development of a platform that analyse and classify metagenomic 16S data) a blend of commercial expertise and biological expertise is necessary. Knowledge allied to specific technical expertise in semantic technology, parallel computing, big data management, animal science, machine learning and software development is required. NSilico does have commercial software development and informatics skills and expertise but does not have parallel computing, semantic technology, and machine learning or animal science knowledge. This means that the researchers involved in this project are exposed to both breadth and depth of technical expertise as well as complementary work-practices and decision-making processes which pertain in each of the academic and commercial worlds.

**Synergies.** *Meta-Plat* also leverages the synergies, which exist between the partners. For example, combining NSilico's genomic software development expertise, FTK's Big Data expertise, EPCC's parallel computing expertise, and UU's machine learning expertise will provide a development environment that will allow accelerated development to take place that otherwise would not be possible if development of each component were to be undertaken in individual silos. The gathering together of the multidisciplinary skillsets enables novel thinking and new innovations, which otherwise would not arise, to come from this project. Thus the collaboration between the partners in *Meta-Plat* enables the researchers individually to concentrate on their core areas of expertise and skills whilst collaboratively working together in the context of the efficient completion of specific Work Packages. Thus, through the synergistic effect of this coming together collectively and deploying these expertise and skills, they accelerate the attainment of an overall objective with commercial potential and exchange and acquire valuable new knowledge from each other in the process.

### **5** References

[1] Brulc JM, Antonopoulos DA, Miller ME, Wilson MK, Yannarell AC, Dinsdale EA, Edwards RE, Frank ED, Emerson JB, Wacklin P, Coutinho PM, Henrissat B, Nelson KE, White BA: **Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases**. *Proc Natl Acad Sci U S A* 2009, 106(6):1948-53.

[2] Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, Rubin EM: **Metagenomic discovery of biomass-degrading genes and genomes from cow rumen**. *Science* 2011, 331(6016):463-7.

[3] DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL: Greengenes, a Chimera-Checked 16S RRNA Gene Database and Workbench Compatible with ARB. *Appl Environ Microbiol* 2006, **72**:5069–72.

[4] Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO: **The SILVA Ribosomal RNA Gene Database Project: improved Data Processing and Web-Based Tools**. *Nucleic Acids Res* 2013, **41**(Database issue):D590–6.

[5] Wang Q, Garrity GM, Tiedje JM, Cole JR: Naive Bayesian Classifier for Rapid Assignment of RRNA Sequences into the New Bacterial Taxonomy. *Appl Environ Microbiol* 2007, **73**:5261–7.

[6] Darling AE, Jospin G, Lowe E, Matsen FA 4th, Bik HM, Eisen JA: **PhyloSift: phylogenetic Analysis of Genomes and Metagenomes**. *PeerJ* 2014, **2**:e243.

[7] Brady A, Salzberg S: **PhymmBL Expanded: confidence Scores, Custom Databases, Parallelization and More**. *Nat Methods* 2011, **8**:367.

[8] Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA: **The Metagenomics RAST Server - a Public Resource for the Automatic Phylogenetic and Functional Analysis of Metagenomes**. *BMC Bioinformatics* 2008, **9**:386.

[9] Bengtsson-Palme J, Hartmann M, Eriksson KM, Pal C, Thorell K, Larsson DGJ, Nilsson RH: metaxa2: improved Identification and Taxonomic Classification of Small and Large Subunit RRNA in Metagenomic Data. *Mol Ecol Resour* 2015.

[10] Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C: **Metagenomic Microbial Community Profiling Using Unique Clade-Specific Marker Genes**. *Nat Methods* 2012, **9**:811–4.

[11] Su X, Pan W, Song B, Xu J, Ning K: Parallel-META 2.0: enhanced Metagenomic Data Analysis with Functional Annotation, High Performance Computing and Advanced Visualization. *PLoS One* 2014, 9:e89323.

[12] Ondov BD, Bergman NH, Phillippy AM: Interactive Metagenomic Visualization in a Web Browser. *BMC Bioinformatics* 2011, **12**:385.

[13] Goecks J, Nekrutenko A, Taylor J, Galaxy Team: Galaxy: a Comprehensive Approach for Supporting Accessible, Reproducible, and Transparent Computational Research in the Life Sciences. *Genome Biol* 2010, **11**:R86.

