

The Munich NGS-FabLab for medical sequence data



Sebastian Schaaf^{1,2}, Aarif Mohamed Nazeer Batcha², Sandra Fischer², Guokun Zhang², Ulrich Mansmann^{1,2}

¹ Department of Medical Informatics, Biometry and Epidemiology (IBE), Ludwig Maximilians University (LMU) Munich, Germany

² German Cancer Consortium (DKTK), Heidelberg, Germany

• <u>Fab</u>rication <u>Lab</u>oratory

- <u>Fab</u>rication <u>Lab</u>oratory
- ,Personal genomics': improve prediction & precision in diagnostics
- esp. in oncology and rare inheritable diseases

- <u>Fab</u>rication <u>Lab</u>oratory
- ,Personal genomics': improve prediction & precision in diagnostics
- esp. in oncology and rare inheritable diseases
- Truth: rather patient <u>stratification</u> by molecular typing
- very young technique, rarely applied in practice, no approved ,gold standards'

validation via clinical studies <-> personalization

- <u>Fab</u>rication <u>Lab</u>oratory
- ,Personal genomics': improve prediction & precision in diagnostics
- esp. in oncology and rare inheritable diseases
- Truth: rather patient <u>stratification</u> by molecular typing
- very young technique, rarely applied in practice, no approved ,gold standards'

validation via clinical studies <-> personalization

- important: biomedical context knowledge
- <u>legal</u> and insurance matters still in their infancy

- (in a nutshell)
- Holds for any personally identifiable information (PII)
- Data items subject to special protection: racial, political, religious, sexual, ... and health

(in a nutshell)

- Holds for any personally identifiable information (PII)
- Data items subject to special protection: racial, political, religious, sexual, ... and **health**

1. Prohibition of conditional permission:

The collection, processing and use of personal data is strictly prohibited, unless it is permitted by the law or the person concerned gives consent (§ 4 I BDSG).

2. Principle of immediacy

- 3. Priority to special laws
- **4. Principle of proportionality:** [...] a balancing of interests has to take place.

5. Principle of Data Avoidance and Data Economy:

Economy of data: Data that are not absolutely necessary to collect for the purpose of achievement may not be applicable

Data avoidance: If data must be collected, dealing with them must be limited to the necessary minimum

6. Principle of Transparency

7. Principle of Earmarking:

If data can be collected for a particular purpose, dealing with them is bound to this purpose. A new purpose-setting requires a law or consent.

(in a nutshell)

- Holds for any personally identifiable information (PII)
- Data items subject to special protection: racial, political, religious, sexual, ... and **health**

1. Prohibition of conditional permission:

The collection, processing and use of personal data is strictly prohibited, unless it is permitted by the law or the person concerned gives consent (§ 4 I BDSG).

- 2. Principle of immediacy
- 3. Priority to special laws
- **4. Principle of proportionality:** [...] a <u>balancing</u> of interests has to take place.

5. Principle of Data Avoidance and Data Economy:

Economy of data: Data that are not absolutely necessary to collect for the <u>purpose</u> of achievement may not be applicable

Data avoidance: If data must be collected, dealing with them must be limited to the necessary minimum

6. Principle of Transparency

7. Principle of Earmarking:

If data can be collected for a particular purpose, dealing with them is <u>bound to this purpose</u>. A new purpose-setting requires a law or consent.

(in a nutshell)

- Holds for any personally identifiable information (PII)
- Data items subject to special protection: racial, political, religious, sexual, ... and **health**

1. Prohibition of conditional permission:

The collection, processing and use of personal data is strictly prohibited, unless it is permitted by the law or the person concerned gives consent (§ 4 I BDSG).

- 2. Principle of immediacy
- 3. Priority to special laws
- **4. Principle of proportionality:** [...] a <u>balancing</u> of interests has to take place.

5. Principle of Data Avoidance and Data Economy:

Economy of data: Data that are not absolutely necessary to collect for the <u>purpose</u> of achievement may not be applicable

Data avoidance: If data must be collected, dealing with them must be limited to the necessary minimum

6. Principle of Transparency

7. Principle of Earmarking:

If data can be collected for a particular purpose, dealing with them is <u>bound</u> to this purpose. A new purpose-setting requires a law or consent.

