

Digital Humanities and NGS

Muhammad S Sarwar

National eScience Centre
University of Glasgow

m.sarwar@nesc.gla.ac.uk



OUTLINE

- Introduction
- ENROLLER - An Enhanced Repository for Language and Literature Researchers
- Current Work
- Using NGS
- Conclusion

Introduction

- Humanities scholars are producing huge volumes of digital data
- editions, text, text archives, corpora, dictionaries, thesauri, sound and video etc.
- Issues affecting the efficient use of data resources
- interoperability, inaccessibility, access rights, data security, user interface etc.



University
of Glasgow



JISC



ENROLLER

ENROLLER

- ENROLLER - An Enhanced Repository for Language and Literature Researchers
 - A JISC funded project (2009-2011)
 - National e-Science Centre (NeSC) and University of Glasgow English Department
 - Multilingual (English, Scots and Gaelic)
- Data sets participating in the project
 - OED (Oxford English Dictionary), HTE (Historical Thesaurus of English), DSL (Dictionary of Scots Language), SCOTS (Scottish Corpus of Text and Speech), CMSW (Corpus of Modern Scottish Writing), NECTE (Newcastle Electronic Corpus of Tyneside English)



OBJECTIVES

- To develop an interactive, research infrastructure providing seamless access to participating data collections
- A well designed easy-to-use search system, access to digital sound, video and textual data
- Develop tools for linguistic analysis (such as: concordancing, collocation and frequency analysis)
- Seamless secure access to licensed data; by developing automatically enforced access and usage policies
- Support for addition of new data collections
- Building large-scale data indexes for searching and exploiting HPC and e-Science facilities (such as ScotGrid and NGS)



University
of Glasgow



JISC



ENROLLER

Example Usage Scenario

- |

- Investigate concepts in use
- Find out about people being “timid” or “afraid”;
 - How was it discussed in different texts?
 - What words are often found near it?
 - What association it has in different areas?

Example Usage Scenario

- 2

- Discovering more about a word
 - Analyst wants to know more about word “canny”.
 - Lookup definition in OED.
 - Lookup definition in DSL
 - Results for the word in SCOTS and NECTE
 - "...Bit fairmers are canny fowk an they..."
 - "...ye wis a canny hoosewife, ye'd keep..."
 - The places word is found in Historical Thesaurus.



University
of Glasgow



JISC



ENROLLER

Current Work

- Simple Searches
 - Data resides on local (to institution servers)
 - Input: a file containing word(s) or phrases
 - Output: to Display, Save to file.
- Cross-collection Searches
 - Search same word(s) in multiple collections

Current Work ...

- Bulk Searches (over NGS)
 - Data resides on NGS
 - Input: Word(s), Phrases,
 - Output: to Display, Save as file
- Execute Workflows (over NGS)
 - Data, Input and Output same as above.

Current Work ...

- Creation of a Virtual Organization (VO) over NGS
- Implementation of data management solutions using available services (SRB, AFS, etc)
- Testing, bug-fixing and UI improvements
- Alpha version of ENROLLER portal scheduled to go live during first week of March 2010.

Workflow Example



Workflow Example

- Input a word OR upload a file containing the words/phrases (terms) to be searched

Timid



Workflow Example

- Input a word OR upload a file containing the words/phrases (terms) to be searched
- Search the terms in thesaurus (for example: HTE)

Timid

{acolmod, egeful, (ge)forht, forhtful
forhtindlic, forhtig, forhtmod, herebleap,
ungedyrstig, unpriste, blethe<bleap, fey<fæge,
unbold<unbeald, unbold<unbeald, argh<earg,
frightful, feared, ferdy, fearful, ferdful,
g(h)astful, trembling, timorous, cremetous,
cremeuse, craintive, sheepish, meticulous,
timid, tremebund, awful, soft, pale,
timorsome, tremulous, pigeon-hearted,
affrightful, formidolous, pavid, timidous,
unsupported, tender-nosed, scary,
pippin-hearted, kitten-hearted, funky,
tender-footed, fearsome misventurous,
scare, cotton-wool,

} **51 entries**



Workflow Example

- Input a word OR upload a file containing the words/phrases (terms) to be searched
- Search the terms in thesaurus (for example: HTE)
- Search the results from thesaurus-search in Scottish Corpus

Timid

{acolmod,egeful,(ge)forht,forhtful
forhtindlic,forhtig,forhtmod,herebleap,
ungedyrstig,unpriste,blethe<bleap,fey<fæge,
unbold<unbeald,unbold<unbeald,argh<earg,
frightful,feared,ferdy,fearful,ferdful,
g(h)astful,trembling,timorous,cremetous,
cremeuse,craintive,shoopish,meticulous,
timid,tremebund,awful,soft,pale,
timorsome, tremulous, pigeon-hearted,
affrightful, formidolous, pavid, timidous,
unsupported, tender-nosed, scary,
pippin-hearted,kitten-hearted, funky,
tender-footed, fearsome misventurous,
scare,cotton-wool,
} **51 entries**



Workflow Example

- Input a word OR upload a file containing the words/phrases (terms) to be searched
- Search the terms in thesaurus (for example: HTE)
- Search the results from thesaurus-search in Scottish Corpus
- Find concordances for for each of the words

Timid

{acolmod,egeful,(ge)forht,forhtful
forhtindlic,forhtig,forhtmod,herebleap,
ungedyrstig,unpriste,blethe<bleap,fey<fæge,
unbold<unbeald,unbold<unbeald,argh<earg,
frightful,feared,ferdy,fearful,ferdful,
g(h)astful,trembling,timorous,cremetous,
cremeuse,craintive,shoopish,meticulous,
timid,tremebund,awful,soft,pale,
timorsome, tremulous, pigeon-hearted,
affrightful, formidolous, pavid, timidous,
unsupported, tender-nosed, scary,
pippin-hearted,kitten-hearted, funky,
tender-footed, fearsome misventurous,
scare,cotton-wool,
} **51 entries**



Workflow Example

- Input a word OR upload a file containing the words/phrases (terms) to be searched
- Search the terms in thesaurus (for example: HTE)
- Search the results from thesaurus-search in Scottish Corpus
- Find concordances for for each of the words
- Display the thesaurus-search results

Timid

{acolmod,egeful,(ge)forht,forhtful
forhtindlic,forhtig,forhtmod,herebleap,
ungedyrstig,unpriste,blethe<bleap,fey<fæge,
unbold<unbeald,unbold<unbeald,argh<earg,
frightful,feared,ferdy,fearful,ferdful,
g(h)astful,trembling,timorous,cremetous,
cremeuse,craintive,shpeepish,meticulous,
timid,tremebund,awful,soft,pale,
timorsome, tremulous, pigeon-hearted,
affrightful, formidolous, pavid, timidous,
unsupported, tender-nosed, scary,
pippin-hearted,kitten-hearted, funky,
tender-footed, fearsome misventurous,
scare,cotton-wool,
} **51 entries**



Workflow Example

- Input a word OR upload a file containing the words/phrases (terms) to be searched
- Search the terms in thesaurus (for example: HTE)
- Search the results from thesaurus-search in Scottish Corpus
- Find concordances for for each of the words
- Display the thesaurus-search results
- Display the corpus-search results

Timid

{acolmod,egeful,(ge)forht,forhtful
forhtindlic,forhtig,forhtmod,herebleap,
ungedyrstig,unpriste,blethe<bleap,fey<fæge,
unbold<unbeald,unbold<unbeald,argh<earg,
frightful,feared,ferdy,fearful,ferdful,
g(h)astful,trembling,timorous,cremetous,
cremeuse,craintive, sheepish,meticulous,
timid,tremebund,awful,soft,pale,
timorsome, tremulous, pigeon-hearted,
affrightful, formidolous, pavid, timidous,
unsupported, tender-nosed, scary,
pippin-hearted,kitten-hearted, funky,
tender-footed, fearsome misventurous,
scare,cotton-wool,
} **51 entries**



Workflow Example

- Input a word OR upload a file containing the words/phrases (terms) to be searched
- Search the terms in thesaurus (for example: HTE)
- Search the results from thesaurus-search in Scottish Corpus
- Find concordances for for each of the words
- Display the thesaurus-search results
- Display the corpus-search results
- Display the concordances

Timid

{acolmod,egeful,(ge)forht,forhtful
forhtindlic,forhtig,forhtmod,herebleap,
ungedyrstig,unpriste,blethe<bleap,fey<fæge,
unbold<unbeald,unbold<unbeald,argh<earg,
frightful,feared,ferdy,fearful,ferdful,
g(h)astful,trembling,timorous,cremetous,
cremeuse,craintive,shleepish,meticulous,
timid,tremebund,awful,soft,pale,
timorsome, tremulous, pigeon-hearted,
affrightful, formidolous, pavid, timidous,
unsupported, tender-nosed, scary,
pippin-hearted,kitten-hearted, funky,
tender-footed, fearsome misventurous,
scare,cotton-wool,
} **51 entries**



Workflow Example

- Input a word OR upload a file containing the words/phrases (terms) to be searched
- Search the terms in thesaurus (for example: HTE)
- Search the results from thesaurus-search in Scottish Corpus
- Find concordances for for each of the words
- Display the thesaurus-search results
- Display the corpus-search results
- Display the concordances
- Save/Download the results

Timid

{acolmod, egeful, (ge)forht, forhtful
 forhtindlic, forhtig, forhtmod, herebleap,
 ungedyrstig, unpriste, blethe<bleap, fey<fæge,
 unbold<unbeald, unbold<unbeald, argh<earg,
 frightful, feared, ferdy, fearful, ferdful,
 g(h)astful, trembling, timorous, cremetous,
 cremeuse, craintive, sheepish, meticulous,
 timid, tremebund, awful, soft, pale,
 timorsome, tremulous, pigeon-hearted,
 affrightful, formidolous, pavid, timidous,
 unsupported, tender-nosed, scary,
 pippin-hearted, kitten-hearted, funky,
 tender-footed, fearsome misventurous,
 scare, cotton-wool,
 } **51 entries**



talk chapped knocked blate bashful timid drax stretch galluses braces yont

defend inverewe handkerchief trees wavetimid surrender while rhododendrons hurl defiance 1 lawrence wrote of histimidtypist nelly morrison dirty bitch

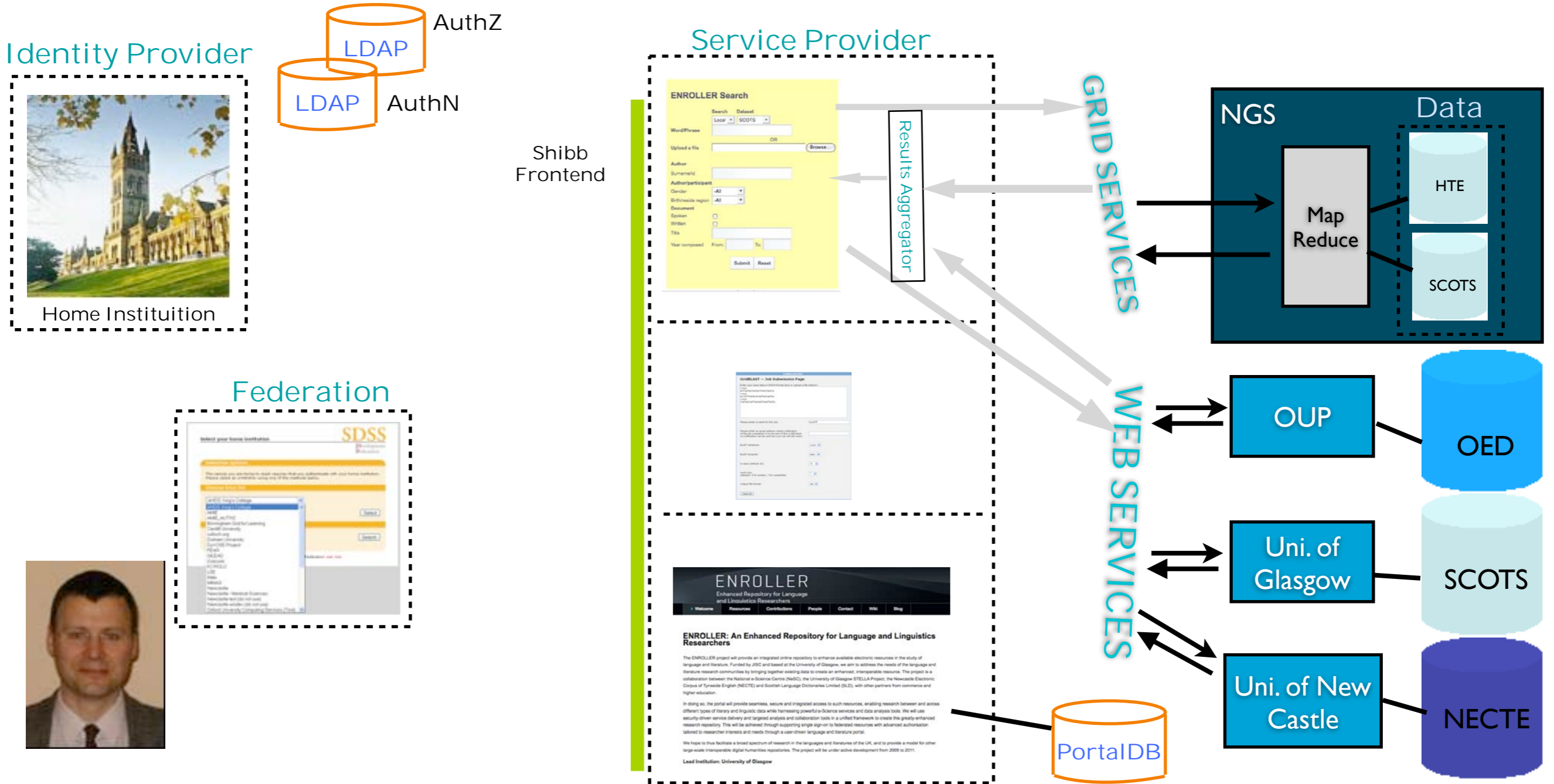
through you i see the timid indeterminate puzzled soul behind that