[14] Yu Z, Morrison M: Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* 2004, 36(5):808-12.

[15] Mande SS, Mohammed MH, Ghosh TS: Classification of metagenomic sequences: methods and challenges. *Brief Bioinform* 2012, 13(6):669-81.

### 6 Capacity of the participating organisations

Name	Location of Research premises (city / country)	Type of R&I activities	No. of full- time employees	No. of Employees in R&I	Website	Annual Turnover (approx., in Euro)
NSilico	Cork/Ireland	Health Informatics & Software Development	7	6	www.nsilico.com	€350k

#### Data for non-academic beneficiaries

Page **37** of **46** 

Organisation	Scotland's Rural College (SRUC)
General Description	High-quality strategic and applied research at SRUC is focused on rural, environmental and land-based activity and underpins our education and consultancy activities. Our work addresses health and productivity in animals, animal welfare and crops, promotes low carbon farming and increases farm output through efficiency and innovation.Our vision is to undertake both leading edge research and application development within 'future beef and sheep farming systems' and for some system components that go from farm into the food chain.
key people (including supervisors)	the Beef and Sheep Research Centre at SRUC. He joined SRUC in 2013, having previous headed leading ruminant research units in Wales, New Zealand and Ireland. He is the author of 95 primary research papers with over 2500 cites and an H-index of 30. He received the Sir John Hammond Memorial Award (British Society of Animal Science) in 2008 and is a former President of the New Zealand Society of Animal Production. Richard's personal research programme has been focussed at the interfaces between ruminant nutrition, product composition and rumen function, and involved many national and international collaborations. Current research is developing biomarkers for feed conversion efficiency and greenhouse gas emissions in ruminants. He is also leading an international collaborative project focussed on rumen stability and long-term effects of early-life nutrition. He has long-standing interests in feed evaluation, ruminant fatty acid metabolism and less invasive techniques to study rumen function.
Key Research Facilities, Infrastructure and Equipment	SRUC has established world-class livestock greenhouse gas (GHG) emission research facilities. These are part of the new SRUC Beef & Sheep Research Centre at the Bush Estate near Edinburgh: Respiration chambers to measure methane output from individual cattle or a cow and calf (or sheep); 44 HOKO feeders enabling individual feed intake in group pens (enabling 130+ cattle to be recorded at a time). HOKO feeders have integrated canopies with 'sniffers' to enable gas sampling from identified cattle whilst they feed.
Independent research premises?	Yes
Previous Involvement in Research and Innovation Projects	Significant contributions at the interfaces between dairy nutrition, milk composition and rumen function - notably modelling of forage composition, dry cow feeding strategies, forages and fatty acids, fatty acids and fertility, and rumen diagnostics. Some work used directly by the animal feed industry – notably feed evaluation systems, feed values and dry cow strategies. Much more of my work has been used by farmers and advisers – notably to develop advice and models to predict and increase forage intakes, improve the utilisation of feed protein (reduce N pollution), improve milk composition, and facilitate the use of forage legumes in conventional and organic farming. Demonstrated practicality and feeding value of baled legume silages, which have been adopted by the industry; Development of 'Acidogenicity Value' concept for feed evaluation, which has been used by French feed manufacturer, COFNA/Evialis; Results from a series of silage-feeding studies incorporated into UK system for predicting silage intake; Results concerning effects of legume silages and organic farming on milk fatty acids used for promotions by UK Organic Milk Supply Company; Results from a long series of dry cow feeding studies incorporated into guidelines on dry cow feeding (UK Milk Development Council).
Current involvement in Research and Innovation Projects	Beef & Sheep Research Centre: just completing two Innovate UK-funded projects and started/about to start 4 further Innovate UK projects – mainly in the area of precision tools for managing beef production. Recent £600k 'Beef Improvement' project secured from DEFRA and two ongoing projects funded by EBLEX. These projects are typically valued at £100-150k per annum, so that total external income for the team is ~£0.5M per annum. Beef Improvement Programme: £600k for SRUC (from DEFRA). BeefMonitor: £302k for SRUC (from Innovate UK).
	PrecisionBeef: £320k for SRUC (from Innovate UK) FCE-MARKER: £90k for SRUC (from Innovate UK).
Relevant Publications and/or innovation products	McCartney, C.A., Dewhurst, R.J. and Bull, I.D. (2014). Changes in the ratio of tetraether to diether lipids in cattle feces in response to altered dietary ratio of grass silage and concentrates. Journal of Animal Science 92: 4095-4098. Wheadon, N.M., McGee, M., Edwards, G.R. and Dewhurst, R.J. (2014). Relationship between Nitrogen isotopic fractionation, measured in plasma, and feed conversion efficiency of growing beef heifers. British Journal of Nutrition 111: 1705-1711. McCartney, C.A., Bull, I.D., Waters, S.M. and Dewhurst, R.J. (2013). Comparison of biomarker and molecular biological methods for estimating methanogen abundance. Journal of Animal Science 91: 5724-5728.