- Genetic information *per se* can be only formally anonymous (and not *de facto*)
- no data export: if, then only minimum information, pseudonymized, with approval and data processing contract

• Gene Center's Galaxy-system as a model

• Gene Center's Galaxy-system as a model

Victoria Stodden, Ross Lazarus, Luca Pireddu: "Reproducible science" ('13)

- Gene Center's Galaxy-system as a model
- storage & processing of human-derived data <u>only</u> in enclosed hospital network

Victoria Stodden, Ross Lazarus, Luca Pireddu:

'Reproducible science" ('13)

 \rightarrow data security approvable & central

- Gene Center's Galaxy-system as a model
- storage & processing of human-derived data <u>only</u> in enclosed hospital network

Victoria Stodden, Ross Lazarus, Luca Pireddu:

(Reproducible science" ('13)

- \rightarrow data security approvable & central
- close interaction to existing environment (their rules, our command roughly)
 - \rightarrow integrative & connective

- Gene Center's Galaxy-system as a model
- storage & processing of human-derived data <u>only</u> in enclosed hospital network
 - ightarrow data security approvable & central
- close interaction to existing environment (their rules, our command roughly)
 - \rightarrow integrative & connective

Larry Helseth: "Customizing Galaxy for a Hospital Environment" ('12)

Sanjay Joshi: "The Clinical Galaxy" ('13)

Victoria Stodden, Ross Lazarus, Luca Pireddu:

'Reproducible science" ('13)

- Gene Center's Galaxy-system as a model
- storage & processing of human-derived data <u>only</u> in enclosed hospital network
 - ightarrow data security approvable & central
- close interaction to existing environment (their rules, our command roughly)

 \rightarrow integrative & connective

• Galaxy for data analysis

 \rightarrow flexible and scalable



Larry Helseth: "Customizing Galaxy for a Hospital Environment" ('12)

Sanjay Joshi: "The Clinical Galaxy" ('13)

Victoria Stodden, Ross Lazarus, Luca Pireddu: "Reproducible science" ('13)

- Gene Center's Galaxy-system as a model
- storage & processing of human-derived data <u>only</u> in enclosed hospital network
 - ightarrow data security approvable & central
- close interaction to existing environment (their rules, our command roughly)

Galaxy

 \rightarrow integrative & connective

- Galaxy for data analysis
 - ightarrow flexible and scalable
- Bii for ontology-based organization
 - \rightarrow structured, understandable and value-increasing



Sanjay Joshi: "The Clinical Galaxy" ('13)



Victoria Stodden, Ross Lazarus, Luca Pireddu: "Reproducible science" ('13)

- Gene Center's Galaxy-system as a model
- storage & processing of human-derived data <u>only</u> in enclosed hospital network

Victoria Stodden, Ross Lazarus, Luca Pireddu:

'Reproducible science" ('13)

- \rightarrow data security approvable & central
- close interaction to existing environment (their rules, our command roughly)



- Gene Center's Galaxy-system as a model
- storage & processing of human-derived data <u>only</u> in enclosed hospital network

Victoria Stodden, Ross Lazarus, Luca Pireddu:

(Reproducible science" ('13)

- ightarrow data security approvable & central
- close interaction to existing environment (their rules, our command roughly)



- cross-connecting IT infrastructure requires respective social structure
 - \rightarrow NGS <u>A</u>dmin <u>R</u>ound<u>t</u>able (NGS-ART)
 - → Teaching Unit (heavily inspired by ICB, Medical College, Cornell University)

- Gene Center's Galaxy-system as a model
- storage & processing of human-derived data <u>only</u> in enclosed hospital network
 - ightarrow data security approvable & central
- close interaction to existing environment (their rules, our command roughly)



- cross-connecting IT infrastructure requires respective social structure
 - → NGS <u>A</u>dmin <u>R</u>ound<u>t</u>able (NGS-ART)
 - → Teaching Unit (heavily inspired by ICB, Medical College, Cornell University)

Victoria Stodden, Ross Lazarus, Luca Pireddu:

'Reproducible science" ('13)



- given general IT infrastructure is nearly completely Windows-driven
- no ,templating' cross-section service
- no working order (= financing) for hospital's central IT to support research
- we are not ,standard' in every respect
- long way from concept to necessary dedicated hardware
- licenses indeed cost money (e.g. SLES & VMware are not for free)
- one full position is not enough and even students cost money
- generic concepts are accepted as long as every particular interest is served first

