

# BioAssist II 2009-2012

Barend Mons, Marcel Reinders

19-10-09

## Preamble

This brief document describes the **aim**, the **organization** and the **approach** of the NBIC support program, **BioAssist**. It will serve as a high abstraction guideline for the Management of NBIC and the dedicated personnel of the support- and professional engineering teams of BioAssist<sup>1</sup>.

## BioAssist: Community-based bioinformatics support for life sciences research

The NBIC BioAssist programme is conceived as a community endeavour in which biologists and (bio)informaticians jointly organise bioinformatics support. BioAssist has a prime role in the support of core NBIC partners, but the programme also actively seeks collaboration with other centres and networks of excellence, in- and outside The Netherlands. The BioAssist programme identifies existing 'code' and 'content' of generic value and assists the original creators in developing these assets for community use. So called 'Support Platforms' integrate tools and databases to provide specific bioinformatics functionalities needed for the analysis, integration and interpretation of experimental data from Genomics, Proteomics, Metabolomics and Epidemiological approaches, including BioBanks. The tools and databases incorporated in these dynamically changing platforms need not be provided as a central service, but can remain at distributed and connected local computer systems (GRID nodes) of the contributors. Consequently, these tools and databases should be continuously updated and maintained by the respective NBIC partners.

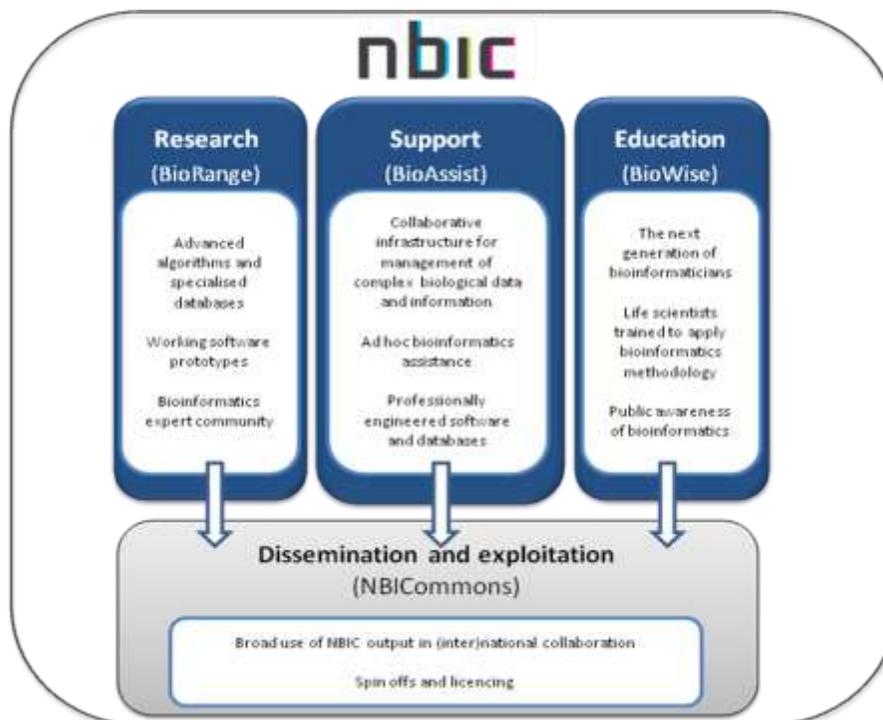


Fig 1. The new overview of NBIC as a whole after 2009

### **BioAssist Mission**

*To realize the perceived role of the NBIC support pillar (see figure 1), and thereby of NBIC as a whole, as an indispensable enabling technology center and network for the informatics needs of biological research and development in The Netherlands, and beyond, by 2013.*

### **Specific aims**

1. **Support** BioAssist platforms (see figure 3) in their mission to develop and implement 'working prototypes' of 'code' and 'content', needed by the community at large, to enable efficient generation, capturing, storage and analysis of complex and massive experimental data and the further biological interpretation of these data to generate new insights in the life sciences.
2. **Co-ordinate** within and between current platforms to ensure effective exchange of information, efficient and regular collaboration and to guarantee a balance of programming expertise available in the overall NBIC supported programming community.
3. **Support collaboration** within and between the Support Platforms and with national and international third parties when appropriate to realize the creation of smooth pipelines from the generation of massive and complex data all the way through to biological interpretation.
4. **Analyze and review** code and content resulting from BioRange, BioAssist platform activities and (potentially) related external Research and Development activities to select elements for further professional engineering into stable, industrial grade Open Source and Open Access products and services.
5. **Develop** selected code and content elements into reliable and user friendly products and services through software engineering
6. **Provide ad hoc support** for practical and experimental early prototypes used in the interim phase from research to development