Organisation	Ulster University (UU), Computer Science Research Institute (CSRI)
General Description	The UU is an innovative research and learning institution. It has an international reputation for research excellence, innovation and regional engagement. Recognized as one of the leading research institutions in the UK, Ulster has consistently achieved international standard ratings in the UK Research Assessment Exercise, demonstrating its commitment to continuous high quality research. The Computer Science Research Institute is one of 17 research institutes at Ulster. Computer Science at Ulster is now in the top 25% in the UK for research power (as measured by the UK RAE2014 national research assessment exercise).
Role and Profile of key people (including supervisors)	<b>Dr. Haiying Wang (male)</b> received a PhD degree on <b>artificial intelligence in biomedicine</b> from the University of Ulster, UK, in 2004. His current research focuses on the development of advanced data integration techniques and network-based approaches for the extraction of disease signatures from heterogeneous biological and medical data. He was appointed as Principal Investigator (PI) since Feb. 2008 for CARDIOWORKBENCH - Drug Design for Cardiovascular Diseases: Integration of in Silico and in Vitro Analysis funded by FP6 STREP program. He was a grant holder of the TSB-ALIP/EPSRC project NOCTURNAL: Night Optimised Care Technology for UseRs Needing Assisted Lifestyles and is a Co-PI of the TSB project: Extraction and modelling of Alzheimer's disease data for patient stratification: a novel integrative, multiplex network approach. Currently he is coordinating a proposal COMPASS: COMPutational Approach to Systems medicine for the understanding of multiple Sclerosis submitted to H2020-PHC-2015-two-stage. Since 2004, he has published more than 100 peer-reviewed papers in journals, books and conference proceedings. He has supervised 4 Research Associates and has successfully supervised 4 PhD students to completion. He will lead WP2 and be responsible for machine learning algorithms development and supervision. He will also support the management, supervison and desemination of the project.
Key Research Facilities, Infrastructure and Equipment	The CSRI has a large-scale Intelligent Environment (approx. 630m2) to support research, deployment and evaluation of Connected Health solutions, data acquisition and semantic analysis of user environments. This facility includes 3 dedicated smart labs (including a smart kitchen and living room), 2 conference rooms, 16 personal offices, and a demonstration area to support local community engagement. The NI Functional Brain Mapping Facility, including a MEG imaging system enables research in neurological modelling to develop pharmaceutical treatments for neurological degenerative diseases. The NETCOM hub is established as the "Research Lab with fastest connections (to USA) in Western Europe". The computational biology insterest group has two high performance computers, GPU and a dedecated server to support the Big Data analytics. It also has a large suite of software products, including Labview, MATLAB, 3ds Max and Adobe CS5 in addition to a number of machine learning and systems biology algorithms and research platform developed in house.
Independent research premises?	Yes
Previous Involvement in Research and Innovation Projects	<ul> <li>SAVASA (multiple video archive search, video semantic analysis, 2011-2014): EU FP7 Security Programme, £350000 to University of Ulster;</li> <li>Self-Management Platform for Connected Health, funded by Invest N. Ireland; £1.09 Million;</li> <li>STAR (STAR: Skills Training and Re-Skilling for Carers of People with Dementia), £35,358, 2010-2012;</li> <li>SMART COPD: Home based management of COPD, funded by Sheffield University from CLAHRC Project via NHS, £51,301;</li> <li>SMART2 project: EPSRC EQAUL5 SMART: Self Management supported by Assistive, Rehabilitation and Telecare Technologies, funded by EPSRC EQUAL5 (EP/F001959/1);</li> <li>NOCTURNAL, Night Optimised Care Technology for UseRs Needing Assisted Lifestyles, funded by TSB-ALIP/EPSRC (TS/G002452/1);</li> <li>DEEPFLOW (Extraction of argument structures from text to support innovation and decision, October 2010-December 2013), a joint research project with SAP and Queen's University Belfast, funded by InvestNI R&amp;D Programme. The total grant is over 1.5 million pound and the University of Ulster grant is 250,000 pound;</li> <li>India-UK Advanced Technology Centre of Excellence in Next Generation Networks Systems &amp; Services: EPSRC, 2009-2012;</li> <li>CARDIOWORKBENCH - Drug Design for Cardiovascular Diseases: Integration of in Silico and in Vitro Analysis, Funded by FP6 STREP program (PL018671). Total project value: €3,038,103, value to Ulster: £187,345, project date: 12/2005-05/2009;</li> <li>Deployment of Sensing Technology in Connected Health Care: DEL RCIF, 2009-2011;</li> <li>Cross-Border Centre for Intelligent Point of Care Sensors: DEL R&amp;D, 2008-2011;</li> </ul>
Current involvement in Research and Innovation Projects	Moneta(Extraction and modelling of Alzheimer's disease data for patient stratification: a novel integrative, multiplex network approach), funded by Technology Strategy Board UK, 2015 – 2018. INTACT (On the Impact of Extreme Weather on Critical Infrastructures, 2014-2017), FP7 SEC-2013.2.1-2. Self-Management Platform for Connected Health, funded by Invest N. Ireland; £1.09 Million Fast Video-based Anomaly Detection on Rail Lines (2014-2016): Royal Society International Exchanges Scheme - 2013 China NSFC Costshare. COBACORE (Community Based Comprehensive Recovery, 2013-2016): EU FP7 SEC-2012.4.3-1.