together in one corner a timid scrum we tried to celebrity

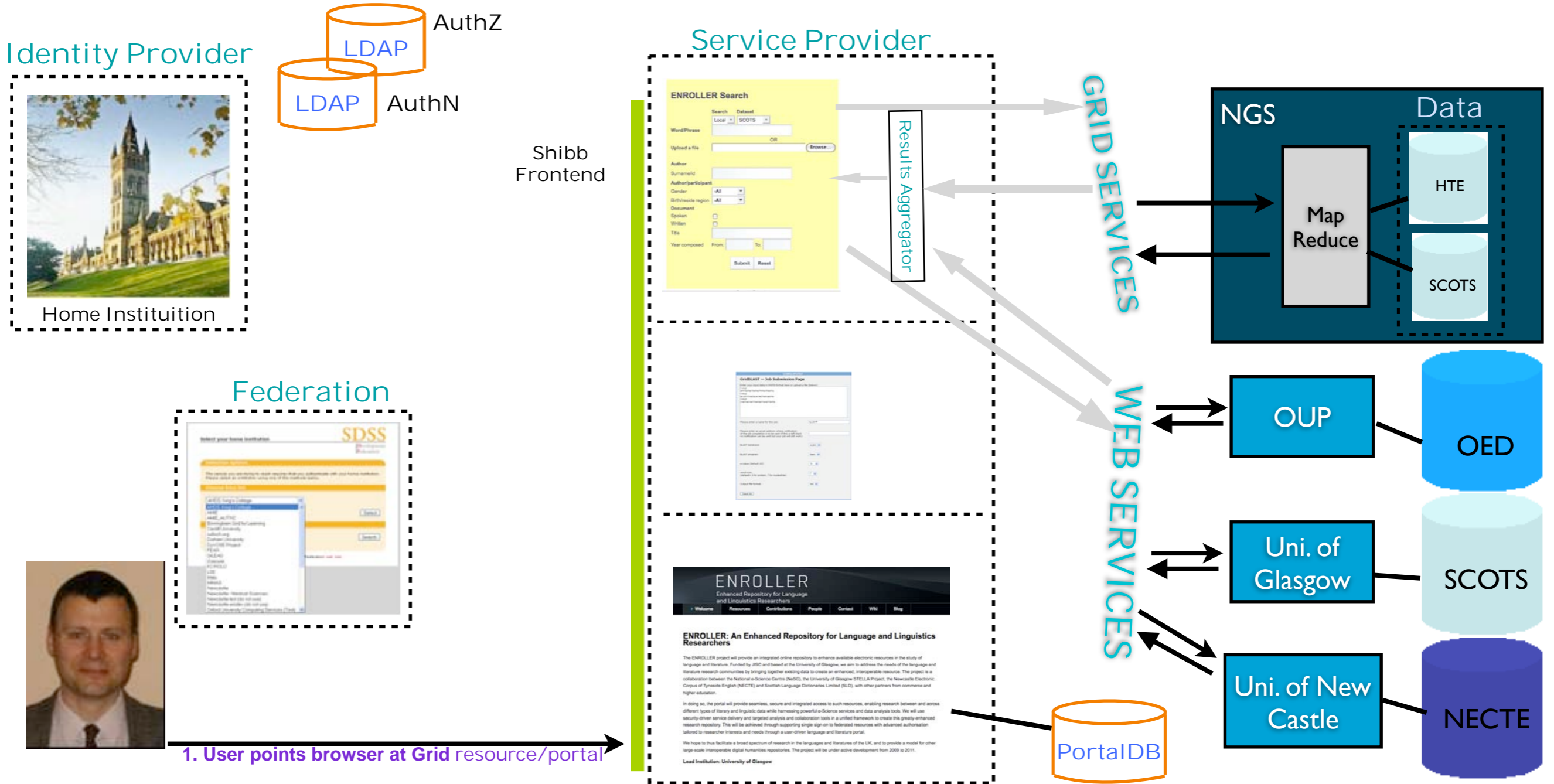
bit rupert wis aye a timid cat it jist hatit e