- given general IT infrastructure is nearly completely Windows-driven
 - \rightarrow web interface (and somehow integrate the server)
- no ,templating' cross-section service
- no working order (= financing) for hospital's central IT to support research
- we are not ,standard' in every respect
- long way from concept to necessary dedicated hardware
- licenses indeed cost money (e.g. SLES & VMware are not for free)
- one full position is not enough and even students cost money
- generic concepts are accepted as long as every particular interest is served first

- given general IT infrastructure is nearly completely Windows-driven
 - \rightarrow web interface (and somehow integrate the server)
- no ,templating' cross-section service
 - \rightarrow be the first (and keep searching even for partial commonalities)
- no working order (= financing) for hospital's central IT to support research
- we are not ,standard' in every respect
- long way from concept to necessary dedicated hardware
- licenses indeed cost money (e.g. SLES & VMware are not for free)
- one full position is not enough and even students cost money
- generic concepts are accepted as long as every particular interest is served first

- given general IT infrastructure is nearly completely Windows-driven
 - \rightarrow web interface (and somehow integrate the server)
- no ,templating' cross-section service
 - \rightarrow be the first (and keep searching even for partial commonalities)
- no working order (= financing) for hospital's central IT to support research
 - ightarrow give some input to the 5-year strategy paper
- we are not ,standard' in every respect
- long way from concept to necessary dedicated hardware
- licenses indeed cost money (e.g. SLES & VMware are not for free)
- one full position is not enough and even students cost money
- generic concepts are accepted as long as every particular interest is served first

• given general IT infrastructure is nearly completely Windows-driven

 \rightarrow web interface (and somehow integrate the server)

- no ,templating' cross-section service
 - \rightarrow be the first (and keep searching even for partial commonalities)
- no working order (= financing) for hospital's central IT to support research

 \rightarrow give some input to the 5-year strategy paper

- we are not ,standard' in every respect
 - \rightarrow sell it as ,interesting' and ,challenging'; include admins!

Olivier Inizan & Mikae

Loaec: "DevOps"

- long way from concept to necessary dedicated hardware
- licenses indeed cost money (e.g. SLES & VMware are not for free)
- one full position is not enough and even students cost money
- generic concepts are accepted as long as every particular interest is served first

• given general IT infrastructure is nearly completely Windows-driven

 \rightarrow web interface (and somehow integrate the server)

- no ,templating' cross-section service
 - \rightarrow be the first (and keep searching even for partial commonalities)
- no working order (= financing) for hospital's central IT to support research

ightarrow give some input to the 5-year strategy paper

• we are not ,standard' – in every respect

 \rightarrow sell it as ,interesting' and ,challenging'; include admins!

long way from concept to necessary dedicated hardware

 \rightarrow stay patient, compromise

- licenses indeed cost money (e.g. SLES & VMware are not for free)
- one full position is not enough and even students cost money
- generic concepts are accepted as long as every particular interest is served first

• given general IT infrastructure is nearly completely Windows-driven

 \rightarrow web interface (and somehow integrate the server)

- no ,templating' cross-section service
 - \rightarrow be the first (and keep searching even for partial commonalities)
- no working order (= financing) for hospital's central IT to support research

ightarrow give some input to the 5-year strategy paper

we are not ,standard' – in every respect

 \rightarrow sell it as ,interesting' and ,challenging'; include admins!

long way from concept to necessary dedicated hardware

 \rightarrow stay patient, compromise

• licenses indeed cost money (e.g. SLES & VMware are not for free)

 \rightarrow stress the requirements

- one full position is not enough and even students cost money
- generic concepts are accepted as long as every particular interest is served first

Olivier Inizan & Mikael Loaec: "DevOps" ('13)

• given general IT infrastructure is nearly completely Windows-driven

 \rightarrow web interface (and somehow integrate the server)

- no ,templating' cross-section service
 - \rightarrow be the first (and keep searching even for partial commonalities)
- no working order (= financing) for hospital's central IT to support research

ightarrow give some input to the 5-year strategy paper

• we are not ,standard' – in every respect

 \rightarrow sell it as ,interesting' and ,challenging'; include admins!

long way from concept to necessary dedicated hardware

 \rightarrow stay patient, compromise

• licenses indeed cost money (e.g. SLES & VMware are not for free)