Page **39** of **46** 

	<b>HARMONISE</b> (Holistic Approach to Resilience and Systematic Actions to make Large Scale UrbaN Built Infrastructure Secure, 2013-2016): EU FP7 SEC-2012.2.1-1.
Relevant Publications and/or innovation products	Bingjing Cai, Haiying Wang, Huiru Zheng, Hui Wang, "Identification of Protein Complexes from Tandem Affinity Purification/Mass Spectrometry Data via Biased Random Walk," IEEE TRANSACTIONS ON Computational Biology and Bioinformatics, 2014, 12(2), pp. 455-466. Haiying Wang, Huiru Zheng, "Organized Modularity in the Interactome: Evidence from the Analysis of Dynamic Organization in the Cell Cycle," IEEE TRANSACTIONS ON Computational Biology and Bioinformatics, 2014, 11(6), pp.1264-1270. Huiru Zheng, Haiying Wang, Hua Xu, Yonghui Wu, Zhongming Zhao, Francisco Azuaje, "Linking Biochemical Pathways and Networks to Adverse Drug Reactions, "IEEE Transactions on NanoBioscience, 2014, 13(2), pp.131-137. Wang, Haiying, Zheng, Huiru, Azuaje, Francisco and Zhao, Xing-Ming, Drug-Domain Interaction Networks in Myocardial Infarction. IEEE Transactions on Nanobioscience, 2013, 12(3), pp.182-188. Cai, Bingjing, Wang, Haiying, Zheng, Huiru and Wang, Hui (2012) Detection of protein complexes from affinity purification/mass spectrometry data. BMC Systems Biology, 6 (Suppl 3). S4.