wis a tom it wis timid i min fine on it

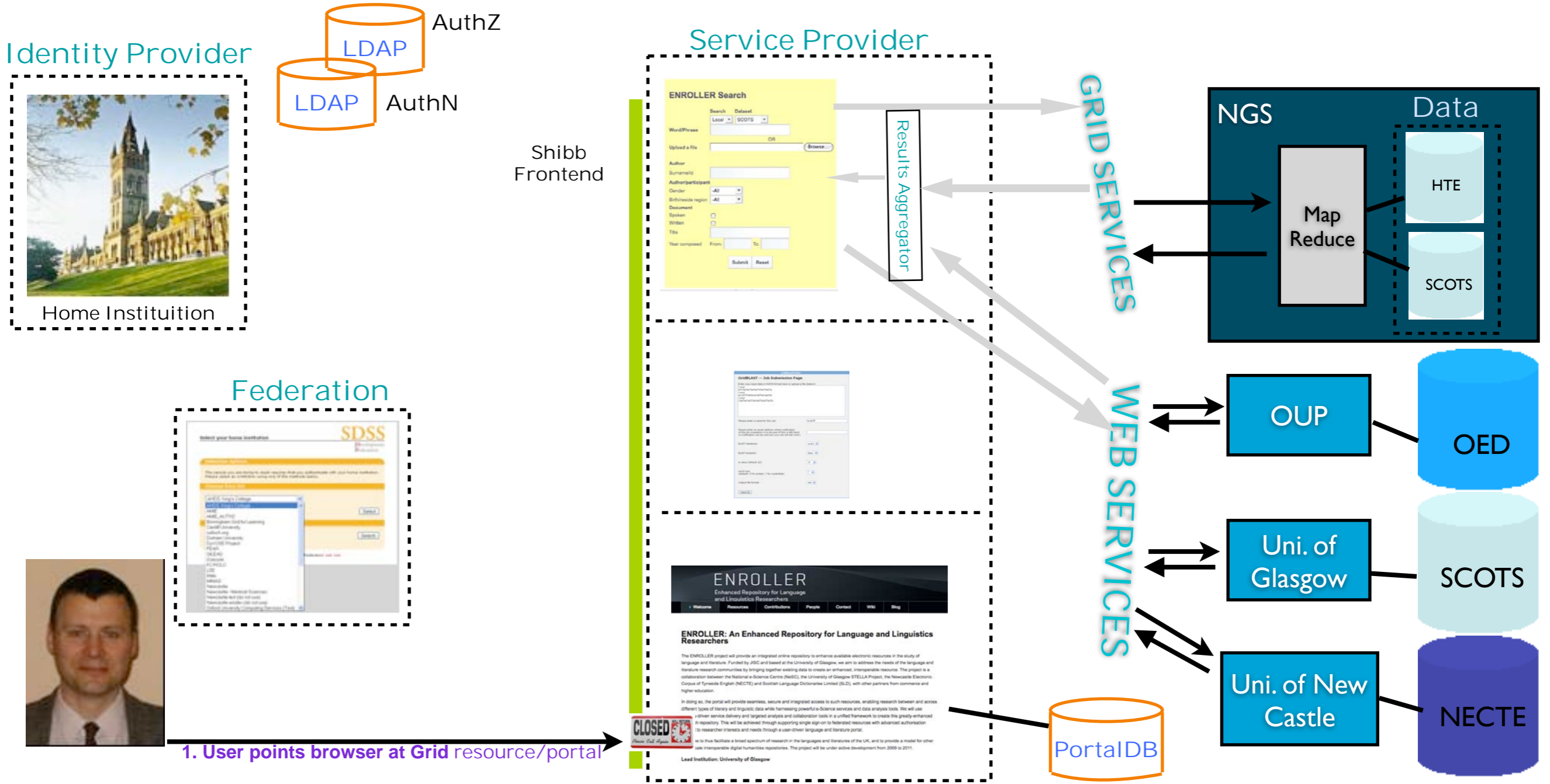
Typical Interaction Flow



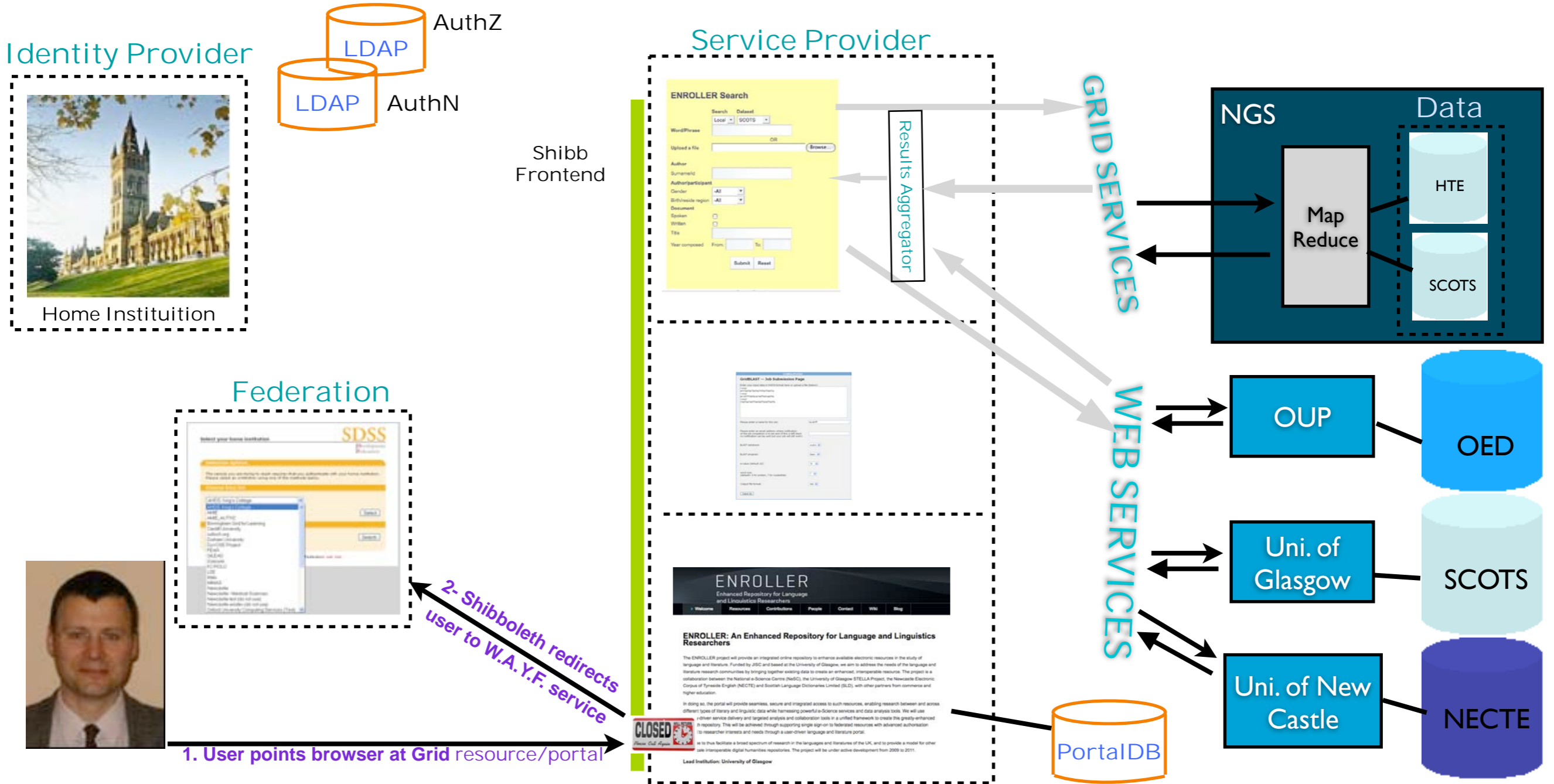
Typical Interaction Flow



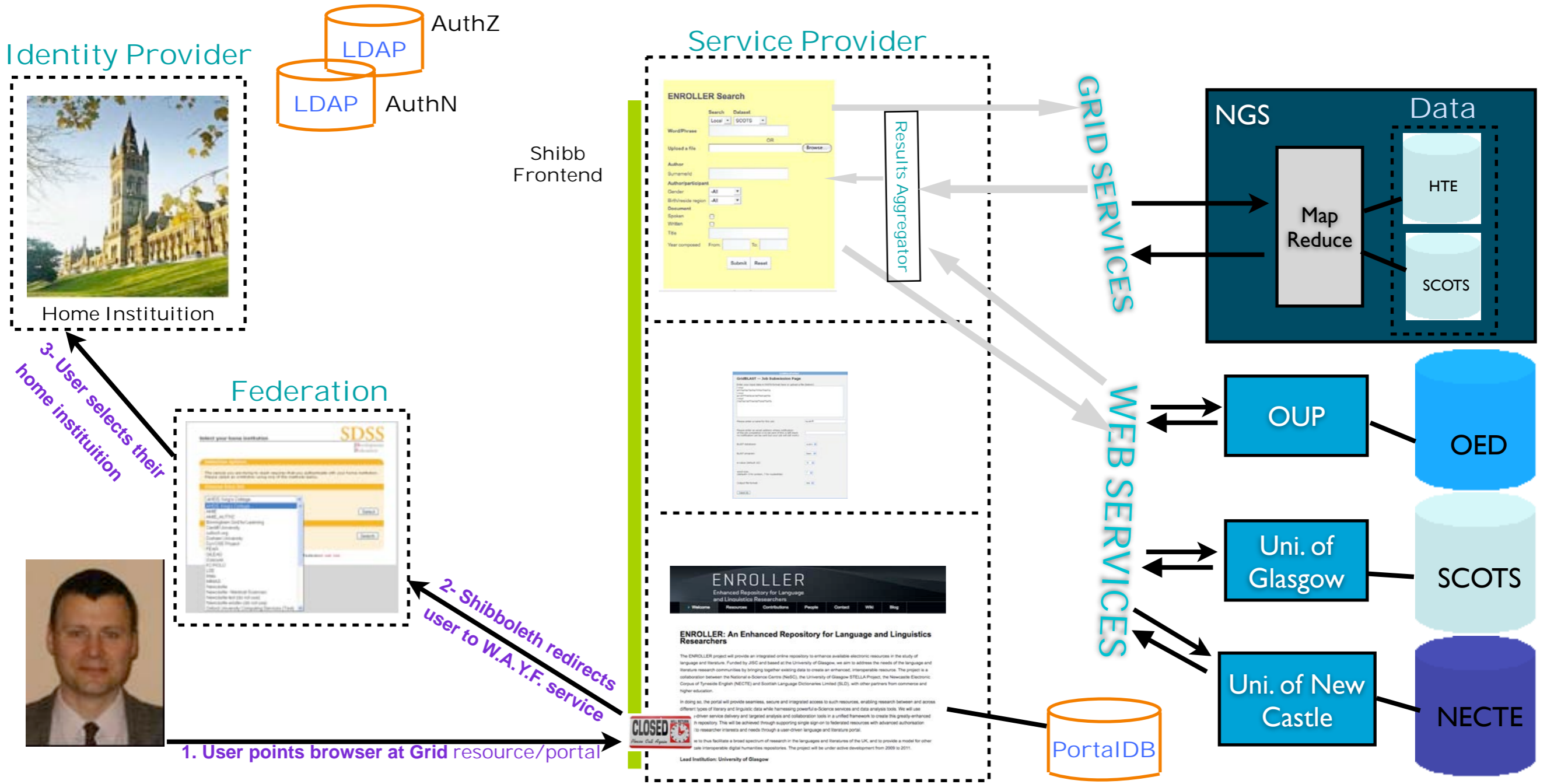
Typical Interaction Flow



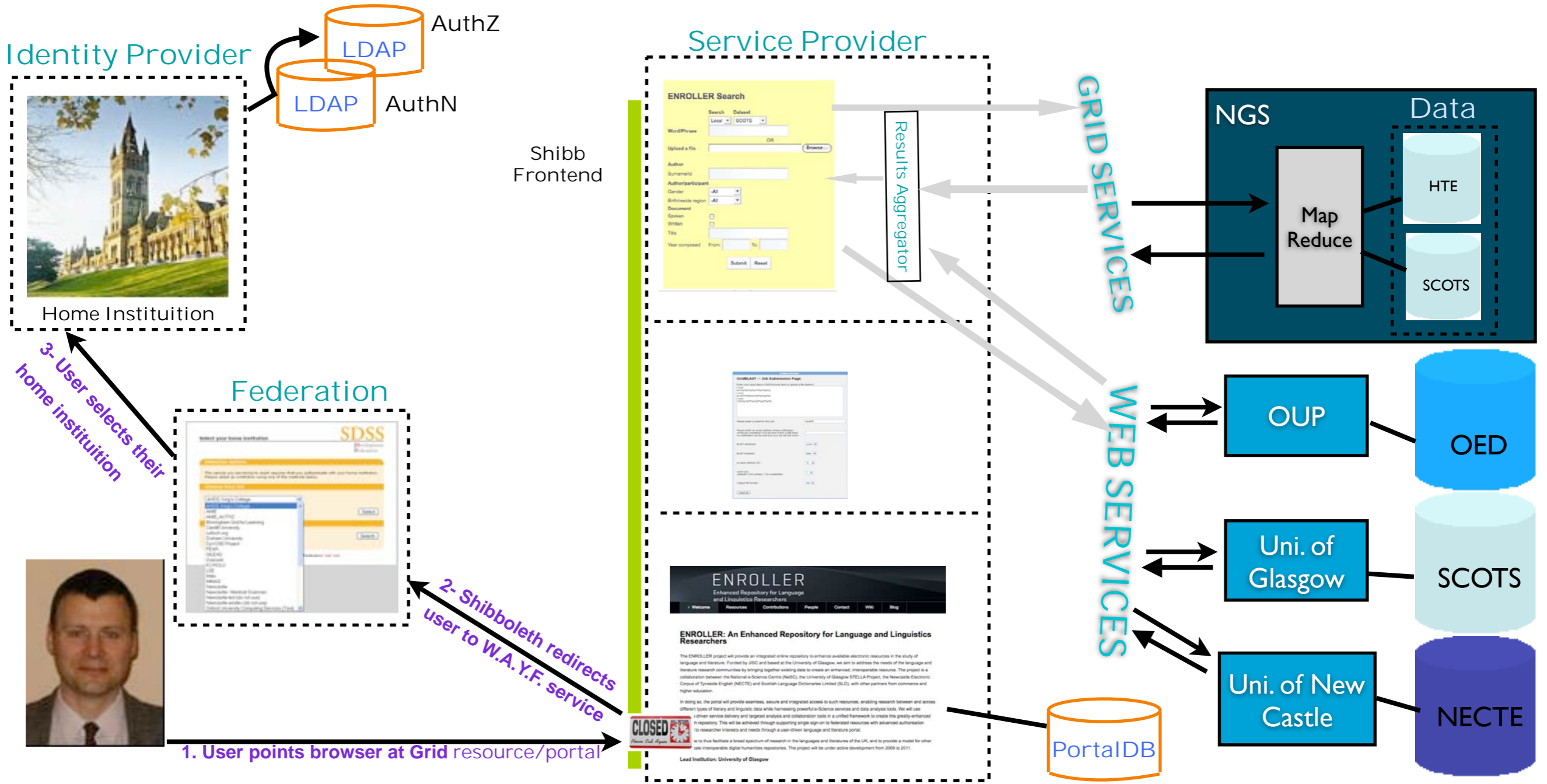
Typical Interaction Flow



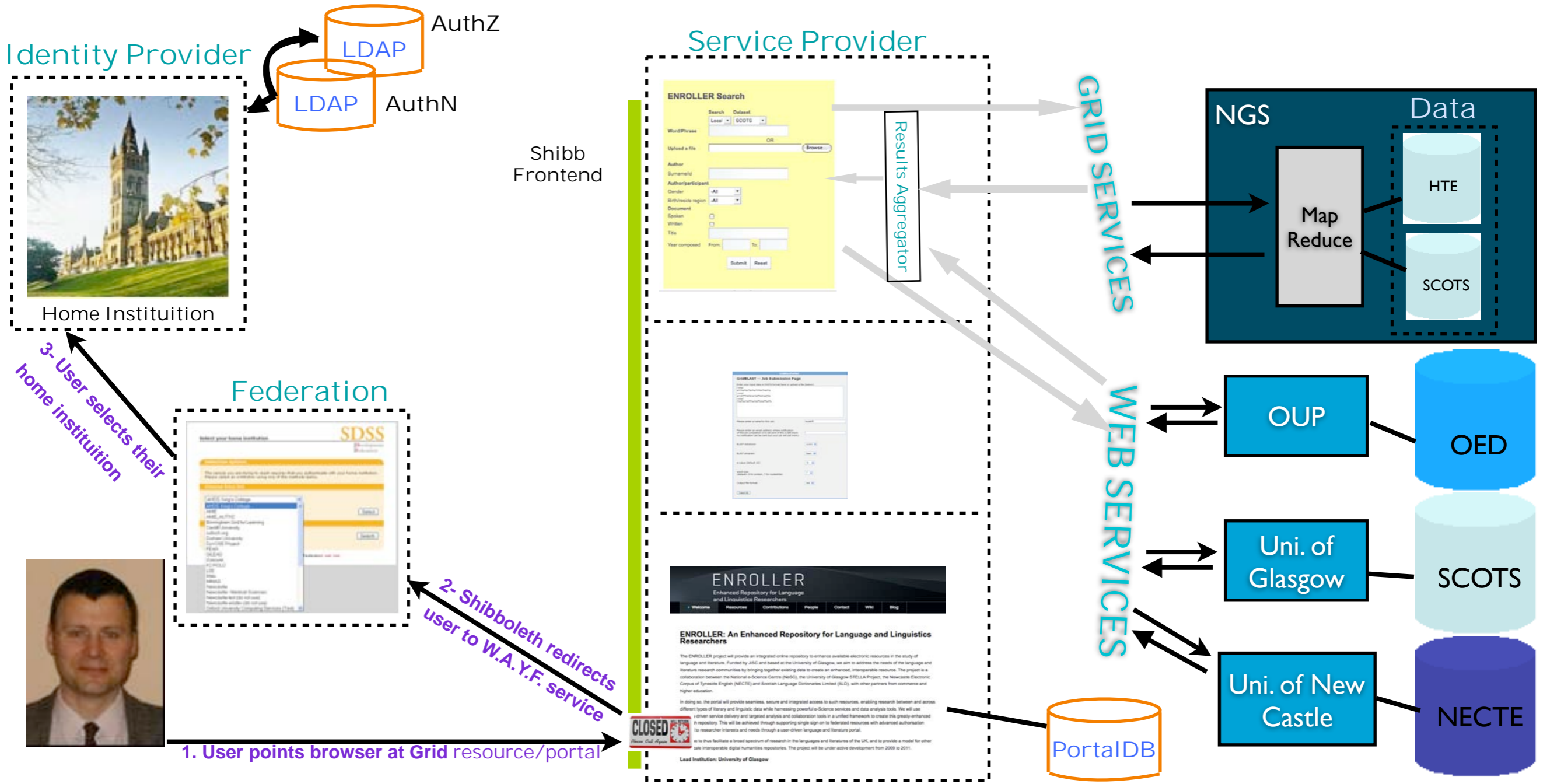
Typical Interaction Flow



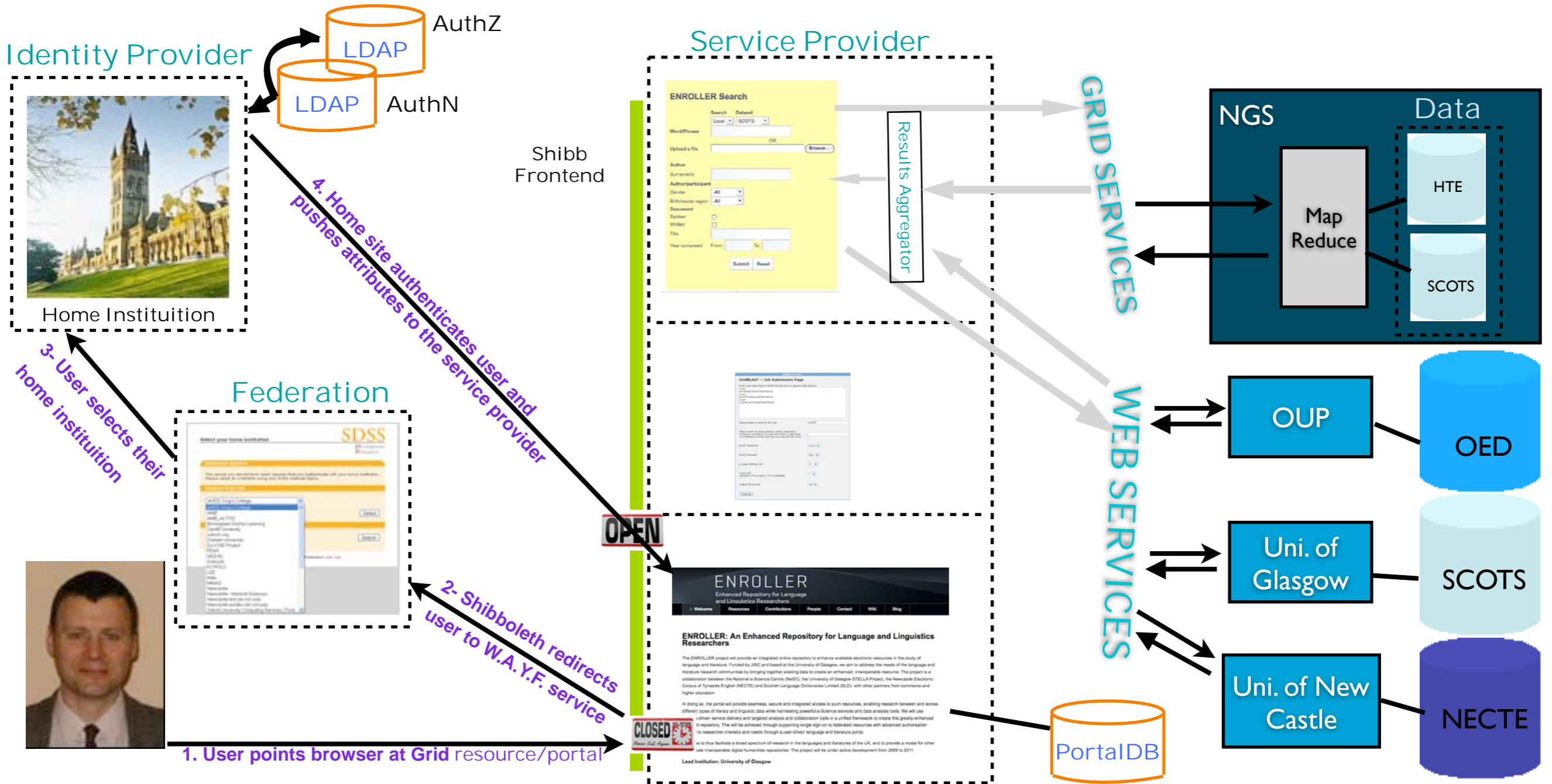
Typical Interaction Flow



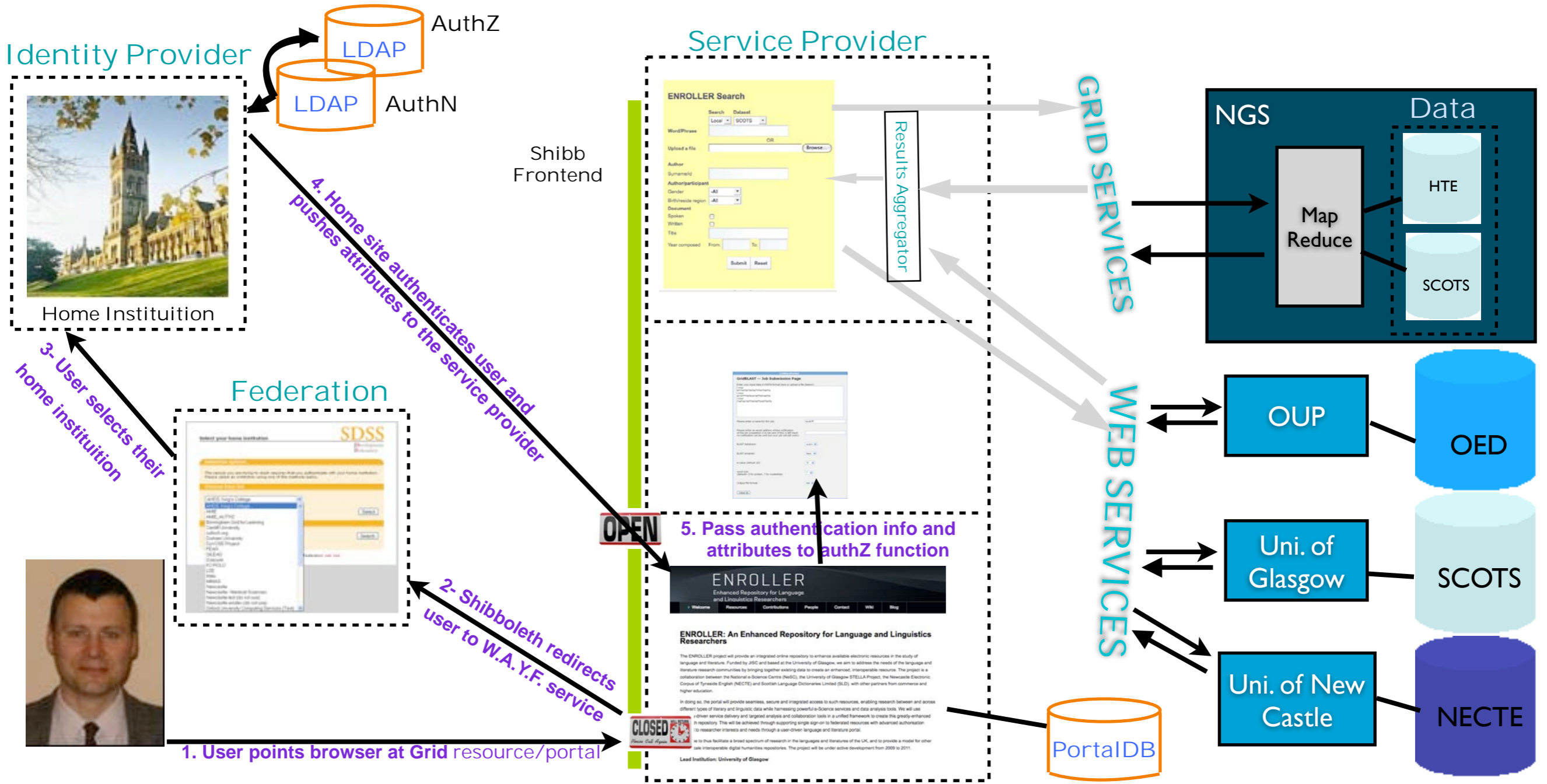
Typical Interaction Flow



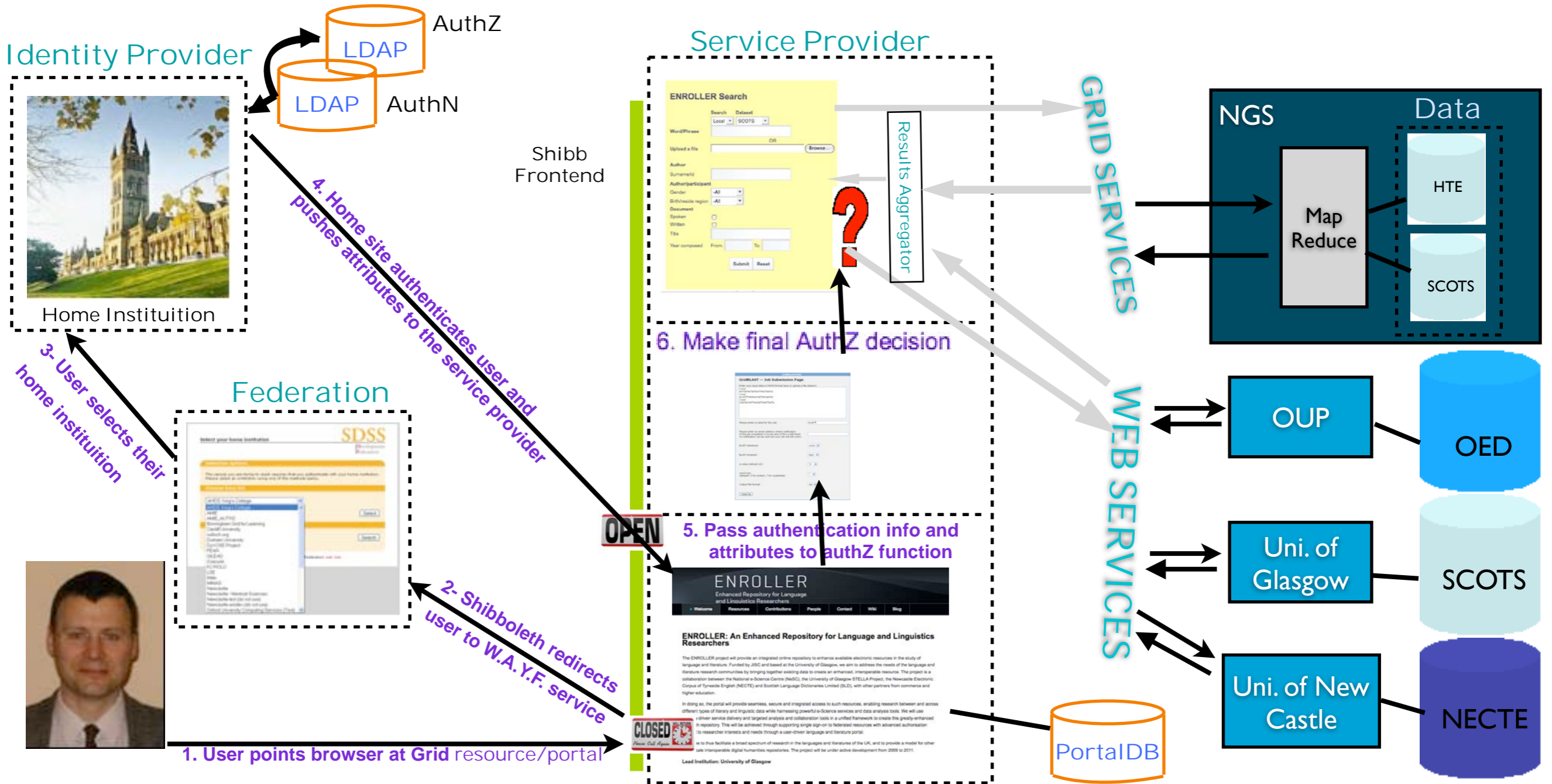
Typical Interaction Flow



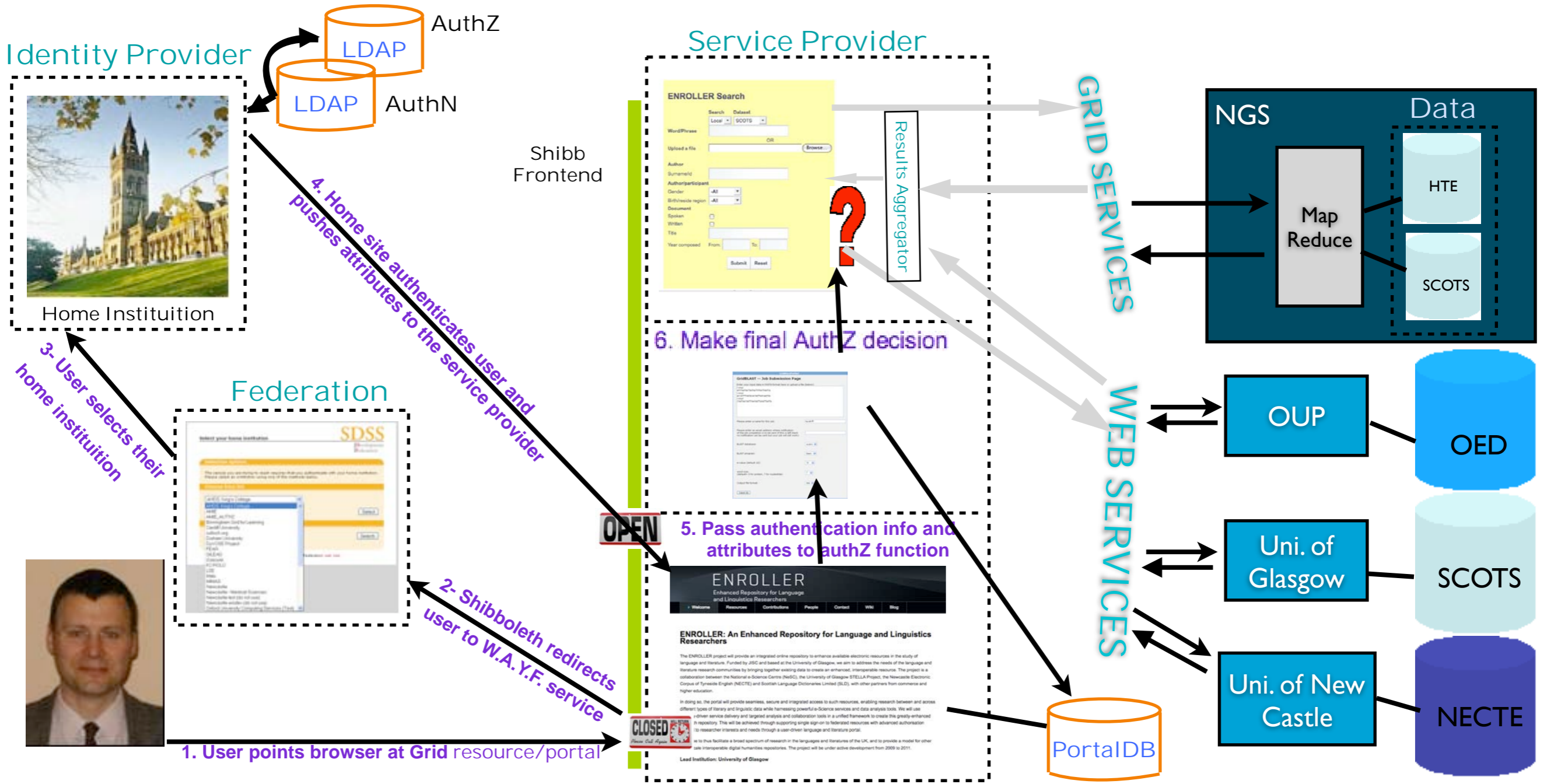
Typical Interaction Flow



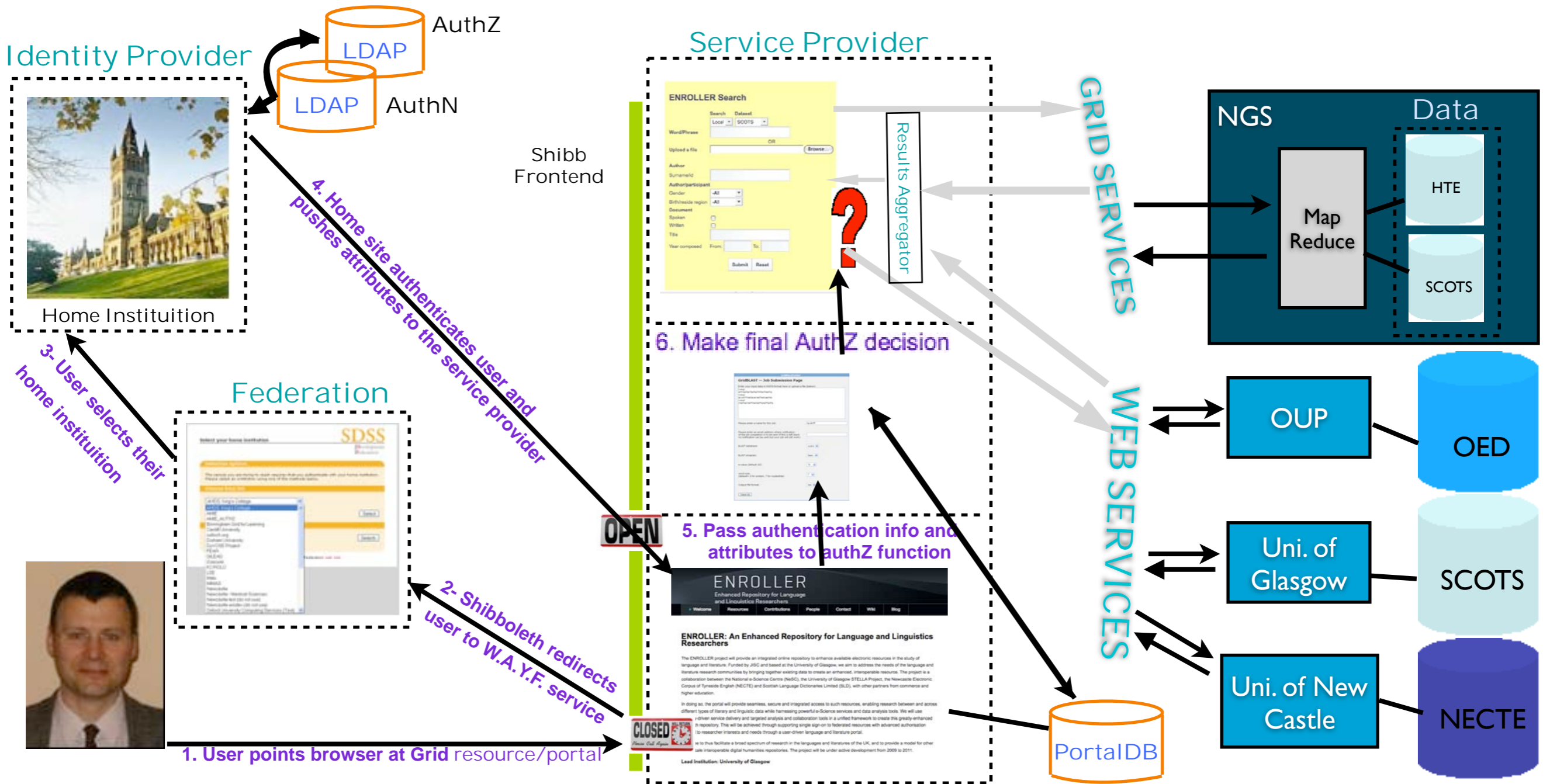
Typical Interaction Flow



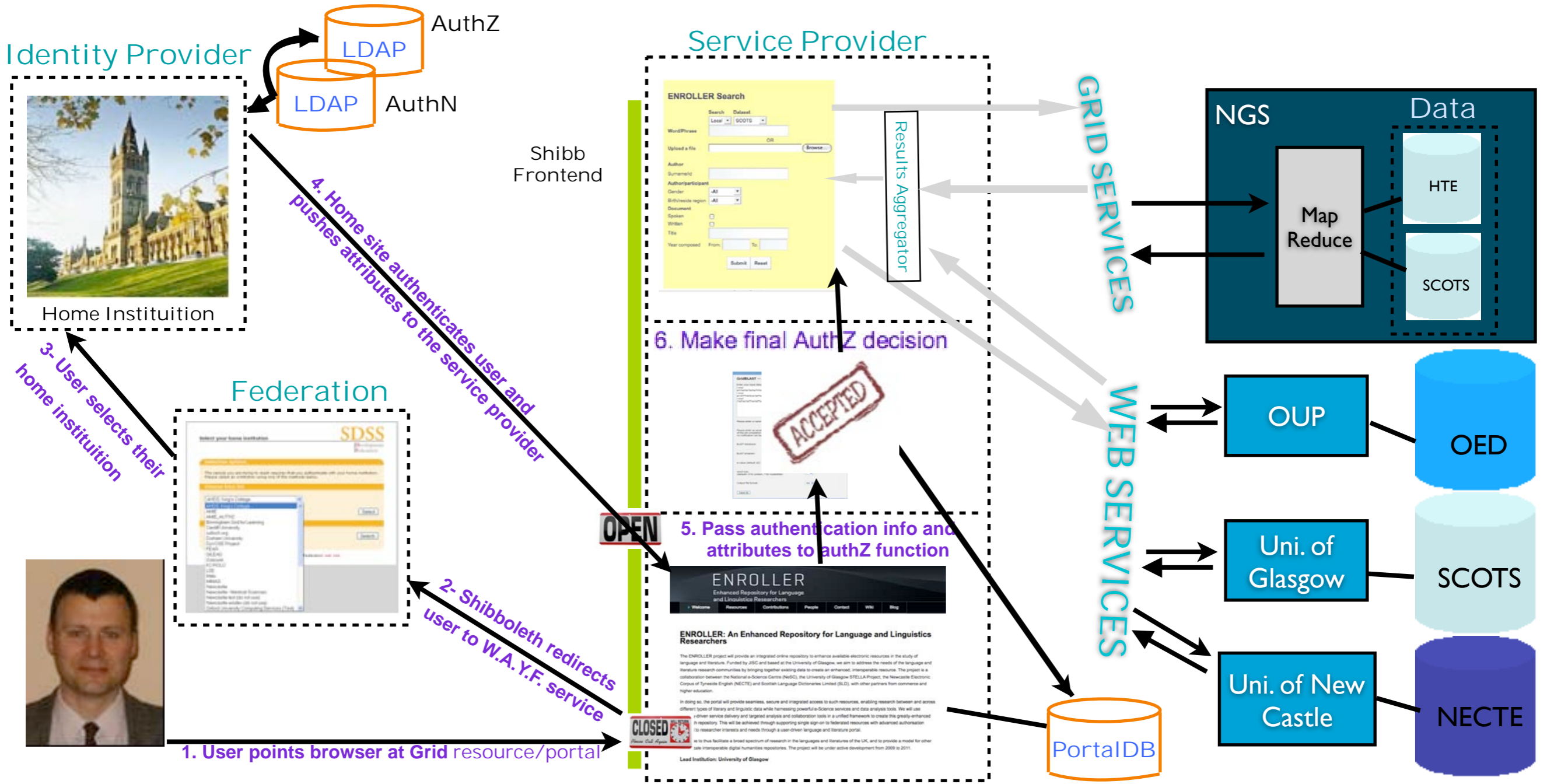
Typical Interaction Flow



Typical Interaction Flow



Typical Interaction Flow



Using NGS

Using NGS



GRID SERVICES



WEB SERVICES



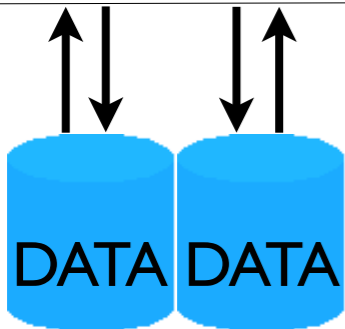
Using NGS



GRID SERVICES



WEB SERVICES



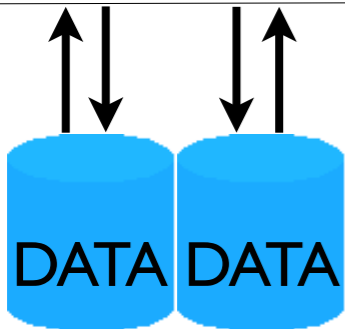
Using NGS



GRID SERVICES



WEB SERVICES



Using NGS

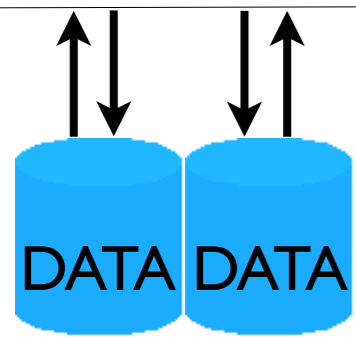


GRID SERVICES

Job Submission Client



WEB SERVICES



```
#!/bin/bash
