Consolidated Comments for Usage Record Document (GWD.xx)

1) Following the practices of GFD.58 is strongly recommended

2) Adding a generalized extensibility capability to the UR definition would be very nice, since it would permit non-resource information to be described much more easily. It is recommended that this be done through updates to the definition of the UsageRecord and JobUsageRecord elements such that each allows an arbitrary number of extra elements not in the UR namespace after the main content of the record. It is not recommended that this be done through updating the UsageRecordType type because that leads to problems with some XML deserializers (apparently; I'm no expert on this aspect). Thus, for example, the JobUsageRecord schema definition would become:

```
<xsd:element name="JobUsageRecord" substitutionGroup="urwg:Usage">
   <xsd:complexType>
      <xsd:complexContent>
         <xsd:extension base="urwg:UsageRecordType">
            <xsd:sequence>
               <xsd:any namespace="##other"
                  minOccurs="0"
                  maxOccurs="unbounded"
                  processContents="lax"/>
            </xsd:sequence>
            <xsd:anyAttribute namespace="##other" processContents="lax"/>
         </xsd:extension>
      </xsd:complexContent>
   </xsd:complexType>
</xsd:element>
```

3) General Comments:

a) Section 3: We did not find a suitable quantity to describe information in an aggregated usage record such as "Number of Jobs".
b) Section 3.14: We find MachineName unsuitable to describe the site on which the job ran. We believe that "ExecutingSite" or "siteName" is more appropriate especially in a Grid computing environment where resources are distributed.
c) Section 3.18: ProjectName fits well in Grid projects like LHC and EGEE, but not in terms of virtual organisations which form a natural grouping. We recommend an additional field called "VirtualOrganisation" or "VO".
d) Section 10.8: CpuDuration type xsd:duration is not consistent with type xsd:positiveInteger listed in Appendix B
e) Section 13: Please provide an additional example for an aggregate record (see below)
f) Appendix D: GlobalUsername described in section 3.6 does not appear in Appendix D

4) Recommendation for additional example: Section 13.3
For an aggregate record describing the total work done at the "RAL-LCG2" computing site on the "EGEE" project for the user "Dave Kant" in the "dteam" VO in December 2005 :-
<A5> Can the editor confirm that the usage record would look as follows?
<A5> Can such an example be provided in the document?

```
<?xml version="1.0" encoding="UTF-8"?>
<JobUsageRecord xmlns="http://www.gridforum.org/2003/ur-wg";
```

```
xmlns:urwg="http://www.gridforum.org/2003/ur-wg";
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance";
xsi:schemaLocation="http://www.gridforum.org/2003/ur-wg
 file:/Users/bekah/Documents/GGF/URWG/urwgf-schema.09.xsd">

  <RecordIdentity urwg:recordID="" urwg:createTime="2005-12-01T02:15:45Z" />
  <aggregate>
   <GlobalUsername>/C=UK/O=eScience/OU=QueenMaryLondon/L=Physics/CN=davekant
   </GlobalUsername>
    <VirtualOrganisation>dteam</VirtualOrganisation>
    <ProjectName>EGEE</ProjectName>
    <ExecutingSite>RAL-LCG2</ExecutingSite>
    <Charge urwg:description="SpecInt2K">800</Charge>
    <NJobs>17423</Njobs>
    <WallDuration>PT3405H</WallDuration>
    <CPUDuration>PT3353H</CPUDuration>
    <StartTime>2005-12-01T02:15:45Z</StartTime>
    <EndTime>2005-12-28T23:54:17Z</EndTime>
  </aggregate>
 </JobUsageRecord>
```

5) The document represents a long-awaited attempt at standardization of resource usage records. We look forward to implement the format that will be agreed by the GGF participants. Still, much work has to be done in order to deliver an unambiguous specification that suits everybody and accommodates most important aspects.

a) The document does not give a clear definition of the "usage record", which leads to rather confusing statements such as "Record identity uniquely defines a record in the usage record" (page 6, definition of RecordIdentity).

b) Furthermore, when talking about the logging & accounting information associated to grid jobs it is crucial to give some kind of definition of a grid job. Many of the UR properties described in the document are ambiguous because it is not clear what constitutes a grid job. An example is StarTime. Is the brokering or stagein phase of a grid job taken into account?

c) The document states that "its main purpose is to outline the basic building blocks of the accounting record"

NorduGrid agrees with this scope, we consider the UR rather as a logging and accounting record of a grid job rather than simply a record of resource consumption. For us the UR should contain information not just about resource consumption but also about job identity, ownership, status, etc. Maybe it would make sense to rename the UR as logging or accounting record?

d) Many of the attributes are inadequately defined or their meaning is deliberately left open. This partly defeats the purpose of the schema since it was designed specifically to be a format for exchanging usage data over grids but if interpretations of the attributes are different at different sites/grids, comparing exchanged data becomes complicated. For example the Charge, Status, StartTime, EndTime, base properties claim that "The meaning of this charge will be site dependent", "the semantic meaning of the status is site

dependent" or "the value of this property may depend on the queuing system", etc.