As of October 2009, a dedicated, centrally managed, but distributed software engineering team supports the scientifically driven platforms to develop professional software and content services. New and emerging technologies from the e-Science, Web and Grid communities increasingly support this interactive and flexible approach. Moreover, it ensures timely adaptation of the platforms to keep pace with the changing needs and increasingly complex questions of the life sciences community.

E-bioscience is the application of e-science approaches in biosciences for multi-disciplinary, distributed collaborations. The BioAssist e-bioscience infrastructure enables seamless incorporation and integration of data producing research facilities, larger data collections, and computing power. In addition this infrastructure will soon be indispensable for complex interpretation through computational analysis and visualisation dispersed data

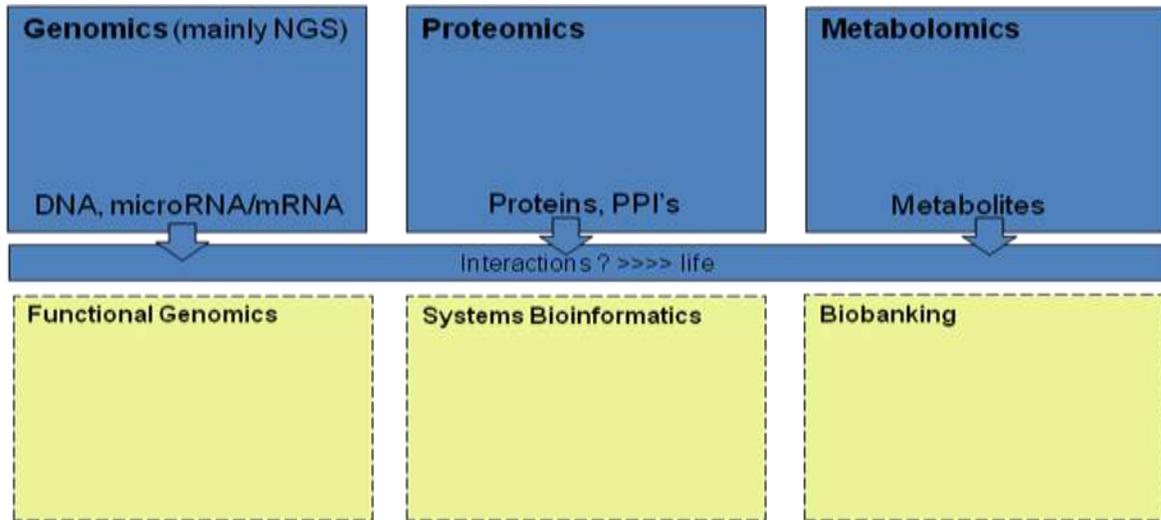


Fig 2: the current NBIC BioAssist platforms. Blue platforms mainly deal with data generation and management; green platforms mainly deal information management and biological data interpretation.

**The 2009 BioAssist Organization (see figure 3)**

The BioAssist pillar of NBIC (fig. 1) is organized in three different, but closely collaborating, components to support both distributed end-user validated prototyping and early development, support with the use of early stage tools and services and finally to guarantee that appropriate code and content is taken 'the last mile' towards professional, user friendly products.