echo "Starting application: scots-app "
echo "Submitting to job-manager at:"
echo $(bin/hostname -f)
echo "with arguments to main "
echo $*

cd /home/ngs0273/java/prog/scots-app

props=/home/ngs0273/java/prog/scots-app/src/main/resources/project_ngs.properties
echo "properties file:" $props

nthreads=4
echo "# threads=" $nthreads

/usr/local/Cluster-Apps/java-1.6.0_03/bin/java -cp target/scots-app-1.0-SNAPSHOT.jar scots.app.App "$@" $props $nthreads
```

Job Submission Script

Using NGS



GRID SERVICES

Job
Sub-
mission
Client



WEB SERVICES



```
#!/bin/bash
echo "Starting application: scots-app "
echo "submitting to job-manager at:"
echo $(bin/hostname -f)
echo " with arguments to main "
echo $*

cd /home/ngs0273/java/prog/scots-app

props=/home/ngs0273/java/prog/scots-app/src/main/resources/project_ngs.properties
echo "properties file:" $props

nthreads=4
echo "# threads=" $nthreads

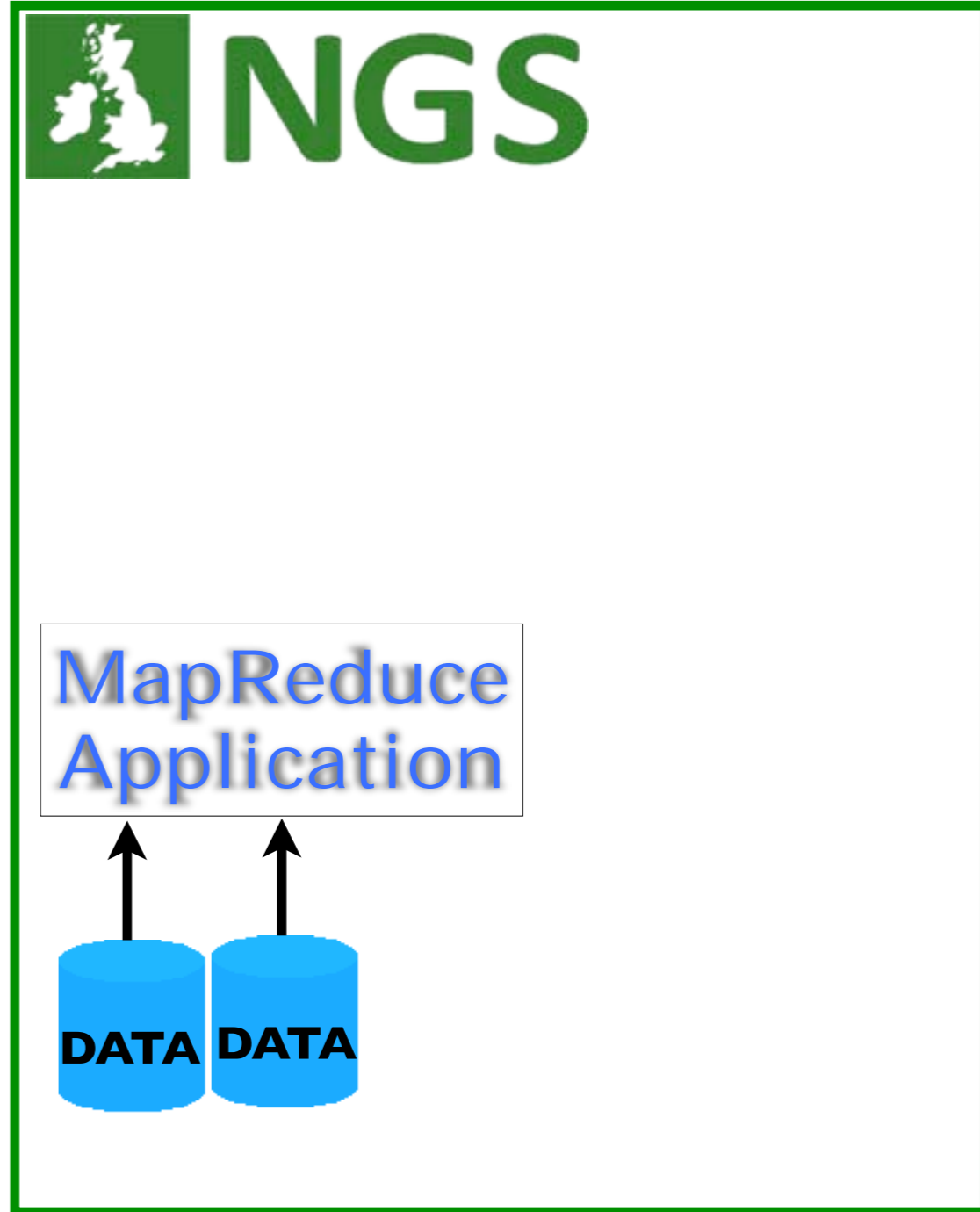
/usr/local/Cluster-Apps/java-1.6.0_03/bin/java -cp target/scots-app-1.0-SNAPSHOT.jar scots.app.App "$@" $props $nthreads
```

Using NGS

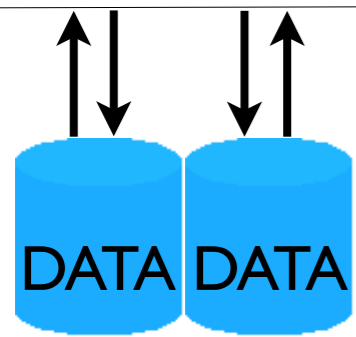


GRID SERVICES

Job Submission Client



WEB SERVICES



```
#!/bin/bash
echo "Starting application: scots-app "
echo "Submitting to job-manager at:"
echo $(bin/hostname -f)
echo "with arguments to main "
echo $*

cd /home/ngs0273/avaproj/scots-app

props=/home/ngs0273/avaproj/scots-app/src/main/resources/project_ngs.properties
echo "properties file:" $props

nthreads=4
echo "# threads=" $nthreads

/usr/local/Cluster-Apps/java-1.6.0_03/bin/java -cp target/scots-app-1.0-SNAPSHOT.jar scots.app.App "$@" $nthreads
```

Using NGS



ENROLLER
Enhanced Repository for Language and Linguistics Researchers

ENROLLER: An Enhanced Repository for Language and Linguistics Researchers

The ENROLLER project will provide an integrated online repository to enhance available electronic resources in the study of language and literature. Funded by JISC and based at the University of Glasgow, we aim to address the needs of the language and literature research communities by bringing together existing data to create an enhanced, interoperable resource. The project is a collaboration between the National e-Science Centre (NeSC), the University of Glasgow STELLA Project, the Newcastle Electronic Corpus of Tyndale English (NECTE) and Scottish Language Dictionaries Limited (SLD), with other partners from commerce and higher education.