 \rightarrow stress the requirements

• one full position is not enough and even students cost money

→ cooperate heavily (social network)

- \rightarrow go for some good students (and ensure payment for at least 6 months)
- generic concepts are accepted as long as every particular interest is served first

Olivier Inizan & Mikael Loaec: "DevOps" ('13)

• given general IT infrastructure is nearly completely Windows-driven

 \rightarrow web interface (and somehow integrate the server)

- no ,templating' cross-section service
 - \rightarrow be the first (and keep searching even for partial commonalities)
- no working order (= financing) for hospital's central IT to support research

 \rightarrow give some input to the 5-year strategy paper

• we are not ,standard' – in every respect

 \rightarrow sell it as ,interesting' and ,challenging'; include admins!

long way from concept to necessary dedicated hardware

 \rightarrow stay patient, compromise

• licenses indeed cost money (e.g. SLES & VMware are not for free)

 \rightarrow stress the requirements

• one full position is not enough and even students cost money

→ cooperate heavily (social network)

- → go for some good students (and ensure payment for at least 6 months)
- generic concepts are accepted as long as every particular interest is served first
 - \rightarrow be consequent reliable partners will support you, the others... well...

What came out?

What came out?



infrastructure:

→ Poster P10

What came out?



infrastructure: → Poster P10 ,experiences': → Poster P3



Hardware

- 2x Xeon 8-core CPU @ 2.7 Ghz
- 256 GB RAM
- 1x 1 TB SSD @ RAID5
- 1x 1 TB HDD @ RAID5
- 1x 40 TB HDD @ RAID 6

Hardware

- 2x Xeon 8-core CPU @ 2.7 Ghz
- 256 GB RAM
- 1x 1 TB SSD @ RAID5
- 1x 1 TB HDD @ RAID5
- 1x 40 TB HDD @ RAID 6

Karen Reddy: "Capacity planning. Go big." ('12)

Hardware

- 2x Xeon 8-core CPU @ 2.7 Ghz
- 256 GB RAM
- 1x 1 TB SSD @ RAID5
- 1x 1 TB HDD @ RAID5
- 1x 40 TB HDD @ RAID 6

Some (maybe remarkable) key features:

- SUSE Linux Enterprise Server (SLES) v11
- Virtualization via VMware vSphere
- NFS-mounted shared /tmp over virtual LAN



Hardware

- 2x Xeon 8-core CPU @ 2.7 Ghz
- 256 GB RAM
- 1x 1 TB SSD @ RAID5
- 1x 1 TB HDD @ RAID5
- 1x 40 TB HDD @ RAID 6

Some (maybe remarkable) key features:

- SUSE Linux Enterprise Server (SLES) v11
- Virtualization via VMware vSphere



Olivier Inizan & Mikael Loaec: "DevOps" ('13)

Larry Helseth: "Customizing Galaxy for a Hospital Environment" ('12)



Hardware

- 2x Xeon 8-core CPU @ 2.7 Ghz
- 256 GB RAM
- 1x 1 TB SSD @ RAID5
- 1x 1 TB HDD @ RAID5
- 1x 40 TB HDD @ RAID 6

Some (maybe remarkable) key features:

- SUSE Linux Enterprise Server (SLES) v11
- Virtualization via VMware vSphere
- NFS-mounted shared /tmp over virtual LAN
- Shell-driven ,whole installation' script for setting up a complete instance
 - environment/network, users/groups/permissions
 - Galaxy & third-party tools
 - reference genomes/indeces, user data
 - config via .ini file (creatable via shell dialogue or Galaxy form)



Olivier Inizan & Mikael Loaec: "DevOps" ('13) Larry Helseth: "Customizing Galaxy for a Hospital Environment" ('12)

 \rightarrow takes ca. 15 min. for everything

Hardware

- 2x Xeon 8-core CPU @ 2.7 Ghz
- 256 GB RAM
- 1x 1 TB SSD @ RAID5
- 1x 1 TB HDD @ RAID5
- 1x 40 TB HDD @ RAID 6

Some (maybe remarkable) key features:

- SUSE Linux Enterprise Server (SLES) v11
- Virtualization via VMware vSphere
- NFS-mounted shared /tmp over virtual LAN
- Shell-driven ,whole installation' script for setting up a complete instance
 - environment/network, users/groups/permissions
 - Galaxy & third-party tools
 - reference genomes/indeces, user data
 - config via .ini file (creatable via shell dialogue or Galaxy form)



