APPENDIX TO THE CHARTER OF THE WORLDWIDE PDB (wwPDB)

1. Current conventions for the PDB accession codes.

PDB accession codes (currently the four-character codes) will be assigned in a coordinated fashion by the Common Deposition and Annotation system at all deposition sites. The use of identifiers by members of the wwPDB resembling PDB accession codes for other purposes is prohibited. The four-character code consists of a number (0-9) followed by 3 letters or numbers. Members of the scientific community will be informed of this guideline and asked to conform.

2. Current file name conventions in the PDB archive

Data files in the PDB archive are distributed with the reserved conventional names as show in Table 1A. These file name conventions are to identify archival PDB entries and should not be used for other purposes.

Archive File Type	File Name Template [1]
PDB format coordinate file	pdb*<entry_id>*.ent
PDBx/mmCIF format coordinate file	*<entry_id>*.cif
PDBML format coordinate file (complete)	*<entry_id>*.xml
PDBML entry file without atom records	*<entry_id>*-noatom.xml
PDBML entry file with alternative format atomic records	*<entry_id>*-extatom.xml
X-ray experimental data file (PDBx/mmCIF)	r*<entry_id>*sf.ent
NMR experimental/constraint data file (deposited format)	*<entry_id>*.mr
NMR experimental/constraint data file (NMRSTAR format)	*<entry_id>*_mr.str
NMR experimental/chemical shifts data file (NMRSTAR format)	*<entry_id>*_cs.str
PDB Resource Description Files (RDF)	*<entry_id>*.rdf

Table 1A. PDB Data File Name Conventions (Note 1. In the file name template *<entry_id>* refers to the PDB 4-character accession code. Files are stored in compressed format on the PDB ftp sites and their file names include an additional ".gz" suffix to reflect this.)

3. Guidelines for the redistribution of PDB data files:

3.1 Data files containing PDB content may incorporate the PDB 4-character entry name (e.g. 1ABC) in standard PDB records only if they are exactly the same as what is residing in the PDB archive. This does not prevent databases from including PDB entry_id's as cross-references here it is clear that they refer to the PDB archive. PDB records refer to the standard PDB format.

3.2 The distribution of modified PDB format data including the records: HEADER, CAVEAT, REVDAT, SPRSDE, DBREF, SEQADV, and MODRES is not allowed.

3.3 The distribution of modified PDBx/mmCIF format data including categories in the dictionary category groups DATABASE and AUDIT is prohibited.

3.4 The data schema for canonical PDBML will be electronically translated from the PDB exchange dictionary. Changes in the schema will automatically track changes in the exchange dictionary. In addition to the canonical PDBML representation, XML data may also be delivered in a content consistent form in which the coordinate records are presented as single record rather than in the fully marked-up canonical form.

3.5 The RDF files will contain RDF schema which will reference back to the wwPDB URL where the necessary schema definition will be available.

4. Current and future guidelines for data exchange supporting the maintenance of the PDB Archive:

4.1 Current update procedure:

Current data exchange between wwPDB member sites is based on the Linux software application RSYNC. The RSYNC application maintains synchronization of file systems between local and remote sites by comparing timestamps and checksums for each file and transferring copies of any outdated files between sites. The overhead of RSYNC synchronization is substantial and updates are currently performed only on a daily or weekly schedule.

Owing to the differences in current data processing procedures used by the wwPDB members, the weekly update of new and modified entries in PDB Archive has not been fully automated. In addition to the exchange of data files among the sites, the weekly PDB update involves a number of tasks including file format conversion, file checking, and data integration. Some human oversight is also required to review the outcomes of these steps. The RCSB PDB has been responsible for these update tasks during the course of the 2003 wwPDB agreement.

4.2 Future update protocol:

Data exchange in the future system will be performed using a protocol supported by

NetApp filer hardware. Using this protocol, each wwPDB site creates two separate file systems: one designated for data import and one designated for data export. These files system hold the data to be *received from* and *sent to* the partner sites. Keeping the file systems distinct and requiring that the imported file system be read-only guards against unintended overwriting of data files. Data transfer within the NetApp system takes place directly between the NetApp filer systems and does not incur any operating system overhead on the Linux servers. As a result, data synchronization can be achieved on a time scale of 5 minutes between wwPDB sites. Using the NetApp hardware protocol allows all the wwPDB sites to maintain a current view of the data exported by any other member site with near real-time currency.