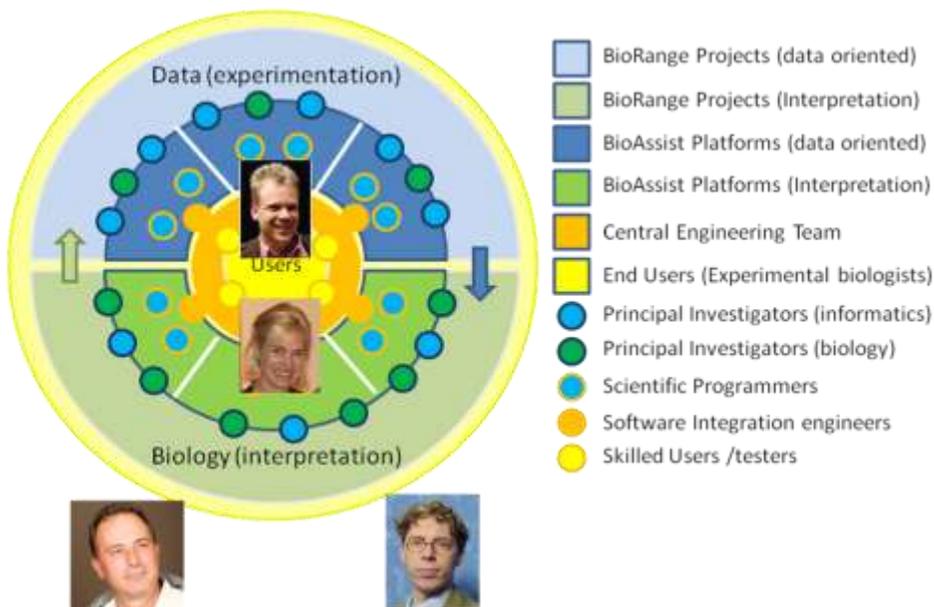


Fig. 3: the new situation of BioAssist after reorganization of platforms into data oriented and biological interpretation oriented platforms. The CTO will mainly focus on software engineering and the support of the data oriented platforms, while the CIO will mainly supervise and co-ordinate the biological interpretation oriented platforms. Please note that the sequence of research, functional design, prototyping, testing and final software engineering to achieve user friendly products goes from periphery to center in this picture

The BioAssist program will function with its primary focus on the partner network of NBIC [URL], but will actively work with national and international parties, to realize its mission.

As of the 1<sup>st</sup> of October of 2009, the BioAssist pillar has been strengthened with a second tranche of scientific programmers as well as with the aforementioned dedicated, centrally managed engineering team headed by a Chief Technology Officer (CTO) and a Chief information Officer (CIO), both with many years of academic as well as industrial experience.

Over 25 **scientific programmers**<sup>ii</sup> will be working in a ‘distributed fashion’ with direct supervision performed by a Principal Investigator of one of the platforms as described in figure 3. However these programmers will regularly meet in groups and interact with the **software integration engineers** of the **Central Engineering Team**. Each platform will have two dedicated Senior Software Integration Engineers, one for 0.5 fte and one ‘back up’ for 0.2 fte as ‘foremen’.

The PI’s of supported platforms define, together with the relevant staff members of the BioAssist Management Team (see below) and the Senior Software Integration Engineers the user/functional design of the systems to be developed and distribute the work amongst the scientific programmers of the platform to develop working prototypes. The Software Integration Engineer is responsible for documenting all specifications and designs. Once the working prototypes have been developed into early applications and tested by skilled users, code and content that is proven to be of generic interest and needs to be supported beyond individual partner institutes will be further engineered and integrated under the sole supervision of the CTO and the CIO of BioAssist. The original developers of the code and content elements will be asked to serve as consultants during this process.

The centrally managed, but physically distributed engineering team supports the distributed programmers’ network, especially in issues related to the effective and efficient development and implementation of code and content of direct use for their respective platforms. In addition, the senior staff of the central engineering team will be in regular and planned contact with all PI’s in the various platforms<sup>iii</sup> The (organically) growing group of software integration engineers will be closely supervised by the CTO for software development. The CIO will work closely with the biologists in the NBIC network and beyond to identify, capture, test, curate and provide the most important content sources for biological interpretation, both originating from NBIC partners and from national and international third parties. Specialized content experts can be seconded by partners to work closely with the BioAssist CIO on content issues for specific platforms such as Biobanking.

The daily collaboration between the CIO and the CTO will ensure optimal co-development of code and content, where code will be dedicated to data generation, storage, analysis and management, while the ‘content’ collected and provided under supervision of the CIO will mainly serve as the substrate for applications developed for the interpretation of the newly generated data in the context of the most comprehensive collection of global information available. The development of the biological interpretation platforms will be closely coordinated with the international standardization developments, such as the Concept Web Alliance and the World Wide Web Consortium (W3C).

The software engineers and content specialists in the centrally managed team will be employed partly directly by NBIC, either from existing budget or through external grants, and partly by third parties that can be qualified as ‘Power Users’.

Power Users are individual institutes or networks (not necessarily restricted to the NBIC narrow constituency, which are mainly the NGI related institutes) in- and outside The Netherlands.