In doing so, the portal will provide seamless, secure and integrated access to such resources, enabling research between and across different types of literary and linguistic data while harnessing powerful e-Science services and data analysis tools. We will use security-driven service delivery and targeted analysis and collaboration tools in a unified framework to create the greatly enhanced research repository. This will be achieved through supporting single sign-on to federated resources with advanced authorisation tailored to researcher interests and needs through a user-driven language and literature portal.

We hope to thus facilitate a broad spectrum of research in the languages and literatures of the UK, and to provide a model for other large-scale interoperable digital humanities repositories. The project will be under active development from 2009 to 2011.

Lead Institution: University of Glasgow

GRID SERVICES

Job Submission Client

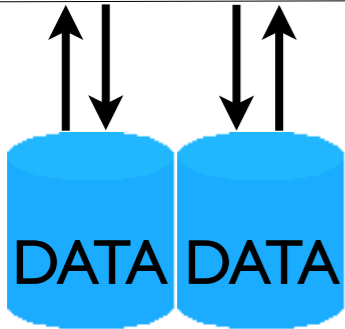
 **NGS**

MapReduce Application

DATA DATA

CE-1
CE-2
CE-3
CE-4
⋮
CE-N

WEB SERVICES



```
#!/bin/bash
echo "Starting application: scots-app "
echo "Submitting to job-manager at:"
echo $(bin/hostname -f)
echo "with arguments to main "
echo $*

cd /home/ngs0273/java/prog/scots-app

props=/home/ngs0273/java/prog/scots-app/src/main/resources/project_ngs.properties
echo "properties file:" $props

nthreads=4
echo "# threads=" $nthreads

/usr/local/Cluster-Apps/java-1.6.0_03/bin/java -cp target/scots-app-1.0-SNAPSHOT.jar scots.app.App "$@" $nthreads
```

Using NGS



GRID SERVICES

ENROLLER
Enhanced Repository for Language and Linguistics Researchers

ENROLLER: An Enhanced Repository for Language and Linguistics Researchers

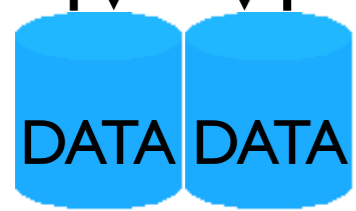
The ENROLLER project will provide an integrated online repository to enhance available electronic resources in the study of language and literature. Funded by JISC and based at the University of Glasgow, we aim to address the needs of the language and literature research communities by bringing together existing data to create an enhanced, interoperable resource. The project is a collaboration between the National e-Science Centre (NeSC), the University of Glasgow STELLA Project, the Newcastle Electronic Corpus of Tyndale English (NECTE) and Scottish Language Dictionaries Limited (SLD), with other partners from commerce and higher education.

In doing so, the portal will provide seamless, secure and integrated access to such resources, enabling research between and across different types of library and linguistic data while harnessing powerful e-Science services and data analysis tools. We will use security-driven service delivery and targeted analysis and collaboration tools in a unified framework to create the greatly enhanced research repository. This will be achieved through supporting single sign-on to federated resources with advanced authorisation tailored to researcher interests and needs through a user-driven language and literature portal.

We hope to thus facilitate a broad spectrum of research in the languages and literatures of the UK, and to provide a model for other large-scale interoperable digital humanities repositories. The project will be under active development from 2009 to 2011.

Lead Institution: University of Glasgow

WEB SERVICES



Job Submission Client

```
#!/bin/bash
echo "Starting application: scots-app "
echo "Submitting to job-manager at:"
echo $(bin/hostname -f)
echo "with arguments to main "
echo $*

cd /home/ngs0273/java/prog/scots-app

props=/home/ngs0273/java/prog/scots-app/src/main/resources/project_ngs.properties
echo "properties file:" $props

nthreads=4
echo "# threads=" $nthreads

/usr/local/Cluster-Apps/java-1.6.0_03/bin/java -cp target/scots-app-1.0-SNAPSHOT.jar scots.app.App "$@" $props $nthreads
```