The adoption of both common data processing procedures and software at all wwPDB member sites will permit greater automation in the preparation and delivery of updates of the PDB archive. Each site will be able to prepare and to export all of the data files to be released or revised by the site. Because the new data exchange protocol allows for the synchronization of data exported by all sites, all of the data files required for a weekly update are automatically available to every site. The responsibility for maintaining the master copy of the PDB ftp archive will be assigned to an "archive keeper".

4.3 Transition and implementation of the future update procedure:

During the course of the 2013 wwPDB Agreement, the procedure for updating the PDB archive will transition from the present to the future technologies described in sections 4.1 and 4.2, respectively. Support for the current update system will be maintained as the new software system is tested and current systems are retired. The RCSB PDB will continue to act as the archive keeper and perform the current weekly update procedure, and the RCSB PDB will perform the steps necessary to integrate data from the current and future update procedures during the transition period. The RCSB PDB will continue to be the "archive keeper" for the first five years of this agreement at which point these arrangements can be renegotiated if necessary.

Larger updates related to remediation will take advantage of the data exchange and common software described in section 4.2. The coordination of these larger updates will be also be accompanied by: advance notification of the user community (60 day), wwPDB-wide internal review and testing of remediated data, pre-release testing of new features and changes by external users, special pre-release coordination for data loading for wwPDB web resources, final integration of remediated data within a weekly data release.

5. Guidelines for maintenance of the PDBx data dictionary:

5.1 An internal dictionary working group will be responsible for the overseeing the PDBx/mmCIF dictionaries used by the wwPDB. The group will be responsible for maintaining, updating and revising the content of PDBx/mmCIF dictionaries and translation to corresponding XSD and RDF/OWL schema. The internal working group will include representatives from the wwPDB member sites and will meet on at least a

monthly schedule to discuss dictionary content and maintenance related issues. The site representatives will be responsible for setting the agenda for meetings, identifying any required internal or external consultants, and communicating discussion issues within each site. Summaries of discussions within this group should be regularly reported to the wwPDB leadership.

5.2 The site representatives will be responsible for ensuring that the review any new content, changed or updated content are conducted in a timely manner (~1 week). The review of any such changes/updates will follow the procedure:

5.2.1 Proposed changes for new content or content reorganization will be discussed by site representatives before posting it on the internal wwPDB website for review. Any new proposals for addition or changes to the dictionary should include proposed definitions and supporting examples. If necessary these will be further discussed internally with other members at each site The site representatives can invite additional members if they feel necessary.

5.2.2 After internal evaluation, the site representatives will assess the need for external review. External reviewers of community experts will also be used to advise the internal working group on the content requirements for new methods and technologies.

5.3 The internal working group will be responsible for the regular release of any updates or changes to the dictionary. These changes will be made available via the wwPDB dictionary documentation site and will include description of changes in a given release. The site representatives will be responsible to assess the impact of any dictionary updates at the individual wwPDB member site. The dictionary updates will be released based on the following guidelines:

5.3.1 Bi-weekly updates will be made for any corrections and updates to controlled vocabularies. These changes will be identified by an increment in the minor version number of the dictionary and schemas.

5.3.2 Quarterly updates will be made for new content (minimum impact on existing entries). These changes will be identified by an increment in the minor version number of the dictionary and schemas.

5.3.3 Major updates/changes/reorganization of dictionary will be reviewed and announced with appropriate notice and leadership approval. Changes in this category include those with major impact on existing entries or a major impact on an internal or external wwPDB member site. A major update will be identified by an increment in the major version number of dictionary and schemas.

5.3.4 The working group will also conduct regular validation of the content dictionary, associated schemas. Any issues arising from such validation checks will be handled by correcting the dictionary and will be released on a bi-weekly update schedule.

5.4 Detailed documentation of this dictionary maintenance procedure including the working group participation and candidate external reviewers will be maintained by the working group.

6. Guidelines for maintenance of reference data:

6.1 The chemical component dictionary will be managed in a version control system hosted by a wwPDB member. The dictionary data will be accessible to all of the wwPDB members.

6.2 BIRD - guidelines to be established at some future date.

6.3 Validation reference data - guidelines to be established at some future date.

7. Changes to this document:

7.1 The contents of this document can be changed at any time, by mutual agreement of the wwPDB member sites.

7.2 The contents of this document will be reviewed at least once a year to make sure they are up to date and accurate.