The dedicated Research Support Team, is headed by the NBIC research coordinator and provides *ad hoc* support of researchers, industry and other platform and application users who encounter problems or bottlenecks in their bioinformatics related activities.

The BioAssist program as a whole is headed by a full time Program Leader, a senior scientist with both long-term international academic and science management experience as well as industrial experience. The Program Leader is a member of the General Management Team of NBIC.

The Program Leader will be supported by a peer senior scientist for 1 day per week, who is a very experienced in the core informatics issues of BioAssist, including e-science and GRID issues. The peer scientist is also a member of the General Management Team of NBIC and participates on behalf of NBIC in the Executive team of BigGRID

In October 2009, the BioAssist team is composed as follows:

Program leader :	<b>Barend Mons, PhD (CV)</b>
Peer Program Leader :	<b>Marcel Reinders, PhD (CV)</b>
Chief Technology Officer:	<b>Rob Hooft, PhD (CV)</b>
Chief Information officer:	<b>Christine Chichester, PhD (CV)</b>
Research Co-ordinator:	<b>Marc van Driel, PhD (CV)</b>

(link to BioRange and head of the Research Support team):

Current Senior Software Integration Engineers:

**Bharat Singh** (cv) partner: Gen2Phen,  
**Dmitry Katsubo** (partner: BBMRI(?)) (cv),  
**Leon Mei**, PhD(NBIC, (cv)  
**Kees Burger** (software integration engineer)

The Two Program Leaders, the CTO, the CIO and the NBIC Research co-ordinator form the BioAssist Management Team.

The Management Team will be supported by an advisory group in which the senior software engineers appointed in the platforms will participate, as well as several specialists on hardware, and content issues.

### **Operational issues**

It is the principle task of the designated platform leader(s) (PI's) to formulate the initial goals of each specialized platform, essentially resulting in a rough 'functional design'. The BioAssist Program Leaders will actively engage at the PI level in the process of functional specifications. The advice of the CTO, the CIO and the foremen will be asked and will be highly valuable, but the central engineering team has a consulting rather than a decisive role in promoting the coherence of the platform at the 'scientific' level. The closer code or content will come to a repetitive, subsidiary application level, the more intensively the central engineering team will be involved in the technical design and the application development. Once the technical design moves from working prototype phases into agile development cycles, based on a crystallized functional specifications document, the CTO or the CIO (respectively for code and content) will take over process responsibility from the PI's and the original developers will function in a consultative role

The processes involved are schematically represented in figure 4.