Organisation	Forschungsinstitut für Telekommunikation und Kooperation e.V. (FTK)
General Description	Since its foundation as an interdisciplinary research organisation closely related to University of Hagen (FernUniversität in Hagen, FUH) in 1991, the work of FTK - Research Institute for Telecommunication and Cooperation (FTK e.V. Forschungsinstitut für Telekommunikation und Kooperation, FTK) concentrates on the area of Information and Communication Technologies (ICT). An interdisciplinary team with well-founded research background paired with an expanded network to all important stakeholders in the field of ICT makes FTK a leading institution for the development and implementation of innovation strategies for business and administration in the ICT area. The institute is organized as a non-profit membership organisation and is statutory pursuing objectives in the public interest.
Role and Profile of	Matthias Hemmje (Male) is affiliated as full professor for Computer Science with the Distance University
key people (including supervisors)	in Hagen, Department of Mathematics and Computer Science, where he holds the Chair of Multimedia and Internet Applications. His primary research interests include information retrieval, multimedia databases, virtual environments, information visualization, visual interaction, and multimedia. He also guides research in the area of content engineering, knowledge technologies, peer-to-peer based systems, collaboration support systems and evaluation of interactive systems. Since 2009, Matthias Hemmje is a member of the board of FTK. Having worked in many international R&D projects with research and industrial partners, he has experience in managing projects with very large consortia. <b>Dominic Heutelbeck (Male)</b> (Ph.D. in Computer Science), the General Manager of FTK, received his Dr. rer. nat. in Computer Science at the FernUniversität in Hagen. His current research covers cross- organisational and collaborative IPR management, as well as location based services. He has extensive experience in research projects includes EC funded projects like CargoScoreCard (CSC), Virtual Information and Knowledge Environment Framework (VIKEF), Digital Preservation Europe (DPE), Nestor II, Sustaining Heritage Access through Multivalent ArchiviNg (SHAMAN). In addition he has experience in coordinating large groups of partners within projects.
Key Research	The FTK research facility is located in Dortmund, Germany. FTK provides workstations, a lot of online
Facilities, Infrastructure and Equipment	services, document sharing, source code version control systems and wikis to every employee.
Independent research premises?	FTK has one office in Dortmund, Germany.
Previous Involvement in Research and Innovation Projects	<ul> <li>FP7 Network of Excellence APARSEN (Alliance for the Permanent Access to the Records of Science Network and Virtual Center of Excellence: formal qualifications, online training portal, and interactive map)</li> <li>FP7 CP-CSA SCIDIP-ES (SCIence Data Infrastructure for Preservation with focus on Earth Science: data preservation infrastructure, interactive platform and training)</li> <li>FP7 SHAMAN (Sustaining Heritage Access through Multivalent ArchiviNg: digital preservation infrastructure and services, scenarios/integration subprojects and training)</li> </ul>
Current involvement in Research and Innovation Projects	H2020 RIA RAGE (Realising an Applied Gaming Eco-system: leading the Ecosystem Development workpackage)
Relevant Publications and/or innovation products	1. Wim Jansen, Roberto Barbera, Michel Drescher, Antonella Fresa, Matthias Hemmje, Yannis Ioannidis, Norbert Meyer, Nick Poole, Peter Stanchev (2013). e-Infrastructures for Digital Libraries the Future. In Trond Aalberg, Christos Papatheodorou, Milena Dobreva, Giannis Tsakonas, Charles J. Farrugia (Eds.): Research and Advanced Technology for Digital Libraries - International Conference on Theory and Practice of Digital Libraries, TPDL 2013, Valletta, Malta, September 22-26. Proceedings. Springer 2013 Lecture Notes in Computer Science ISBN 978-3-642-40500-6, Volume 8092, 2013, pp 480-481. 2. Matthias Hemmje (2013). Scientific Communities in Knowledge Societies. In Shantanu Ganguly and P K Bhattacharya (eds): Proceedings of International Conference on Digital Libraries (ICDL) 2013: Vision 2020: Looking back 10 years and forging new frontiers, 27-29 November 2013, New Delhi 3. Matthias Hemmje (2014). Towards a Virtual Centre of Excellence supporting Digital Preservation of the Records of Science. In Dinesh Katre and David Giaretta (eds.): Proceedings of APA/C-DAC International Conference on Digital Preservation and Development of Trusted Digital Repositories. 5-6 February 2014, New Delhi, India (New Delhi: Excel India Publishers) 4. Wolfgang Wilkes, Jörg Brunsmann, Dominic Heutelbeck, Andreas Hundsdörfer, Matthias Hemmje, Hans-Ulrich Heidbrink, (2011). "Towards support for long-term digital preservation in product life cycle management". In Int. Journal of Digital Curation, Vol. 6, No 1, 282-296. 5. Dominic Heutelbeck. "Preservation of Enterprise Engineering Processes by Social Collaboration Software". In Proc. of the 2nd Int. Conf. on Collective Intelligence (COLLIN 2011), June 14, 2011 Jeju, South Korea.