e) The GGF UR proposal is still too 'site' or batch system specific; in many places the document assumes data exchange among computing centers (sites) and not among Grids. Seems that the grid layer is not really taken into account, the UR at many places resembles as a data exchange format between batch systems (or sites) and not Grids. For example, the Current Practices Survey (Appendix A) covers only supercomputing centers and not a single Grid. The UR document should make it clear it is a Grid UR proposal, enabling grid job logging & accounting data exchange between different Grids.

f) The document states: "The document does not attempt to dictate the format in which the accounting records are stored at a local site, instead it meant to be a common exchange format"

NorduGrid fully supports this approach (we assume "Grid" everywhere we read "site"), we consider the UR as an exchange format between Grids. This implies that the GGF UR should be powerful enough to "accommodate" other Grid/local Usage Records. Later in these comments, a mapping of the NorduGrid UR attributes to the GGF UR attributes is presented. Unfortunately, not all of the planned NorduGrid UR information can be described in the GGF UR without relying on the GGF UR Extensions.

g) Appendix B lists the "Common Usage Record Properties"; it is not clear what is the purpose of this incomplete table? Many of the base properties are not listed in the table (e.g. RecordIdentity, GlobalJobID, LocalJobID, LocalUserID, GlobalUserId, etc..)

h) Two properties, TimeDuration and TimeInstant, described in the XML schema (11.7, 11.8 on page 33) are missing from Section 3.

i) The Introduction distinguishes between Base properties and Differentiated properties; nevertheless, the "natural language form" part of the document contains only "Base properties" section and no "Differentiated properties", all the attributes are described under Section 3 "Base properties". On the other hand, the XML representation part lists many of the attributes (starting from Network) as Differentiated properties in Section 11.

j) The term "unique" is used rather ambiguously in several places in the draft. Most of the time it is not specified on which level the uniqueness is required. An example: page 12 Section 4: "Each occurrence of the Host property should contain a unique value". Unique within a cluster, a site, a Grid or globally unique?

k) Experience shows that the record format should be optimized for usage in databases, e.g., avoiding string variables, introducing strictly defined enumerations etc.

6) Detailed discussion of different GGF UR properties

a) 3.1 RecordIdentity

The definition of this property is rather strange, "defines a record in the usage record". This necessitates the definition of the Usage Record.

What is the scope of the uniqueness of this property?

The property name is not consistent with the other ID type attributes, those are called GlobalJobID, ProcessId, LocalUserID while this attribute is called RecordIdentity. It would make sense to rename it as RecordId, or GlobalRecordId, depending on the scope of the uniqueness.

b) 3.6 GlobalUsername

What about renaming this property as GlobalUserID, since there is already GlobalJobID, LocalJobID, LocalUserID and this could be GlobalUserID.

c) 3.9 Status
"This property represents the completion status of the job" and goes on to say that "it must support the following values:
   a) completed
   b) failed
   c) aborted
   d) held
   e) queued
   f) started
   g) suspended"

This is a rather batch system specific approach. In a grid usage record we are talking about 'final' or completion status of a grid job. In this sense differentiating between "held, queued, started, suspended" has not much relevance. It would be better to merge these 4 active states into one "unfinished" state.

BTW, NorduGrid uses the Finished, Failed, Killed, Deleted labels for completed job states. The NorduGrid "Finish, Failed, Killed" corresponds to the GGF "completed, failed, aborted", respectively while the NorduGrid "Deleted" does not exists in th GGF states (see page 26 about nordugrid-job-status in [3])

The statement "This property must contain data of type integer" implies that the proposed job states above also should carry an associated number, in this case.

h) 3.10 WallDuration, 3.11 CpuDuration

The definition has to make clear what belongs to these durations. Is this only the time the grid job spends actively in the batch system, e.g. running on a execution node? What about time the job spends in waiting in the batch queue, doing grid stage-in/out, sitting in a grid-broker? See the general comment about the definition of a grid job.

i) 3.12 EndTime

"The time at which the job completed"

Completed where? In the batch system? In the grid layer? Is grid stage-out included? Maybe there is a need for several EndTime-like properties? The timestamp when the job finished in the batch system, the other timestamp the job completed with e.g. the grid-stageout?

In the case of NorduGrid, "EndTime" refers to the grid stageout as well, as the job is not considered to be completed until its output is off-loaded to a grid-storage and grid-registered. Again, there is a need for unambiguous common grid job definition.

Another small thing: the definitions of Endtime given in 3.12 (page 8) and 10.9 (page 27) are not exactly the same.

j) 3.13 StartTime

"The time at which the job started"

Again, we are missing the definition of a grid job here. What does the "grid job started" mean here? The job was submitted to a grid broker or registered in a job pool? Or the job has reached the selected grid compute element front-end? Or the job got submitted to the batch system? Or the job started to run on a computing node in the batch system?