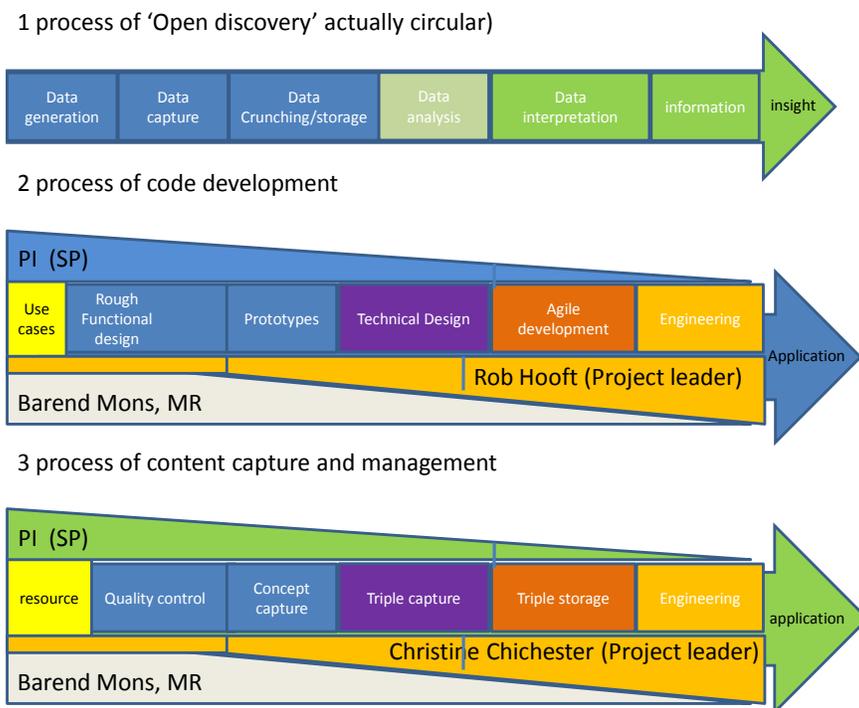


Fig. 4: (1) the general process from data generation to biological insight (actually frequently a circular process) is depicted first. Please note that in many -omics driven discovery processes the hypothesis (and thus 'queries') are frequently not well defined or even largely absent, the so called 'open discovery' process. Data oriented platforms covering -omics research usually have as a half-product to discovery lists of genes, RNA's, proteins and metabolites. In the midst of the process (the transition from blue to green in the picture) the emphasis switches from (newly generated) data analysis to 'projection' of these new data on previously generated information to reach biological insight. These information resources for the vast majority globally spread and maintained and therefore the 'green' phase of discovery is much more internationally oriented than the first set of steps. (2,3) In terms of code development, the picture shows the relative and changing involvement of the program leaders, the PI's and the CTO, respectively CIO of BioAssist. The vertical line in the orange triangular, indicates the switch point where further downstream agile development circles become the responsibility of the central engineering team staff.

#### Further measures to ensure coherence within and between platforms.

- The mandate of the BioAssist programme is to deliver broad bioinformatics applications in The Netherlands by creating workable solutions that are needed by a wide community of bioinformaticians and biologists bioinformaticians and 'platforms' are an organizational means, not a goal in themselves.
- Newly assigned scientific programmers as well as software engineering positions will be subject to a reasonable level of consensus in the platform specific development groups. Platforms that cannot agree on code and content issues will thus be put under 'healthy pressure' by the Bio-Assist Management Team to reach consensus on one or in exceptional cases multiple approaches for their specializations, before the platform PI's can ask the assistance of the central engineering team. During the negotiations of the distributed scientific programmers' positions of BioAssist II, the level of consensus within one platform will be part of the evaluation and negotiation before the position is actually awarded and filled. If a platform

cannot agree on even the basics of an approach, that platform will be reorganized, split or eventually could lose the support and the subsidy of BioAssist.

- Please also note that the platform composition and definition is a purely organizational matter. Several platform activities (and increasingly so as a matter of policy) will be performed in close collaboration with institutes and power users that are not directly under the NBIC Management Teams' influence.
- The CTO and the CIO have created criteria for both code and content to be supported by the engineering team beyond what can be done by the individual NBIC partners. These criteria are described on the website of NBIC BioAssist [URL]. In as far as the candidate code and content to be further developed by the engineering team do not naturally emerge from the activities and procedures agreed upon by the platforms and the foremen (the default route), individual researchers or NBIC partners (as well as third parties) can offer code and content to the BioAssist program for 'further development'. The CTO and/or the CIO will make a first selection and the final decision on support for any code or content will be made by the Program Leaders of BioAssist in regular or *ad hoc* meetings of the BioAssist Management Team.
- NBIC has the following policy for the management of its BioAssist projects, and this is a requirement for obtaining support from the NBIC BioAssist Program:

-Each major project has a group in MyExperiment (hyperlink).

-The project management (priority setting, ticketing etc. is done in the on-line system RT (hyperlink)

-Each participant in the project has a profile in MyExperiment and has subscribed to the project group (example in figure 5)

-Each participant in the project has access to the RT project management system to the extent deemed necessary by the project oversight group.

- The central engineering team will actively support and promote wider use of code and content in conjunction with the original developers or other third parties, a process coined in Europe as 'valorization' and closely equivalent to what would be coined as 'dissemination and exploitation' in the USA.
- However, the valorization process is delegated specifically to a NBIC subgroup called NBICCommons, headed by the managing director of NBIC (Ruben Kok, PhD) and supported by the program leader of BioAssist Barend Mons, PhD). NBICCommons will decide and support the initial valorization steps of code and content related to NBIC in any way, in agreement with the organization that is the IP holder. Actual commercial use of code and content under the auspices of NBIC may be delegated at the discretion of NBICCommons and the original IP holder (and decided by the managing director) to an external management/service company that will in turn make decisions about licensing, spin-off creation and related commercial issues. Such a management/service company will necessarily be independent from NBIC management. The management company will be charged with the creation of revenues to support its own costs, and with the specific aim to return part of its revenues to NBIC to make the BioAssist program and potentially other programs of NBIC partly self sustained.