Virtualization:

- Maximum flexibility
- multi-system env on one server
- Resource priority settings for prod instance
- development cycle on OS level
- Scalability: decoupling of hardware and system

Olivier Inizan & Mikael Loaec: "DevOps" ('13) Larry Helseth: "Customizing Galaxy for a Hospital Environment" ('12)

 \rightarrow takes ca. 15 min. for everything



HOME

DATA REGISTRY

ANALYIS PLATFORM

GALAXY MAIN RESOURCES

WIKI

GUIDELINES

PUBLICATIONS

NGS-ART

HOME

Welcome to NGS infrastructure! Please click the following link for <u>Galaxy-Analysis platform</u> and <u>Data Registry</u>.



....

Impressum | Datenschutz | Kontakt



Affliated Institutions

- IBE
- Medical clinic III
- Pathology

Environment

- MIT
- Children's hospital
- Gene Center



- To Do's
 - Galaxy version upgrade (June 2013...)
 - launch SGE cluster (managing issue, computing power)
 - get more 'Galaxy'ish,
 - contribute...
 - automatic references setup
 - visualizations!

- To Do's
 - Galaxy version upgrade (June 2013...)
 - launch SGE cluster (managing issue, computing power)
 - get more 'Galaxy'ish.
 - contribute...

Dannon Baker: Contribute to Galaxy ('13)

- automatic references setup
- visualizations!

- To Do's
 - Galaxy version upgrade (June 2013...)
 - launch SGE cluster (managing issue, computing power) _



- To Do's
 - Galaxy version upgrade (June 2013...)
 - launch SGE cluster (managing issue, computing power) ____



- To Do's
 - Galaxy version upgrade (June 2013...)
 - launch SGE cluster (managing issue, computing power)



- To Do's
 - Galaxy version upgrade (June 2013...)
 - launch SGE cluster (managing issue, computing power)
 - ISAtools get more 'Galaxy'ish. (Desktop software) Dannon Baker: Contricontribute... Pathway bute to Galaxy ('13) Analysis automatic references setup Galaxy Bii visualizations! (BioInvestigation Index) future? launched \rightarrow end of guerilla movements Ontologies secondary instance Sequencer **Biomedical** financing? (Illumina miSeq, Ion Torrent) Metadata extensions? organization? responsibility? Convey HC-2^{ex} (FPGA **Beowulf Cluster** CONVEY (PC clients @ SGE) Hybrid Core)

10

THANKS

- open source projects: Galaxy, isa-tools.org
- funding: m⁴ + DKTK
- Prof. Ulrich Mansmann (IBE)
- Prof. Klaus Kuhn (IMSE, TUM)
- IBE:
 - Tobias Schleinkofer
 - Klaus Rüschstroer
 - Lother Weiker
 - Dr. Thomas Müller
 - Amy Werner-Allen, PhD
- hospital's IT (MIT) guys:
 - Gregor Pickert
 - Jens Kather
 - Frank Hülle
 - Sammy Simba
 - Simon Leutner
 - Sebastian Haaske
 - Dr. Kurt Kruber









- Dep. of Internal Medicine 3:
 - Prof. Wolfgang Hiddemann
 - Prof. Volker Heinemann
 - Prof. Sebastian Stintzing
 - Dr. Klaus Metzeler
 - Dr. Philipp Greif
 - Sebastian Vosberg
 - Maja Rothenberg-Thurley
- Institute of Pathology:
 - Prof. Andreas Jung
 - Matthias Frölich
- LAFUGA group (Gene Center Munich):
 - Alexander Graf
 - Dr. Stefan Krebs
 - Dr. Helmut Blum

CLUSTER

• members of the NGS-ART





THANKS²



The NGS-FabLab core team

- > Aarif Mohamed Nazeer Batcha
- Guokun Zhang
- Sandra Fischer
- Ashok Varadharajan
- Lena Rüschstroer

THANKS²



The NGS-FabLab core team

- > Aarif Mohamed Nazeer Batcha
- Guokun Zhang
- Sandra Fischer
- Ashok Varadharajan
- Lena Rüschstroer

...and thanks to all of you for listening!

visualization

visualization



visualization



isatools/isasoftwaresuite