Organisation	University of Edinburgh, Edinburgh Parallel Computing Centre (EPCC)
General Description	A part of the University of Edinburgh, the Edinburgh Parallel Computing Centre (EPCC) is an International centre for excellence in high-performance computing for over 20 years, with World- class systems, data storage and support services for industry and science. It has a global reputation for innovative and leading-edge high-performance computing research, expert-led training in high-performance computing for engineers and scientists.
Role and Profile of key people (including supervisors)	<b>Terry Sloan (Male),</b> Group manager EPCC, Terry originally joined EPCC in 1994 and holds the position of Group Manager within EPCC's Software Development Group. He has extensive experience of managing novel, HPC and Grid projects for Scottish SMEs, UK corporations, European and global collaborations. In 2006, Terry became one of four Group Managers in EPCC's Applications Group with responsibility for resourcing and development of staff as well as continuing to lead my own projects. These included a role as an Activity Leader from 2004 to 2008 within the EU-funded DEISA (Distributed European Infrastructure for Supercomputing Applications) consortium where he managed a software development team distributed across 4 European sites with the resulting software being supported across the 11 leading European supercomputing sites. Between 2008 and 2009, Terry also managed the development of DEISA's open source Life Sciences portal by a European-wide team of software engineers. From 2005 to 2011, Terry was the Project Manager for the £1.5m Scottish Funding Council eDIKT2 (e-Science Data, Information and Knowledge Transformation) project. This involved managing a diverse group of researchers, software developers and systems staff located in various Schools and research groups within the University of Edinburgh. eDIKT2 produced research proposals that generated more than £15m and 400 publications. This experience
Key Research Facilities, Infrastructure and Equipment	makes him ideal as a supervisor for incoming researchers. He will act as WP Leader for WP 2 EPCC is the lead partner in the HPCx consortium, supporting the national supercomputing service for UK academic research, funded to the tune of £115 million. EPCC houses an exceptional range of supercomputers, with 75 staff committed to the solution of real-world problems. The facility is based around a Cray XC30 supercomputer that provides the central computational resource. It is supported by a number of additional components including: high-performance parallel file systems, pre- and post-processing facilities, external login nodes, and a large, resilient, long-term data facility.
Independent research premises?	Yes. EPCC is based in James Clerk Maxwell Building (JCMB) in the University of Edinburg, and includes access to supercomputing frameworks, ample research desk space, large study rooms, an open area of informal, catered study space, a refurbished foyer area with study pods, WiFi and open access PCs.
Previous Involvement in Research and Innovation Projects	A large number of national and EU funded projects, a small sample EU projects of which is listed below: ADMIRE: Advanced Data Mining and Integration for Europe BEINGRID: Business Experiments in Grid DEISA 2 JRA2: Cosmological applications
Current involvement in Research and Innovation Projects	A large number of projects are running currently, including the Marie Skłodowska-Curie IAPP FP7 project ClouDx-I and the following <b>DEISA 2 JRA3</b> : Tools for scientific data services <b>DEISA</b> : Distributed European Infrastructure for Supercomputing Applications <b>EGEE</b> : Enabling Grids for E-SciencE <b>EUFORIA</b> : HPC for EU fusion research community
Relevant Publications and/or innovation products	Parallel Optimisation of Bootstrapping in R, M. Sloan, T., Piotrowski, M., Forster, T. & Ghazal, P. 24 Jan 2014 ArXiv, Troup, E., Forster, T., Sloan, T., Cebamanos, L. & Ghazal, P. Nov 2013. Embedded systems for global e-Social Science: Moving computation rather than data, Lloyd, A., Sloan, T., Antonioletti, M. & McGilvary, G. Jul 2013 In : Future Generation Computer Systems. 29, 5, p. 1120- 1129. SPRINT: taking biomedical analysis from the desktop to supercomputers and the cloud, Sloan, T. Jun 2013 Edinburgh : EPCC, University of Edinburgh. Exploiting Parallel R in the Cloud with SPRINT, Piotrowski, M., McGilvary, G. A., Sloan, T. M., Mewissen, M., Lloyd, A., Forster, T., Mitchell, L., Ghazal, P. & Hill, J. Jan 2013 In : Methods of Information in Medicine. 52, 1, p. 80-90. Parallel classification and feature selection in microarray data using SPRINT, Mitchell, L., Sloan, T. M., Mewissen, M., Ghazal, P., Forster, T., Piotrowski, M. & Trew, A. 2012 In : Concurrency and Computation: Practice and Experience.