Make a Group Announcement
Invite People
Edit Group
Delete Group

### Group: BioAssist BioBanking Platform (BBMRI-related)

[Members \(29\)](#) | [News](#) | [Shared Items \(1\)](#) | [Citations \(0\)](#) | [Manage Memberships](#) | [Requests Pending \(0\)](#) | [Invites Pending \(0\)](#) | [Tags \(0\)](#) | [Announcements](#) | [Comments \(0\)](#)

This group has been set up to exchange ideas, documents, workflows and news about the [BioBanking Platform](#) of NBIC in the The Netherlands. However, international members of the [Concept Web Alliance](#) or the [BBMRI](#) initiatives are very welcome to join. It will keep you updated on all technical developments regarding data harmonization and other aspects of BioBanking.

Colleagues involved in the [P3G datashaper project](#) are also invited to join this group as conceptual interoperability of BioBanks and comparability of samples based on the manner in which samples were collected are very related, yet distinct challenges in [BBMRI](#).

[More later](#)

**Created at:** Friday 31 July 2009 @ 11:53:06 (BST)

**Unique name:** NBICommonsBBMRI

You have set this Group to send you all Membership Requests for confirmation first

**Administrator**



[Barend](#)

---

**29 members**

**1 shared item**

**2 announcements**

0 requests pending

0 invites pending

---

**This group has been credited 0 times**

**Members**

 <a href="#">Barend</a>	 <a href="#">Segolene Ayme</a>	 <a href="#">Peter 't Hoen</a>	 <a href="#">M. Scott Marshall</a>	 <a href="#">Xiaoliang</a>	 <a href="#">Mgros</a>
 <a href="#">Margreet Brandsma</a>	 <a href="#">Antoine van Kampen</a>	 <a href="#">Susanna A Sansone</a>	 <a href="#">Mark Wilkinson</a>	 <a href="#">Bharat Singh</a>	 <a href="#">Daniel Kinzler</a>
 <a href="#">Christine Chichester</a>	 <a href="#">Denis costello</a>	 <a href="#">Kristina Hettne</a>	 <a href="#">Albert Mons</a>	 <a href="#">Joseph</a>	 <a href="#">Gjvo</a>
 <a href="#">Hailiang Mei</a>	 <a href="#">Rafaelsidi</a>	 <a href="#">Andrew Su</a>	 <a href="#">Alejandra Gonzalez Beltran</a>	 <a href="#">Bvanbre...</a>	 <a href="#">Dmitry Katsubo</a>

**Tags (0)**

None [\[ edit \]](#)

Add Tags

---

**Announcements**  
(private & public)

[Visit Paris to discuss Wikipedia Scholar with Segolene and Denis](#)  
47 days ago by [Barend](#)

[wikiwords \(for the fanatics\)](#)  
73 days ago by [Barend](#)

[\[ New \]](#) [\[ See All \]](#)

Fig. 5, the biobanking interest group (currently 29 members) a link to RT for specific projects will be added soon

**The research Support team (Marc)**

Initial, experimental hosting and services beyond the capacity of individual institutes will be provided by BioAssist. Professional services in terms of support (SLA) and hosting (24/7) will, in principle. NOT

be provided by NBIC BioAssist, but will be delegated to one or more commercial specialized parties, at the discretion of either the BioAssist program leaders, CTO or CIO or of the management company (for commercial services)

Some important policy issues of BioAssist.

Code should be Open Source (different licenses available) by default. In BioRange and BioAssist projects that start as of 2009, partners will have to agree up front with this policy. In rare cases, and only based on solid arguments, code can be kept from Open Source licensing, at the discretion of an advisory panel to the Scientific Director of NBIC (Antoine van Kampen PhD).

1. Content should be Open Access and Interoperable<sup>iv</sup> by default In BioRange and BioAssist projects. NBIC partners will have to agree up front with this policy. In rare cases, and only based on solid arguments, code can be kept from Open Source licensing, at the discretion of an advisory panel to the Scientific Director of NBIC (Antoine van Kampen PhD).
2. Code and Content developed within NBIC or its partner groups will only be considered for further support and development if considered 'best of breed' in international context. If this criterion is not met, other solid arguments should be brought forward to consider support for the code or content by BioAssist. Important criteria for code might be that the current software on the 'market' is not Open Source. For content, similar arguments will be considered, such as that certain databases are not Open Access, or should be repackaged or combined for specific purposes.
3. NBIC BioAssist will avoid [Despautism](#) and will thus not engage in repetitive or redundant code or content development and support, and as a consequence, an international comparison of each code and content element, including a survey using all available scientific literature should always be part of any request for support by BioAssist. For highly specialized code and not previously published biological and medical information, even a small country such as The Netherlands may assume to have unique experience that warrants de novo code or content development beyond regular prototyping and data production. Most information is by definition universal and thus global and that is why BioAssist and NBICommons will actively participate in and heavily rely on international activities and networks for code, infrastructural issues and content management. Such networks are e.g. ESFRI, ELIXIR, BBMRI, Gen2Phen and W3C/CWA. For the judgment of added value of code and content presented to BioAssist and NBICommons the close European partners, the Swiss Institute for Bioinformatics and OMII UK will be involved as consulting partners in case any doubt exists within the NBIC management teams responsible for decisions.