NGS

Head Node

Job Manager

MapReduce Application

DATA DATA

CE-1

CE-2

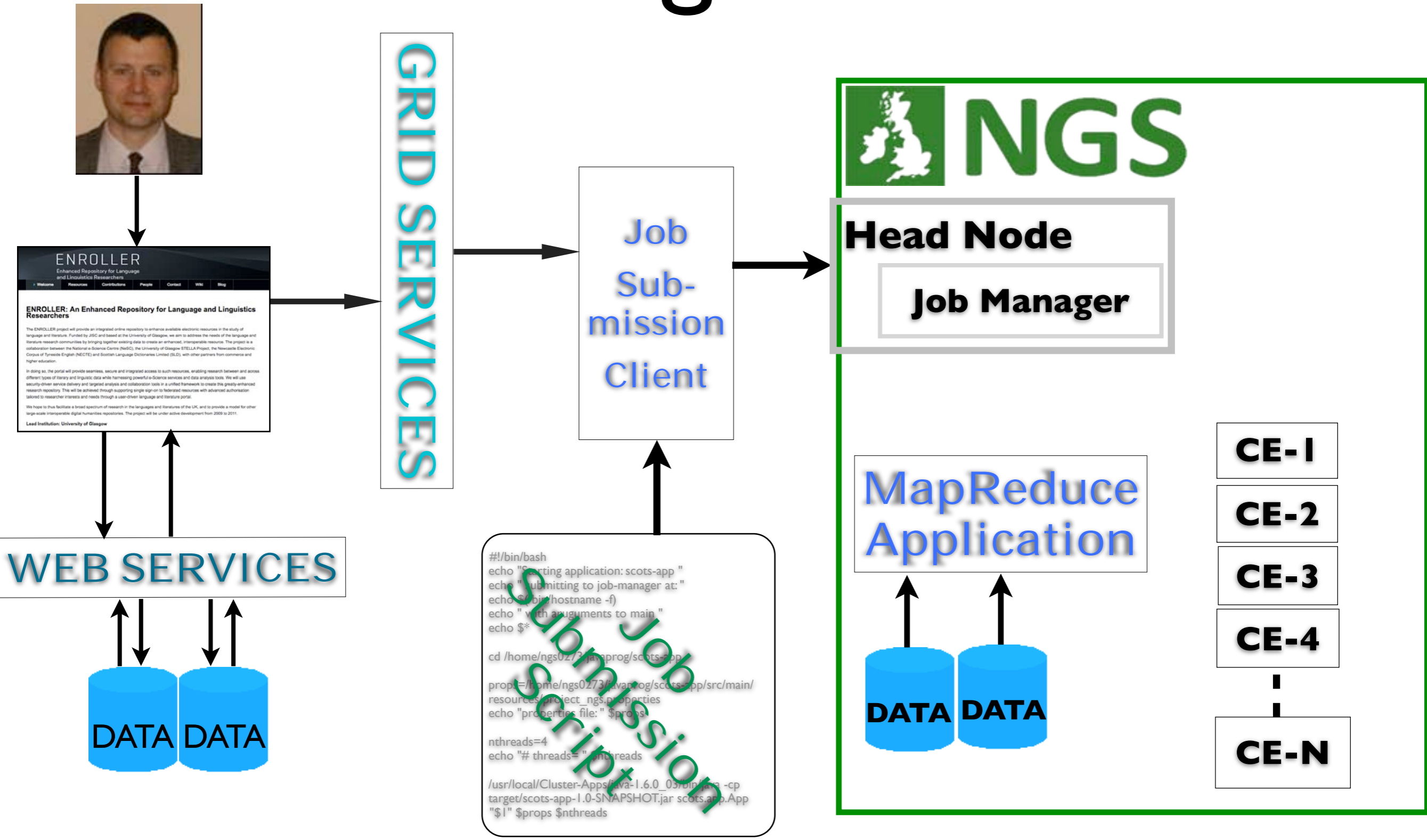
CE-3

CE-4

⋮

CE-N

Using NGS

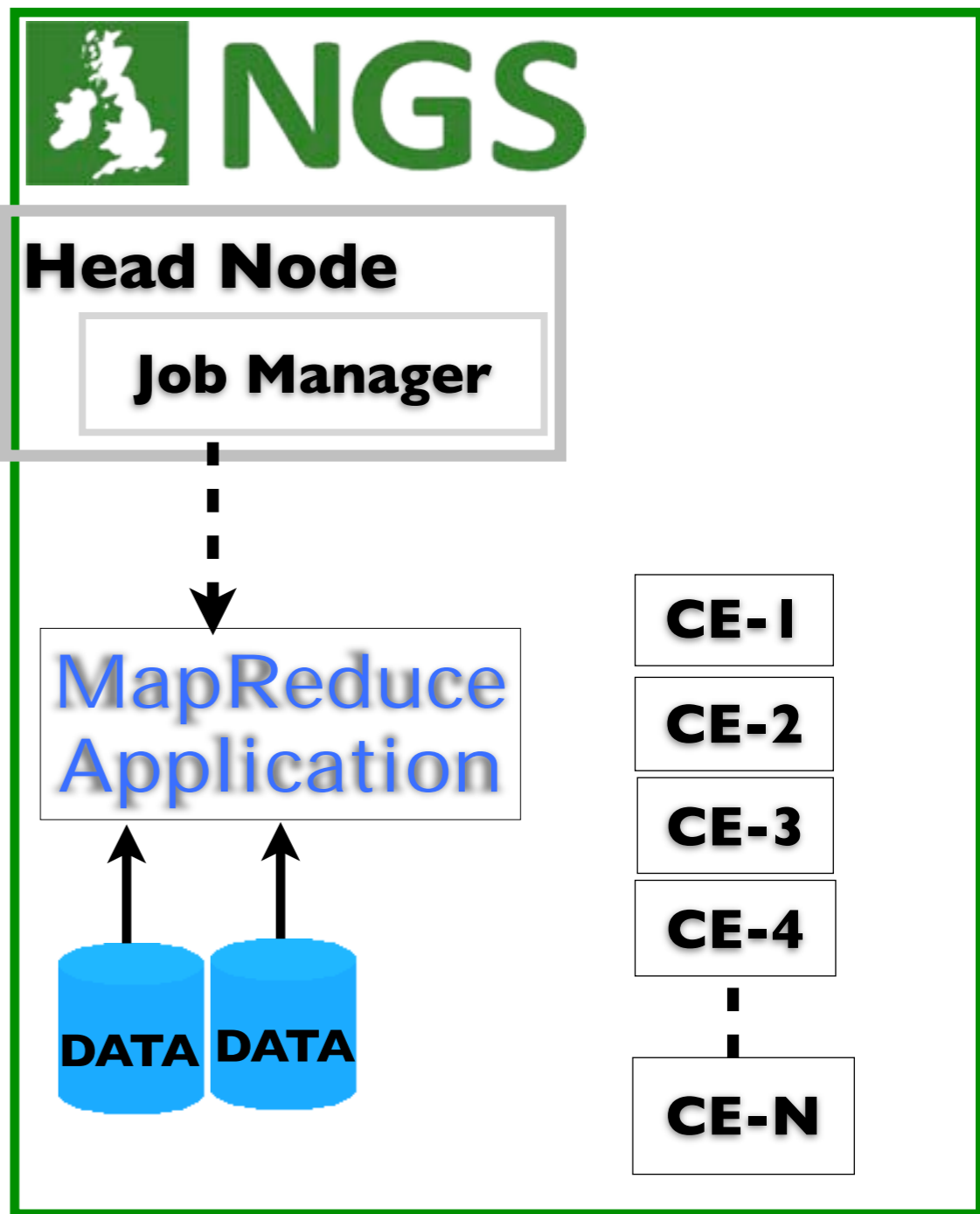


Using NGS

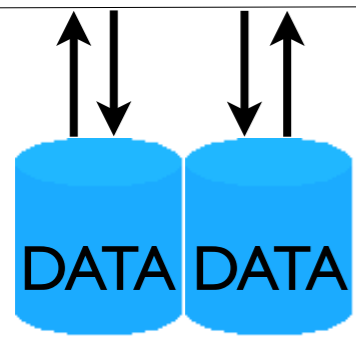


GRID SERVICES

Job Submission Client



WEB SERVICES



```
#!/bin/bash
echo "Starting application: scots-app "
echo "submitting to job-manager at:"
echo $(bin/hostname -f)
echo "with arguments to main "
echo $*

cd /home/ngs0273/java/prog/scots-app

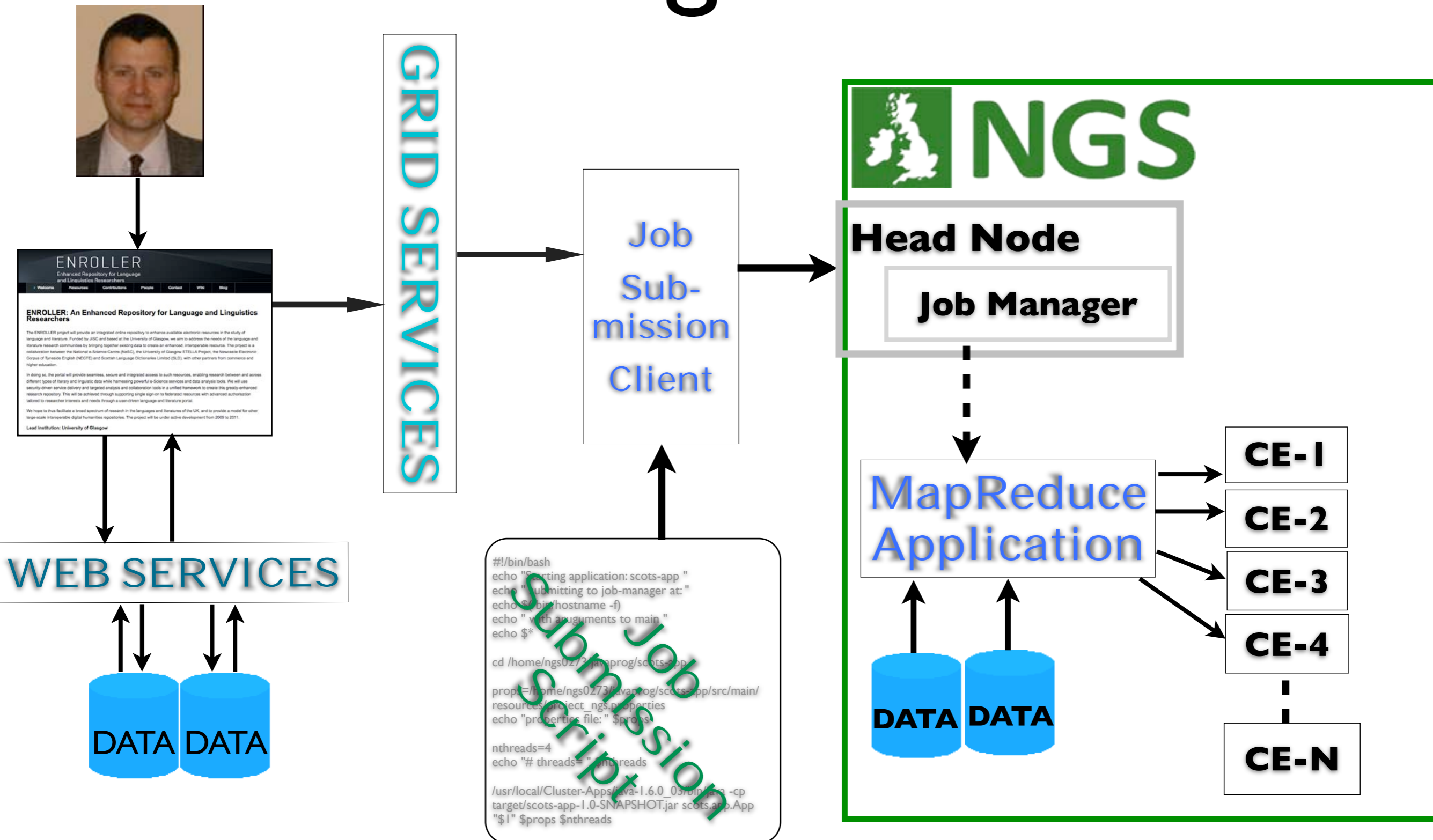
props=/home/ngs0273/java/prog/scots-app/src/main/resources/project_ngs.properties
echo "properties file:" $props

nthreads=4
echo "# threads=" $nthreads

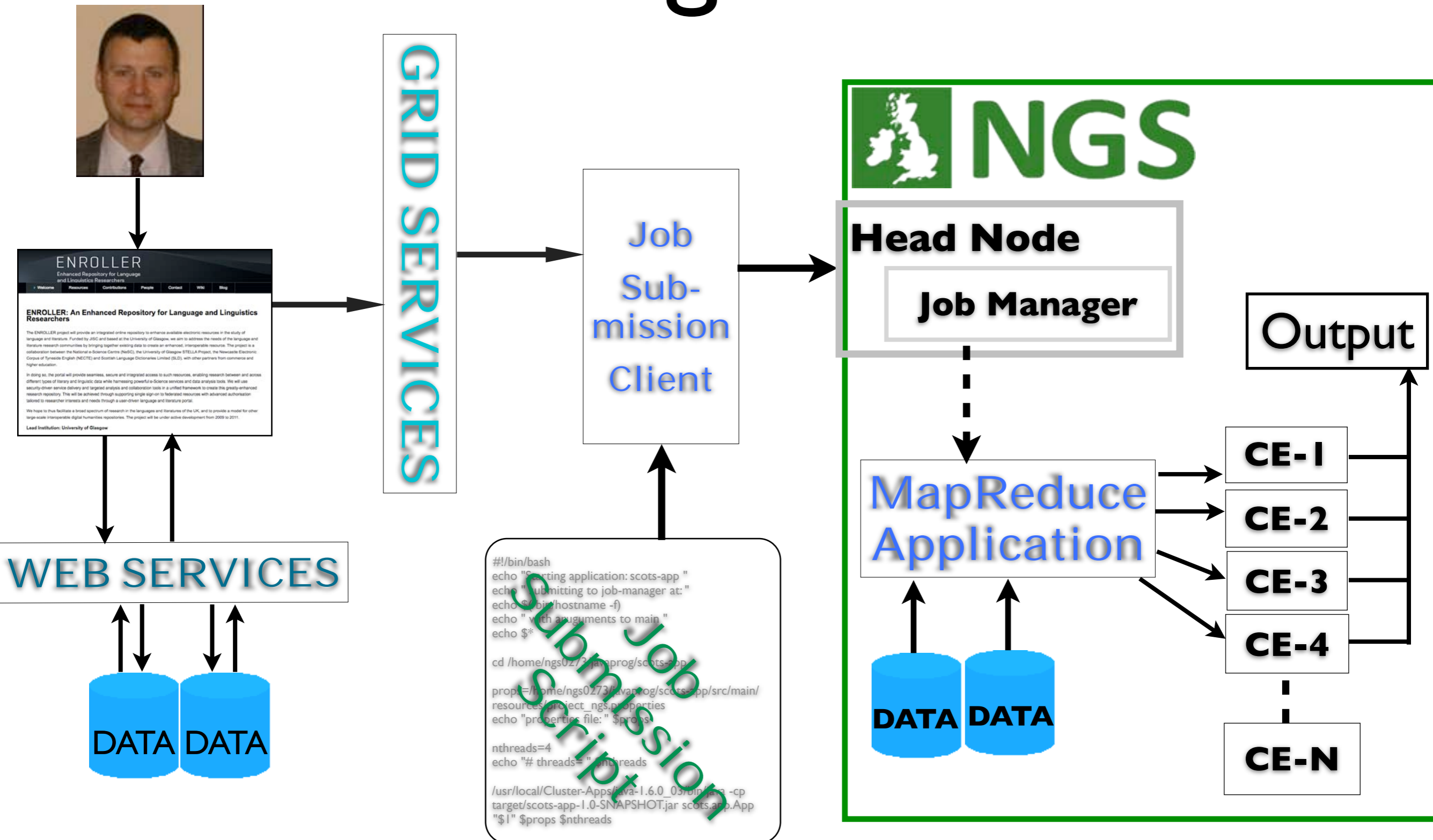
/usr/local/Cluster-Apps/java-1.6.0_03/bin/java -cp target/scots-app-1.0-SNAPSHOT.jar scots.app.App "$@" $nthreads
```

Submit Job

Using NGS



Using NGS

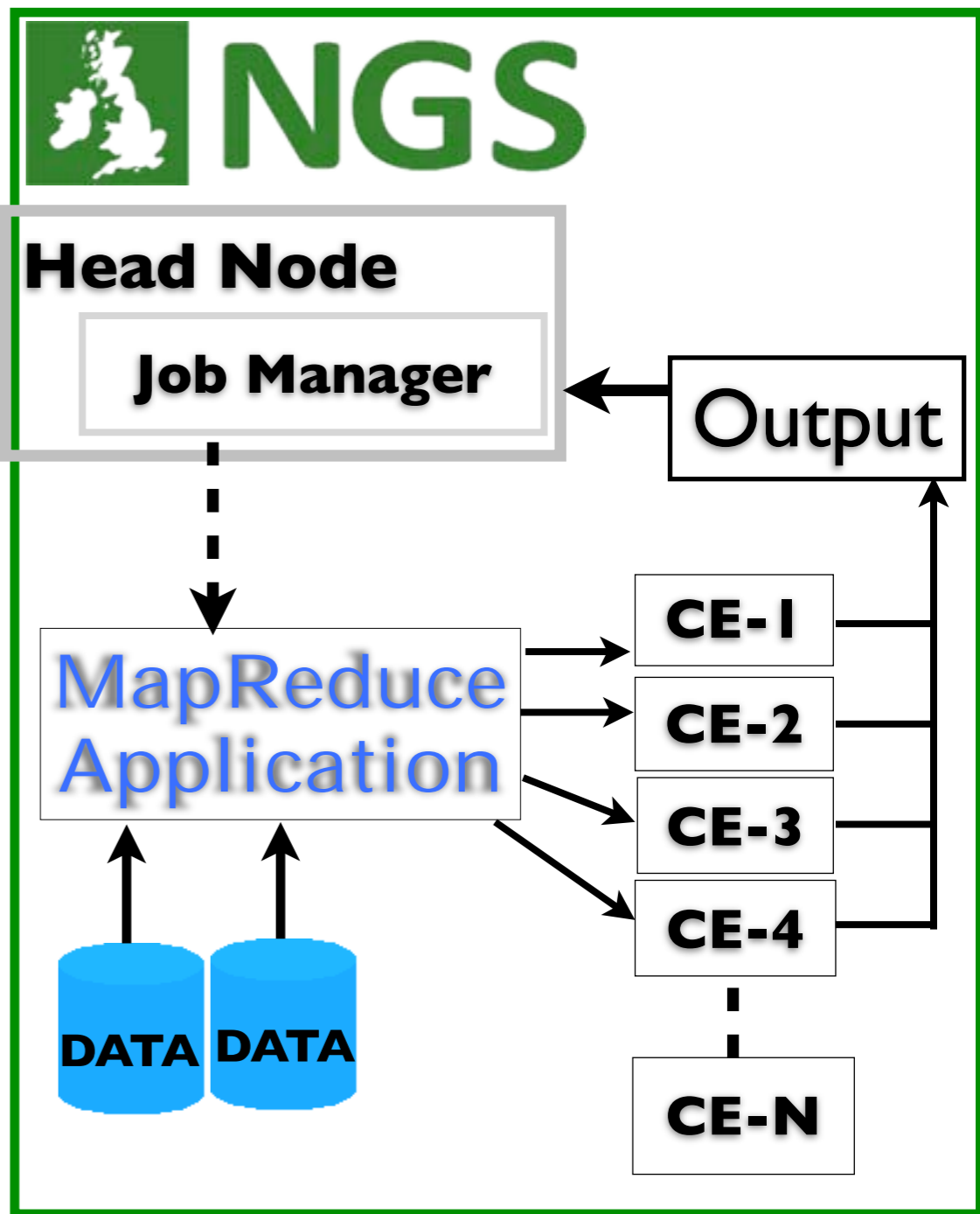


Using NGS

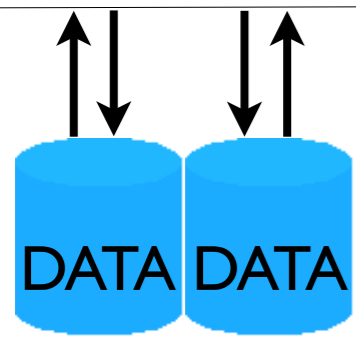


GRID SERVICES

Job Submission Client



WEB SERVICES



```
#!/bin/bash
echo "Starting application: scots-app "
echo "submitting to job-manager at:"
echo $(bin/hostname -f)
echo "with arguments to main "
echo $*

cd /home/ngs0273/java/prog/scots-app

props=/home/ngs0273/java/prog/scots-app/src/main/resources/project_ngs.properties
echo "properties file:" $props

nthreads=4
echo "# threads=" $nthreads

/usr/local/Cluster-Apps/java-1.6.0_03/bin/java -cp target/scots-app-1.0-SNAPSHOT.jar scots.app.App "$@" $props $nthreads
```