Organisation	NSilico Lifescience Ltd (NSilico)
General Description	NSilico is the provider of easy-to-use data management and analytics software for the life sciences and healthcare industries. The company's offerings are based upon a unique and unrivalled blend of biological, computing, software-development and clinical experience and expertise which enables NSilico to provide customers with solutions which significantly increase the efficiency and accuracy of their work. NSilico is officially partnered with major technology providers such as Microsoft, IBM and Dell. NSilico is officially supported by Ireland's Health Innovation Hub, which is a health service scheme to work with innovation companies to provide technology for health care.
Role and Profile of key people (including supervisors)	<b>Dr Paul Walsh (Male)</b> CTO: Dr Paul Walsh is an accomplished entrepreneur and a leading technical architect with extensive experience in designing and implementing a range of biopharma and medical IT systems and is a certified Project Management Professional. He has won numerous research awards including best papers in diverse fields ranging from machine learning to entrepreneurship. He also holds a lectureship position at Cork Institute of Technology. Thus, his industry and academic experience make him ideal to be part of the Knowledge Sharing Oversight Group. He will be the WP leader for WPs 1 and 5. <b>Xiangwu Lu (Male)</b> holds an MSc in Computer Science and has published research papers on bioinformatics. Alfie Keary (Male) has over 30 years ICT (Information & Communications Technologies) and holds an MSc in Computer Science from National University Ireland. He is an experienced EU researcher having participated in the FP4 TALENT project. Tim Manning (Male) is an ER researcher with an MSc in Computer Science and a wide range of academic and industry outputs including beer reviewed papers on software systems for breast cancer detection. Robin Deegan (Male) is an ER with an MSc in computer science and a wide range of academic and industry outputs including best paper awards.
Key Research Facilities, Infrastructure and Equipment	Suite of high end work stations, high throughput compute instances, multiple servers of (8 X 1.6GHZ CPU, 14GB RAM, 2,040GB STORAGE), 10 TByte RAID storage, server with over 100 compute cores designed in partnership with Dell, latest state of the art development and test environments, access to office space of 3,000m <sup>2</sup>
Independent research premises?	Yes, access to office space of 3,000m <sup>2</sup>
Previous Involvement in Research and Innovation Projects	Enterprise Ireland: Bioinformatics management framework, short listed Innovation award http://www.itcork.ie/
Current involvement in Research and Innovation Projects	<b>FP7 Project ClouDx-i</b> a Marie Skłodowska-Curie actions IAPP project for the use of cloud computing based molecular diagnostics for sepsis management in neonatal units, see www.cloudxi.eu. Health Innovation Hub supprt to bring well developed healthcare projects access to real-world experience in the health sector, see www.hih.ie.
Relevant Publications and/or innovation products	Improving the Performance of CGPANN for Breast Cancer Diagnosis using Crossover and Radial Basis Functions, Timmy Manning, Paul Walsh The European Conference on Evolutionary Computation, Machine Learning and Data Mining in Computational Biology 2012. Accelerating in silico research with workflows: A lesson in Simplicity. Walsh P, Carroll J, Sleator RD. Comput Biol Med. 2013 Dec 1;43(12):2028-35. Naturally selecting solutions: the use of genetic algorithms in bioinformatics. Manning T, Sleator RD, Walsh P. Bioengineered. 2013 Sep-Oct;4(5):266-78. Biologically inspired intelligent decision making: A commentary on the use of artificial neural networks in bioinformatics. Manning T, Sleator RD, Walsh P., Bioengineered. 2013 Dec 16;5(2). Metagenomics and novel gene discovery: Promise and potential for novel therapeutics. Culligan EP, Sleator RD, Marchesi JR, Hill C. Virulence. 2013 Apr 1;5(3). Review.