### External Relations

The national and international embedding of the BioAssist activities will be the responsibility of the Program Leader(s) of BioAssist in close co-ordination with the directors of NBIC. Within the Netherlands, active collaborations are being set up with other 'enabling technology centres, such as the Netherlands Proteomics Centre and the Netherlands Metabolomics Centre. These centers also actively participate in the BioAssist Platforms.

NGI centers and networks of Excellence, such as CBSG, Kluyver Centre, CGC, CMSB and the top-institutes for Pharma and Green Genetics as well as non-NGI funded centers and networks such as CTMM, VIRGO and Cyttron, will be actively engaged in the generic e-science and bioinformatics infrastructure development lead by NBIC. NBIC is also actively participating in BigGrid and other infrastructural initiatives.

Internationally, NBIC plays an increasingly important role, both in the development of code and workflows (mainly with OMII UK, headed by Neil Chue Hong) and in the management of globally important content (mainly with the Swiss Institute of Bioinformatics, SIB), headed by Ron

Appel. With OMII UK a formal Memorandum of Understanding is in place and with SIB a similar agreement is in preparation.

For the interoperability of content needed to make comprehensive interpretation efforts for massive data, NBIC plays a leading role in the Concept Web Alliance [URL] and is collaborating actively with the Semantic Web Consortium (W3C).

The international embedding of the BioAssist activities is the responsibility of the Program Leader of BioAssist

#### Current BioAssist support platforms

##### **A. Data production and analysis**

1. High-throughput sequencing data
2. Analysis and data management for Proteomics
3. Analysis and data management for Metabolomics

##### **B. Biological Interpretation of data**

4. Integration of clinical and genomics data (biobanking)
5. Systems bioinformatics
6. Functional analysis of genomics data
7. Infrastructure / e-Science

Please note, some of these platforms will be merged and/or renamed as a result of the current reorganization process.

#### **Getting involved**

Please contact:

For General issues:

Dr. Barend Mons, Program Leader

[Barend.mons@nbic.nl](mailto:Barend.mons@nbic.nl)

For technical issues: Dr. Rob Hooft

Chief technical Officer

[Rob.hooft@nbic.nl](mailto:Rob.hooft@nbic.nl)

For content issues: Dr. Christine Chichester

Chief information Officer

[Christine.chichester@nbic.nl](mailto:Christine.chichester@nbic.nl)

For *ad hoc* research support: Dr. Marc van Driel

[title]

[marc.van.driel@nbic.nl](mailto:marc.van.driel@nbic.nl)

Website: <http://www.nbic.nl/support/bioassist/>

---

<sup>i</sup> Some parts of this text may be used for the description of BioAssist on the Web Site of NBIC, but in principle this is a document for internal purposes.

<sup>ii</sup> Some 'programmers' may actually have PostDoc level, but in that case should not be under pressure of 'first author' publication output requirements.

<sup>iii</sup> (at request for the platforms/positions of BioAssist I and as a pre-condition for funding scientific programmers in platforms for BioAssist II).

<sup>iv</sup> Interoperability of content means that **data** are stored in formats and systems that makes them accessible and understandable by other groups and systems. For previously analyzed data (now **information**) that is published in text or databases, this is even more important and the concepts (genes, micro-RNA's, proteins, small molecules, organelles, cell types, tissues, organisms, diseases, people etcetera) should be unambiguously identified and computer readable. NBIC is a leading partner in the international a Concept Web Alliance (URL to section in NBIC site] with the mission to: **"enable an open collaborative environment to jointly address the challenges associated with high volume scholarly and professional data production, storage, interoperability and analyses for knowledge discovery."**