The current definition does not take into account the grid layer and probably simply assumes the last option. This might not be necessarily adequate, as resources spent for e.g. matchmaking, brokering, data staging etc may not be negligible.

k) 3.14 MachineName / 3.15 Host

This is a very batch system specific view, again. The relation of MachineName & Host should be more accurately defined. We assume MachineName here is actually meant to be 'site' or 'computing resource'.

l) 3.17 Queue

Is this a queue of a grid scheduler (broker) or the execution queue of the local batch system? Has to be clarified. Actually, both can be interesting in an accounting record.

m) 3.19 Network

The document defines this as "The network used by the job" where the default metric is "total volume of data transferred in the specified unit".

Within NorduGrid, there is a need to be able to specify the amount of data moved as a

part of the grid stagein, job execution and stageout process separately.

7) Missing information from the GGF UR

a) No information is recorded in connection with the original grid job request. The (part of the) grid job specification request should be recorded in an accounting record.

Within NorduGrid, the complete job description is stored as a part of the job usage record. There are also separate attributes to record the "requestedcputime, requestedmemory, requesteddisk".

b) No information is recorded in connection with the grid stagein/stageout process. Download/Upload resource consumption, downloaded/uploaded data volume (see comment under the Network property) would be useful.

c) No information is recorded in connection with job failures. Job exit codes, error codes, error messages are required by most user communities.

NorduGrid records "exitcode" and "failurestring".

d) Information about the local resource management system executing the grid job is missing. NorduGrid records the hosting "lrms" type for each grid job.

e) Information about requested/used application environments are also missing.

8) Multiple levels of definition are confusing; sections 10 and 11, repeat and sometimes alter the definitions in sections 1 through 9.

9) The appended schema mandates the inclusion of the “status” field. This is described as a must in section 3.9 but is not specifically included as a must in section 10.6. Further the allowed values in 10.6 are not enforced in the Schema (this might be okay as other values are allowed). Suggest that this is removed as a mandatory record as it is not applicable in the general case (what about data records?)

10) Queue fields are provided for, but not the batch system on which the queue resides. Suggestion to add <Backend> tag or equivalent for systems that run multiple batch systems. e.g., PBS, SunGrid Engine, Fork. Note: The “Resource” field could be used to describe this but for consistency the queue should probably be described here to and not as a separate entity.

11) A Resource Broker is different from a meta-scheduler (provided for by Global-job-id), suggest an additional field to capture jobs from a resource broker (which may themselves use a meta-scheduler)

12) How ambitious is the scope of a Usage Record? For job accounting it’s likely to be successful. However the current extensibility mechanism and pre-defined fields, I

suspect, will make life very difficult to try to capture a Usage Record for a database resource or file store (whatever form they may take…)

13) Concerning Sections 3.14 and 3.15 of the specifications:

MachineName and Host may not be sufficient where, for example, there is more than one cluster at a site. We propose to have an additional element "SiteName" (see also the proposed name change from MachineName to SiteName by Dave Kant, keeping both elements instead of changing names might allow to include more detailed information in the UR, specifying both the cluster name as MachineName and the site ID as SiteName).

Furthermore in some grids the grid resource ID is different from Host, MachineName or SiteName (for example, for LCG/EGEE, computing element IDs such as "t2-ce-01.lnl.infn.it:2119/jobmanager-lcglsf-cms" are used as resource IDs and there may be several (logical) computing elements on a single cluster). Thus there should be an additional element, "ResourceID".

14) Concerning Section 3.18 (ProjectName) of the UR specification: Since a single Virtual Organization (VO) may be involved in several projects and several VOs may be involved in the same project (thus having only the ProjectName is not sufficient), the usage record should contain an element indicating the "UserVO" (see also the equivalent proposal by Dave Kant).

Additionally, the role (such as VO admin, student ...) of a user might influence issues like charging or quota enforcement. It would therefore be good to have an additional element "UserRole" (might also be a part of the element UserVO that we proposed above).

15) The OGSA WG is working on an information model for Grid resources. It would be nice if the UR work and the OGSA work agreed on terminology, etc. But this may entail unwelcome delay in the standardisation process for UR.

16) The document seems addresses the usage of raw compute facilities when running a computatation. Can it also handle the case where the resource provided is raw storage, perhaps with various levels of QoS. See e.g. the SRM work of the GSM WG.

17) The document addresses the usage of raw compute facilities - processors, storage, network - when running a "job". Would it make sense to address other cases, such as tracking the usage of a particular application? This is important e.g. for licensing models - see the discussion panel on licensing at GridWord 2005.

I see two sub-cases. In the first, the computing resource provides the application. In the second, it is staged in as part of the "job". These may need to be addressed differently.