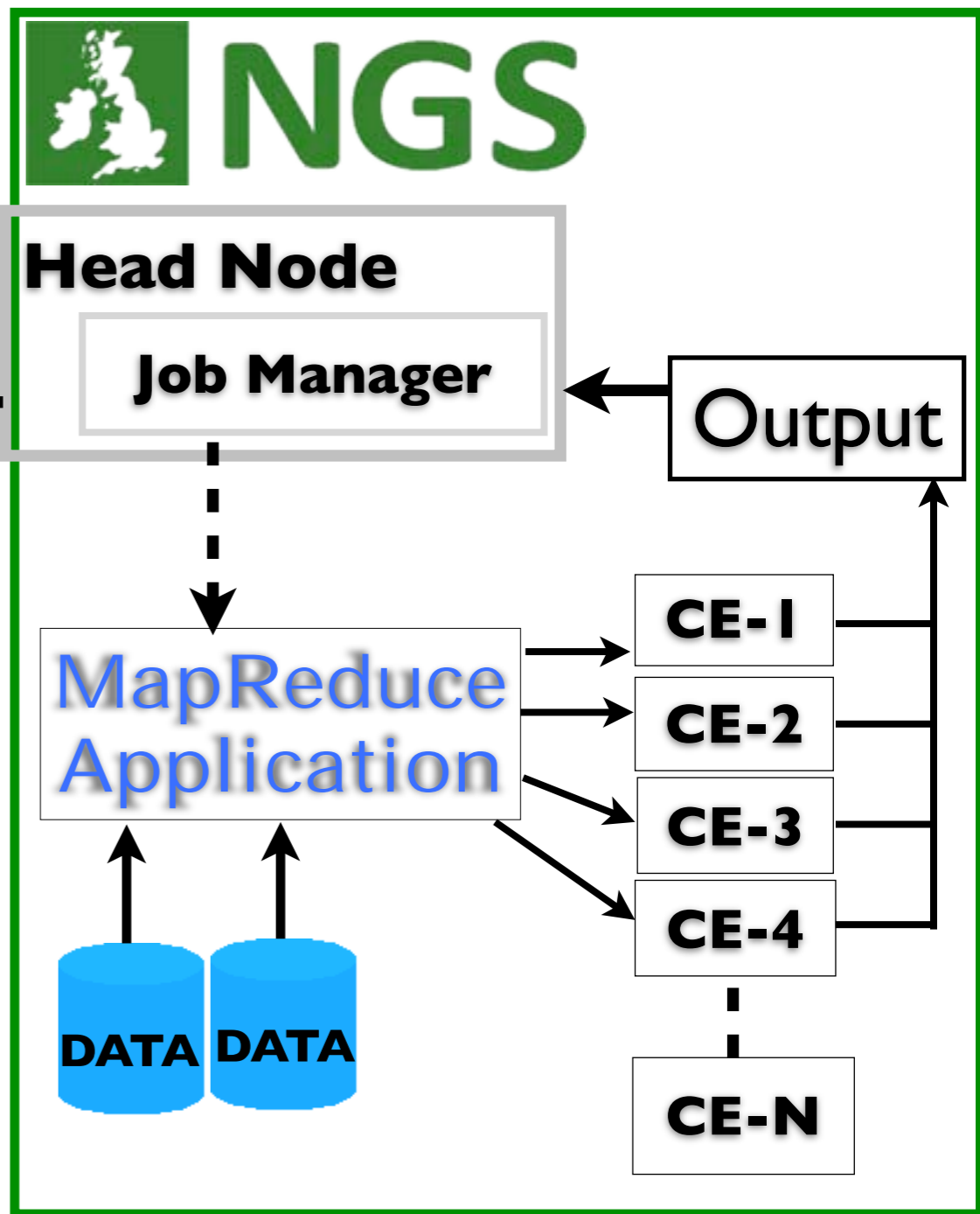
Submit Job

Using NGS

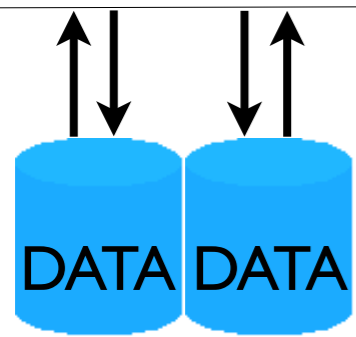


GRID SERVICES

Job Submission Client



WEB SERVICES



```
#!/bin/bash
echo "Starting application: scots-app "
echo "submitting to job-manager at:"
echo $(bin/hostname -f)
echo "with arguments to main "
echo $*

cd /home/ngs0273/java/prog/scots-app

props=/home/ngs0273/java/prog/scots-app/src/main/resources/project_ngs.properties
echo "properties file:" $props

nthreads=4
echo "# threads=" $nthreads

/usr/local/Cluster-Apps/java-1.6.0_03/bin/java -cp target/scots-app-1.0-SNAPSHOT.jar scots.app.App "$@" $props $nthreads
```

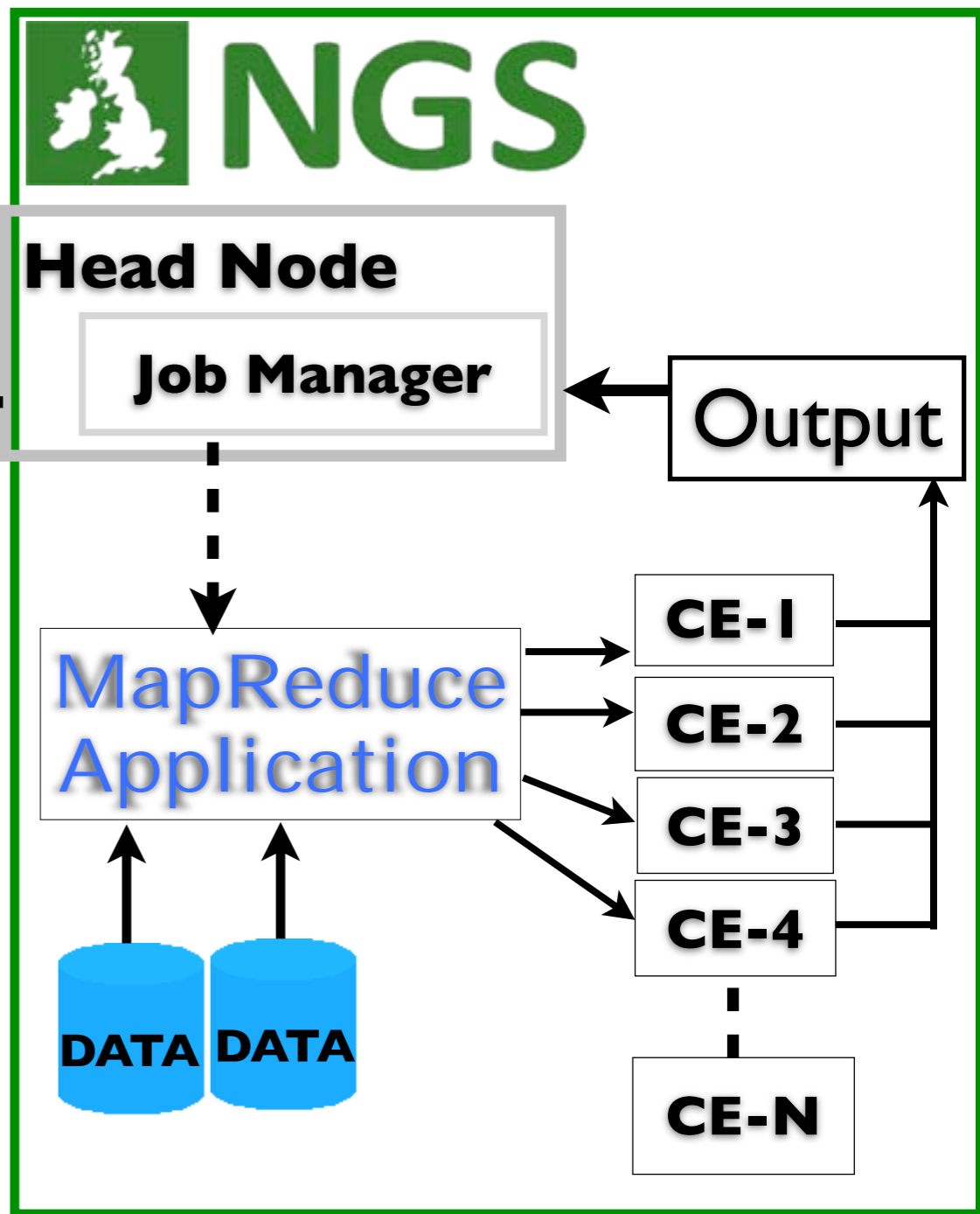
Submit Job

Using NGS

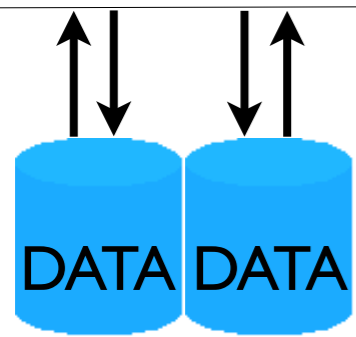


GRID SERVICES

Job Submission Client



WEB SERVICES



```
#!/bin/bash
echo "Starting application: scots-app "
echo "Submitting to job-manager at:"
echo $(bin/hostname -f)
echo "with arguments to main "
echo $*

cd /home/ngs0273/java/prog/scots-app

props=/home/ngs0273/java/prog/scots-app/src/main/resources/project_ngs.properties
echo "properties file:" $props

nthreads=4
echo "# threads=" $nthreads

/usr/local/Cluster-Apps/java-1.6.0_03/bin/java -cp target/scots-app-1.0-SNAPSHOT.jar scots.app.App "$@" $props $nthreads
```

Submit Job

Using NGS



ENROLLER
Enhanced Repository for Language and Linguistics Researchers

ENROLLER: An Enhanced Repository for Language and Linguistics Researchers

The ENROLLER project will provide an integrated online repository to enhance available electronic resources in the study of language and literature. Funded by JISC and based at the University of Glasgow, we aim to address the needs of the language and literature research communities by bringing together existing data to create an enhanced, interoperable resource. The project is a collaboration between the National e-Science Centre (NeSC), the University of Glasgow STELLA Project, the Newcastle Electronic Corpus of Tyne and Wear English (NECTE) and Scottish Language Dictionaries Limited (SLD), with other partners from commerce and higher education.

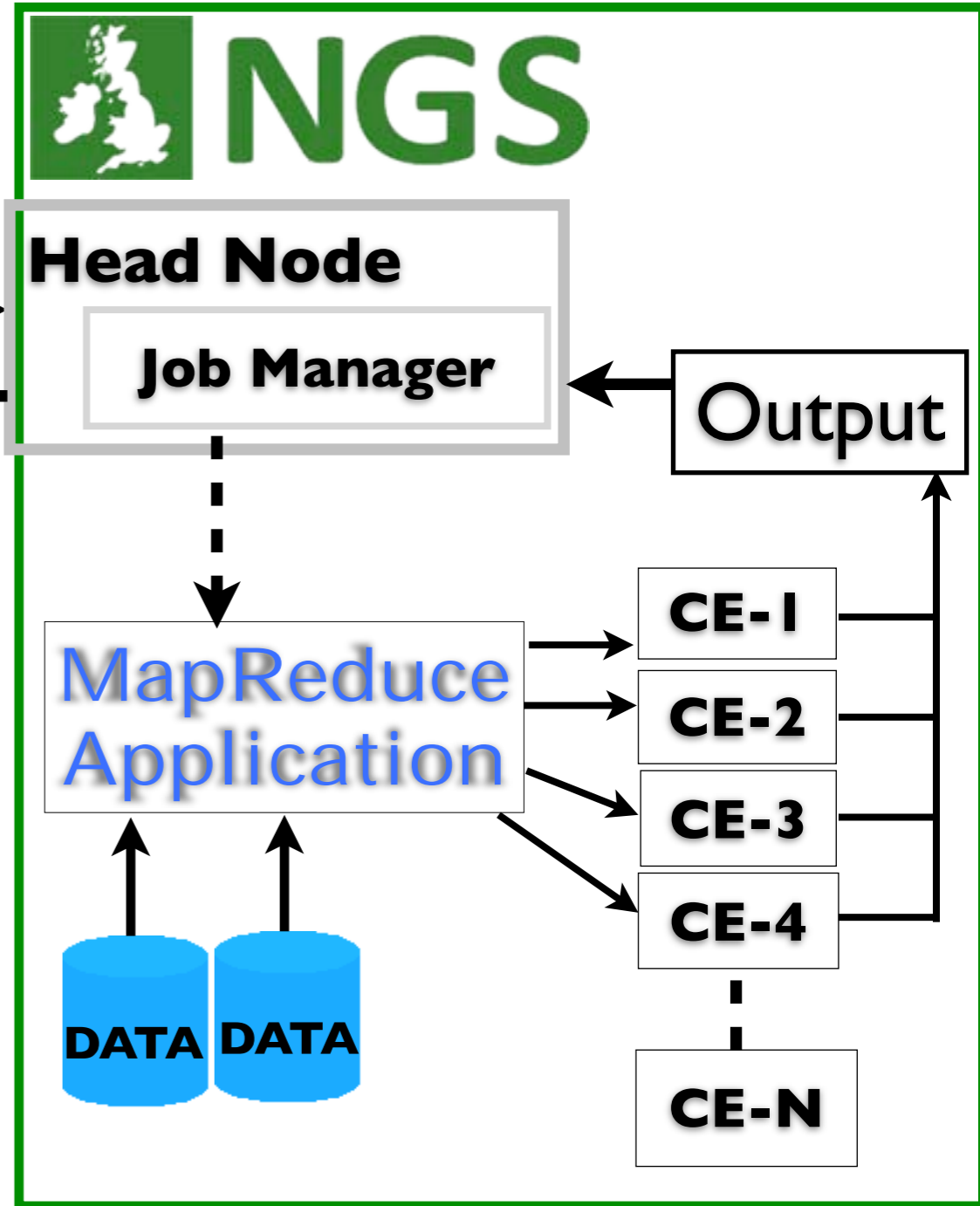
In doing so, the portal will provide seamless, secure and integrated access to such resources, enabling research between and across different types of library and linguistic data while harnessing powerful e-Science services and data analysis tools. We will use security-driven service delivery and targeted analysis and collaboration tools in a unified framework to create the greatly enhanced research repository. This will be achieved through supporting single sign-on to federated resources with advanced authorisation tailored to researcher interests and needs through a user-driven language and literature portal.

We hope to thus facilitate a broad spectrum of research in the languages and literatures of the UK, and to provide a model for other large-scale interoperable digital humanities repositories. The project will be under active development from 2009 to 2011.

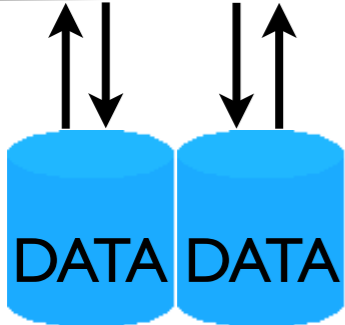
Lead Institution: University of Glasgow

GRID SERVICES

Job Submission Client



WEB SERVICES



```
#!/bin/bash
echo "Starting application: scots-app "
echo "Submitting to job-manager at:"
echo $(bin/hostname -f)
echo "with arguments to main "
echo $*

cd /home/ngs0273/java/prog/scots-app

props=/home/ngs0273/java/prog/scots-app/src/main/resources/project_ngs.properties
echo "properties file:" $props

nthreads=4
echo "# threads=" $nthreads

/usr/local/Cluster-Apps/java-1.6.0_03/bin/java -cp target/scots-app-1.0-SNAPSHOT.jar scots.app.App "$@" $props $nthreads
```

Future Work

- Development of Language Analysis Tools
- Addition of new data collections
- Addition of UI features in portal for better user experience
- Working towards the development of a VRE for Language and Literature community

THE END.

- Thank you.
- Questions?