### 7 Ethics Issues

Any sampling (of bacterial samples in cattle gut material) and physical recording of the animal will be reviews and approved by SRUC Animal Ethic Committee. The ethical review process is a statutory requirement (since 1999) at all UK establishments designated under the Animals (Scientific Procedures) Act 1986. All staff that work with animals must agree to abide by SRUC's Code of Practice on the use of animals in research.

The purpose of the AEC is to oversee, assess, monitor and ensure the health and well-being of all experimental animals within SRUC or those used outside of SRUC that involve its staff or students. The SRUC's AEC is to review all proposed research and provide advice to scientists, students and consultants on how to best design their studies, taking into account the 3Rs (replacement, reduction, refinement) in accordance with the EU Directive 86/609 and Irish legislation by Statuary instrument 17/94; to ensure that all animals used receive the best care and attention to their welfare, while also being compliant with the relevant legislation and codes of practice; and to review, assess and approve all personal and project licence applications to the UK Home Office and to maintain close contact with the UK Home Office Inspector.

### 8 Letters of Commitment of partner organisations



An Lárionad um Thaighde agus Nuálaíocht Maldir le hAinmhithe agus Talamh Féaraigh AN GHRÁINSEACH Dún Samhnai, Co na Mi Anúmal & Grassland Research, and Innovation Centre GRANGE: Dunsany, Co. Meath Tel: +353 (0)46 906 1100 Fax: +353 (0)46 902 6154

Dr Haiying Wang Senior Lecturer Computer Science Research Institute School of Computing and Mathematics University of Ulster Jordanstown Campus Shore Road Newtownabbey Co. Antrim BT37 0QB

Dear Haiying

I am writing to confirm that Teagasc will participate as an associate in the RISE Project Meta-Plat and that we fally support the objectives of the proposed research programme.

Teagasc is the agriculture and food development authority in Ireland. Its mission is to support sciencebased innovation in the agri-food sector and the broader bioeconomy that will underpin profitability, competitiveness and sustainability. Teagasc has been working directly on application of next generation sequencing (NGS) technologies (Illumina Hiseq and Miseq and Roche 454 platforms)to agriculture and food for over 10 years and has published on use of NGS in metagenomics, whole genome sequencing and gene expression (mRNA and miRNA seq). We have significant expertise in both data generation (animal and plant models, extraction of RNA and DNA, library preparation and sequencing) and high throughput computational analysis. We also own a number of next generation sequencing platforms (Illumina Miseq, Illumina Next seq 500, Life technologies Ion Torrent, Roche 454).

Teagase will assist in the project in the following ways:

1. Advise on sequencing activity

2. Use case study input

3. Review of scientific goals and progress

Yours sincerely

Matthew McCabe

TEAGASC- The Agriculture and Food Development Authority www.teagast.ie

V.A.T. BEO NO. IE 0050202 O

Page 45 of 46
Meta-Plat RISE

## **END PAGE**

Marie Skłodowska-Curie Actions

## Research and Innovation Staff Exchange (RISE)

## Call: H2020-MSCA-RISE-2015

PART B

"MetaPlat"

Page **46** of **46** 



This electronic receipt is a digitally signed version of the document submitted by your organisation. Both the content of the document and a set of metadata have been digitally sealed.

This digital signature mechanism, using a public-private key pair mechanism, uniquely binds this eReceipt to the modules of the Participant Portal of the European Commission, to the transaction for which it was generated and ensures its full integrity. Therefore a complete digitally signed trail of the transaction is available both for your organisation and for the issuer of the eReceipt.

Any attempt to modify the content will lead to a break of the integrity of the electronic signature, which can be verified at any time by clicking on the eReceipt validation symbol.

More info about eReceipts can be found in the FAQ page of the Participant Portal. (<u>http://ec.europa.eu/research/participants/portal/page/faq</u>)